Knowledge management issues in the workflow of translation memory systems

Timmy Oumai Wang (Imperial College London) and Mark Shuttleworth (UCL)

1. Background

Translation memory systems (TMSs) are generally believed to be the most important type of computer-assisted translation (CAT) tool. Today, TMSs feature many different functions, ranging from providing term bases to facilitating project management, and are able to handle both commercial and open-source file types. The development of TMSs requires not only the use of new technology, but also a framework that explains how TMSs work and why they are successful in the first place. Using such a framework, we may be able to see into the future of TMSs. This paper reports on a new perspective for conceptualising TMS workflow within the framework of knowledge management (KM). Despite the variety of functions available within a modern TMS, these tools can be seen as a platform within which translators can process various types of knowledge. Some concepts from knowledge management are used to construct this framework.

Knowledge management is a generic concept that refers to the process of creating, sharing and applying knowledge. A knowledge management system (KMS) is an information system that supports or enables all these processes. We argue that a TMS should be seen as a type of knowledge management system. This perspective explains several phenomena relating to the use of knowledge in translation and its relation to CAT technology. The use of a knowledge management framework for translation memory systems contributes to our understanding of how to harness vast translation resources and how to deploy new technologies for the development of the TMSs of the future. This paper is based on a major research project that explores the possibility of merging KMS technology with TMS.

1.1. The state of the art in commercial TMS technology

TMSs have been commercially available for more than twenty years now. During that time a typical TMS has progressed dramatically in terms of its specification and the number of features that it offers the user. In its core, however, the technology has remained largely unchanged: typically, the tool offers a window in which to edit the text being translated, one or more windows offering details of hits from the translation memory (TM) and, usually, a further window with hits from the terminology resources. What is of interest is the different sources of information that a TMS has at its disposal. Traditionally, these have been purely linguistic assets: one or more private and/or shared TMs for sentence-level suggestions, and one or more private and/or shared terminology databases for hits at the word or phrase level. Besides this, most tools allow the user to search the TM manually for suggestions on sentence fragments of any size. Certain individual tools offer further possibilities: Déjà Vu, for example, allows the user to create and populate a new, project-specific resource known as the Lexicon on the fly.

In addition, a more recent trend has been to allow the user to consult on-line machine translation services to fill the translation with draft-quality hits. A typical scenario here would be for the TMS to consult the TMs first and then turn to the on-line MT system to supply content for every segment for which nothing was found in the TMs. All MT-sourced hits would require careful review and/or post-editing before being added to the TM.

What is common to all these resources is that they are all purely linguistic. This means, for example, that in the case of a TM fuzzy match two sentences such as 'I live in a house' and 'I live in a skyscraper', which a human would intuitively recognise as very similar, would only be likely to register as a 60% or 70% match because the edit distance between them would be calculated to be relatively high. Even in tools such as Déjà Vu and memoQ, which offer a feature respectively known as 'Assemble from portions' and 'Assemble from fragments', all that effectively occurs is the matching of strings on the basis of character-by-character similarity. Only tools such as Similis, which are programmed with some of the grammatical parameters of a limited number of languages, would be capable of the kind of intelligent parsing that would enable them to understand the intrinsic similarity of the two sentences cited above.

This section has focused on commercially available tools; approaches that are currently being developed as not included here although two such initiatives are listed in Section 3.2 below. Importantly, however, in the context of this paper, few if any tools known to the authors have yet been developed that try to draw on any kind of real-world rather than linguistic knowledge.

1.2. Knowledge from the Perspective of Translation Studies

Translators obviously need different types of knowledge to translate texts. However, it is difficult to identify precisely what knowledge translators need and how this knowledge should be employed during translation, or, most importantly, in what way a TMS could help. In this section, an attempt is made to examine knowledge in the context of translation studies.

