

Translation skill-sets in a machine-translation age

Anthony Pym
Intercultural Studies Group
Universitat Rovira i Virgili
Tarragona, Spain

Version 1.2 (May 2012)

Abstract: The integration of data from statistical machine translation into translation memory suites (giving a range of TM/MT technologies) can be expected to replace fully human translation in many spheres of activity. This should bring about changes in the skill sets required of translators. With increased processing done by area experts who are not trained translators, the translator's function can be expected to shift to linguistic postediting, without requirements for extensive area knowledge and possibly with a reduced emphasis on foreign-language expertise. This reconfiguration of the translation space must also recognize the active input roles of TM/MT databases, such that there is no longer a binary organization around a "source" and a "target": we now have a "start text" (ST) complemented by source materials that take the shape of authorized translation memories, glossaries, terminology bases, and machine-translation feeds. In order to identify the skills required for translation work in such a space, a minimalist and "negative" approach may be adopted: first locate the most important decision-making *problems* resulting from the use of TM/MT, and then identify the corresponding skills to be learned. A total of ten such skills can be identified, arranged under three heads: learning to learn, learning to trust and mistrust data, and learning to revise with enhanced attention to detail. The acquisition of these skills can be favored by a pedagogy with specific desiderata for the design of suitable classroom spaces, the transversal use of TM/MT, students' self-analyses of translation processes, and collaborative projects with area experts.

My students are complaining, still. They have given up trying to wheedle their way out of translation memories (TM); most have at last found that all the messing around with incompatibilities is indeed worth the candle: all my students have to translate with a TM all the time, and I don't care which one they use. Now they are complaining about something else: machine translation (MT), which is generally being integrated into translation memory suites as an added source of proposed matches, is giving us various forms of TM/MT. These range from the standard translation-memory tools that integrate machine-translation feeds, through to machine translation programs that integrate a translation memory tool. When all the blank target-text segments are automatically filled with suggested matches from memories or machines, that's when a few voices are raised: "I'm here to translate," some say, "I'm not a posteditor!"

"Ah!", I glibly retort. "Then turn off the automatic-fill option..."

Which they can indeed do. And then often decide not to, out of curiosity to see what the machine can offer, if nothing else.

The answer is glib because, I would argue, statistical-based MT, along with its many hybrids, is destined to turn most translators into posteditors one day, perhaps soon. And as that happens, as it is happening now, we will have to rethink,

yet again, the basic configuration of our training programs. That is, we will have to revise our models of what some call translation competence.¹

Reasons for the revolution

MT systems are getting better because they are making use of statistical matches, in addition to linguistic algorithms developed by traditional MT methods. Without going into the technical details, the most important features of the resulting systems are the following:

1. The more you use them (well), the better they get. This would be the “learning” dimension of TM/MT.
2. The more they are online (“in the cloud” or on data bases external to the user), the more they become accessible to a wide range of public users, and the more they will be used.

These two features are clearly related in that the greater the accessibility, the greater the potential use, and the greater the likelihood the system will perform well. In short, these features should create a virtuous circle. This could constitute something like a revolution, not just in the translation technologies themselves but also in the social use and function of translation. Recent research (Pym 2009, García 2010, Lee and Liao 2011) indicates that, for Chinese-English translation and other language pairs², statistical MT is now at a level where beginners and Masters-level students with minimal technological training can use it to attain productivity and quality that is comparable with fully human translation, and any gains should then increase with repeated use. In more professional situations, the productivity gains resulting from TM/MT are relatively easy to demonstrate.³

Of course, as in all good revolutions, the logic is not quite as automatic as expected. When free MT becomes ubiquitous, as could be the case of Google Translate, uninformed users publish unedited electronic translations with it, thus recycling errors that are fed back into the very databases on which the statistics operate. That is, the potentially virtuous circle becomes a vicious one, and the whole show comes tumbling down. One solution to this is to restrict the applications to

¹ Here I refer more readily to “skill sets” rather than “competence” because the latter has been polluted as a term in translation pedagogy. In full spread, “competence” should refer to a set of interdependent and isolable skills, knowledge and attitudes (or indeed “virtues”, in the classical sense). Too often, however, it is being used to name each and every level of all those things, both with and without a developmental aspect (for which “expertise” is proving to be a superior concept anyway). For further discontent with the term, see Pym 2003, 2011a. Yamada (2012) calculates that this point should be reached for English-Japanese translation within two to three years.

³ In most experiments, the productivity gain is a direct result of the database used and the type of text to be translated, thus making general comparisons an almost banal affair. When Plitt and Masselot (2010: 10) report that “MT allowed translators to improve their throughput on average by 74%”, this is because their MT system had been fed the company’s previous translations, and the text translated was normal for that same company (cf. productivity gains with the same research set-up reported at Autodesk 2011). Christensen and Schjoldager (2010: 1) state that “[m]ost practitioners seem to take for granted that TM technology speeds up production time and improve translation quality, but there are no studies that actually document this.” That no longer seems to be true. What is remarkable in all the research, however, is the high degree of inter-subject variation, which might be a feature of the learning curves and degrees of resistance associated with any new technology.

which an MT feed is available (as Google did with Google Translate in December 2011, making its Application Program Interface a pay-service, and most companies should do, by developing their own in-house MT systems and databases). A more general solution could be to provide short-term training in how to use MT, which should be of use to everyone. Either way, the circles should all eventually be virtuous.

