

Scientific Report of Short Term Scientific Mission

COST STSM Reference Number: COST-STSM-IS1305-23044 Period: 01-04-2015 to 01-07-2015 Duration: 92 working days COST Action: IS1305	STSM type: Regular (from Portugal to Belgium) STSM Title: <i>Analysis of linguistic problems connected with Legal English in legal texts of the UK, Ireland and the EU related to Law of Space</i> Guest/STSM applicant: Justyna Zdanowska, PhD student of Terminology, Lexicology, Lexicography, at Universidade Nova de Lisboa, FCSH- Faculdade de Ciencias Sociais e Humanas Host: Prof. Dr Frieda Steurs, Faculteit Letteren, KU Leuven Campus Antwerpen Host Contact: Frieda.Steurs@kuleuven.be
---	--

1. Purpose of the STSM

The main purpose of the STSM internship was to create methodology which would be the basis for the «Analysis of linguistic problems connected with Legal English in legal texts of the UK, Ireland and the EU related to Law of Space» which is the topic of my PhD project. Participating in the STSM internship at Faculteit Letteren at KU Leuven University Antwerpen Campus allowed me firstly to decide and choose suitable methodology and then to imply it to create corpora, which in future will lead to the analysis of linguistic problems of Legal English. Also, I got the opportunity to broaden my horizons and expand knowledge in the field of Terminology, Lexicology and Lexicography by not only using the resources available at the university, but also knowledge and experience of professors who were always available and immensely helpful in case of any problems.

2. Description of the work carried out during the STSM

During the STSM internship, I was very kindly welcomed by Prof. Dr Frieda Steurs at KU Leuven Campus, Faculteit Letteren in Antwerp. Also, I met Prof. Dr Hendrik Kockaert and Mr Kris Heylen who is Postdoctoral Research Fellow at KU Leuven University as well as all the other professors from the Department.

I received the opportunity of using the latest software SDL Trados available at university as well as the most recent publications in the field of Terminology, Lexicology and Lexicography, access to the library and online resources which were very crucial for improving my knowledge and background for the project.

Furthermore, I have also contacted and had consultations with Mr Kris Heylen who very kindly helped me with advancing my project from the methodological side. Thanks to Mr Kris Heylen's suggestions, I was able firstly to choose software which might have been valid for the project and then tested it. As Mr Kris Heylen is a specialist in Corpus Linguistics, I managed to test the most appropriate tools (mentioned below) and chose those which are the most suitable for the corpora. Below, I presented the stages and topics which I

investigated, tested, learnt and advanced the knowledge and tools (software) that I used during my internship in Antwerp.

- I. Background knowledge
 - a) Using the web as a source of corpora
 - b) Characteristic features of a new type of corpus and potential problems to be solved
 - c) Size and content of corpora

At this stage of my investigation, I strengthened my knowledge of methodology for creating corpus, background information related to Corpus Linguistics and Computational Linguistics. Also, it was necessary to reconsider characteristic features of a corpus due to the fact that texts, for creating corpora for this project, are downloaded from the web.

Furthermore, it was crucial to predict potential problems and disadvantages which were expected to appear while creating the corpora, i.e. issues related to size, content, representativeness and authenticity which might be influenced by unlimited sources of the web. Also, the investigation included learning about the World Wide Web, its statistical linguistic and computational data that are constantly changing and which potentially can influence the corpora.

- II. Defining the criteria
 - a) Criteria for selecting texts – content of the corpora
 - b) Criteria for selecting sources of the corpora- institutions and organizations
 - c) Criteria for selecting software- tools to verify, store and maintain texts

At this stage, I defined criteria for texts, sources and software to limit the scope of the project, obtain the best results and fulfil the main aim of this research (title of the STSM internship). Also, it turned out that defining the criteria is crucial because of unlimited resources and information available on the web and market.

- III. Methodology
 - a) Initial methodology for selecting texts
 - b) Advanced methodology

The methodology of collecting the data was problematic because of the field of the PhD project. Due to the fact that there still has not been described any similar methodology for collecting and treating data, it was very problematic to find, collect and store texts. Our first assumption was to use search engine tools (search boxes, usually in the right upper corner of a website to search information) provided on individual websites of organizations/institutions but results were not satisfactory and number of acquired texts was not sufficient for the analysis.

Consequently, it was necessary to change the methodology. At this point of the STSM internship, the consultation meetings with Prof. Dr Frieda Steurs and Mr Kris Heylen were crucial. After discussion, we decided to test: (1) BootCat, (2) XML Sitemap Generator, (3) SketchEngine, (4) Extension DownThemAll for Mozilla Firefox, (5) KwicFinder, (6) Google advanced search tool. In order to be able to test all of the tools, it was necessary to learn methodology of using them as well as reading about the studies which have already been using the above-mentioned software to be able to verify if they might be useful (and if yes, how) for the

case of this project. Also, in this case we decided to define the criteria which a given software had to fulfil. The table below presents the comparison and results of the tools according to the chosen criteria.

