

Knowledge management issues in the workflow of translation memory systems

Click to edit Master subtitle style

Timmy Oumai Wang (Imperial)
Mark Shuttleworth (UCL)

Introduction

- Inspiration of the project: producing a new kind of translation resource, or improving on old ones.
- Now the aim is to find a way to supplement what TMSs already do by providing them with a new kind of translation resource based on the semantic power of knowledge management.

Commercial TMSs: the state of the art

- The technology is now 20 years old.
- Still essentially the same technology: provides suggestions based on string matching rather than some measure of semantic similarity.
- This means that the two sentences *I live in a house & I live in a skyscraper* may not be recognised as similar.

More sophisticated approaches

- Déjà Vu ‘Assemble from portions’ and memoQ ‘Assemble from fragments’.
- Supplementing TM hits by the use of on-line MT.
- Similis: grammatical pattern matching.
- Research currently being conducted on methods of semantic matching.
- This is the context in which this present work is being carried out.

Knowledge management and knowledge management systems

- Knowledge management: a generic concept that refers to the process of creating, sharing and applying knowledge (Stevens et al. 2010:131-132).
- Knowledge in the KM context.
- A knowledge management system is an information system that supports or enables activities of managing knowledge (Alavi & Leidner, 2001).
- KMSs should serve the general objectives of knowledge management.

TMS as a type of KMS

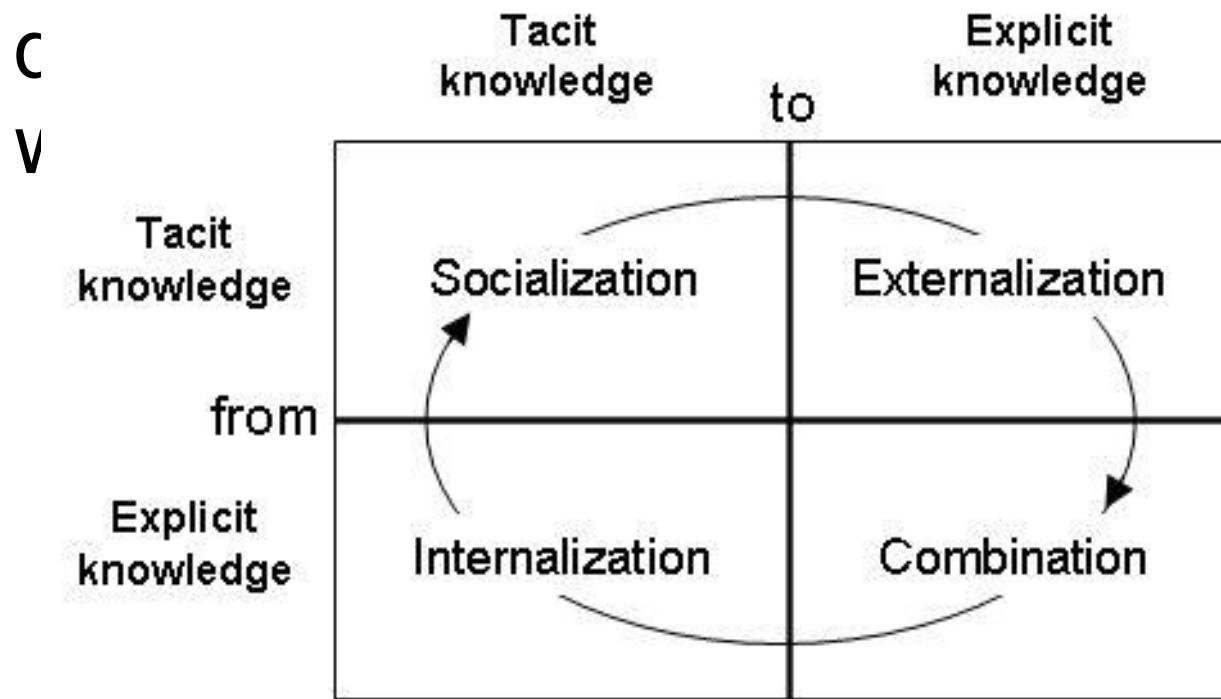
- Most CAT tools are designed for two purposes:
 - i. Improving translation quality
 - ii. Improving translation efficiency
- A TMS can be seen as a type of KMS that aims to serve a translation purpose.