Even superficial pursuit of this logic should reach the point that most irritates my students: postediting, the correction of erroneous electronic translations, is something that “almost anyone” can do, it seems. When you do it, you often have no constant need to look at the foreign language; for some low-quality purposes, you may have no need to know any foreign language at all, if and when you know the subject matter very well. All you have to do is say what the translation seems to be trying to say. So you are no longer translating, and you are no longer a translator. Your activity has become something else.

But what, exactly, does it become? Is this really the end of the line for translators?

Models of translation competence

Most of the currently dominant models of “translation competence” are multi-componential. That is, they bring together various areas in which a good translator is supposed to have skills and knowledge (“know how” and “know that”), as well as certain personal qualities, which remain poorly categorized. An important example is the model developed for the European Masters in Translation (EMT) (Figure 1), where it is argued that the “translation service provider” (since this mostly concerns market-oriented technical translation) needs competence in business (“service provision”), languages, subject matter (“thematic”), text linguistics and sociolinguistics (“intercultural”), documentation (“information mining”), and technologies (“technological”).



Figure 1. The EMT model of translation competence (EMT Expert Group 2009: 7)

There is nothing particularly wrong with such models. In fact, they can be neither right nor wrong, since they are simply lists of training objectives, with no particular criteria for success or failure. How could we really say that a particular component is unneeded, or that one is missing? How could we actually test to see whether each component is really distinct from all the others? How could we prove that one of these components is not actually two or three stuck together with watery glue? Could we really object that this particular model has left out something as basic and important as “translating skills,” understood as the set of skills that actually enable a person to produce a translation (i.e. what some other models term “transfer skills”)? There is no empirical basis for these particular components, at least beyond teaching experience and consensus. At best, the model represents coherent thought about a particular historical avatar of this thing called translation.⁴ The EMT configuration is nevertheless important precisely because it is the result of significant consensus, agreed to by a set of European experts and now providing the ideological backbone for some 54 university-level training programs in Europe, for better or worse.

So what does the EMT model say about machine translation? MT is indeed there, listed under “technology,” and here is what they say: “Knowing the possibilities and limits of MT” (2009: 7). It is thus a knowledge (“know that”), not a skill (“know how”), apparently – you should know that the thing is there, but don’t think about doing anything with it.

Admittedly, that was in 2009, an age ago, and no one in the EMT panel of experts was particularly committed to technology (Gouadec, perhaps the closest, remains famous for pronouncing, in a training seminar, that “all translation memories are rotten”). As I predicted some years ago (finding inspiration in Wilss), the multi-componential models are forever condemned to lag behind both technology and the market (Pym 2003).

What happens to this model if we now take TM/MT seriously? What happens if we have our students constantly use tools that integrate statistical MT feeds? Several things might upset multi-componential competence:

- For a start, “information mining” is no longer a visibly separate set of skills: much of the information is there, in the TM, the MT, the

⁴ In the same vein, it is intriguing to consider previous models as expressing the technologies and communication systems of their day. For example, when Étienne Dolet pronounced, in 1547, that the good translator needs extensive knowledge of both languages involved (“La seconde chose, qui est requise en traduction, c’est, que le traducteur ait parfaite congnoissance de la langue de l’auteur, qu’il traduit: & soit pareillement excellent en la langue, en laquelle il se met a traduire”), he was saying something that had not been obvious for most medieval theories of translation, where teams of source-language and target-language experts would tend to work together around the one manuscript version. Similarly, the “three requirements” famously pronounced by Yan Fu (1901/2004) – faithfulness (*xin*), comprehensibility (*da*) and elegance (*ya*) – would appear in his practice to be heavily weighted in favour of target-side considerations of what kind of language to write in, and what kind of examples and terms should convey the general ideas of the foreign text, as was fitting for an age of limited possibilities for foreign-language expertise.

established glossary, or the online dictionary feed. Of course, you may have to go off into parallel texts and the like to consult the fine points. But there, the fundamental problems are really little different from those of using MT/TM feeds: you have to know what to trust. And that issue of trust would perhaps be material for some kind of macro-skill, rather than separate technological components.

- The “languages” component must surely suffer significant asymmetry when TM/MT is providing everything in the target language. It no doubt helps to consult the foreign language in cases of doubt, but it is now by no means necessary to do this as a constant and obligatory activity (we need some research on this). Someone with strong target-language skills, strong area knowledge, and weak source-language skills can still do a useful piece of postediting, and they can indeed use TM/MT to learn about languages.⁵
- Area knowledge (“thematic competence”) should be affected by this same logic. Since TM/MT reduces the need for language skills, or can make the need highly asymmetrical, a lot of basic postediting can be done by area experts who have quite limited foreign-language competence. This means that the language expert, the person we are still calling a translator, can come in and clean up the postediting done by the area expert. That person, the translator, no longer needs to know everything about everything. What they need is great target-language skills and highly developed teamwork skills.
- The one remaining area is “intercultural”, which in the EMT model turns out to be a disguise for text linguistics and sociolinguistics (and might thus easily have been placed under “language”). Yes, indeed, anyone working with TM/MT will need tons of these suprasentential text-producing skills, probably to an extent even greater than is the case in fully human translation.