Translation and knowledge can be related along the lines of many different parameters. For example, translation is often regarded as intercultural knowledge transfer (Schubert, 2005:p.125; Bedeker & Feianauer, 2006); the communication of extra-linguistic knowledge is also an important purpose of translation in language for special purposes (Bajaj, 2003:pp.81-85). The relationship between knowledge and translation is often studied from the perspectives of translator training or descriptive translation studies. Wilss (1994:p.133) argues that translation is a knowledge-based activity designed to solve translation problems, and that translators need two types of knowledge: declarative knowledge (knowing what) and processual knowledge (knowing how). Translators' works are formed by using processual knowledge as a set of skills to process the semantic information contained in the material being translated. Wilss' notion can be backed up by Kim's (2006) research. Kim (2006:p.287) conducted thinking-aloud protocol research on three groups of Korean speakers: translation students, professional translators and English-language learners. These subjects participated in a translation test and were required to give an oral description of their progress of translating the text (Kim, 2006:p.287). Kim found that translation students who had better awareness of the subject matter outperformed professional translators who mainly relied on dictionaries in terms of presenting meanings of source texts (2006:pp.291-293). Translation students preserved rhetorical styles at a nearly professional level.

However, the understanding of knowledge in translation is not directly applied in the study of TMSs. The definition of knowledge is vastly different according to different possible contexts within translation studies. Many translation scholars do not consider the functionality of TMSs when they study knowledge in the context of translation studies. The term ‘knowledge’ has sometimes been used interchangeably with other concepts such as ‘intelligence’, ‘valuable information’ and ‘problem-solving skills’. Translation scholars often address only a certain perspective of knowledge such as the usefulness of knowledge of a particular subject. Consequently, the lack of a clear definition of knowledge may cause many conceptual obstacles. In order to cover all aspects of knowledge that we discussed, it is necessary to involve the understanding of knowledge from other disciplines. The next section provides an overview of knowledge management and knowledge management systems.

2.Translation Memory Systems from a Knowledge Management Perspective

Although KM practices are most often found in professional service contexts, e.g. typically consulting and accounting firms, translators themselves need to manage different types of knowledge in order to optimise the efficiency of their work. However, it is possible to use a knowledge management framework to conceptualise translation memory systems.

2.1 Brief introduction of knowledge management and knowledge management systems

Knowledge management is a generic concept that refers to the process of creating, sharing and applying knowledge (Stevens et al., 2010:pp.131-132). A generic definition of knowledge management is given by Dalkir (2005:p.3) as follows:

> The deliberate and systematic coordination of an organisation’s people, technology, processes, and organisational structure in order to add value through reuse and innovation. This value is achieved through the promotion of creating, sharing, and applying knowledge as well as through the feeding of valuable lessons learned and best practices into corporate memory in order to foster continued organisational learning.

A KMS is an information system that supports or enables activities of managing knowledge (Hall, 2009; Alavi & Leidner, 2001). Dalkir (2005) defines KMSs as follows:

> Centralized databases in which employees enter information about their jobs and from which other employees can seek answers. This system often relies on groupware technologies, which facilitate the exchange of organizational information, but the emphasis is on identifying knowledge sources, knowledge analysis, and managing the flow of knowledge within an organization—all the while providing access to knowledge stores (p.352)

KMSs should serve the general objectives of knowledge management, namely ‘knowledge reuse to promote efficiency and innovation to introduce more effective ways of doing things’ (Dalkir 2005:p.166). Different technologies are also employed for KM purposes, such as data mining and content management systems (Dalkir, 2005:p.217).

2.2 TMS as a type of KMS

The basic notion of CAT is generally regarded as 'the process whereby human translators use computerised tools to help them with translation-related tasks' (Bowker, 2002:144). Despite different types of new CAT software being released every year, most CAT tools are designed for two purposes:

- Improving translation quality
- Improving the efficiency of translation

Some translation tools or resources, such as machine translation software, may provide rough translations that are used as references by translators (Shei, 2005). And they can be used by translators to produce better translations. On the other hand, TMSs are primarily designed to improve translation efficiency. The core function of a TMS is an information retrieval platform that searches a database of previously translated text fragments (i.e. translation units or 'TUs') to retrieve translation units similar to the one currently being translated (Trujillo, 1999:pp.60-61). A TMS offers a relief from laborious works by providing translation suggestions based on previously stored translation when translating repetitive content. In the workflow of a TMS, different types of knowledge are involved and processed. Therefore, a TMS can be seen as a type of KMS that aims to serve a translation purpose.

The knowledge in knowledge management literature is extremely complex, but is generally understood in a pragmatic way, rather than being theoretical or epistemological. Most KM researchers have reached a consensus that knowledge is a valuable, intangible object and a manageable factor that brings benefits such as the improved process of decision making (Dalkir, 2005), improved skills for work (Singh et al., 2006) and innovation (Davenport & Prusak, 2005).

