

Words in #LancsBox: Frequency lists and keywords

T **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				
2				
3				
4				
5				

T **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?
- Which of the following adverbs is most evenly distributed? Place the adverbs on the Julland's D dispersion scale.

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

most unevenly distributed 0 ← → 1 most evenly distributed

Julland'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. Julland's D was originally developed for use in frequency dictionaries (Julland et al. 1964, 1970; Leech et al. 2001; Davies & Gardner 2010).

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

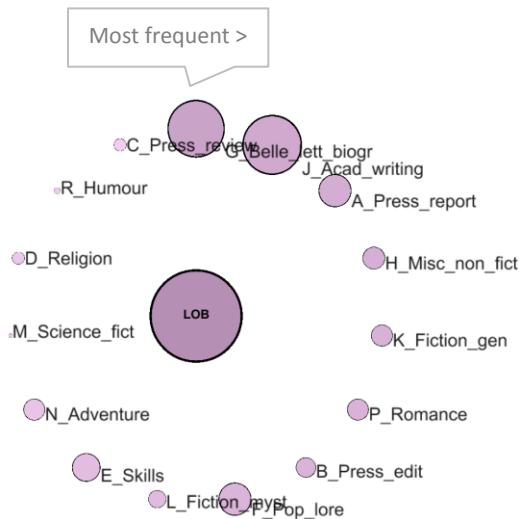


Figure1. *Already in LOB*



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				
2				
3				
4				
5				