

Lexical Variation

Seminar 'Lexicology'

Quirin Würschinger, LMU Munich

July 16, 2025

Outline

- **The importance of variation in lexicology**
- **Dimensions of variation:** Types and categories of lexical variation
- **Speaker variation:** Regional and social variation
- **Situational variation:** Register and text type variation
- **Case study: Modal verbs:** Lexical change and variation in English modals
- **Practice:** Corpus analysis of modal verb variation
- **Summary:** Key takeaways

The importance of variation in lexicology

Discussion question (3 minutes in pairs):

Why is lexical variation important for understanding language?

However, it should be stressed here that lexical variation and the restriction of items to specific varieties must not be neglected in lexicology, as has so often been the case in the past. ([Lipka 1992, 9](#))

Today's focus: How does variation manifest across different dimensions and what can we learn from studying it?

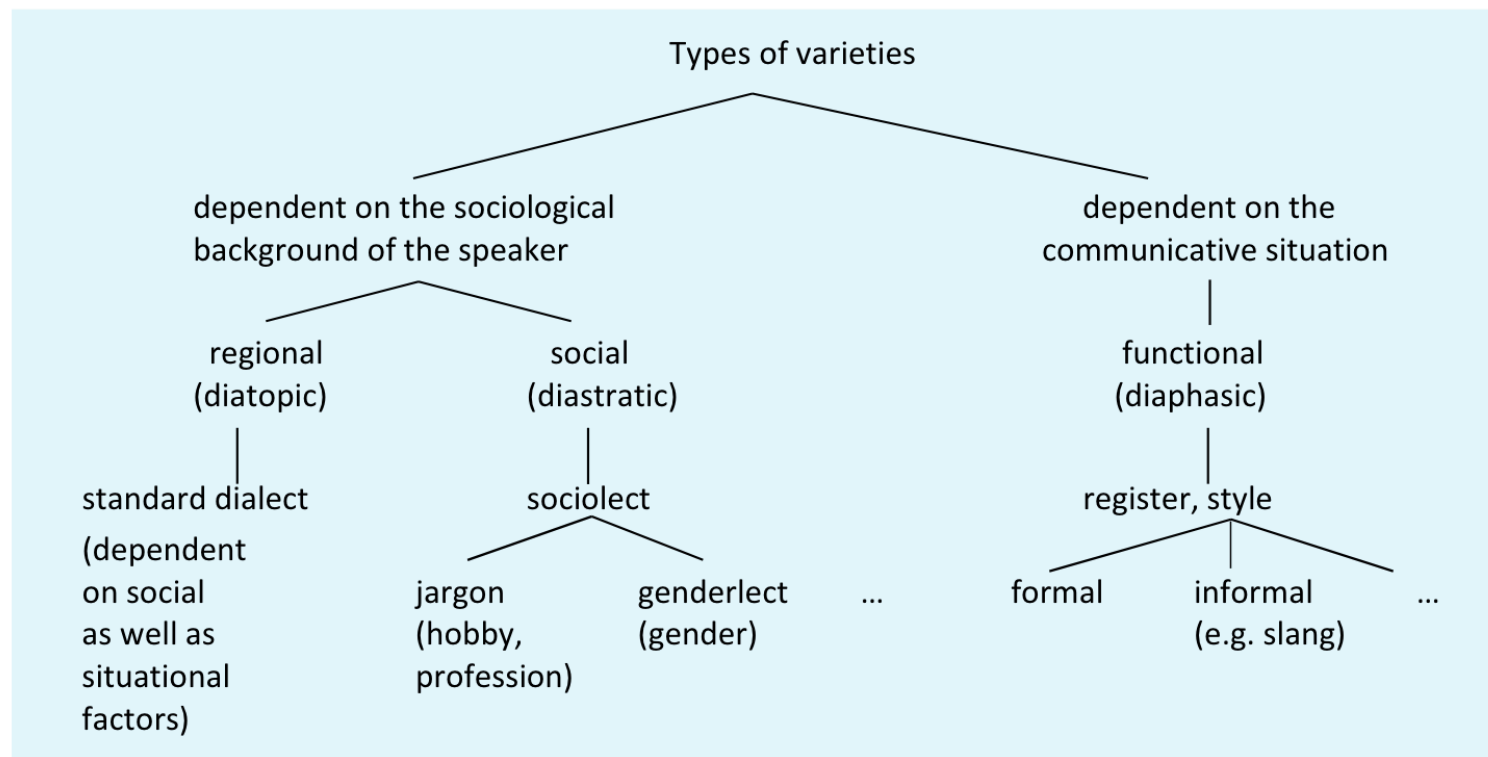
Dimensions of variation

Types of variation

- **Diatopic:** Geographic/regional variation
- **Diastratic:** Social variation (class, education, age)
- **Diaphasic:** Situational variation (register, style)