For the purposes of conceptualising TMSs, the knowledge involved in the workflow of such systems should be understandable to both human beings and computers. 'Understandable' knowledge means machine-readable information for computers; it is also the information used to assist the translation process. Therefore, three categories of knowledge are involved in the workflow of a TMS:

1) The knowledge that is manipulated directly by the TMS;

2) The knowledge that is used within the TMS to enhance its performance;

3) The knowledge that is used by translators to employ translation suggestions.

Different categories of knowledge have different functions: the first category of knowledge is the useful information that a TMS collects, stores and presents; the second category of knowledge is information that can enhance the performance of the information retrieval component in a TMS (e.g., linguistic data, ontologies, etc.); the third category of knowledge refers to a translator's competences such as a set of skills for solving linguistic, cultural, terminological and text-related problems. These three categories of knowledge should be interrelated in the use of TMSs.

A TMS is a type of KMS that helps translators to process and manipulate these categories of knowledge. The interactions of these three categories of knowledge can be seen as a knowledge process that is defined as a practical model that specifies activities implemented in the practices of KM (Anand and Singh 2011:pp.934-935; Dalkir, 2005:pp.25-26). In this study, we employ Nonaka and Takeuchi's (1995) Knowledge Spiral Model to analyse the interactions of these three categories of knowledge in the workflow of a TMS. The next section explains the Knowledge Spiral Model and the work of a TMS from a KMS perspective.

2.3 The workflow of TMSs as a type of KMS

The Knowledge Spiral Model (Nonaka & Takeuchi, 1995) is a simple but robust KM process that recognises that knowledge can be categorised into two types: explicit knowledge and tacit knowledge. Explicit knowledge is composed of 'formal and systematic', 'quantifiable data, codified procedures, [and] universal principles'(Nonaka & Takeuchi, 1991:pp.91-93). This type of knowledge was defined by Dalkir (2005: p.334) as being 'rendered visible (usually through transcription into a document); typically, captured and codified knowledge'. Tacit knowledge is fundamentally different from explicit knowledge, which corresponds to our common understanding of knowledge. Tacit knowledge is embedded in individual experiences in forms such as insights, intuitions and hunches; it is knowledge that is hard to express and is internalised by people and is usually concerned with the process of performing particular skills or demonstrating expertise (Nonaka & Takeuchi, 1991:p.95). In addition, Nonaka and Takeuchi recognise that 'tacit knowledge is highly personal', and that it is 'hard to formalize and, therefore, difficult to communicate to others' (Nonaka & Takeuchi, 1991:p.96). This model can be described in a four-step knowledge management process (Nonaka & Takeuchi, 1995: p.57):

1) Socialisation: one shares the tacit knowledge with others;

2) Externalisation: tacit knowledge is articulated as explicit knowledge by the individual;

3) Combination: the discrete pieces of explicit knowledge are organised into new systematic and codified knowledge;

4) Internalisation: the formalised explicit knowledge becomes an individual's own new knowledge, and can also be used as a source for creating new knowledge.

(See Figure 1.1 below.) Some important features of the Knowledge Spiral Model make it a framework that can easily describe all activities in the workflow of TMSs. Its simplicity makes it more flexible to use with other theoretical frameworks and allows it to have technical extensions. Its robustness means that one does not need to follow all the steps presented in the model and that it can be modified easily in response to new situations. The four steps of the Knowledge Spiral Model are also broad enough to cover most activities in the KM process. Therefore, the Knowledge Spiral Model is a suitable model for managing the knowledge involved in the workflow of TMSs.