So much for a traditional model of competence. The basic point is that “technology” is no longer just another add-on component. The active and intelligent use of TM/MT should eventually bring significant changes to the nature and balance of all other components, and thus to the professional profile of the person we are still calling a translator.

Reconfiguring the basic terms of translation

Of course, you might insist that the technical posteditor is no longer a translator – the professional profile might now be one variant of the “technical communicator,” a range of activities that is indeed seeking a professional space. Such a renaming of

⁵ This may be what is happening when Lee and Liao (2011) find that the use of MT reduces the gap between different degrees of foreign-language proficiency in student groups. In part, the MT suggestions replace deficiencies in knowledge of the foreign language, which is another way of saying that the MT acts as a surrogate (and hopefully mistrusted) instructor.

our profession would effectively protect the traditional models of competence, bringing comfort to a generation of translator-trainers, even while it risks reducing the employability of graduates. Yet careful thought is required before we throw away the term “translator” altogether, or restrict it to old technologies: our professionalization may be faulty, but it is still more institutionally sound, at least in Europe and Canada, than is that of the “technical communicator.”

If we want to retain our traditional name but move with the technology, then a good deal of thought has to be given to the cognitive, professional, and social spaces thus created.

For example, translation theory since the European Renaissance has been based on the binary opposition of “source text” versus “target text” (with many different names for the two positions). For as long as translation theory – and research – was based on comparing those two texts, the terms were valid enough. Now, however, we are faced with situations in which the translator is working from a database of some kind (a translation memory, a glossary or at least a set of bitexts), often sent by the client or produced on the basis of the client’s previous projects. In such cases, there is no one text that could fairly be labeled the “source” (an illusion of origin that should have been dispelled by theories of intertextuality anyway); there are often several competing points of departure: the text, the translation memory, the glossary, and the MT feed, all with varying degrees of authority trustworthiness. Sorting through those multiple sources is one of the new things that translators have to do, and which we should be able to help them with. For the moment, though, let us simply recognize that the space of translation no longer has two clear sides: the game is no longer played between source and target texts, but between a foreign-language text, a range of databases, and a translation to be used by someone in the future (a point well made in Yamada 2012).

In recognition of this, I propose that the thing we have long been calling the “source text” should no longer be called a “source.” It is a “start text” (we can still use the initials ST) – an initial point of departure for a work flow, and one among several criteria of quantity for a process that may lead through many other inputs. As for “target text,” there was never any overriding reason for not simply calling it a “translation,” or a “translated text” (TT), if you must, since the actual “target” concept moved, long ago, downstream to the space of text use.

Reconfiguring the social space of translation

An even more substantial reconfiguration of this space involves situations where language specialists (translators or other technical communication experts) work together with area specialists (experts in the particular field of knowledge concerned). This basic form of cooperation was theorized long ago (most coherently in Holz-Mänttari 1984); it now assumes new dimensions thanks to technologies.

Figure 2 shows a possible workflow that integrates professional translators and non-translator experts (shoddily named the “crowd,” although they might also be in-house scientists, Greenpeace activists, or long-time users of Facebook). Follow the diagram from top-left: texts are segmented for use in translation memories (TM); the segments are then fed through a machine translation system (MT); the

output is postedited by non-translators (“crowd translation”); the result is then checked by professionals, reviewed for style, corrected, and put back with all layout features and graphical material that might have been removed at the initial segmentation stage, resulting in the final “localized content.” The important point is that the machine translation output is postedited by non-translators but is then revised by professional translators and edited by professional editors.

There are many possible variations on this model. In most of them, I suggest, translators will need skills that are a little different from those contemplated in the traditional models of competence.

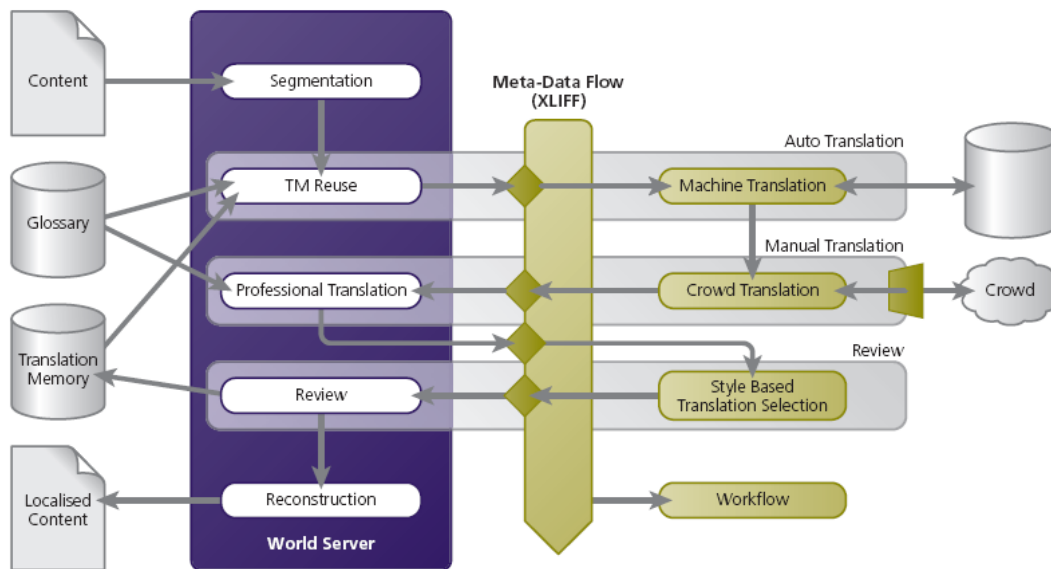


Figure 2. Possible localization workflow integrating volunteer translators (“crowd translation”), from Carson-Berndsen et al. (2010: 60)⁶

New skills for a new model?

