

## **Copyright aspects of feeding terminology databases Barcelona speaking points**

Mikael Johansson, Legal adviser, Unit DGT-03, Interinstitutional relations and General affairs  
Directorate-General for Translation of the European Commission

### **1. INTRODUCTION**

I represent the Directorate-General for Translation (**DGT**) of the European Commission, one of the institutions of the European Union. The others are the Council, the European Parliament, the Court of Justice, the Court of Auditors, the Committee of the Regions and the Economic and Social Committee. I am here today representing DGT, not the European Union.

As an institution, DGT was more or less spared copyright considerations until recently. Since we almost exclusively translate laws, reports and other texts that have been drafted within the European Commission, we never really had to worry about copyright in the source texts. And a peculiarity is that once our translations have left DGT, they are considered as original documents not translations.

This is reflected in the fact that DGT, that translates 1,5 million pages per year in 23 languages, doesn't really have a dedicated copyright lawyer. DGT has several lawyers, but their brief concerns mostly issues relating to the languages regime of the Commission, and institutional issues regarding the functioning of DGT. This year, the lawyers have even drafted the rules and instructions for the translation contest, *Juvenes Translatores*, that will culminate on 14 November, when more than 2000 pupils around Europe will sit down simultaneously to a translation contest.

However, with the opening to the public of **IATE**<sup>1</sup>, the European Union's interinstitutional database for terminology, copyright issues are a priority.

I will first give a background to the IATE project, and then move on to the legal challenges ahead.

### **2. THE BACKGROUND TO IATE**

Each of the institutions has its own translation service. There are good reasons for this: the institutions have different roles in the legislative process and in general in the democratic

---

<sup>1</sup> IATE is the abbreviation for InterActive Terminology for Europe.

machinery of the Union, so their independence must be ensured. Historically, the contacts between the translation services have been sparse, so when the time came to compile computer based term databases, each service built their own: Eurodicautom at the Commission, TIS at the Council, and Euterpe at the Parliament<sup>2</sup>.

But at the Commission, much of the straightforward everyday terminology work is carried out by the translators themselves as they translate their way through a document. Eurodicautom was slightly cumbersome to handle, and I don't believe all translators had writing rights, so the translators used other formats, often simple word documents, or Multiterm databases, to make their own term collections. In some language departments, the translators used almost exclusively these more informal bases.

And remember, the term bases were created for exclusively internal use. And there were first 12, then 22 and now 23 languages.

The troika of TIS, Euterpe and Eurodicautom, the "official" term bases, were managed by terminologists according to guidelines with a certain control. For Eurodicautom, the Commission often bought batches of terms from outside companies. But the informal termcollections and term bases, that had become very important for the everyday work of the translators, were often handled without any real supervision.

**That was the situation when the institutions decided to merge all the term bases into one, the mother of all term databases, IATE.**

It was soon realized that as much as possible of the informal term bases or termcollections had to be added too, if IATE was going to fulfil its purpose.

The informal termcollections were in many different formats, that probably only had one thing in common: that they were not immediately importable into IATE.

This was an enormous conversion project, there was no time to analyze every term entry in every respect. The terminologists settled, logically and out of necessity, for term reliability as criteria: reliable terms in, unreliable out. This meant that references, contexts etc. often were entered together with the term with little investigation. So also e.g. personal data of sources found its way into IATE, together with extracts from external sources.

All in all, around 1,4 million entries with almost 9 million terms were entered in this terminological tidal wave. This is what we call **the legacy data**.

This was a problem, but a problem we could hide, more or less, until the political decision was made that IATE should be made accessible to the public via the Internet. No more hiding.

At this stage, the lawyers were consulted. They immediately spotted two things: the legacy data might contain infringements of copyright and personal data protection, and there were no provisions for sharing between the partners any liability that might result.

That more or less brings us up to date.

---

<sup>2</sup> The other institutions also had their own term databases.

### 3. LEGAL ISSUES

Before I outline the problems that lie ahead, I will bore you with some figures, to give you a hint of the size of the work:

IATE is the chosen receptacle for the terminology work of the 10 IATE-partners: 7 EU institutions, that I have mentioned, 2 EU banks (the European Investment Bank and the European Central Bank), and one EU agency, the Translation Centre<sup>3</sup>.

The database covers 23 official languages.

In the first quarter of 2007, almost 17 000 terms were added, 35 000 terms modified, and 38 000 terms validated.

