

Corpus Analysis Activity 6.5c: Pronoun variation across registers

by Robert Poole

It is suggested that you complete Module C. Introduction to Corpus Analysis before doing this activity.

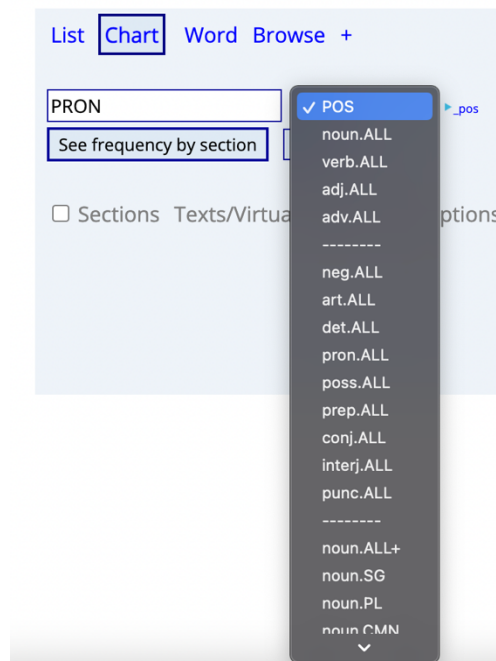
In this corpus activity, we will explore how grammar, in this case, pronouns, varies across registers. In applied linguistics, a register is a functional variety of language shaped by external contextual variables. For example, the academic register is comprised of a multitude of genres such as conference proposals, abstracts, proposals, lectures, and more. All of these genres are shaped by the situational variables of academia. In [COCA](#), the language data are organized into 8 registers, e.g., newspaper, film/television, spoken, academic, etc. If we consider a register as a functional variety of language shaped by the contexts in which it is produced, then we would expect each register to display variation in how various language features are employed. Further, if such variation exists within such a social register of language use, we, as applied linguists and discourse analysts, should be able to make statements concerning variation across contexts and the rhetorical and functional factors contributing to such divergence.

In our analysis, we should remember that the use of a feature such as pronouns certainly is present across all registers. In register analysis, we do not attempt to show that one feature is present in one register but absent in another—that is unlikely. However, we attempt to demonstrate that registers employ features from the language system at greater or less probability than other registers. In this activity, let's explore pronoun use across the eight registers included in the Corpus of Contemporary American English. We will first collect empirical data and then offer qualitative interpretations for the existence of variation.

1. Go to English-Corpora.org and select the Corpus of Contemporary American English: <https://www.english-corpora.org/coca/>
2. Click the Chart search function
3. Select pron.ALL from the drop down menu next to the search bar. After you select pron.ALL from the menu, you will see PRON in the search bar.
4. Click “See frequency by section” and answer the following questions.

A. Which three registers have the highest per million (per mil) use rates? [Note: The per million score is a useful way to compare frequencies across corpora. While the registers are balanced in the corpus, they are not exactly the same size. Thus, as is regularly practice in corpus linguistics, we must “normalize” the data to make comparisons possible. Data is typically normalized to per million for large corpora, but researchers analyzing smaller corpora may choose lower rates such as per 1,000 or per 10,000.]

- 1.
- 2.
- 3.



B. Which three registers have the lowest per million use rates?

- 1.
- 2.
- 3.

C. List three reasons why you think Television/Movies would have the highest frequency of pronoun use and Academic would have the lowest.

- 1.
- 2.
- 3.

5. Let's explore pronoun use more deeply in these two registers. In the COCA, complete the following sequence:
1. Click the List search function
 2. Enter PRON in the search bar
 3. Click Sections
 4. In column 1, select TV/Movies; in column 2; select Academic
 5. For Sorting, select Frequency
 6. Click Find matching strings

Fill in the following tables with the top 5 most frequent pronouns for each register.

Pronouns (TV/Movies)	Per Million	Pronouns (Academic)	Per Million

This module is part of [Critical Language Awareness: Language Power Techniques and English Grammar](#), a freely available open resource offered by the [Clarify Initiative](#), a privately funded project with the goal of raising critical language awareness and media literacy throughout society.

Answer Key

4A. 1. Television/Movies, 2. Fiction, 3. Spoken

4B. Academic, News, Magazines

5. *Note: numbers were accurate as of September 2022; as the corpus is added to, numbers may differ.*

Pronouns (TV/Movies)	Per Million	Pronouns (Academic)	Per Million
You	39,303	It	4,811
I	37,502	They	2,883
It	19,063	We	2,244
We	9,671	I	2,010
me	8,359	He	1,896

Here again it is important to revisit the point that registers select from the language system at varying probabilities. It is not that one feature is absent or present in a certain register, but that it is used at a higher or lower frequency than another register. In this case, we can see that pronouns are indeed present in academic writing, but it is interesting to observe the divergence both in per million use and in pronoun preference. It is perhaps unsurprising that personal pronouns *you* and *I* are so frequently used in television and film. After all, television and film involves conversation centering on humans in their unfolding drama, humor, romance, etc. The dialogue present in such a register between characters dictates a high frequency of personal pronouns. In contrast, *I* is used much less frequently in the academic register and *you* is nearly outside of the top 10 most frequent pronouns. This pattern in the academic register seems reflective of the notion that language use in academic genres is generally impersonal and author-absent. Such impersonal language use is imagined to indicate greater objectivity from the academic writer/speaker as they present themselves as unbiased and impartial.