I have suggested elsewhere (Pym 2003) that we should not be spending a lot of time modeling a multicomponential competence. It is quite enough to identify the cognitive process of translating as a particular kind of expertise, and to make that the centerpiece of whatever we are trying to do, be it in professional practice or the training of professionals. If we limit ourselves to that frame, the impact of TM/MT is relatively easy to define (cf. Pym 2001b): whereas much of the translator’s skill-set and effort was previously invested in *identifying* possible solutions to translation problems (i.e. the “generative” side of the cognitive process), the vast majority of those skills and efforts are now invested in *selecting* between available solutions,

⁶ My thanks to the journal *Localisation Focus* and the Centre for Next Generation Localisation (CNGL) for permission to reproduce this graph.

and then adapting the selected solution to target-side purposes (i.e. the “selective” side of the cognitive processes). The emphasis has shifted from generation to selection. That is a very simple and quite profound shift, and it has been occurring progressively with the impact of the Internet.

At the same time, however, some of us are still called on to devise training programs and fill those programs with lists of things-to-learn. That is the legitimizing institutional function that models of competence have been called upon to fulfill. The problem, then, is to devise some kind of consensual and empirical way of fleshing out the basic shift, and for justifying the things put in the model.

The traditional method seems to have been abstract expert reflection on “what should be necessary.” You became a professor, so you know about the skills, knowledge and virtues that got you there, and you try to reproduce them. Or your institution is teaching a range of things in its programs, you think you have been successful, so you arrange those things into a model of competence. An alternative method, explored in recent research by Anne Lafeber (forthcoming) with respect to the recruitment of translators for international institutions, is to see what goes *wrong* in current training practices, and to work back from there. Lafeber thus conducted a survey of the specialists who revise translations by new recruits; she asked the specialists what they spend most time correcting, and which of the mistakes by new recruits were of most importance. The result is a detailed weighted list of forty specific skills and types of knowledge not of some ideal abstract translator but of the things that are *not* being done well, or are not being done enough, by current training programs. From that list of shortcomings, one should be able to sort out what has to be done in a particular training program, or what is better left for in-house training within employer institutions. In effect, this constitutes an empirical methodology for measuring “negative” competence (i.e. the things that are missing, rather than what is there), and thus devising new models of what has to be learned.⁷

It should not be difficult to apply something like this negative approach to the specific skills associated with TM/MT. Anyone who has trained students in the use of any TM/MT tool will have a fair idea of what kinds of difficulties arise, as will the students involved. That is an initial kind of practical empiricism – a place from which one can start to list the possible things-to-teach. However, there is also a small but growing body of controlled empirical research on various aspects of TM/MT, including some projects that specifically compare TM/MT translation with fully human translation. Those studies, most of them admittedly based on the evaluation of products rather than cognitive processes, also give a few strong

⁷ Lafeber’s research, I hasten to admit, actually finds that intergovernmental institutions currently do *not* require new recruits to have any great expertise in TM/MT (which comes in at number 33 in her list of 40 skills ordered according to the impact of errors). In such employer organizations, the consensus seems to be that specific tools and techniques are best learned in-house, rather than in an academic training program. That, however, represents the state of technological advance and specific language requirements in just one sector of the translation market. Most localization companies will give a very different weighting of technological skills. For example, Ferreira-Alves’s survey of translation companies in Portugal (2010) finds that expertise in “software and MT” is considered more important than having a degree in translation or being specialized in any particular sector.

pointers about the kinds of problems that have to be solved.⁸ From experience and from research, one might derive the things to watch out for, bearing in mind that those things then have to be tested in some way, to see if they are actually missing when graduates leave to enter the workplace targeted by any particular training program.

Here, then, is a suggested initial list of the skills that might be missing or faulty; it is thus a proposal for things that might have to be learned somewhere along the line:

1. Learn to learn

This is a very basic message that comes from general experience, current educational philosophies of life-long learning, and the recent history of technology: whatever tool you learn to use this year will be different, or out-of-date, within two years or sooner. So students should not learn just one tool step-by-step. They have to be left to their own devices, as much as possible, so they can experiment and become adept at picking up a new tool very quickly, relying on intuition, peer support, online help groups, online tutorials, instruction manuals, and occasionally a human instructor to hold their hand when they enter panic mode (the resources are to be used probably more or less in that order). Specific aspects of this “learning to learn” might include:

- 1.1. Ability to reduce learning curves (i.e. learn fast) by locating and processing online resources.
- 1.2. Ability to evaluate the suitability of a tool in relation to technical needs and price.
- 1.3. Ability to work with peers on the solution of learning problems.
- 1.4. Ability to evaluate critically the work process with the tool.

The last two points have important implications for what happens in the actual classroom or workspace, as we shall see below.