All commission translators (around 2600) have writing rights; in the other institutions, the right is mostly restricted to terminologists, which means that altogether, we have almost 3000 people who can enter information, few of whom, I may add, are lawyers.

Consultations via the website is around 200 000 per day.

The legal challenges I see are personal data protection, that I won't mention again, since it is not the subject of today's sermon, and copyright.

And here I mean only the risk that we are infringing somebody else's copyright. We are not, or not yet, worrying about the protection of our copyright arising from the base.

In our analysis we have divided the issues into the internal division of responsibility, dealing with any copyright infringements that might be inherent in the legacy data, and minimizing the risk of future infringements.

#### 3.1. Internal division of responsibility.

This is dealt with in a new IATE agreement, currently under negotiation.

In the agreement, we specify that each partner remains liable for the information it has entered into the database; we share liability for information that cannot be attributed to a certain partner. The last provision targets consolidation projects for individual entries. When the old databases were merged, we ended up with many doubles, triples etc. Terminologists from all partners now work together to get rid of as many of these as possible. The idea is that once an entry or a term is consolidated, the partners will share the responsibility for it.

This construction may seem strange, since the institutions are not legal persons<sup>4</sup>. We chose this construction because some of the partners do have legal personality (the banks, the Translation Centre), and we believe that a joint responsibility could be

---

<sup>3</sup> EU agencies are fairly independent entities that are created to deal with technical aspects of a certain issue. There is e.g. the European Environment Agency in Copenhagen, or the European Aviation Safety Agency in Cologne. There are about 20 such agencies. The Translation Centre was created to cater to their translation needs.

<sup>4</sup> To put it simply, they do not speak for themselves in legal matters, and can neither be defendant or complainant in a court proceeding. If there is a claim on any of the institutions, the defendant would be the *European Community* – that has status of legal person – with the Legal Service of the Commission as legal representation.

perceived as diminishing the pressure on the individual partners to manage the copyright issue.

### 3.2. The legacy data – can it come back to haunt us?

We were asked to ensure that no copyright dispute could arise from the legacy data. Our simple answer was that the only chance to do that would be to not open IATE up to the public. This was, however, not an option.

Now, do we infringe copyright, or don't we? Can we be sure that we are not in breach? Simple answer: No.

One worry is that some sources may have been heavily raided. We suspect that some sources have been quoted a large number of times. We want to avoid a situation where somebody can claim that they can't sell their product because crucial parts of it are available free in IATE.

Of course, the problem is mostly about quotes. Certainly IATE contains many "cut and paste" from various sources. Most of them are very short, but some are longish, and lack a reference to the source.

**Which risks do you take when you quote?** First, let's look at what a terminologist does. He or she

- (1) investigates the meaning of the term in the source language;
- (2) finds the equivalent concept in the target language;
- (3) finds the closest term in the target language;
- (4) draws conclusions and defines the level of equivalence between the source language term and the proposed target language term.

This is, of course, independent, creative and original work, that would probably be copyright protected in its own right.

A quoted definition or context becomes background material. In fact, we add information to the quoted definition by connecting it to its equivalent in other languages.

The *fair use* doctrine has been claimed by many to cover IATE. I don't think that it goes without saying that *fair use* could be claimed as a general protective shield for IATE.

- (1) It is not certain that e.g. the Berne Convention's *fair practice*<sup>5</sup>, the European copyright Directive's *fair practice*<sup>6</sup>, or the American *fair use*<sup>7</sup> would cover the compilation of a term base.

---

<sup>5</sup> Article 10 para. 1 of the *Berne Convention for the Protection of Literary and Artistic Works*: "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, including quotations from newspaper articles and periodicals in the form of press summaries."

- (2) There is no coherent *fair use* principle applied universally.
- (3) Copyright law, while based on a number of international conventions, is still adopted nationally, with national variations.
- (4) Therefore, we would need to investigate the national law of the country where the source is located.
- (5) It may be very difficult to know what country a source is from, especially on the Internet.

Anyway, we hope that copyright owners may be reluctant to raise a case, since any compensation awarded would probably be small.

**In case of dispute, our first line of defence will be to immediately offer to withdraw** information when there is a complaint.

We would point out that we make no money from the database, in fact, it costs us quite a lot to keep it up (more than 627 000 Euros for 2007). So anything we quote is strictly for non-commercial use.

The transformative nature of terminology work, according to my model above, may work in our favour.

We would maybe even argue that it is a sign of recognition and authority to be referred to as a source in IATE, so the author does get something out of it.