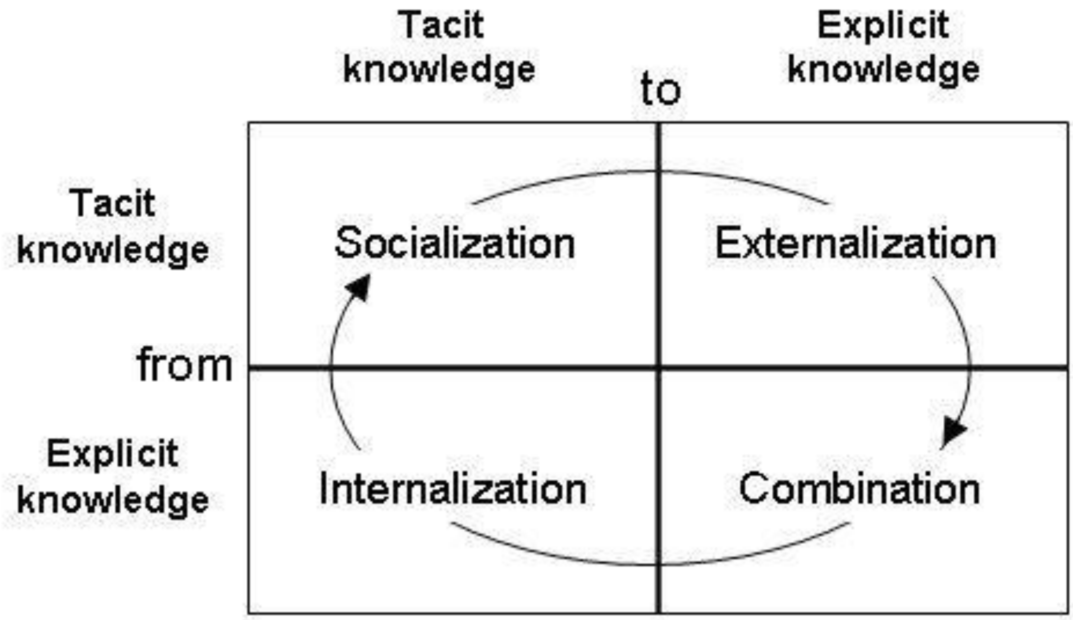

Figure 1.1: Nonaka's Knowledge Spiral Model (Nonaka and Takeuchi, 1995:p.62).

The knowledge involved in the workflow of TMS is different from the knowledge required for the translation process. The explicit knowledge manipulated by a TMS is fairly simple, and consists of translation suggestions in the form of target texts aligned with source texts. Technically, the explicit knowledge that a TMS processes is stored mainly in various machine-readable formats such as Translation Memory Exchange (TMX), which is an XML-based format (GALA, 2011). These translation memory files belong to the first category of knowledge within the TMS workflow.

Although a TMS does not directly manipulate tacit knowledge, this type of knowledge is also involved in TMS workflow. This tacit knowledge is the knowledge that assists translators to assimilate, to analyse and to adopt translation suggestions in different contexts. (For example, TMS users should have the ability to rephrase translation suggestions for new contexts.) Tacit knowledge is always the knowledge that cannot be directly shared or used by other translators. This tacit knowledge is the third category of knowledge in the TMS workflow. The use of tacit knowledge should depend on explicit knowledge and the knowledge capture process, which converts tacit knowledge into explicit knowledge.

One difference between the tacit knowledge and explicit knowledge involved in the TMS workflow is that users benefit directly from explicit knowledge, i.e. bilingual aligned translation suggestions retrieved by the TMS, while tacit knowledge refers to how translators use translation suggestions.

The TMS workflow can be analysed using the Knowledge Spiral Model, which focuses on the conversion between tacit knowledge and explicit knowledge (Nonaka & Takeuchi,1995). The Knowledge Spiral Model should be modified when it is used to analyse the use of knowledge in the TMS workflow. Our focus on TMS workflow is from the perspective of how individual translators use TMSs. Therefore, TMS workflow can be explained in KM terms as follows:

Knowledge Capture

When translators use a TMS, human-produced translations should be seen as tacit knowledge stored in TUs captured by the TMS. A TMS does not process tacit knowledge directly, but manipulates TUs that contain the tacit knowledge about translation. Each TU is formed as a bilingual aligned text fragment that contains tacit knowledge about translation. The tacit knowledge embedded in the newly generated TU is captured by the TMS as it updates translation memory files.

Knowledge Codification

The codification step involves converting the tacit knowledge into explicit knowledge. It is a relatively simple step conducted in most TMS. Once the tacit knowledge is captured, it is codified, which means the newly captured translation unit is stored in the TM. By doing so, the tacit knowledge is saved and the structured explicit knowledge can be used.

Knowledge Application

Knowledge application in TMSs means the codified explicit knowledge is reused to improve the productivity and quality of translation. The TUs are retrieved as translation suggestions by the TMS according to various similarity measure methods.