2. Learn to trust and mistrust data

Many of the experiments that compare TM/MT with fully human translation pick up a series of problems related to the ways translators evaluate the matches proposed to them. This involves not seeing errors in the proposed matches (Bowker 2005, Ribas 2007), working on fuzzy matches when it would be better to translate from

⁸ Here I do not follow Christensen (2011: 140) when she insists on focusing on “mental” studies only, discounting the studies that compare products (translations done under different conditions) and that thus make hypotheses about the kinds of cognitive processing that could have given rise to the products – Christensen explicitly excludes Bowker 2005, Guerberof 2009, Yamada 2011. In a situation where there are so few studies, on very small groups of subjects, we can scarcely afford to ignore any of the data available. And we must recognize, I suggest, that data on products constitute a legitimate source of clues about the translators’ cognitive processes.

scratch (a possible extrapolation from O'Brien 2008, Guerberof 2009, Yamada 2012), or not sufficiently trusting authoritative memories (Yamada 2012). There is also a tendency to rely on what is given in the TM/MT database rather than search external sources (Alves and Campos 2009). We might describe all three cases as situations involving the distribution of trust and mistrust in data, and thus as a special kind of risk management. This general ability derives from experience with interpersonal relations in different cultural situations, more than from any strictly technical expertise (cf. Pym 2012). Teixeira (2011) picks up some of this risk management when he finds, in a pilot experiment, that translators who know the provenance of proposed matches spend less time on them than translators who do not. That is, translators do assess the trustworthiness of proposed matches, and they seem to need to do so. The specific skills would be:

- 2.1. Ability to check details of proposed matches in accordance with knowledge of provenance and/or the corresponding rates of pay ("discounts"). That is, if you are paid to check 100% matches, then you should do so; and if not, then not.
- 2.2. Ability to focus cognitive load on cost-beneficial matches. That is, if a proposed translation solution requires too many changes (probably a 70% match or below⁹), then it should be abandoned quickly; if a proposed match requires just a few changes, then only those changes should be made¹⁰; and if a 100% match is obligatory and you are not paid to check it, then it should not be thought about.¹¹
- 2.3. Ability to check data in accordance with the translation instructions: if you are instructed to follow a TM database exactly, then you should do so (Yamada 2012¹²); if you are required to check references with external sources, then you should do that. And if in doubt, then you should remove the doubt (i.e. transfer risk by seeking clarifications from the client, which is a skill not specific to TM/MT).

⁹ Yamada (2012) calculates this "baseline" as a GTM score of 0.46, which would correspond to the 70% fuzzy-match level below which O'Brien (2008) finds that translators' stress level increases (cf. O'Brien 2007a, 2007b).

¹⁰ On one level, this involves a logic of simple efficiency: Yamada (2012) actually penalizes one translator for making too many unjustified changes, not because the translation is wrong but because the translator was required to follow the TM as closely as possible. This also concerns the possibility of learning from MT. Lee and Liao find that "the more words from the MT text a student uses, using sentence as a unit, the less likely a student would make a mistake in translating that particular sentence" (2011: 128), although this may depend on a particular level of prior language skills.

¹¹ Cf. The general "do's and don'ts" for postediting outlined by Belam (2003).

¹² Yamada (2012) actually adopts the quite radical position of assessing translation quality on the basis of how well translation memories are respected, which would be the view of the client who has previously established the validity of the memory. Interestingly, Martín-Mor (2011: 310) finds that in-house professionals have a greater propensity to produce lexical interferences (i.e. adopting the lexical solutions proposed in the databases) than do novices and other professionals, presumably because they are more given to accepting the matches proposed to them. Aesthetic surrender might thus become a minor capacity to be acquired.

3. *Learn to revise translations as texts*

Some researchers report effects that are due not to the use of databases but to the specific type of segmentation imposed by many tools. Indeed, the databases and the segmentation are two quite separate things, at least insofar as they concern cognitive work. Dragsted (2004) points out that sentence-based segmentation can be very different from the segmentation patterns of fully human translation, and the difference may be the cause of some specific kinds of errors; Lee and Liao (2011) find an over-use of pronouns in English-Chinese translation (i.e. interference in the form of excessive cohesion markers); Vilanova (2004) reports a specific propensity to punctuation errors and deficient text cohesion devices; Martín-Mor (2011) concurs with this and finds that the use of a translation memory tends to increase linguistic interference in the case of novices, but not so much in the case of professionals (although in-house professionals did have a tendency to literalism). At the same time, he reports cases where TM segmentation heightens awareness of certain microtextual problems, improving the performance of translators with respect to those problems. As for the effects of translation memories, Bédard (2000) pointed out the effect of having a text in which different segments are effectively translated by different translators, resulting in a “sentence salad.” This is presumably something that can be addressed by post-draft revision. At the same time, Dragsted (2004) and others (including Pym 2009, Yamada 2012) find that translators using TM/MT tend to revise each segment as they go along, allowing little time for a final revision of the whole text at the end. This may be a case where current professional practice (revise as you go along) could differ from the skills that should ideally be taught (revise at the end, and have someone else do the same as well). The difference perhaps lies in the degree of quality required, and that estimation should in turn become part of what has to be learned here.

