

## **Report 1 – Instructions – Corpora and Statistics**

Please write a short report to count as part of your assessment for this course Corpora and Statistics. It will count as either 40% or 50%. This decision will be made after consultation with the members of the class.

### **Aim -**

In the age of computers, the hardest things about statistics are:

1. Knowing which technique to use (and it's strengths and weaknesses) and
2. Knowing how interpret the results it gives you.

The aim of this project is assess your competence in these two regards.

### **Marks –**

Professionalism: Language and Style guide accuracy – 10/50

Stats: choosing the right technique for the right question – 20/50

Interpretation: Interpreting the results correctly – 20/50

### **Instructions –**

Choosing whatever data / research question appropriate from the lexical semantics data or the discourse analysis data from 1<sup>st</sup> semester, perform the following statistical analyses and interpret the results.

Present the results and interpretations in a short report with one section devoted to each statistical technique.

Make sure to insert any graphics into your report in an appropriate manner and present any number results in appropriate tables.

1. Chi<sup>2</sup> Test for independence and Pearson Residuals of its results
2. Multiple Correspondence Analysis and a hierarchical clustering of its results
3. Hierarchical Agglomerative Clustering and a *K*-Medoid clustering of its results.
4. Binary Multiple Fixed-effects Logistic Regression including the statistics of the model and a test for orthogonality and parsimony.

Please, if any of that does not make sense to you, let me know and we will go through it individually or in class.

dsg.up8@gmail.com

[www.dsglynn.univ-paris8.fr/corpus\\_uam\\_2020-.html](http://www.dsglynn.univ-paris8.fr/corpus_uam_2020-.html)

stylesheet - [www.dsglynn.univ-paris8.fr/Style\\_Guide.pdf](http://www.dsglynn.univ-paris8.fr/Style_Guide.pdf)