Language change = language variation over time = **diachronic variation**

Variation based on “user” and “language use”



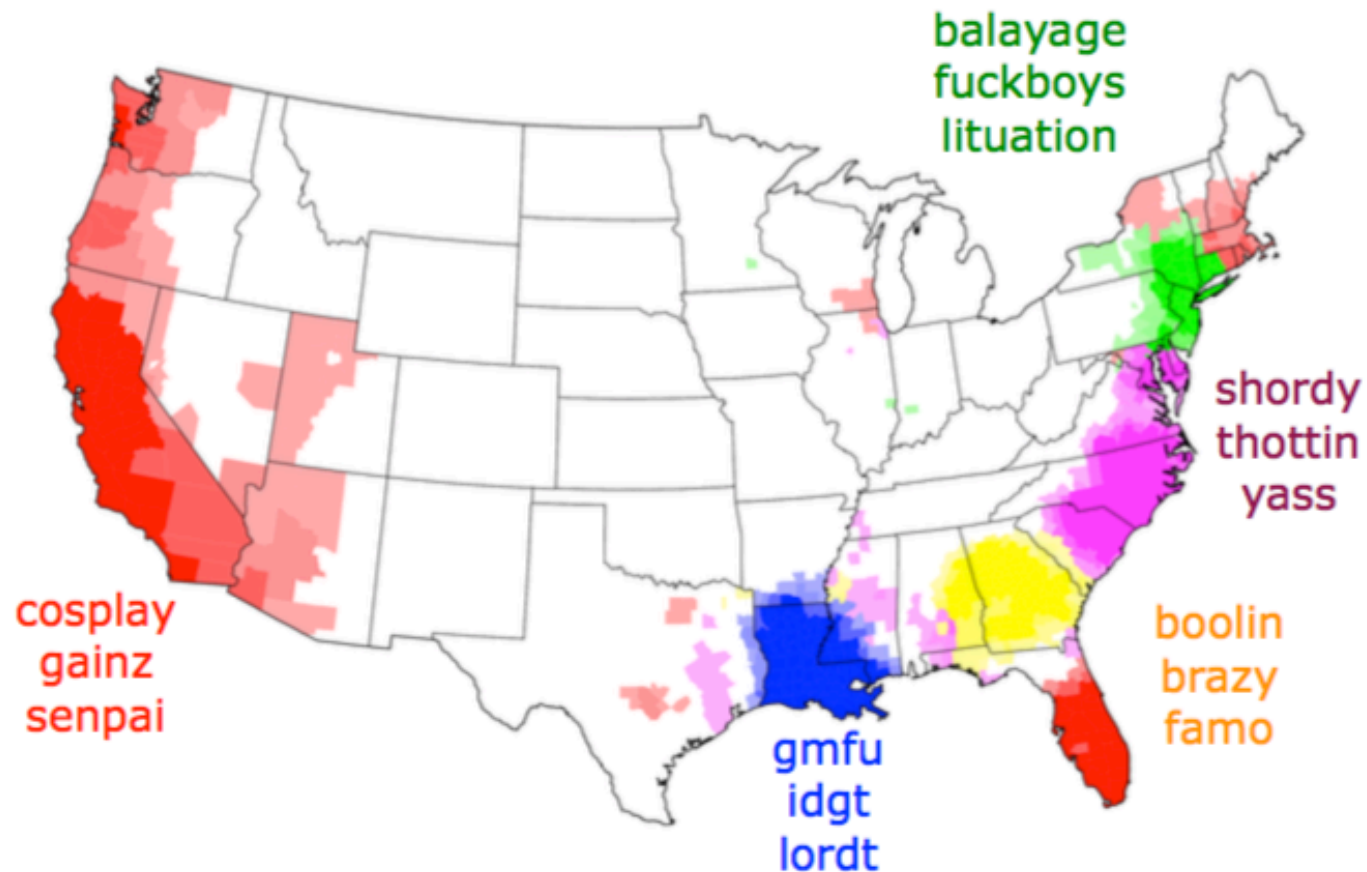
Dimensions of variation in English. (Kortmann 2020, 204)

Speaker variation

Regional variation

Regional variation in **lexical innovation**

Hubs of Lexical Innovation: Examples (Factor Loadings)



Mapping lexical innovation across regions.

