5th Media for All – Q&A with Lindsay Bywood and Yota Georgakopoulou

LB: Our companies have been working together on the SUMAT project for over two years now. Can you give the readers a brief overview?

YG: SUMAT is an EU-funded project that we began working on in 2011, as part of a consortium consisting of nine companies in total, four of them being subtitling companies. Deluxe and VSI are, of course, two of them and the other two are Titelbild (Germany) and inVision (The Netherlands). The aim of the project is to build an online service that provides machine translation for subtitles, much like Google Translate does for generic text. The key difference from Google Translate is that the Statistical Machine Translation (SMT) engines behind SUMAT have been trained using professional quality subtitling data provided by the subtitling partners in the project. The subtitling companies' role in the project has not only been to make available the data needed for training the SMT engines, but also to act as evaluators of the MT output by gathering and providing extensive feedback to the Research and Technology Development (RTD) partners, so that they can, in turn, refine the system and enhance its accuracy. We are now in the last year of the project; most of the work has already been completed and the final SUMAT systems will be ready for our last evaluation round in September.

LB: What were your initial impressions of the project and what drew you to become involved?

YG: My background is in text translation as well as subtitling, so I am naturally interested in Machine Translation (MT) as well as any technology that can help a translator do his/her job better and faster. As you know, MT is not in widespread use yet in the subtitling industry, as the matter of fact, not even translation memories are, which has always puzzled me – our industry seems to be over a decade behind the traditional text translation industry in this respect. For years I have been following all attempts to bring MT to the subtitling workflow, such as the MUSA project, eTITLE, etc., so when we were approached to participate in the SUMAT project back in 2010 I jumped at the opportunity. The translation side of things is the only aspect in subtitling that does not benefit from any type of automation yet, unless you count spell checking! I consider MT to be one of the most promising developments in our industry, as it is the only way to really make vast volumes of content accessible in any language, and I am very excited to be part of it from the very beginning!

3. What are your hopes for the outcome of the project?

As I mentioned before, I think there is much room for automation in the translation portion of a subtitler's work. At the same time, the subtitling market has been under constant pressure for the past decade, both in terms of pricing and turnaround times; subtitling service providers are continuously asked to do work faster and at lower rates. This has become more of a necessity as a result of globalisation and the content volume explosion. Obviously the interest of all subtitling partners in SUMAT is to roll out the service internally in their organisations. We are also looking at ways to commercialise the service, so other subtitling service providers and subtitlers themselves that do not have the volume of parallel data or the resources necessary to build such an MT system can take advantage of the technology to improve their workflows and productivity. Of course this will change the landscape of the

subtitling world quite significantly, resulting in consequences for the resources and skills needed to produce large-scale subtitling. I believe that post-editing skills will become increasingly important in future; something that both subtitlers and universities should be aware of.

4. What have been the positives and negatives for you?

On a personal level, the most positive effect of the project is that it has given me the opportunity to work with very experienced computational linguists and academics that specialise in MT. The learning process has been a unique experience in itself, I almost feel like I am back at university again, researching and working on an amazing project with great people. SUMAT has also made it possible for knowledgeable people that I know and respect from competitive subtitling companies to work together for the benefit of all of us, instead of competing with each other, which does not happen often in our industry! I guess we have the EC to thank for that. Naturally though, a three-year long project involving nine different companies means an awful lot of organisation and paperwork the like of which is not usual in a subtitling company, where the longest possible project usually does not last more than a couple of months. So planning resources can be an issue, while it can also be hard at times to get consensus among so many partners. I think we are very fortunate in this regard though; despite the fact we have had many staff changes within companies, and hence the project, new staff join and responsibilities are re-arranged to an extent. Collaboration has been excellent among all partners and all hiccups have been overcome without any major disruptions to the project. I would certainly do it all over again and I feel I have made friends for life as a result of this project.

5. Tell me about the workshop at Media for All

Part of the planned work in the project is to disseminate what we do and, more importantly, the results of our work. This has a dual purpose: (a) to make the academic community aware of our progress and share some of the tools we have developed, so research can continue outside the project and technology can be improved for the benefit of the subtitling market, and (b) to make the industry stakeholders aware of our technology in order to support its uptake in subtitling production. So we have been continuously presenting papers and publishing articles on our work throughout the project life cycle. The workshop at *Media for All* is the culmination of our dissemination activities, as it will give the conference attendees the opportunity to test the live system and give us their feedback, which is key to us in terms of finalising our commercialisation plans.

6. We are also presenting a paper - what will that cover?

Apart from the workshop, we will also be giving a paper on the evaluation work we have completed in SUMAT to date. I view the evaluation of the systems as the most interesting part of the project. Here we can objectively measure how effective the system is and how successful we've been in terms of achieving our goals. There has been a lot of discussion in the media and at subtitling events as to what MT has to offer to subtitlers and whether MT is even suited to the subtitling domain. With our presentation, we plan to offer answers to these questions which are not based on generic impressions of MT tools that are widely known and accessible, but rather on a scientific examination of the systems we build, which themselves are a result of laborious and time-consuming testing carried out by our own linguists and analysed by the technical and academic staff of the project.