Knowledge Creation

Ideally, the KM process can be continued as a mutually beneficial relationship: translators keep updating translation memory files and the TMS assists translators more effectively as the scale of translation memory knowledge grows.

The TMS workflow is a KM process during which explicit and tacit knowledge is reciprocally converted at every stage and different categories of knowledge can also be involved. The conversion of different types and categories of knowledge in the TMS workflow is as displayed in Figure 1.2 below.

The KM process as set out in the figure below suggests that as a KMS, a TMS primarily assists the translation process by retrieving codified explicit knowledge. A TMS employs a successful technical approach to capturing translators' tacit knowledge, and it presents the explicit knowledge properly in a bilingual parallel form for translators. The technical implementation is relatively easy, as the use of translation suggestions also depends on the translator's professional proficiency, which makes good use of reference information.

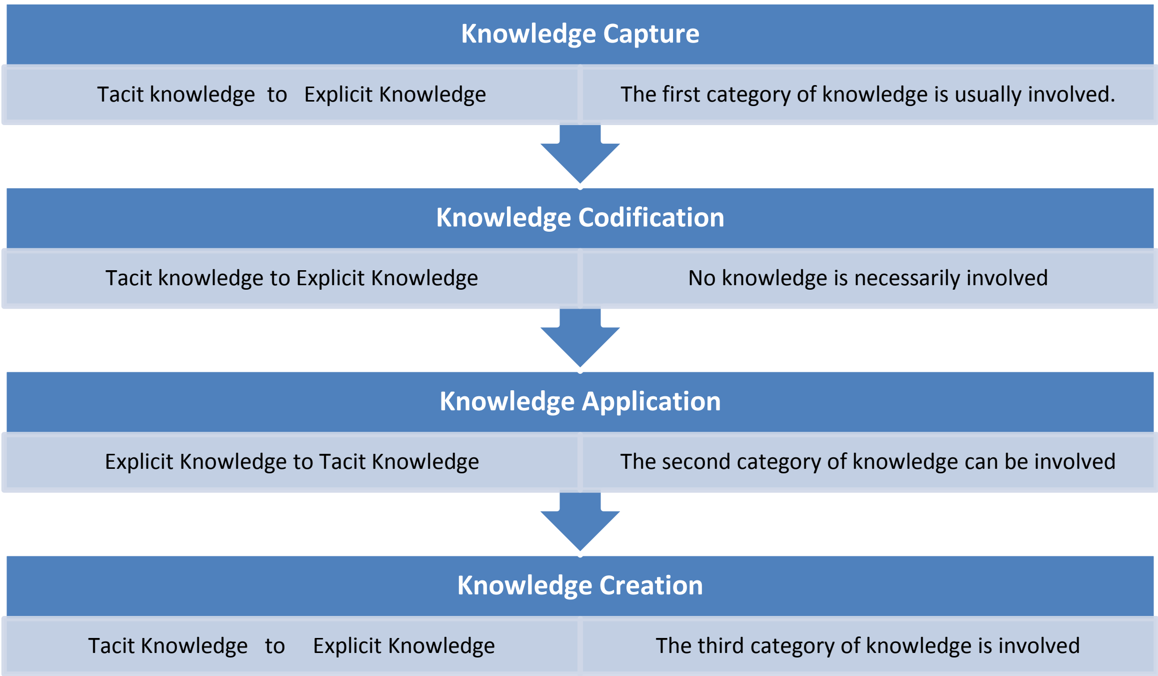

Figure 1.2: The transfer and conversion of knowledge in the workflow of a TMS seen as a type of KMS.

2.4 Current KM bottlenecks in TMSs

TMSs do not perform equally effectively at every step of the KM process. As Figure 1.2 indicates, current TMSs do not actively involve the second category of knowledge that is used within them in order to enhance their performance. As a result, most TMSs may perform less effectively at the knowledge codification step. In KM terms, a TMS can be severely affected by two technical problems: the small size of translation memory files and the low efficiency of TU retrieval (Macken, 2010:p.197). The size of translation memory files can be increased fairly easily by using bilingual alignment tools (Macken, 2010). However, the second problem, which is associated with the TMS workflow, may require great effort in order to improve it.