Mapping lexical innovation across regions (Grieve, Nini, and Guo 2018)

Differences between British and American English

Examples:

British	American
autumn	fall
lift	elevator
lorry	truck

Speaker variation: Social variation

Sociolect

Sociolect: Social dialect - variation depending on education and social standing.

Examples of social variation:

- **Education level:** *ain't* vs *am not/is not/are not*
- **Social class:** *pardon* vs *what* vs *excuse me*
- **Age groups:** *cool* vs *awesome* vs *lit*
- **Professional jargon:** *Sprachwissenschaft* vs *Linguistik*

Key principle: Lexical choices often signal social identity and group membership.

Situational variation: Register

Register variation

Register: Variation that depends on language use, not the user. Includes lexical, phonological, and grammatical features.

Examples:

- **Academic:** *lexical item, corpus analysis, frequency distribution*
- **Informal:** *word, text study, how often*
- **Technical:** *morphological process, phonological rule*
- **Everyday:** *word formation, sound pattern*

Key insight: The same speaker uses different vocabulary in different situations.

Case study: Modal verbs in English

Lexical change and variation in English modal verbs

Research foundation: ([Hilpert and Mair 2015](#))

“Another domain of English grammar that is currently undergoing change is the domain of modality, specifically the modal auxiliaries. In the most general of terms, the situation is that several of the core modal auxiliaries are declining in text frequency (Leech 2003; Mair 2006), while at the same time new quasi-modal elements are undergoing grammaticalization (Krug 2000).” ([Hilpert and Mair 2015, 185–86](#))

Research questions:

- Why are certain forms declining while others increase?
- Is there a relation between these developments?
- How do we assign cause and effect roles?

Core and peripheral modals

Core modals

- *will, would*
- *can, could*
- *may, might*
- *shall, should*
- *must*

Peripheral modals

- *BE going to*
- *have to*
- *got to*
- *need to*

Pattern: Core modals show frequency decline while peripheral modals show increase.

Frequency changes over time

Corpus evidence

Brown family of corpora: Brown (1961), LOB (1961), Frown (1992), FLOB (1992)