Three Categories of Knowledge involved in the workflow of a TMS

1. The knowledge that is manipulated directly by the TMS (e.g. translation memory files)
2. The knowledge that is used within the TMS to enhance its performance (e.g. linguistic data, ontologies, etc.)
3. The knowledge that is used by translators to employ translation suggestions (e.g. translator's competences)

Workflow of TMSs as a type of KMS

- Using the Knowledge Spiral Model (Nonaka & Takeuchi 1995) to



Workflow of TMSs as a type of KMS

Two types of Knowledge:

1. Explicit knowledge: ‘formal and systematic information, such as quantifiable data, codified procedures, [and] universal principles’.
2. Tacit knowledge: hard to express and is embedded in individual experiences in forms such as insights, intuitions and hunches.

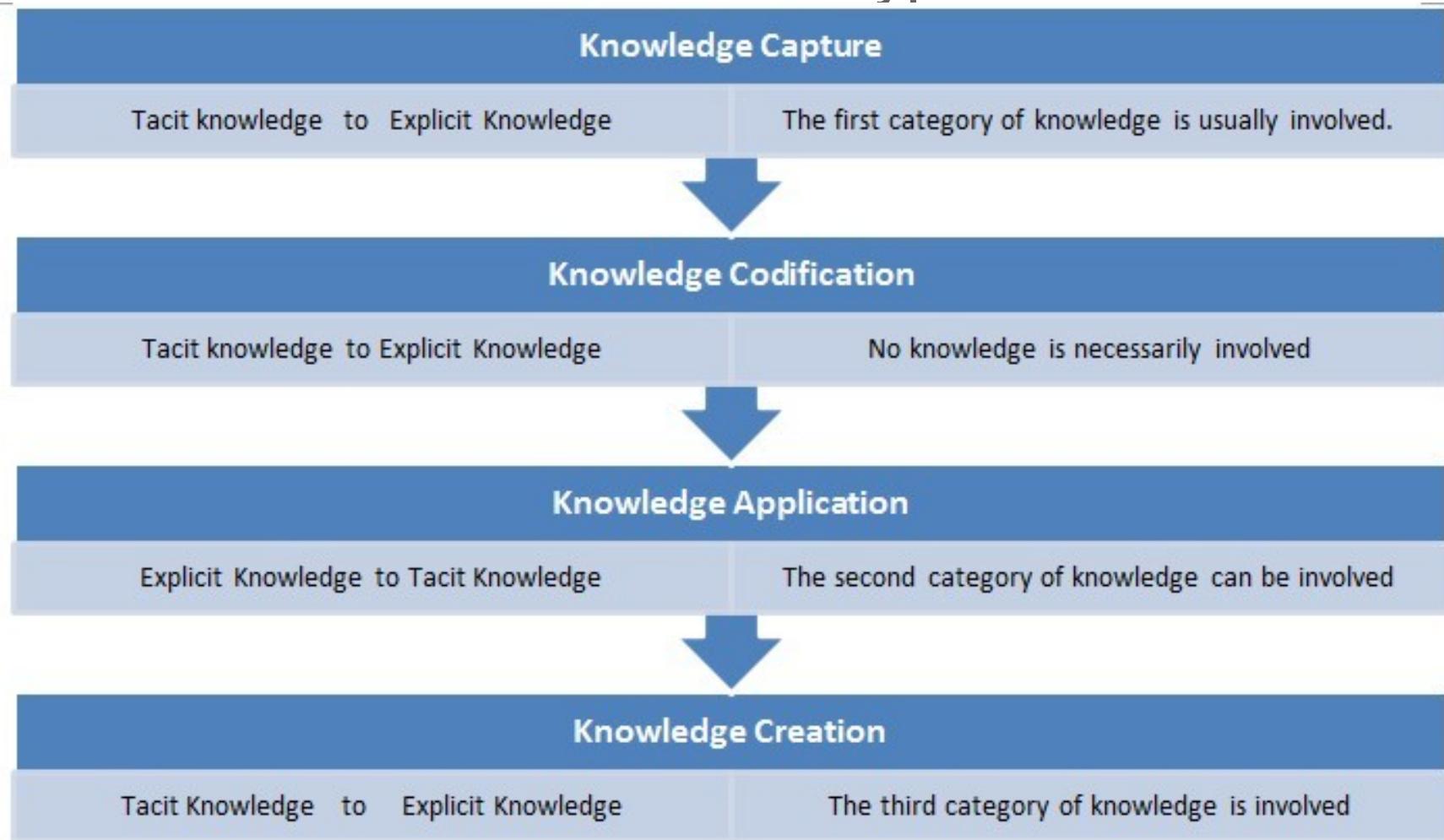
Workflow of TMSSs as a type of KMS

- Knowledge Capture
- Knowledge Codification
- Knowledge Application
- Knowledge Creation

Current KM Bottlenecks in TMS

- Low efficiency of using TM files
- XML-based formats such as TMX or XLIFF do not store semantic information or other more descriptive features of translation units.