The low efficiency of using the explicit knowledge captured within TUs by a TMS is mainly caused by the oversimplistic technical approach used in knowledge codification and application. In the knowledge codification and application steps, XML-based formats such as TMX do not store semantic information or other more descriptive features of translation units in the TMS repository. This is because TMSs can only retrieve TUs by using the relatively simple similarity measure methods such as edit distance, which is essentially based on calculating the number of words that the source and target segments have in common (Trujillo, 1999:p.64). These methods are employed because they do not require additional resources such as linguistic knowledge or world knowledge. However, as a result, only key words found in sentences or phrases are used as the basis of retrieval; this is only effective for retrieving highly repetitive text segments contained in long-term projects (Trujillo 1999:pp.64-69). A TMS cannot match TUs that are similar on the semantic level but substantially different in terms of the wording.

3.The Future Development of TMSs

For the most part, TMSs lack the capability to process the semantics contained in translation units, and in particular the capability to retrieve the required text fragments based on semantic similarity between sentences. Seen as a type of KMS, a TMS requires a set of technologies to enable this capability. We need to find a method that can store and represent knowledge from some domains in a computable format and can support KM to compute semantic similarity for the effective retrieval of TUs. Fortunately, current bottlenecks can be solved by some newly available resources and techniques.

3.1 New Resources and functions

Those translation resources that can be used are normally easily accessible for confirming terms and phrases or retrieving background information (Austermühl, 2001:p.85-86). However, Austermühl does not foresee the possibility that a translation resource can also consist of significant amounts of machine-readable information, which can be used to augment particular applications such as TMSs. It should be noted that the new techniques make many new resources available for the development of new TMSs that include new functions to perfect their KM functionality.

Recently, techniques for processing the semantics of natural language have become a very active and dynamic field, especially in improving the efficiency of building and using large-scale resources as well as converging techniques from knowledge engineering (Tokunaga et al., 2008; Gurevych & Zesch, 2013). It is a reasonable expectation that many new techniques from natural language processing (NLP) and machine learning (ML) can be used to augment a KM-based approach to TMS design. Such techniques may enable the use of large-scale machine-readable information within applications to improve the efficiency of using translation memory files. Knowledge bases such as WordNet and ConceptNet have been used to achieve semantic similarity measures in information retrieval tasks. TMSs also perform information retrieval tasks to match newly input translations and previously stored translation units. In fact, many ML techniques that rely on large-scale knowledge bases or corpora can probably be applied to a KM-based TMS. CAT software, especially TMSs, may improve the efficiency of using translation resources if it is able to utilise the semantics of texts. For translation tasks, the advantage of such approaches is that they may save the considerable amount of manual work involved in developing language-dependent systems for each translation task, as manually created linguistic rules are less important than corpora. Similar research approaches have tried to enhance the performance of TMSs with semantic processing techniques, such as finding semantically equivalent sentences through identification of rhetorical predicates (Mitkov & Corpas, 2008) or employing ontology-based resources (Yao, 2010).

3.2 Practical implementation of KM in TMSs

As was discussed in Section 1.1, the interface of commercially-available TMSs has traditionally involved a TM that presents the user with segments of text that have either been copied verbatim from a database or, in the case of some tools, assembled from asset fragments located in a number of different databases. In addition to this, some tools are now also available that implement a certain amount of grammatical analysis of source text segments to enable them to identify some highly similar matches that would otherwise be overlooked.

The emphasis within a KM-based approach would be entirely different. At this stage, it is not envisaged that hits from such a KM-based resource would provide the sole source of information, but would rather overlap with and/or complement the existing types of linguistic resource discussed above.

For the experiment that underlies this article, three corpora, consisting of sets of aircraft accident reports, *Scientific American* articles and articles from the Reuters news agency, and each comprising approximately 200,000 words, were used as the source for potential matches. One of the most important aims of the investigation was to determine if KM-type data could usefully supplement matching based on edit distance and, if so, precisely what kind of assistance it could offer. For each separate corpus, it was found that two distinct types of hit emerged: firstly, fuzzy matches of the kind provided by a TM, and secondly, matches where the similarity emerged on the semantic level. While the former could potentially be pasted into a translation and edited by the translator like TM hits, the help afforded by the latter would be in terms of conceptual clarification of what might be an unfamiliar subject area and also, as an incidental effect, in the form of terminological suggestions. In other words, even at this relatively early stage in the investigation, it seems to be the case that by no means all the hits generated by such an approach would offer wording that was very close to that of the original, although this does not signify that such hits would be of no use to the translator.