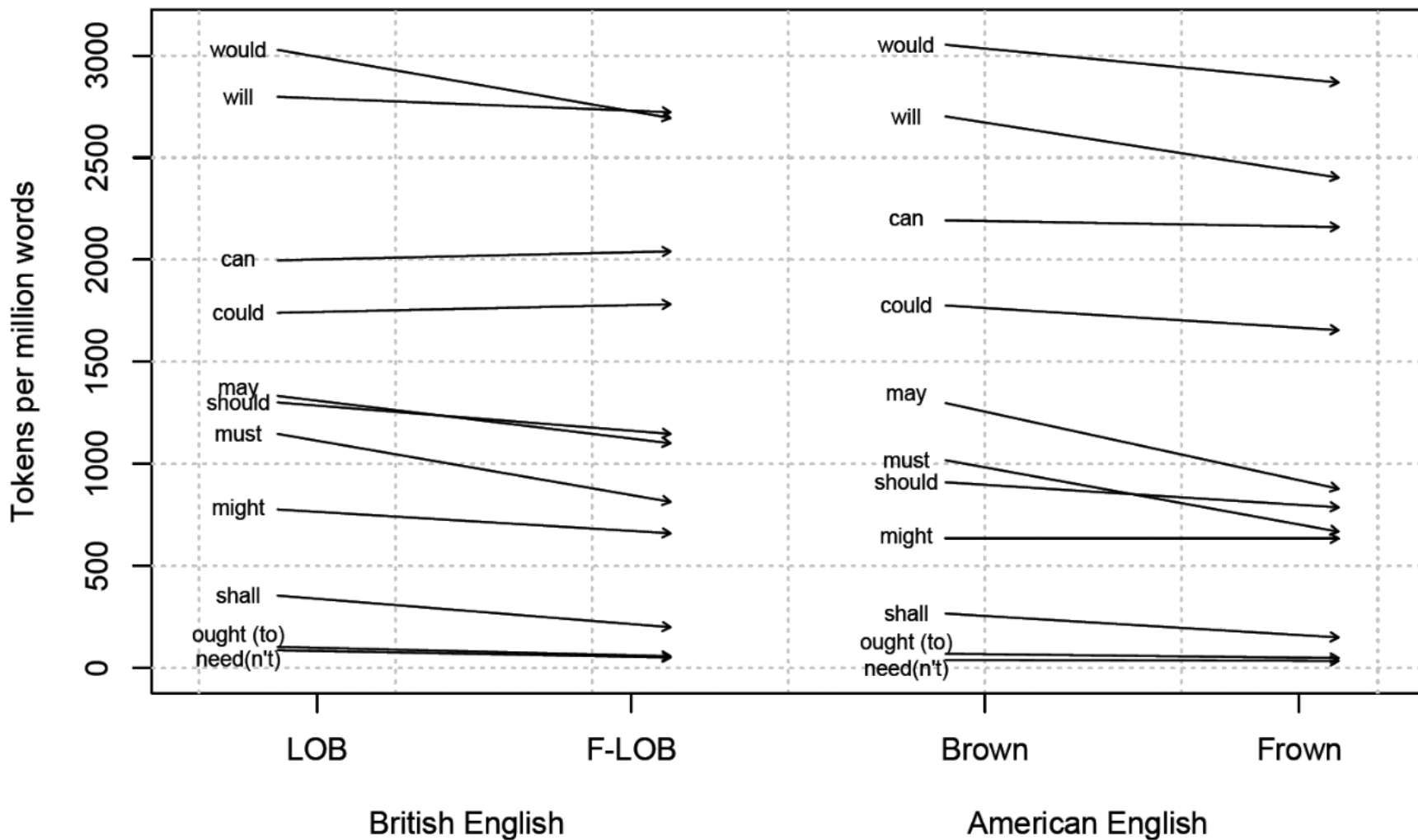


Figure 10.2 The decline of the modal auxiliaries (based on Leech 2003: 228, table 3)

Brown family of corpora structure. ([Hilpert and Mair 2015](#))

“Parallel cross-variety declines are particularly in evidence for the modals *would, may, should, must, and shall*.” (Hilpert and Mair 2015, 186)

Declining core modals:

- *would, may, should, must, shall*

Increasing peripheral modals:

- *BE going to, have to, got to, need to*

“This is further corroborated by Hilpert (2008: 37), who finds that typical verbal collocates of *shall* in the British National Corpus are the explicitly metalinguistic verbs *consider, examine, discuss, and argue*. The decline

Text type variation

Genre-specific patterns

“In a study that is based on the Time magazine corpus (Davies 2007), Millar (2009) tracks the frequency of the modal auxiliaries in American English press writing. He finds that *shall*, *must*, and *ought* are declining between the 1920s and the 2000s, but that interestingly, *can*, *could*, and *may* are undergoing substantial frequency increases and *will*, *might*, and *should* show at least small increases (Millar 2009: 205).” (Hilpert and Mair 2015, 187)

Time magazine corpus analysis (Millar 2009):

- *shall*, *must*, *ought* declining (1920s-2000s)
- *can*, *could*, *may* increasing substantially
- *will*, *might*, *should* showing small increases

“One explanation for the discrepancies between the tendencies in the Brown family of corpora and in the Time corpus is the composition of the respective corpora. Whereas the Brown corpora represent a balanced set of genres, the Time corpus represents a single text type.” (Hilpert and Mair 2015, 187)

Key finding: Text type variation can explain frequency discrepancies across different corpora.