The transfer and conversion of knowledge in the workflow of a TMS seen as a type of KMS



A New Development for TMSs

- New resources that can be exploited for a second category of knowledge: WordNet, ConceptNet, Wikipedia.
- Initial stage of experimentation:
 1. Three corpora based on different text types (each 230,000 words)
 2. Terrier as the information retrieval platform

The screenshot shows an IDE interface with several tabs at the top: ComputeESASimilarity, CountWords.java, demo_context, ComputeESASimilarity, ComputeESASimilarity, ConceptMatch..., and a minimized tab. The left pane displays Java code for the ComputeESASimilarity class. The right pane shows a file tree with a 'demo' folder containing 'ComputeESASimilarity.java'. Below the code editor is a toolbar with icons for Problems, Javadoc, Declaration, and Console. The bottom half of the screen is a terminal window titled 'terminated' showing the output of the Java application. The output lists various similarity scores and corresponding text snippets from two documents.

```
62     conceptIndexBean = config.getString( "concept_index_beans" );
63 }
64
65 IConceptIndex index = (IConceptIndex) context.getBean( conceptIndexBean );
66 logger.info( "size of source index: " + index.size() );
67
68 IConceptExtractor esaExtractor = index.getConceptExtractor();
69
70 TextDocument docA = new TextDocument( "text_a" );
71 docA.setText( "content", Language.EN,"Neither pilot saw the other aircraft in time to take effective avoiding ac
72
73 // TextDocument docB = new TextDocument( "text_b" );
74 // docB.setText( "content", Language.EN,"Tomorrow, Japan is over ");//, config.getString( "text_b" ) );
```

```
<terminated> ComputeESASimilarity10Results [Java Application] H:\Program Files\Java\jre7\bin\javaw.exe (3 Sep 2013 17:58:22)
ESA Similarity Score 相似值排序[0.0, 1.2919576798473847E-8, 4.634776545472566E-8, 2.1526830367310173E-7, 2.468570324695118E-7, 3.264552223844128E-7]
Output Corresponding Similarity Score(相似度). ESA Similarity Score Ranked #Key#*#
Similarity Score. NaN. Corresponding Results. ACJ 145.[END]
Similarity Score. 0.8356392810653204. Corresponding Results. Whilst in the Base workshop he saw the fitter consulting the AWM about the torques
Similarity Score. 0.8179988525974382. Corresponding Results. Other witnesses saw a fireball descending rapidly out of the low cloud base and one
Similarity Score. 0.8062253538797748. Corresponding Results. It should namely the cardiac incident, is happening to the handling pilot at the ti
Similarity Score. 0.7908466452872632. Corresponding Results. 12 Another saw 'shiny' ice along the rear half of the wing and saw it subsequently
Similarity Score. 0.7562324274746258. Corresponding Results. Some witnesses saw flames issuing from No 1 engine and a cabin attendant, seated by
Similarity Score. 0.7549396897705759. Corresponding Results. He did not think 1375 that Captain Key saw the writing.[END]
Similarity Score. 0.7522043038420549. Corresponding Results. The evidence was that so far as that Captain Key saw the writing.[END]
Similarity Score. 0.7500145526113141. Corresponding Results. P3 immediately saw what was happening and reversed the 1499 movement.[END]
Similarity Score. 0.73364088771344491. Corresponding Results. He saw the droop gauge moving up and immediately pushed the stick 2449 forward to in
Similarity Score. 0.7210674361992825. Corresponding Results. The Controller had recognised that, because of the scarcity of opportunities, he wa
Similarity Score. 0.7120747503906596. Corresponding Results. anything offensive, there is no doubt that the P2 was very upset, and that his confi
Similarity Score. 0.6982781722230897. Corresponding Results. It was not until late December 1970 or early January 2500 1971 that he saw the confi
Similarity Score. 0.6961956893625504. Corresponding Results. 1 CVR strip and examination To gain access to the recording tape it was necessary
Similarity Score. 0.6823586214410015. Corresponding Results. 2414 So far as the ASB was concerned Mr Gordon Burge saw this memorandum.[END]
Similarity Score. 0.6501875820670137. Corresponding Results. Some of the passengers on the ATP witnessed the formation of ice on the aircraft im
```

First type of translation suggestion

- Semantically Similar Suggestions :

Query The cause of the **braking loss** could not be positively established.

Result The cause of the **failure** could not be assessed.

Second type of translation suggestion

- Conceptual explanation or clarification :

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.

Conclusion

Thank you very much!

Timmy Oumai Wang
oumai.wang10@imperial.ac.uk

Mark Shuttleworth
m.shuttleworth@ucl.ac.uk