An example of the first type would be the following:

Query	The cause of the braking loss could not be positively established.
Result	The cause of the failure could not be assessed.

(NB At this early stage of experimentation all examples are monolingual; if bilingual data were to be used then the target segment, or possibly both the source and target, would be presented to the translator. Each example was obtained using Terrier, an information retrieval plaftform which can be adjusted to measure semantic similarity based on selected knowledge sources.)

This is in fact the closest match retrieved for any of fifteen match results obtained for each corpus in turn, although given the size of the corpora it is perhaps not surprising that no closer hits were identified. Be that as it may, in this example there is clearly some material that could be reused as-is, even though it would probably only correspond to a 40-50% TM match calculated in terms of edit distance.

Even in the case of fuzzier hits terminological assistance can still be provided:

Query	A somatogravic illusion is a non-visual illusion that produces a false sensation of helicopter attitude.
Result	False sensations about the pitch attitude of the aircraft are caused by a misinterpretation of the gravity vertical, known as ‘somatogravic illusion’.

This hit, for example, would in a bilingual context offer the translator the exact equivalent of ‘somatogravic illusion’ and a hint at the translation of ‘helicopter attitude’. In this way, the process of submitting an enquiry to the KM-style database is akin to a corpus search on the TM.

The second type of hit is exemplified by the following:

Query	Dark energy will continue to push galaxies ever faster away until they fade completely from view.
Result	But if the dark energy density decreases and matter becomes dominant again, our cosmic horizon will grow, revealing more of the universe.

Interestingly, the similarity here lies almost entirely on the semantic level as there is very little resemblance between the Query and the Result in terms of either terminology or syntax. Such a hit may conceivably offer a translator a kind of explanation of the information contained in the Query should one be needed. At the same time it would also provide the translation equivalent of 'dark energy'.

It would be the responsibility of the translator to decide what use each suggestion should be put to since at this stage in its development the system would not distinguish between these two different types of hit.

This research is of course still at a very experimental stage, so that any examples provided are simply intended to indicate what potential such an approach might possess. However, it seems clear that the closeness of matches identified may be increased a) through use of a larger corpus and b) by modifying the segmentation rules so that matches on the sub-sentential level could be searched for, possibly in conjunction with some kind of assembling from segment fragments.

4. Conclusion

It is easy to see the knowledge management issues within the development of TMSs, because translation is a knowledge management practice that processes different types and categories of knowledge. This paper reports on a new conceptual framework of TMS from the knowledge management perspective in which different types and categories of knowledge interact in the TMS workflow. The TMS framework proposed reveals several problems in current TMSs that have different levels of performance at each step of the KM process. These can cause a TMS's workflow to be less effective. In the meantime, we also foresee that some techniques from NLP offer the possibility of improving the workflow of TMSs in the future.

References

Alavi, M. &. L., D. (2001) Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly.* 25 (1), 107-136.

Anand, A. & Singh, M. (2011) Understanding Knowledge Management: a Literature Review. *International Journal of Engineering Science and Technology.* 3 (2), 926-939.

Austermühl , F. (2001) *Electronic Tools for Translators (Translation Practices Explained).* Manchester, St Jerome Publishing.

Bajaj, B. (2003) *A Terminological Approach to the Investigation of Temporal Ordering Relations in English and German Aircraft Accident Reports.* PhD. Unversity of Surrey.

Bedeker, L. & Feinauer, I. (2006) The Translator as Cultural Mediator. *Southern African Linguistics and Applied Language Studies.* 24 (2), 133-41.

Bowker, L. (2002) *Computer-aided Translation Technology: A Practical Introduction.* Ottawa, University of Ottawa.

Dalkir, K. (2005) *Knowledge Management in Theory and Practice.* USA, Elsevier.

Davenport, T. &. P., L. (2005) Working Knowledge: How Organizations Manage What They Know. *Ubiquity.* [Online], 2nd June 2012. Available from: http://wang.ist.psu.edu /course/05/IST597/papers/Davenport_know.pdf.

Globalization and Localization Association (GALA). *LISA OSCAR Standards.* [Online] Available from: http://www.gala-global.org/lisa-oscar-standards [Accessed 13th April 2013].