Practice: Corpus analysis of modal verbs

Collaborative analysis

Microsoft Excel spreadsheet:

<https://1drv.ms/x/s!AvkgNVI9yS6aoi84illjaZZQfHzZ>

Task 1: Study modal verb frequency changes in COHA

Search queries:

Core	Peripheral
would *_vv	BE going to *_vv
may *_vv	HAVE to *_vv
should *_vv	got to *_vv
must *_vv	NEED to *_vv
shall *_vv	

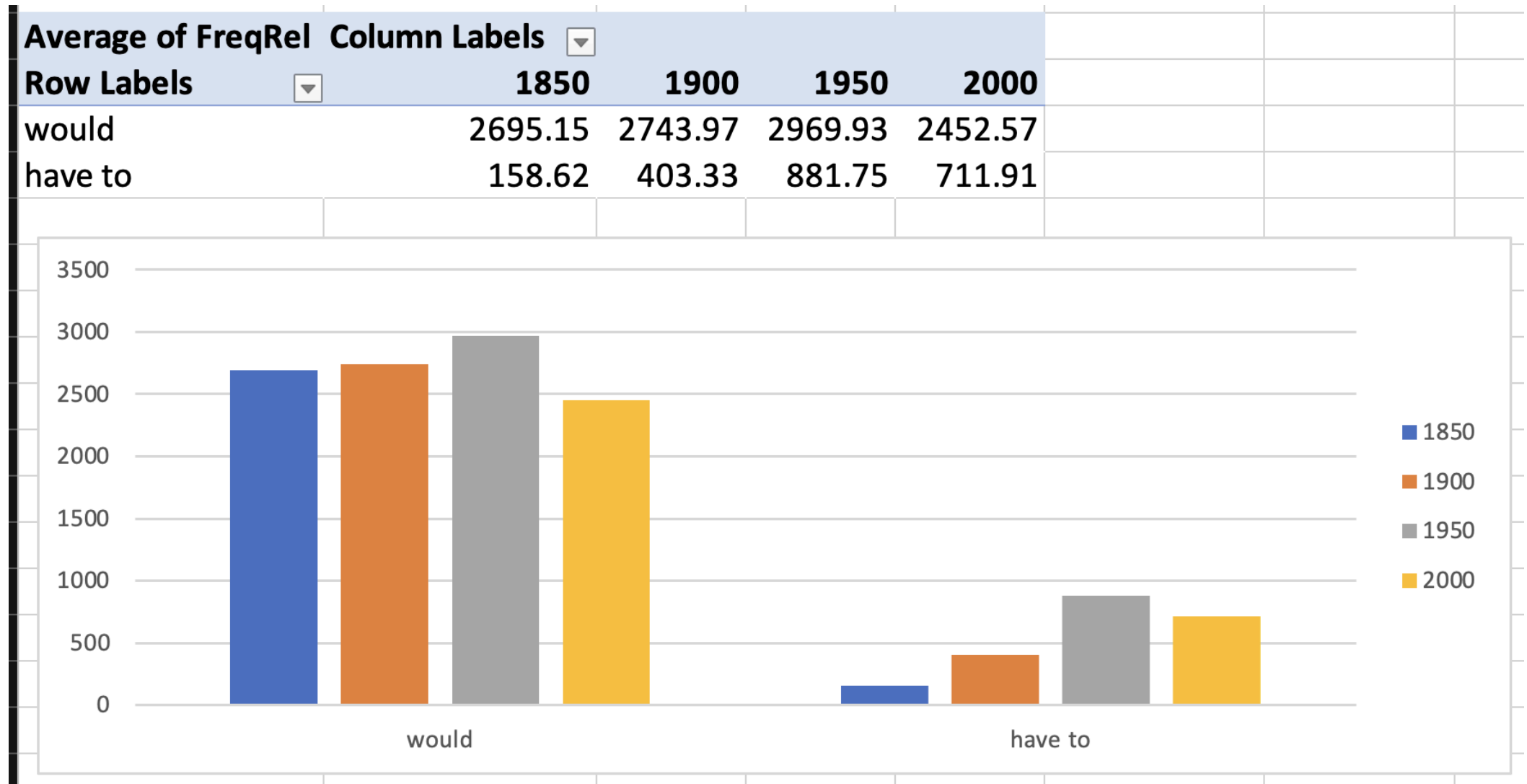
Time intervals: 1850, 1900, 1950, 2000

Lexeme	Type	Period	FreqAbs	FreqRel
would	core	1850	44567	2,695.15
would	core	1900	60305	2,743.97
would	core	1950	85122	2,969.93
would	core	2000	85403	2,452.57
have to	peripheral	1850	2623	158.62
have to	peripheral	1900	8864	403.33
have to	peripheral	1950	25272	881.75
have to	peripheral	2000	24790	711.91

Time interval data format.

