

# **Translation Technology (LC412) Assignment**

**Option 1: Translation Memory**

**Module Co-ordinator:  
Minako O'Hagan**

**Pablo Muñoz Sánchez  
Student Number: 54166161  
Erasmus Student**

## Report of my experience

As the purpose of this assignment was to edit the source texts in order to make them more translation memory friendly, i.e. to obtain more matches, it was necessary to keep in mind the different types of matches of a TM. Once I realized that I would obtain, firstly, fuzzy matches and secondly, full matches, I read the first five forecasts trying to find similar expressions with different words.

First of all, I noticed there were always references to temperature, especially to the highest and lowest reached during a day. However, although I expected there would be the same structures and placeables (numbers in this case) with different word order, I came across with different ways of writing the same thing. For instance, *degrees Celsius* was written as *Celsius* or *C*, with or without the word *degrees*. Moreover, to specify the range of the temperatures, *and* was used (14 *and* 17) as well as *dash* (9-12) or *to* (7 to 12). Another important characteristic of the structure was that before the placeables there were many expressions such *as will range*, *will reach* and *of*.

Secondly, I realized that while information about cool and cloudy weather came before or after the lowest temperatures, information about winds came before or after the highest temperatures. I also realized that winds tended to be referred to in relation to their force (e.g. *light* and *moderate*) and the direction (e.g. *north* and *south*).

I concluded, therefore, that I had to use a convention to write sentences referring to temperatures and winds, so I decided to use the following: for temperatures, *Lowest/Highest temperatures of* (placeable1) *to* (placeable2) *Celsius*; for winds, (adjective referring to its force) + (area1 *to* area2) + *winds*, for example *Freshening south to southwest winds*. While it is true that one could use two slightly different sentences in order to avoid constraints in language, I thought it was better to use a prototypical sentence in order to achieve the best results of text reusability, which was my goal.

Keeping in mind all this information, I made the necessary edits to the 1-5 forecast file and analyzed it with TW. Despite I obtained 18 repetitions, I expected more, so I thought it would be a good idea to shrink similar segments. For instance,

*Highest temperatures of 15 to 18 Celsius with light variable west winds* became *Highest temperatures of 15 to 18 Celsius. Light variable west winds.* Consequently, I now achieved a better result: 22 repetitions. Furthermore, another idea came to my mind: as in the source text there were two ways to specify the area of winds, such as *east* and *easterly*, I decided to simply ignore the *-ly* suffix. As a result, I achieved 23 repetitions – one more than before.

Next I deleted those sentences from the source text to enable me to see more clearly the common elements in the remaining sentences. It was rather difficult to find more similarities. Then I thought that using the same word order could improve repetitions. Thus, I thought that using an [adjective] + (weather) + (time) order might work, so I rephrased the sentences shrinking also similar segments as I did before. After that, I ran the analysis tool of TW again with the edited text and got only one more repetition. This did not surprise me at all, as the vocabulary was broader than before – it was the time to begin the translation of the edited text.

During the translation I got several full matches and fuzzy matches around 95-98% not only with the temperature segments (as I expected), but also with the wind segments after translating 3 sentences (I also got a 75% fuzzy match). I also observed that restraints in the vocabulary could improve the matches. To give some examples, *variable* is similar to *patchy*, *sunny* is like *bright* and *showers* can sometimes be changed to *rain(y)*. Although I didn't obtain many matches in the first five forecasts, I knew that I would obtain more matches if I translated more sentences.

Next I studied the 6-10 forecasts file. As in the previous file, vocabulary could be restrained and long sentences shortened (i.e. made simpler) keeping in mind the purpose of its future reusability. Therefore, I edited the text trying to use the same term where applicable. For instance, *period* became *spell* and *Connaught* became *west*. In this way, not only would I obtain more matches, but also more consistency in the appropriate terminology.