SOFTWARE	BootCatwebcra wler	XML Sitemap Generator	SketchEngine	Extension DownThemAll for Mozilla Firefox	KwicFinder	Google Advanced SearchTool
Free	YES	NO	Only 1 month free (beta version), then 57,99E per year + extra costs*	YES	YES	YES
User-friendly	NO	NO	NO	NO	NO	YES
Available online	YES	YES	YES	YES	YES	YES
Cloudcomputing	NO	NO	NO	NO	NO	NO

*extra costs only in case of overpassing the limit of the 1 000 000 words processed

Surprisingly, it turned out that unfortunately none of the mentioned applications have fulfilled all of the criteria, but it was the Google Advances Search Tool that met most of them (3) and was the most useful application to be used.

- IV. The current and future corpora
 - a) Results
 - b) Potential future enlargement of the corpora

Having chosen the most suitable software, then it was possible to select, download and store the texts. Although the process of collecting and storing the texts is still being in progress, it was successful that we chosen the methodology and tools. Untill now, we created 11 singular corpora including 1137 files. The number of the downloaded texts is influenced by the character and field of the research project, as well as criteria that were chosen and time-consuming process of acquiring the texts.

In conclusion, finishing successfully the STSM internship, allowed me not only to define and apply the methodology and broden knowledge from the field, but also it was possible to initially start working and analyzing the gathered corpora.

3. References

1. COTSOES Recommendations for Terminology Work, <https://termcoord.files.wordpress.com/2013/07/here.pdf> available 14/07/2015 6:00pm
2. De Groot, G.R, Carrera, S. European Citizenship at a Crossroads Enhancing European Cooperation on Acquisition and Loss of Nationality, No. 72/November 2014 (<http://www.ceps.eu/system/files/No%2072%20European%20Citizenship%20at%20Crossroads.pdf> available on 10/04/2015) , Liberty and Security CEPS
3. De Groot, Laer C.J.P. The Dubious Quality of Legal Dictionaries. International Journal of Legal Information, the Official Journal of the International Association of Law Libraries, Volume 34, Issue 1 Spring 2006, Article 6

4. De Groot, G.R, Wautelet, P. Reflections on quasi-loss of nationality in comparative, international and European perspective No. 66/August 2014 (http://www.ceps.eu/system/files/No%2066%20ILEC%20-%20Quasi-loss%20of%20nationality_0.pdf available on 10/04/2015), Liberty and Security CEPS
5. Gambier, Y., & van Doorslaer, L. (Eds.). (2011). Handbook of translation studies (Volume 2). Amsterdam: John Benjamins.
6. Gatto, M. The 'body' and the 'web': The web as corpus ten years on. Universita di Bari. http://clu.uni.no/icame/ij35/Maristella_Gatto.pdf available on 13.07.2015, 8:30 am.
7. Heylen, K. , Steurs, F. Translating legal and administrative language: How to deal with legal terms and their flexible meaning potential. Available 14/07/2015 8:00pm http://www.ling.arts.kuleuven.be/qvl/prints/IdYouss_Steurs_2015abs_Concept_mapping_legal_documents.pdf
8. Kilgarriff, A. et al. The Sketch Engine: ten years on. Springer Berlin Heidelberg 10 July 2014, Lexicography ASIALEX (2014) 1:7-36, DOI 10.1007/s40607-014-0009-9
9. Kilgarriff, A. 2001. Web as corpus. In P. Rayson, A. Wilson, T. McEnery, A. Hardie and S. Khoja (eds.). Proceedings of the Corpus Linguistics 2001 Conference, UCREL, Lancaster, 342–344. Available at <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/kilgarri.pdf>
10. Kilgarriff, A., Grefenstette, G.. 2003. Introduction to the special issue on the web as corpus. Computational Linguistics 29 (3): 333–347. Available at <http://acl.ldc.upenn.edu/J/J03/J03-3001.pdf>.
11. Kockaert H.J. , Steurs F. Handbook of Terminology volume 1, John Benjamin Publishing Company, 2015
12. LaHoussine Id-Youss, Steurs F. Concept mapping in legal documents. Case study: The translation of family law for Moroccan migrants. Universite de Savoie, Chambéry, 4-5 June 2015
13. Leech, G. 2007. New resources or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf and C. Biewer (eds.). Corpus linguistics and the web, 133–150. Amsterdam and New York: Rodopi.
14. LISE Legal Language interoperability Services. Guidelines for collaborative legal/administrative terminology work. Eurac Research (http://webfolder.eurac.edu/EURAC/Publications/Institutes/autonomies/commul/Guidelines_for_collaborative_legal_administrative_terminology_work.pdf) available on 11/04/2015 Published in Bolzano/Bozen in July 2013, ICT PSP.
15. McEnery, Tony and Andrew Wilson. 2006. Corpus linguistics. 2nd edition. Edinburgh: Edinburgh University Press.
16. Picht, H. (Ed.). (2006). Modern approaches to terminological theories and applications (Vol. 36). Bern: Peter Lang.
17. Sinclair, John. 1991. Corpus, concordance, collocation. Oxford: Oxford University Press.
18. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014DC0207&from=EN>