Sub-segment TM recall

With Translation Memory (TM), sub-segment recall – recalling a fragment of a Translation Unit (TU) segment – should work just like segment-level TM recall:

- Recall matching content even if only found once in the TM (and even if the TM has only one TU in it)
- Retrieve the corresponding translation (so find the fragment translation for you – not just show you the whole TU translation)

Sub-segment alignment

Sub-segment recall can be enabled by aligning words or word spans in a TU, so that partial translations can be retrieved.

Alignment without a large TM

Lift’s aligner calculates probabilities that different source and target word spans are aligned, then selects the most probable alignments until none remains. All such automated alignment processes are subject to error. During recall, Lift can optimise the precision of translations recalled by applying a probability threshold for alignments used – analogous to the ‘fuzzy match’ threshold in segment-level matching.

SDL Trados Studio integration

Lift’s functionality can easily be added to SDL Trados Studio using the Lift plug-in, so that sub-segment matches are found for segments that otherwise show no matches. This integration will be used for real-world testing with professional translators.

Performance

Lift’s aligner calculates probabilities that different source and target word spans are aligned, then selects the most probable alignments until none remains. All such automated alignment processes are subject to error. During recall, Lift can optimise the precision of translations recalled by applying a probability threshold for alignments used – analogous to the ‘fuzzy match’ threshold in segment-level matching.

Lift’s aligner can invoke shallow or deep parsers and use the results to constrain the word spans considered, which in some cases can improve alignment quality. Similarly, parsers can constrain recall so that only the most complete ‘chunks’ are sought.

Linguistic information

Significant sub-segment translation recall can be provided with this approach, using standard bilingual dictionaries, without a requirement for a large TM. Where such dictionaries are not available, the approach can still be used with a large TM. In both cases, results are far less ‘lossy’ than other currently-available sub-segment recall implementations.

Conclusion

Unlike some systems, Lift can robustly align the words in the source text and target text of a TU straight away, regardless of TM size, using standard bilingual dictionaries. Lift can query MT systems for word translations as an alternative or supplement to bilingual dictionaries.

For language combinations where such resources are not available, Lift can derive word translations from a large TM. Once TUs are aligned, sub-segment recall is then available without the high levels of ‘lossiness’ found in statistical, phrase-extraction approaches.