After saving the new edits, I ran the analysis tool of TW again to compare the two files I had edited with the current TM. I was quite happy with the results, as I only got 23% of no matches. I translated the 6-10 forecasts file and realized that in order to

achieve more matches, given the characteristics of the retrieval of previous translated sentence, “variable elements” such as the period of the day (as *this evening*) or the cardinal points (as *in the south*) had to always be put in the same position, usually at the end of the sentence. Sometimes, even writing unnecessary information (e.g. writing *afternoon* again even if it is already known) could increase matches.

Finally, I made the appropriate changes to the 11-15 forecasts file being aware that I would possibly obtain many more matches than before not only because my TM was larger, but also because I had found the best way to write a more translation memory friendly weather forecast. Therefore, I tried to preserve some “new” vocabulary, that is, I did not follow all the rules about constraints in vocabulary as I did in the other forecast files because I knew I would get many fuzzy matches. So, I analyzed the edited 11-15 forecasts file with the current TM. Fortunately, I was right – I got 45% fuzzy matches of 50-74%! Besides, as I only got 4% of no matches I thought it was not necessary to edit more texts. However, as I wanted to check whether the 50-74% fuzzy matches could be useful, I translated the file. Indeed, despite 2 retrieved translations were totally misleading, the others were rather useful even without vocabulary restraints, thus proving the possibility to use new vocabulary after translating several sentences so as not to always write the same sentences.

## Guide for weather forecasters to write translation memory friendly weather forecasts

In order to translate weather forecasts obtaining more benefits of a translation memory, you should write them according to the following rules:

- Keep your sentences as short as possible. Do not write about two different ideas in the same sentence and try to avoid the use of verbs (especially the verb *to be*).

Example: *Dry tonight. Some clear spells.* (NOT ~~*Tonight will be dry with...*~~)

- Restrain the vocabulary. You must always use the same terms instead of synonyms (e.g. **either** *Connaught or west*) and **sometimes** use one term for two similar concepts (e.g. either *rain* or *showers* for water). Moreover, this will improve terminology consistency. Nevertheless, **you may add both synonyms and terms** after a few weather forecasts (about 10) have been translated.

Example: (1) *Some sunny spells.*

(2) *Some sunny spells/periods.* (NOT ~~*Occasional bright periods.*~~)

- Always write using the same word order. An [adjective] + (weather) + (time) order is recommended, although you may change it sometimes after a few weather forecasts (about 10) have been translated **if** there are no extra information.

Example: *Patchy rain this evening.* (NOT ~~*Patches of rain this evening.*~~)

- Use a convention to write information about temperatures. You must always use the same structure. In a case where you have to add extra information (e.g. the period of the day), it is better to put it at the end of the sentence. Adjectives about the temperature must be put at the beginning after a comma.

Example: *Highest temperatures of 16 to 20 Celsius tomorrow afternoon.*

*Cool, lowest temperatures of 6 to 9 Celsius.* (NOT ~~*between 6-9 degrees.*~~)

- Always specify the period of the day using the same words (*this/tomorrow evening*) whether possible. Note you may **occasionally** ignore this rule after a few weather forecasts (about 10) have been translated.

Example: *Cloudy this morning. Dry this afternoon.* (NOT ~~*Dry during the evening.*~~)

- Use a convention to write information about winds. You must always use the same structure because there may be many adjectives referring to its force and several cardinal points.

Example: *Light to moderate west to north winds.* (NOT ~~*Winds will be light and moderated and westerly to northerly.*~~)

Remember that although sentences will be very concise, the weather forecasts you write will be read by a newsreader who will have your texts as a reference, **but will use a more natural language to link ideas.**

## **Bibliography**

- Hutchins, J.W. and Somers, H.L. (1992) *An Introduction to Machine Translation* London /San Diego: Academic Press.
- Somers, H.L. (ed.) (2003) *Computers and Translation: A translator's guide.* Amsterdam/Philadelphia: John Benjamins Publishing Co.
- Austermühl, Frank (2001) *Electronic Tools for Translators.* Manchester: St. Jerome.