All these reports concern problems for which the solution should be, I propose, heightened attention to the revision process, both self-revision and other-revision (sometimes called “review” in its monolingual variant). The specific skills would be:

- 3.1. Ability to detect and correct suprasentential errors, particularly those concerning punctuation and cohesion.
- 3.2. Ability to conduct substantial stylistic revising in a post-draft phase (and hopefully to get paid for it!)
- 3.3. Ability to revise and review in teams, alongside fellow professionals and area experts, in accordance with the level of quality required.

Note that all these items, under all three heads, concern skills (“knowing how”) rather than knowledge (“knowing that”). This might be considered a consequence of the fast rate of change in this field, where all knowledge is provisional anyway – which should in turn question the pedagogical boundary between skills and knowledge (since “knowing how to find knowledge” becomes more important than internalizing the knowledge itself).

One might also note that the general tenor of these skills is rather traditional. There is a kind of “back to basics” message implied in the insistence on punctuation, cohesive devices, revision, and the following of instructions (in 2.1 and 2.3). While foreign-language competence may become less important, rather exacting skills in the target language become all the more important. Indeed, “attentiveness to target-language detail” might be the one over-arching attitudinal component to be added to this list of skills. Issues of cultural difference, rethinking purpose, and effect on target reader are decidedly less important here than they have become in some approaches to translation pedagogy.

Research using the “negative skills” approach could now take something like this initial list (under all three heads) and check it against the failings of recent graduates, as assessed by their revisers or employers in the market segment targeted by a specific program. This may involve deleting some items and adding new ones. It will hopefully produce a weighted list, telling us which skills we should emphasize in each specific training program.

For a pedagogy of TM/MT

In an ideal world, fully completed empirical research should tell us what we need to teach, and then we start teaching. In the real world, we have to teach right now, surrounded by technologies and pieces of knowledge that are all in flux. In this state of relative urgency and hence creativity, there has actually been quite a lot of reflection on the ways MT and postediting can be introduced into teaching practices.¹³ O’Brien (2002), in particular, has proposed quite detailed contents for a specific course in MT and postediting, which would include the history of MT, basic programming, terminology management, and controlled language (cf. Kenny and Way 2001). In compiling the above list, however, I have not assumed the existence of a specific course in MT; I have thought more of the minimal skills required for the effective *use* of TM/MT technology across a whole program; I have left controlled writing for another course (but each institution should be able to decide such things for itself).

The initial list of skills thus suggests some pointers for the way TM/MT could be taught in a transversal mode, not just in a special course on technologies. That is, we are envisaging a general pedagogy, the main traits of which must start from the reasons why a specific course on TM/MT may not be required.

The technologies should be used everywhere

Since we are dealing with skills rather than knowledge, the development of expertise requires repeated practice. For this reason alone, TM/MT should ideally be used in as much as possible of the student’s translation work, not only in a special

¹³ For recent general overviews of research on TM/MT, see Christensen 2011 and Pym 2011a. Harold Somers has an outdated bibliography on the teaching of MT available at: <http://personalpages.manchester.ac.uk/staff/harold.somers/teachMTbibl.html>. Most of the links do not work, but many of the papers can be found in the Machine Translation Archive (<http://www.mt-archive.info>).

course on translation technologies. This is not just because TM/MT can actually provide additional language-learning (cf. Lee and Liao 2011), nor do I base my argument solely on the supposition that any particular type of TM/MT will necessarily configure the students' future employment (cf. Yuste 2001). General usage is also advisable in view of the way the technologies can diffusely affect all other skill sets (cf. my comments above on the EMT competence model). In many cases, of course, any general usage will be hard to achieve, mostly because some instructors either do not know about TM/MT or see it as distracting from their primary task of teaching fully human translation first (which does indeed have some pedagogical virtue – you have to start somewhere). Our markets and tools are not yet at the stage where fully human translation can be abandoned entirely, and TM/MT should obviously not get in the way of classes that require other tools (many specific translation skills can indeed still be taught with pen and paper, blackboard and chalk, speaking and listening). That said, at the appropriate stage of development, students should be encouraged to use their preferred technologies as much as possible and in as many different courses as possible. This means 1) making sure they actually have the technologies on their laptops, 2) teaching in an environment where they are using their own laptops online, and 3) using technologies that are either free or very cheap, of which there are several very good ones (there is no reason why students should be paying the prices demanded by the market leader).

Appropriate teaching spaces

From the above, it follows that no one really needs or should want a “computer lab,” especially of the kind where desks are arranged in such a way that teamwork is difficult and the instructor cannot really see what is happening on students' screens. The exchanges required are more effectively done around a large table, where the teacher can move from student to student, seeing what is happening on each screen (see Figure 3) (cf. Pym 2006).



Figure 3. A class on translation technology (Ignacio García teaching in Tarragona)

Work with peers

The worst thing that can happen with any technology is that a student gets stuck or otherwise feels lost, then starts clicking on everything until they freeze up and sit there in silence, feeling stupid. Get students to work in pairs. Two people talking stand a better chance of finding a solution, and a much better chance of not remaining silent – they are more likely to show they need help from an instructor.