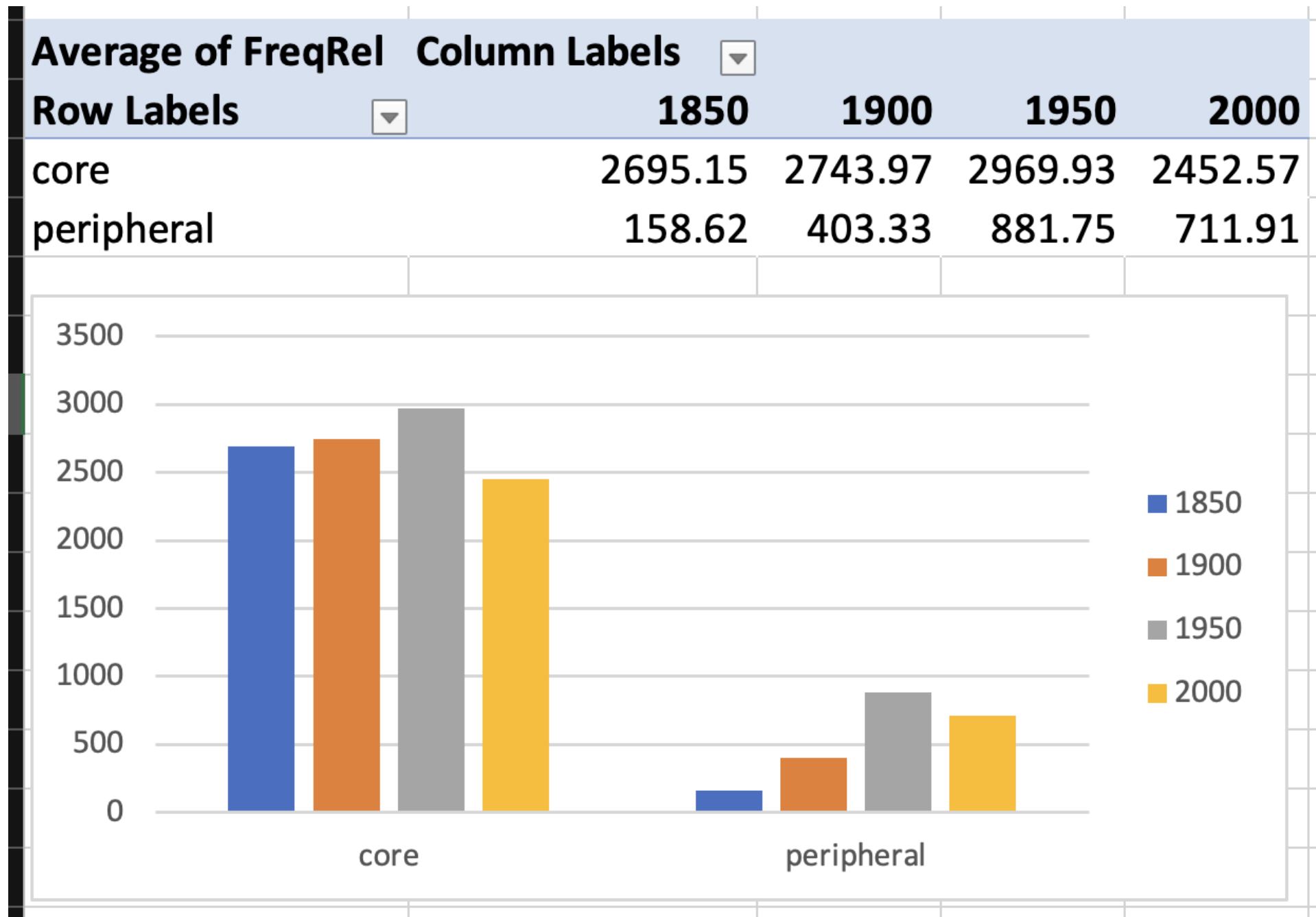
Practice: Frequency analysis

Individual modal analysis



Individual modal verb frequency changes over time.

Group analysis



Core vs peripheral modal verb frequency changes on aggregate.

Practice: Text type specificity

Coefficient of Variation (CV)

Definition: Statistical measure of relative variability.

$$CV = \frac{\sigma}{\mu} \times 100$$
$$= \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$$

Application: Measures variability of word frequencies across different text types.

