Title of dataset: Supplementary materials to *Intensifiers in Late Modern English*

Authors/Investigators (in alphabetical order):

Claudia Claridge (University of Augsburg)
claudia.claridge@philhist.uni-augsburg.de

Ewa Jonsson (Mid Sweden University)
ewa.jonsson@miun.se

Merja Kytö **(**Uppsala University)
merja.kyto@engelska.uu.se

Related publication: Claudia Claridge, Ewa Jonsson, and Merja Kytö. 2024. *Intensifiers in Late Modern English: A sociopragmatic approach to courtroom discourse.* Cambridge: Cambridge University Press.

Date of data collection: 2016

Description of files:

1. Search strings.docx
2. R code.r
3. Model output.docx
4. Data files for regression analyses
    4.1 Intensifiers.xlsx
    4.2 Maximizers.xlsx
    4.3 Boosters.xlsx
    4.4 Downtoners.xlsx
    4.5 *Very*.xlsx
    4.6 Boosters without *very*.xlsx
    4.7 Intensifiers without *very*.xlsx
    4.8 *A little*.xlsx
    4.9 Downtoners without *a little*.xlsx
    4.10 Diminishers.xlsx
    4.11 Minimizers.xlsx
    4.12 Moderators.xlsx
    4.13 Diminishers without *a little*.xlsx

The files listed are supplementary to the publication and relate to the data retrieval and regression analyses for the study; for further information on the corpus methodology, screening criteria, etc., see the book, especially Chapters 4 and 8.

The file **1. Search strings.docx** offers comprehensive lists of the search terms used for the retrieval of intensifiers from the Old Bailey Corpus 2.0. The searches yielded a total of 193,349 hits in the corpus. By way of manual screening, 64,975 of the hits were found to be relevant examples of intensifiers.

The regression analyses were run in R, and the results of these are documented in the book. The R code for the analyses is given in file **2. R code.r**; the code was provided by Lukas Sönning (University of Bamberg). The output of the models is documented in the file **3. Model output.docx**.

The data files for the regression analyses are available in the folder **4. Data files for regression analyses**. File **4.1 Intensifiers.xlsx** comprises the full dataset of 64,975 intensifiers and files **4.2–4.13** various subsets of this, each file specifying the raw frequency per speaker ID and the gender, role, social class, and total word count of the speaker. The regression analyses were run on the speakers with a known value for each of the three sociopragmatic variables (gender, role and class), altogether 74,445 speakers (speaker IDs).

The speaker ID information (ID, gender, role, class, word count, year) derives from an OBC 2.0 speaker list provided by Magnus Nissel for the purposes of the study. Speaker IDs with question marks have been edited to show 'X' instead of the question mark, for better Excel functionality. 'Maximizers' is used as a generic term in the R code, and column B thus has the heading 'Maximizers' in each data file. However, what the content of the column specifies is the raw frequencies for the relevant dataset of the file in question (for instance, for 4.1, the frequency of intensifiers overall; for 4.3, the frequency of boosters, etc.). In column D, speakers with the role 'Interpreter' are retained, but these were excluded from the analyses (see the R code). The frequencies in files 4.12 and 4.13 are too small for meaningful statistical analysis and are provided here only for the sake of documentation.

Like the Old Bailey Corpus, http://fedora.clarin-d.uni-saarland.de/oldbailey/, our supplementary materials are intended for non-commercial use only. Reference line to our materials:

Claudia Claridge, Ewa Jonsson, and Merja Kytö. 2024. Supplementary materials to *Intensifiers in Late Modern English: A sociopragmatic approach to courtroom discourse.* Cambridge: Cambridge University Press. www.cambridge.org/intensifiers