Self-analysis of translation processes

Once relative proficiency has been gained in the use of a tool, students should be able to record their on-screen translation processes (there are several free tools for doing this), then play back their performance at an enhanced speed, and actually see what effects the tool is having on their translation performance. This should also be done in pairs, with each student tracking the other's processes, calculating time-on-task and estimating efficiencies. Students themselves can thus do basic process research, broadly mapping their progress in terms of productivity and quality (see Pym 2009 for some simple models of this). The time lag between research and teaching is thus effectively annulled – they become the one activity, under the general head of “action.”

This kind of self-analysis becomes particularly important in business environments – and there are many – where translators will have to negotiate and renegotiate their pay rates in terms of productivity. Simulation of such negotiations can itself be a valuable pedagogical activity (see Hui 2012). Only if our graduates are themselves able to gauge the value of their work will they then be in a position to defend themselves in the marketplace.

Collaborative work with area experts

The final point to be mentioned here is the possibility of having translation students work alongside area experts who have not been trained as translators, on the assumption that the basic TM/MT technologies should be of use to all. Some inspiration might be sought in a project that had translation students team up with law students (Way 2003), exploring the extent to which the different competences can be of help to each other. This particular kind of teamwork is well suited to technologies designed for non-professional translators (such as Google Translator Toolkit or Lingotek), and can more or less imitate the kind of cooperation envisaged in Figure 2.

In sum, the pedagogy we seek cannot operate through fixed recipes. The above list of ten skills, in three categories, should be taken as no more than a possible starting point for creative experimentation.

References

- Alves, Fabio, and Tânia Liparini Campos. 2009. "Translation Technology in Time: Investigating the Impact of Translation Memory Systems and Time Pressure on Types of Internal and External Support." Susanne Göpferich, Arnt Lykke Jakobsen, Inger M. Mees (eds) *Behind the Mind. Methods, Models and Results in Translation Process Research*. Copenhagen: Samfundslitteratur, 191-218.
- Autodesk. 2011. "Machine Translation at Autodesk".
<http://translate.autodesk.com/index.html>. Accessed January 2012.
- Bédard, Claude. 2000. "Translation memory seeks sentence-oriented translator...". *Traduire* 186. http://www.terminotix.com/eng/info/mem_1.htm. Accessed January 2012.
- Belam, Judith. 2003. "'Buying up to falling down'; a deductive approach to teaching post-editing". Paper presented to the MT Summit IX, New Orleans, November 2003.
<http://www.dlsi.ua.es/t4/proceedings.html>. Accessed January 2012.
- Bowker, Lynne. 2005. "Productivity vs. Quality? A pilot study on the impact of translation memory systems". *Localisation Focus* 4:1. 13-20.
- Ferreira-Alves, Fernando Gonçalves. 2011. *As faces de Jano: Contributos para uma cartografia identitária e socioprofissional dos tradutores da região norte de Portugal*. Doctoral thesis. Braga: Universidade do Minho.
- Carson-Berndsen, Julie, Harold Somers, Carl Vogel, Andy Way. 2010. "Integrated language technology as a part of next generation localisation," *Localisation Focus* 8(1): 53-66.
<http://www.localisation.ie/resources/locfocus/vol8issue1.htm>. Accessed January 2012.
- Christensen, Tina Paulsen. 2011. "Studies on the Mental Processes in Translation Memory assisted Translation – the State of the Art". *trans-kom* 4(2): 137-160.
<http://www.trans-kom.eu/>.
- Dolet, Étienne. 1547. *La manière de bien traduire d'une langue en aultre*.
<http://www.gutenberg.org/ebooks/19483>. Accessed January 2012.