Lexeme ▾	Type ▾	BLOG ▾	WEB ▾	TV/M ▾	SPOK ▾	FIC ▾	MAG ▾	NEWS ▾	ACAD ▾
would	core	1,056.31	1,000.86	594.31	1,076.83	1,215.72	806.82	928.19	696.04
got to	periphery	37.76	35.65	317.79	239.39	71.91	33.67	52.74	5.03

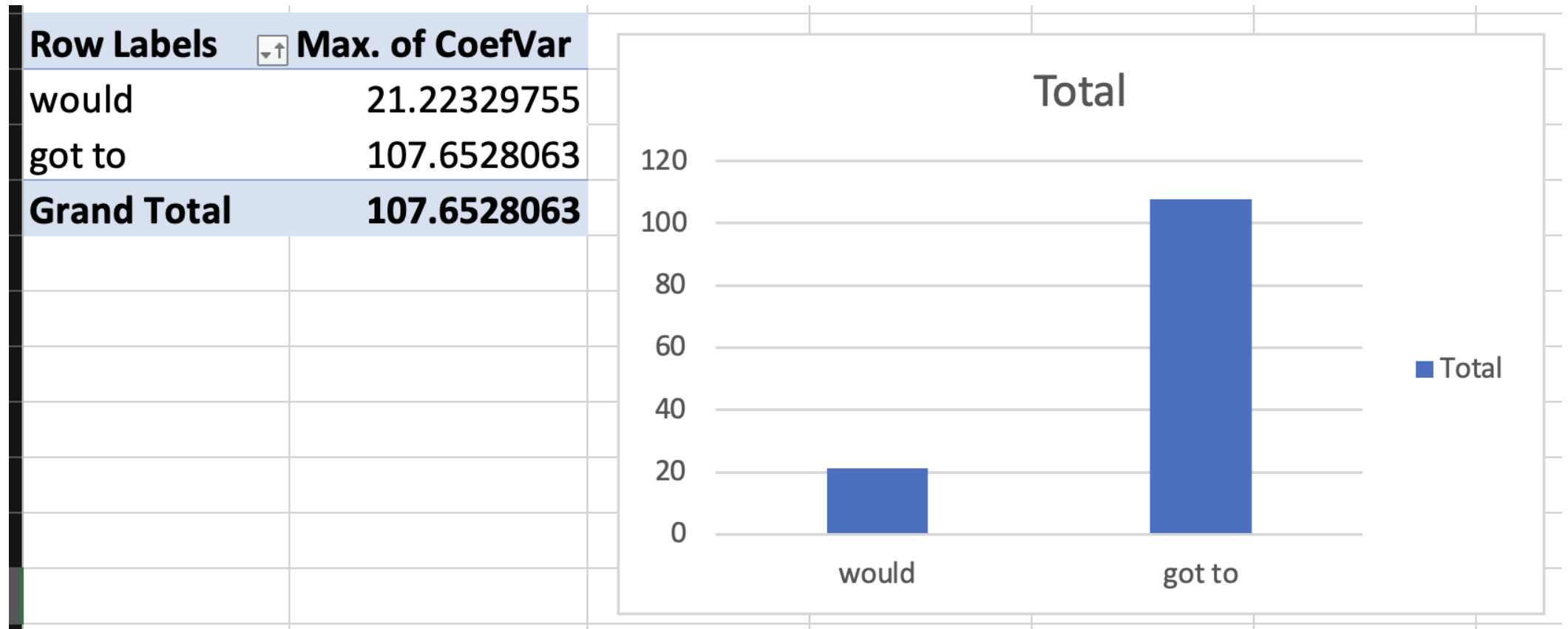
Gathering text type data for analysis.

