STST report:
COST Action: IS1312

STSM Applicant: Karin Sim Smith, Department of Computer Science, University of
Sheffield, UK
STSM Topic: Structuring Discourse in Multilingual Europe
STSM Period: 21.02.2016 – 27.02.2016
STSM Host :Dr Christian Hardmeier, Uppsala University, Sweden

1. Purpose of the visit

The purpose of this mission was to integrate my coherence model(s) into Docent, the document level decoder developed by Dr Hardmeier in Uppsala, in an
attempt to increase the discourse awareness of Machine Translation in the matter of coherence. We also hoped it would ultimately also allow us to validate the coherence corpus which we are currently creating [https://aclweb.org/anthology/W/W15/W15 2500.pdf]. So, in addition to generally relating to the goal of enhancing discourse in the field of MT, this also falls directly under one of the goals of the action, namely discourse annotated corpora.

Current SMT decoders work on one sentence at a time, largely disregarding discourse. This
prevents MT from correctly identifying discourse related links that go beyond sentence level. Docent is a document level decoder which allows for some elements of document level discourse to be integrated, in that it provides access to the document as a whole.

The aim of this STSM was to attempt to integrate our coherence models into Docent, which
will mean that the MT output would be evaluated on various aspects of coherence. This will be the first attempt to do this directly, in the decoding process.

2. Description of the work carried out

During the visit, Christian firstly spent some time explaining the architecture of the Docent system. We also discussed the performance issues, and options for integrating my models. He helped me resolve some of the issues which frequently occur when setting up Docent, particularly with dependency libraries and the Moses dependency.

With Christian's help I managed to get Docent up and running with Moses, which is more of an involved process than the normal Moses install. I also

benefited from Christian's expertise in advising me of various configuration options.
I ran the system without a Moses initialisation to verify it was working correctly. I then ran it integrated with Moses as the baseline translation.

3. Description of the main results obtained

While it was just one week, the trip allowed me to get the decoder up and running, and begin the process of integrating coherence models and assessing their ability to influence the output.
I adapted my input text into the requisite format, and was able to run the decoder on it. This is the source text (French) for the previously mentioned coherence corpus.
I also had time to start the integration of one of my coherence models. I adapted the Docent decoder so that my model was integrated, and managed to make a call from the Docent system to my feature function (ultimately calling the EntityGraph metric which we have developed).

I was able to appreciate the complexities of the system, and the influence that any model can have. Moreover, I have an understanding of the workings, the parameters and options.

Ideally we hope to influence the decoder to identify more coherent versions of MT output.
We still need to run the experiment with an enhanced translation model, to fully assess whether errors made by Docent before integrating the coherence models correlate with the MT coherence errors in our artificially created corpus. We also hope that if the models perform similarly independently on the corpus, as when integrated into Docent using the source side of the corpus (ie French source side, translated by Docent) that this validates the errors, and thus our process of automatically inserting them.

4. Future collaboration with host institution (if applicable)

I also presented my work (on coherence in MT) to the group in Uppsala, and had some valuable suggestions for improvements.
I hope to continue to consult Christian for other options of integration within Docent.

5. Projected publications/articles resulting or to result from the STSM

We hope to have results to include in publications later this year.

6. Confirmation by the host of the successful execution of the mission

7. Financial summary

Flights:

Total paid today

Item		Description	Price
	One-way flight (including payment card fees)	Glasgow to Stockholm	GBP 182.28
	One-way flight (including payment card fees)	Stockholm to Glasgow	GBP 190.61
			Total GBP 372.89

Hotel:

Your reservation	6 nights, 1 room Change
Check-in	(14:00 - 22:00)
Check-out	(until 11:00)

Single Room	**SEK 4,960.71**
12 % VAT is included.	**SEK 595.29**
Total price	**SEK 5,556**

Flights:	£372.89
Hotel: £461.85+ transaction fee	£12.70 = £474.55
Food	£100 approx
train	166Skr x2=£28

TOTAL £975