- Dragsted, Barbara. 2004. *Segmentation in Translation and Translation Memory Systems: An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process*. Doctoral dissertation, Copenhagen Business School: Samfundslitteratur.
- EMT Expert Group. 2009. "Competences for professional translators, experts in multilingual and multimedia communication".
http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf.
- García, Ignacio. 2010. "Is Machine Translation Ready Yet?" *Target* 22 (1): 7-21.
- Guerberof, Ana. 2009. "Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation." *Localisation Focus* 7(1): 11-21
- Holz-Mänttari, Justa. 1984. *Translatorisches Handeln. Theorie und Methode*. Helsinki: Academiae Scientiarum Fennicae.
- Hui, Maggie Ting Ting. 2012. *Risk Management by Trainee Translators. A Study of Translation Procedures and Justifications in Peer-Group Interaction*. PhD thesis. Tarragona: Universitat Rovira i Virgili.
- Kenny, Dorothy, and Andy Way. 2001. "Teaching Machine Translation & Translation Technology: A Contrastive Study". Paper presented to the Machine Translation Summit VII, Teaching MT Workshop, Santiago de Compostela, Spain, 17-22 September 2001. <http://www.dlsi.ua.es/tmt/proceedings.html>. Accessed January 2012.
- Lafeber, Anne. Forthcoming. "Translation: The Skill Set Required. Preliminary Findings of a Survey of Translators and Revisers Working at Inter-governmental Organizations." *Meta*.
- Lee, Jason, and Posen Liao. 2011. "A Comparative Study of Human Translation and Machine Translation with Post-editing". *Compilation and Translation Review* 4(2): 105-149. <http://ej.nict.gov.tw/CTR/v04.2/ctr040215.pdf>.
- Martín-Mor, Adrià. 2011. *La interferència lingüística en entorns de Traducció Assistida per Ordenador*. PhD thesis. Bellaterra: Universitat Autònoma de Barcelona.
- O'Brien, Sharon. 2002. "Teaching Post-editing, a proposal for course content". Paper presented to the *6th International Workshop of the European Association for Machine Translation*, 14-15 November 2002, Manchester, UK. <http://www.mt-archive.info/EAMT-2002-OBrien.pdf>. Accessed January 2012
- O'Brien, Sharon. 2005. "Methodologies for Measuring Correlations between Post-Editing Effort and Machine Translatability". *Machine Translation*. Dordrecht: Springer. 37-58.
- O'Brien, Sharon. 2007a. "Pauses as Indicators of Cognitive Effort in Post-Editing Machine Translation Output." *Across Languages and Cultures*, 7 (1):1-21.
- O'Brien, Sharon. 2007b. "An Empirical Investigation of Temporal and Technical Post-Editing Effort." *Translation And Interpreting Studies*, II, 1
- O'Brien, Sharon. 2008. "Processing fuzzy matches in Translation Memory tools: an eye-tracking analysis". Susanne Göpferich, Arnt Lykke Jakobsen, Inger M. Mees (eds) *Looking at eyes. Eye-tracking studies of reading and translation processing*. Copenhagen: Samfundslitteratur, 79-102.
- Plitt, Mirko, and François Masselot. 2010. "A Productivity Test of Statistical Machine Translation. Post-Editing in a Typical Localisation Context". *The Prague Bulletin of Mathematical Linguistics* 93 (January 2010): 7-16. <http://ufal.mff.cuni.cz/pbml/93/art-plitt-masselot.pdf>. Accessed January 2012.
- Pym, Anthony. 2003. "Redefining translation competence in an electronic age. In defence of a minimalist approach". *Meta* 48(4): 481-497.
- Pym, Anthony. 2006. "Asymmetries in the teaching of translation technology". Anthony Pym, Alexander Perekrestenko, Bram Starink (eds) *Translation technology and its*

- teaching (with much mention of localization)*. Tarragona: Intercultural Studies Group. http://isg.urv.es/publicity/isg/publications/technology_2006/index.htm. Accessed January 2012.
- Pym, Anthony. 2009. "Using process studies in translator training. Self-discovery through lousy experiments", Susan Göpferich, Fernando Alves and Inge Mees (eds) *Methodology, Technology and Innovation in Translation Process Research*. Copenhagen: Samfundslitteratur. 135-156.
- Pym, Anthony. 2011a. "Translation research terms: a tentative glossary for moments of perplexity and dispute". Anthony Pym (ed.) *Translation Research Projects 3*. Tarragona: Intercultural Studies Group. 75-110.
- Pym, Anthony. 2011b. "What technology does to translating". *Translation & Interpreting* 3(1): 1-9. <http://www.trans-int.org/index.php/transint>
- Pym, Anthony. 2012. "Democratizing translation technologies. The role of humanistic research". Valeria Cannavina and Anna Fellet (eds) *Atti del convegno Language and Translation Automation Conference Roma, 5 - 6 aprile 2011*. Rome: The Big Wave. 14-30. http://usuaris.tinet.cat/apym/online/research_methods/2011_rome_formatted.pdf
- Ribas, Carlota. 2007. *Translation Memories as vehicles for error propagation. A pilot study*. Minor Dissertation. Tarragona: Intercultural Studies Group, Universitat Rovira i Virgili.
- Teixeira, Carlos. 2011. "Knowledge of provenance and its effects on translation performance in an integrated TM/MT environment." In Bernadette Sharp et al. (eds) *Proceedings of the 8th International NLPCS Workshop - Special theme: Human-Machine Interaction in Translation*. Copenhagen: Samfundslitteratur. 107-118. Available at: http://dl.dropbox.com/u/7757461/TPR/CSL_41complete.pdf
- Vilanova, Sílvia. 2007. *The impact of translation memories on the target text: interferences and shifts*. Minor Dissertation. Tarragona: Intercultural Studies Group, Universitat Rovira i Virgili.
- Way, Catherine. 2003. "Making theory reality. An example of interdisciplinary cooperation". In Georges Androulakis, ed. *Translating in the 21st Century. Trends and Prospects. Proceedings*. Thessaloniki: Aristotle University. 584-592.
- Yamada, Masaru. 2011. "The effect of translation memory databases on productivity". In Anthony Pym, ed. *Translation research projects 3*. Tarragona: Intercultural Studies Group. 63-73.
- Yamada, Masaru. 2012. *Revising text: An empirical investigation of revision and the effects of integrating a TM and MT system into the translation process*. PhD dissertation, Rikkyo University, Tokyo.
- Yan Fu. 1901/2004. Preface to *Tiyanlan* (Evolution and Ethics and other essays). Trans. C. Y. Hsu. In Tak-hung Leo Chan, ed. *Twentieth-century Chinese translation theory: modes, issues and debates*. Amsterdam and Philadelphia: John Benjamins. 69-71.
- Yuste Rodrigo, Elia. 2001. "Making MT commonplace in translation training curricula - too many misconceptions, so much potential". Paper presented to the Machine Translation Summit VII, Teaching MT Workshop, Santiago de Compostela, Spain, 17-22 September 2001. <http://dx.doi.org/10.5167/uzh-19088> Accessed January 2012.