# Words in #LancsBox: Frequency lists and keywords

**Task 1.** Find out the frequencies of words and word classes in the one-million-word LOB corpus. Fill in the table below.

| ID | Top 5 types | Top 5 POS tags | Top 5 nouns | Top 5 adjectives |
|---|---|---|---|---|
| 1 | the | NN | time | other |
| 2 | of | IN | year | good |
| 3 | and | DT | man | more |
| 4 | to | JJ | mr | first |
| 5 | a | NP | way | new |

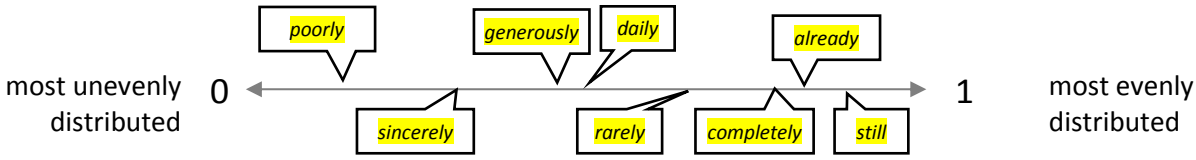
**Task 2.** Find out the dispersions of words in the one-million-word LOB corpus. Answer the following questions.

- Which is the most evenly dispersed verb in the corpus?

It is the verb *make* if we look at the CV and Juilland's d measures (CV =0.11), d(0.97); it is the verb *be* if we look at the DP measure (DP 0.044).

- Which of the following adverbs is most evenly distributed? Place the adverbs on the Juilland's d dispersion scale.

Adverbs to consider: *already, completely, daily, generously, poorly, rarely, sincerely* and *still.*




**Juilland's *D*** is a measure of dispersion that builds on the Coefficient of variation. It is a number between 0 and 1, with 0 signifying extremely uneven distribution and 1 perfectly even distribution. Juilland's D was originally developed for use in frequency dictionaries (Juilland et al. 1964, 1970; Leech et al. 2001; Davies & Gardner 2010).

- Which of these adverbs is used most frequently in adventure (fiction)? still
- Which of these adverbs is used most frequently in newspaper editorials? sincerely

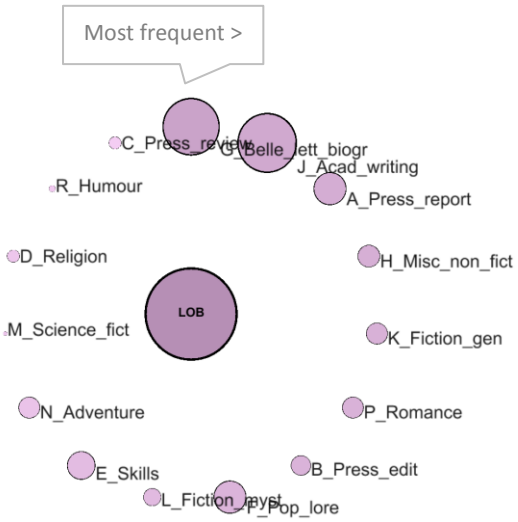



Figure1. *Already* in LOB

**Task 3.** Find the keywords in LOB when compared to Brown. Fill in the table below:

| ID | Top 5 keywords (types) | Top 5 keywords (nouns) | Top 5 keywords (adjectives) | Top 5 keywords (verbs) |
|---|---|---|---|---|
| 1 | labour | labour | British | realise |
| 2 | london | london | 6d (occasionally written Lsd) is the symbol for the pre-decimal currency | recognise |
| 3 | sir | sir | main | marry |
| 4 | towards | colour | further | ensure |
| 5 | round | centre | grey | round |