**But our conclusion is that it is best to wait and see** what comes up. To mount a pre-emptive strike against our most valuable terminology resource to clean out any suspected elements seems to be overkill.

### 3.3. Minimizing the risk of future infringements

**To minimize the risk of future infringements**, we are in the process of drafting guidelines for the people who have writing rights. This is not easy.

One big challenge is the workload for our terminologists. At the Commission, there is usually one per language, that is, for about 65 translators. Since our translators always have to work on very short deadlines, the terminologists usually work on even shorter deadlines to satisfy their needs. Consequently, they have to look for fast answers. This does not mean, however, that IATE is fed in a hurry. After having provided the translator with the answers, the terminologist has to find the time to properly validate the information that he or she intends to enter into IATE.

---

<sup>6</sup> According to article 5 para. 3 subpara. d) of the *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society*, Member States may exempt from copyright "quotations for purposes such as criticism or review, provided that...their use is in accordance with fair practice".

<sup>7</sup> *Copyright act of 1976*, 17 U.S.C. § 107: "...the fair use of a copyrighted work...for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright."

So they may copy a definition if they find an authoritative source; they may be reluctant to rewrite a given definition, since every change carries a risk of making a mistake, and it takes time. And definitions are usually very concise, so you can't really edit out any text. But already many terminologists are adapting to the new situation, and writes their own concise definitions.

What they need last of all is to have to make a copyright decision every time they need to quote.

Perhaps the one priority in drafting the guidelines is, however, to avoid making the terminologists so unsure of what they can do that we hinder their work, because DGT couldn't work for one day without their efficient and professional terminology support. If DGT stops, the entire EU stops functioning, believe it or not, because no legislation enters into force unless it is available in all languages, no court case can be settled without translation, and no institution could work. That's reality with 23 official languages.

Consequently, we will have to keep the guidelines very concrete.

Starting point, very briefly: the terms as such are not under copyright; term pairs would probably not be under copyright<sup>8</sup>; text (definition or context) will probably be under copyright. There is no blanket exception from copyright that covers term databases.

A working group with representatives from all the IATE partners is drafting the guidelines; so my personal opinion that the guidelines could look like something like this **neither reflects the opinion of the working group, nor can it be regarded as my final proposal. This is a work in progress:**

- (1) Check the copyright notice. If it says that you can quote, you are home free.
- (2) Don't copy info from dictionaries that are readily available to the translators.
- (3) If you have to quote from a protected source, ask permission.
- (4) Quote as little as possible, and always state the source.
- (5) Use your common sense; don't overquote, don't overuse a source. If in doubt, contact the lawyers.

I admit that this list<sup>9</sup> will not make any terminologist happy. I hope that the working group<sup>10</sup> may make it more palatable. You may see a contradiction in that I say: don't quote! and still give advice on what to do if you quote. No contradiction; what I want to express is that ideally, we shouldn't quote, but I am realistic enough to understand that we will.

---

<sup>8</sup> Size matters here: it would definitely be illegal to quote all the term pairs from a dictionary where the rights have been reserved.

<sup>9</sup> As I have pointed out, the list **neither reflects the opinion of the working group, nor can it be regarded as my final proposal.**

<sup>10</sup> Did I mention that the list **neither reflects the opinion of the working group, nor can it be regarded as my final proposal?**

#### 4. THE END

But in the end, the real challenge is not legal uncertainty, but the complexity of the structure, the sheer size of IATE, the number of people with writing rights, the tight deadlines that DGT always work to, the urgency of the work of the terminologists. All this we have to deal with for any measure that concerns IATE, and all this may conspire to the early death of any guidelines born out of the working group's efforts.

So this short introduction to the copyright management of IATE will have illustrated that while copyright and IATE is about law, it is even more about human resource management.

A quote – yes, I will risk it – from the pessimistic Irishman Yeats comes to mind:

Turning and turning around in the widening gyre  
The falcon cannot hear the falconer;  
Things fall apart; the centre cannot hold  
Mere anarchy is loosened upon the world

But I would like to end on a cautiously optimistic note, with the wisdom of the poor theatre producer Henslowe in the movie *Shakespeare in love*:

*Henslowe*: Mr. Fennyman, allow me to explain something about the theatre business. The natural condition is one of insurmountable obstacles on the road to imminent disaster.

*Fennyman*: So what do we do?

*Henslowe*: Nothing. Strangely enough, it all turns out well.

*Fennyman*: But how?

*Henslowe*: I don't know. It's a mystery.

Barcelona, 13.10.2007