Excel calculation:

1. Calculate mean: `=AVERAGE(A1:A10)`
2. Calculate standard deviation: `=STDEV.S(A1:A10)`
3. Calculate CV: `=(B2/B1)*100`

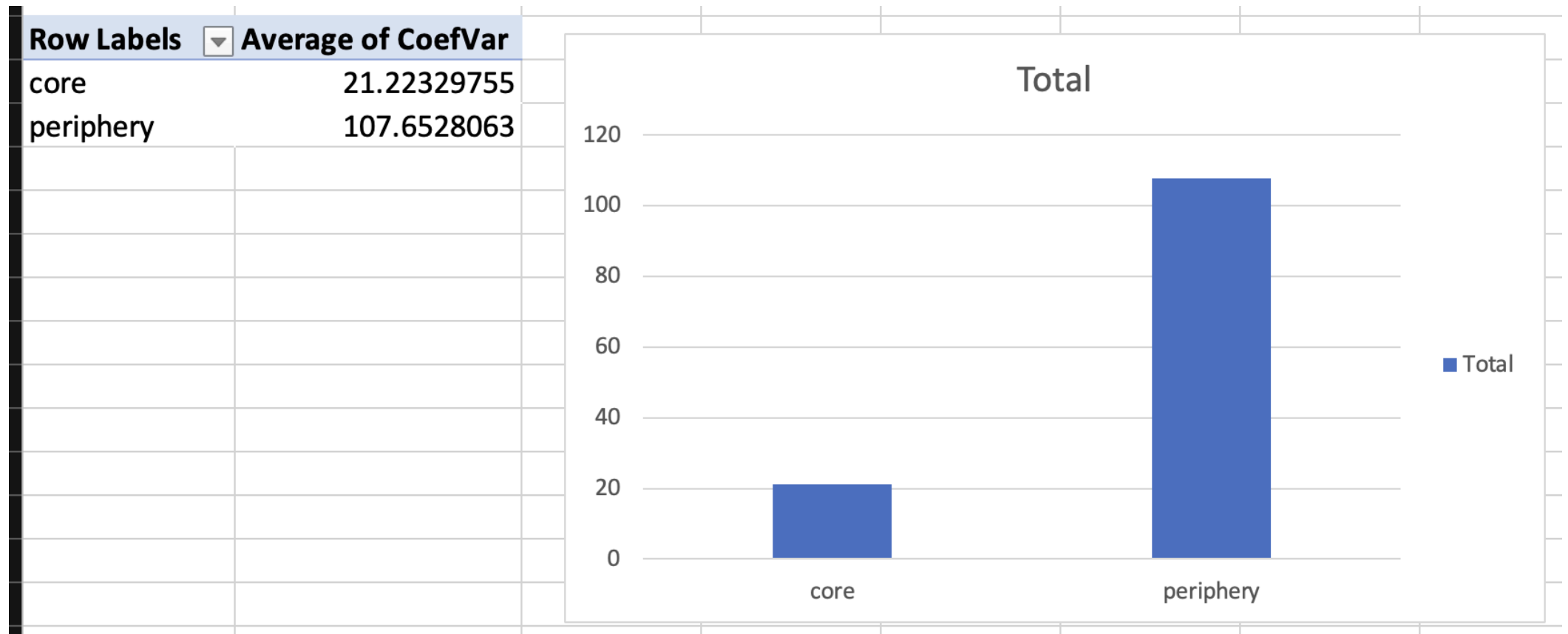
Practice: Text type results

Individual modal differences



Coefficient of Variation for individual modal verbs.

Group differences



Coefficient of Variation for core vs peripheral modal groups.

Key finding: Core modals show higher text type specificity (higher CV) than peripheral modals.

Summary

Key takeaways

General:

- **Variation is fundamental:** Lexical variation occurs across multiple dimensions
- **Speaker variation:** Regional (dialect) and social (sociolect) differences
- **Situational variation:** Register differences based on context and use

Modal verbs:

- **Diachronic patterns:** Modal verbs show systematic frequency changes over time
- **Text type specificity:** Different modals show varying degrees of genre preference
- **Corpus methods:** Quantitative analysis reveals patterns not visible through intuition alone

References

- Grieve, Jack, Andrea Nini, and Diansheng Guo. 2018. "Mapping Lexical Innovation on American Social Media." *Journal of English Linguistics* 46 (4): 293–319.
- Hilpert, Martin, and Christian Mair. 2015. "Grammatical Change." In *The Cambridge Handbook of English Corpus Linguistics*, edited by Douglas Biber and Randi Reppen, 180–200. Cambridge: Cambridge University Press.
- Kortmann, Bernd. 2020. *English Linguistics: Essentials*. J.B. Metzler. <https://books.google.com?id=0Bd4zQEACAAJ>.
- Lipka, Leonhard. 1992. *An Outline of English Lexicology*. Forschung Und Studium Anglistik. Tübingen: Niemeyer.