London Interdisciplinary Social Science Doctoral Training Partnership
Advanced Research Methods in Social Sciences

Course Code & Title	**LISS326 Researching Digital Texts with Corpus Tools**				
Convenor(s)	Dr Clyde Ancarno, King's College London				
Institution	King's College London		**Department**	School of Education, Communication & Society	
Academic Year	2018-19		**Term**	Autumn	
Number of sessions	2	**Research Platform**	Digital & Archival Research Methods	**Length of Session(s)**	1.5 Days
Day, Date			**Start : End**	**Room Location**	
Tuesday 13 November 2018 Wednesday 14 November 2018			1000 : 1700 1300 : 1700	B20, James Clerk Maxwell Building Waterloo Campus, SE1 8WA	
Enrolment Link:	You may be prompted to log into SkillsForge Details on when courses open for booking can be found here: https://liss-dtp.ac.uk/our-training-programme/				

Course Description:

This skills-based course will introduce you to corpus tools which you can use to analyse collections of digital texts/*corpora* (e.g. using traditional corpora but also corpora of answers to open-ended questions in a questionnaire, of legal texts relevant to your research, or of transcripts of news broadcast). We will illustrate how corpus methods can be used to support researchers in the social sciences and humanities, helping them to enhance their research in a range of ways. You will, for example, become aware of patterns of language use in your textual data which may otherwise remain invisible and be able to reduce researcher bias in your analysis of language use. By the end of the course, you will know what to consider when collecting and preparing textual data for analysis with corpus tools and you will be capable to produce a range of corpus outputs - e.g. lists of words which tend to co-occur together (*collocate* lists) - to support your analyses.

Course Outline:

This 1.5 days course will be split into 5 sessions.

SESSION 1 - Introduction to corpus linguistics (Clyde Ancarno)

This initial session will introduce the field of corpus linguistics and three key concepts for corpus exploration: word lists, concordances and collocations. To prepare you for the following session, we will also introduce the basic rules underpinning corpus building.

SESSION 2 - Building a corpus from scratch: tools and approaches (Anda Drasovean)

In this session, you will learn how to format a purpose-built corpus (e.g. using a collection of texts gathered using Nexis). We will focus on the process of building and designing a purpose-built specialised corpus and discuss the key stages of corpus collection: (i) selecting and extracting relevant texts, (ii) cleaning up the data and (iii) annotating corpora. For each of these stages various tools that might assist you will be introduced.

SESSION 3 - Keyness (Clyde Ancarno)

A commonly used corpus technique involves the comparison of the word frequencies in your corpus and in a reference corpus (usually a larger corpus). This comparison identifies the words that are "key" for your corpus based on their statistical prominence. In this session, you will learn how to compile and analyse keyword lists and explore some of the pitfalls and opportunities these processes provide.

SESSION 4 - Exploring and comparing specific words: word sketches and sketch differences (Clyde Ancarno)

During this session, we will explore ways in which the grammatical and collocational behaviour of specific words (including comparisons of two chosen words) can help us gain insights into our objects of inquiry. To do so, we will focus on the *Word Sketch* and *Word Sketch Difference* functionalities in Sketch Engine.

SESSION 5 - Corpus surgery (Clyde Ancarno and Anda Drasovean)

At the end of the course, you will be encouraged to work with digital texts relevant to your research (you will need to bring those along with you – you will receive instructions about this before the start of the course) to discuss the possibility of using corpus software to explore them. You will also have the opportunity to practice compiling corpus outputs using corpora made available to you.

Reading List:

Sinclair, J., 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Simpson, J., 2011. *The Routledge handbook of applied linguistics*. Abingdon: Taylor & Francis [Chapter 42 - Corpus linguistics – by Adolphs and Ungerer]. Available at:
https://kclidp.kcl.ac.uk/idp/profile/SAML2/Redirect/SSO?execution=e1s1 [last accessed 14.09.18]

Kilgarriff A, Baisa V, Bušta J, Jakubíček M, Kovář V, Michelfeit J, Rychlý P, Suchomel V. The Sketch Engine: ten years on. Lexicography. 2014 Jul 1;1(1):7-36. Available at: https://www.sketchengine.co.uk/wp-content/uploads/The_Sketch_Engine_2014.pdf [last accessed 14.09.18]

Rayson, P. 2012. Corpus Analysis of Key Words. In *The Encyclopedia of Applied Linguistics*. Iowa: US. Blackwell Publishing. Available at:
http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0247/full [last accessed 14.09.18]

Eligibility:

Students must be registered PhD students at either King's College London, Imperial or Queen Mary University of London.

Pre-course preparation:

Students will be given preparatory work at the end of each session.

Number of students: 15