Gurevych, I. & Zesch, T. (2013) Collective Intelligence and Language resources: Introduction to the Special Issue on Collaboratively Constructed Language Resources. *Language Resources and Evaluation.* 47 (1), 1-7.

Hall, H. (2010) Principles of Knowledge Management: Theory, Practice, and Cases. *Journal of the American Society for Information Science and Technology.* 61 (2), 430-432.

Kim, R. (2006) Use of Extralinguistic Knowledge in Translation *Meta: Translators' Journal.* 51 (2), 284-303.

Macken, V. (2010) In search of the recurrent units of translation. In: Daelemans, W. &. H.,V. (ed.). *Evaluation of translation technology.* Brussel, Academic & Scientific Publishers (ASP). pp. 195-212.

Mitkov, R. & Corpas, G. (2008) Improving Third Generation Translation Memory Systems Through Identification of Rhetorical Predicates. *Proceedings of LangTech. 2008, Rome.*

Nonaka, I. (1991) The Knowledge-Creating Company. *Harvard Business Review.* 69, 96-104.

Nonaka, I. & Takeuchi, H. (1995) *The Knowledge-Creating Company.* New York, Oxford University Press.

Schubert, K. (2005) Translation Studies: Broaden or Deepen the Perspective? In: Dam, H., Engberg, J. & Gerzymisch-Arbogast, H. (eds.). *Knowledge Systems And Translation.* , Mouton de Gruyter.

Shei, D. (2005) Translation Commentary: A Happy Medium Between Translation Curriculum and EAP. *System.* 33 (2), 309-325.

Singh, M. & Shankar, R. & Narain, R. & Kumar, A. (2006) Survey of knowledge management practices in Indian manufacturing industries. *Journal of Knowledge Management.* 10 (6), 110-128.

Stevens, R. & Millage, J. & Clark,S. (2010) Waves of Knowledge Management: The Flow between Explicit and Tacit Knowledge. *American Journal of Economics and Business Administration.* 2 (1), 129-135.

Tokunaga, T. & Huang, C. & Lee, S. (2008) Asian Language Resources: the State-of-the-Art. *Language Resources and Evaluation.* 42 (2), 109-116.

Trujillo, A. (1999) *Translation Engines: Techniques for Machine Translation.* 1st edition. London, Springer.

Wilss, W. (1994) A Framework for Decision-Making in Translation. *Target.* 6 (2), 131–150.

Yao, Y. (2010) An Ontology-Based Translation Memory Model in Localization Translation. *Information Science and Engineering (ISISE), 2010 International Symposium on. 2010, Shanghai.* pp.340-344.

Dr Mark Shuttleworth, UCL
m.shuttleworth@ucl.ac.uk
http://www.linkedin.com/in/markshut
http://tinyurl.com/uclmarkshut
School of European Languages, Culture & Society
University College London
Gower Street
London WC1E 6BT
United Kingdom

Mark Shuttleworth has had a life-long passion for languages. He has been involved in training translators since 1996, first at the University of Leeds, then at Imperial College London and now since October 2013 at UCL. During this time he has launched and run two highly successful Masters programmes in translation and has become an established PhD supervisor. He is the author of numerous academic papers on translation technology and other topics and has spoken at industry events such as Translating and the Computer and the annual conferences of the Institute of Translation and Interpreting and the Association of Translation Companies. He studied at the Universities of Oxford and Birmingham, and has recently completed a part-time PhD at the University of London. He is a fluent speaker of Russian, German, Polish and French and has some knowledge of a number of other languages.

Timmy Oumai Wang, Imperial College
www3.imperial.ac.uk/people/oumai.wang10
Centre for Co-Curricular Studies
Imperial College London
South Kensington Campus
London SW7 2AZ

Timmy Oumai Wang is a third-year PhD student at Imperial College London. Previously, he earned his MSc in *Scientific, Technical and Medical Translation with Translation Technology* from the Translation Studies Unit after having received a BA in Historical Studies from The University of Bristol. He is very keen on open-source software and collaboratively edited content, and the technical methods for exploiting them for the use of translators. Before conducting his doctoral research, he worked as a research intern at the Institute of Intelligent Multimedia Technologies and Applications, Hisense Electric Co Ltd. In 2013, he became one of the laboratory assistants who help MSc translation students to use computer-assisted translation software. His research interests focus on the impact of natural language processing techniques on computer-assisted translation and the use of online resources in translation.