




A macroscopic re-examination of language and gender

A corpus-based case study in university instructor discourses

Presenter: Zhengxiang (Jack) Wang
Committee: Zhi Li, Veronika Makarova,
Julita Vassileva, David Parkinson

Table of contents

- Research Background
- Research Questions
- Corpus of instructor discourses
- Linguistic features under analysis
- Data analysis
- Results
- Interpretations
- Conclusions
- References

Research Background

- Gender difference in language use:
 - a) a key theme in field of language and gender studies (Motschenbacher, 2012)
 - b) Empirically: inconclusive and inconsistent (e.g., Leaper & Ayres, 2007; Newman et al., 2008)

- Methodological issues (e.g., Crawford, 1995; Jacklin, 1981; Wallentin, 2009):
 - 1) ignorance of effects of other non-gender factors;
 - 2) overgeneralization from narrow database;
 - 3) based in mixed contexts;
 - 4) limited number of linguistic features analyzed.

- Gender differences in instructor discourses:
 - Largely understudied (Howe, 1997; Litosseliti, 2006; Sabbe & Aeltermann, 2007)
 - A very specific social context

Research Questions

- **About the study**
 - Corpus-based case study: university instructor discourses across 4 disciplines
 - Macroscopic: 87 linguistic features (e.g., parts of speech) analyzed

- **Research questions:**
 - RQ1: How do male and female university instructors in the corpus as a whole differ in their use of these 87 linguistic features?

 - RQ2: How do male and female university instructors in the corpus differ in their use of these 87 linguistic features within each academic discipline?

Corpus of instructor discourses

- British Academic Spoken English (BASE) corpus was used to compile a corpus of university instructor discourses (Nesi & Thompson, 2006; Thompson & Nesi, 2001).
- 160 lectures & 4 academic disciplines: Arts and Humanities (AH), Life and Medical Sciences (LS), Physical Sciences (PS) and Social Studies and Sciences (SS).
- The construction of corpus of instructor discourses (Figure 1).

Corpus of instructor discourses

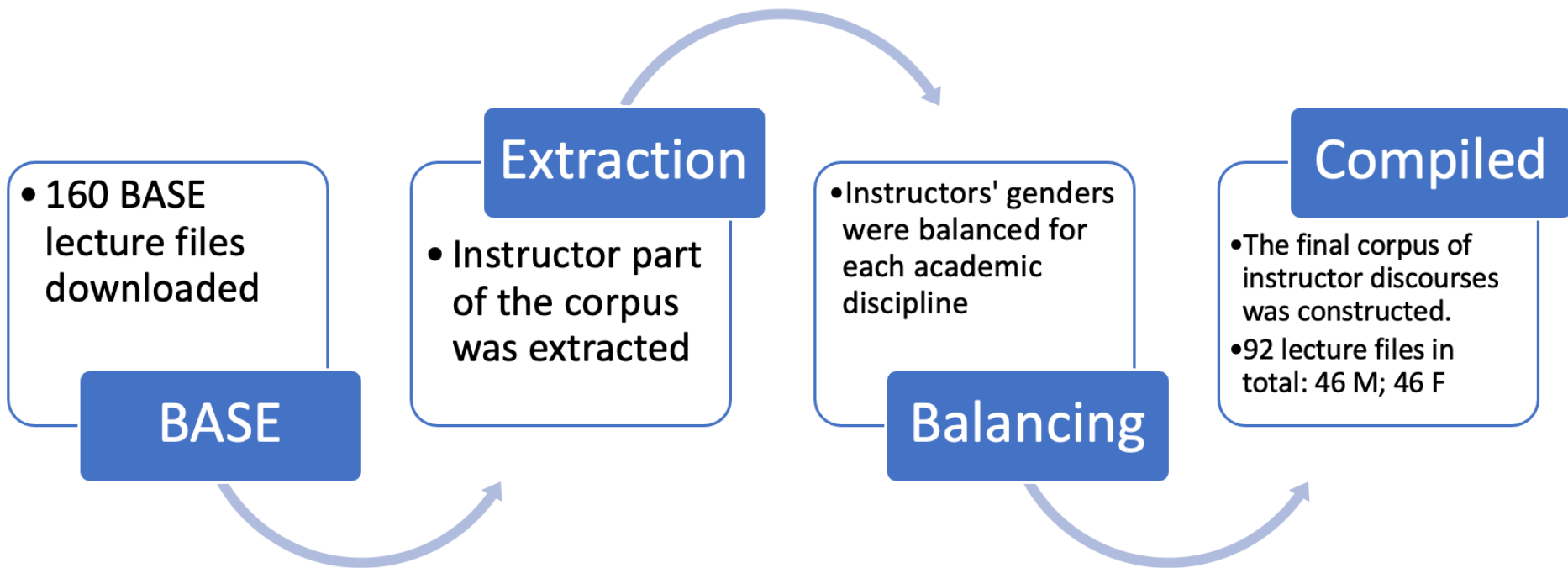


Figure 1. The construction of corpus of instructor discourses

Corpus of instructor discourses

- British Academic Spoken English (BASE) corpus was used to compile a corpus of university instructor discourses (Nesi & Thompson, 2006; Thompson & Nesi, 2001).
- 160 lectures & 4 academic disciplines: Arts and Humanities (AH), Life and Medical Sciences (LS), Physical Sciences (PS) and Social Studies and Sciences (SS).
- The construction of corpus of instructor discourses (Figure 1).
- Basic statistics (Table 1)

Corpus of instructor discourses

Table 1

Basic statistics of the compiled corpus of instructor discourses.

| Academic Discipline | Male Instructor | | Female Instructor | | # of text tokens (raw text) |
|---------------------|--------------------|----------------|--------------------|----------------|-----------------------------|
| | # of lecture files | Average tokens | # of lecture files | Average tokens | |
| AH | 13 | 6935.6 | 13 | 7055.7 | 181887 |
| LS | 13 | 5502.7 | 13 | 4578.7 | 131058 |
| PS | 6 | 5372.5 | 6 | 4993.2 | 62194 |
| SS | 14 | 7682.5 | 14 | 7831.2 | 217192 |
| All | 46 | 6554.1 | 46 | 6322.7 | 592331 |

Linguistic features under analysis

- 98 general linguistic features informed by two sources:
 - a) 73 linguistic features summarized from 35 independent quantitative studies in language & gender;
 - b) 129 linguistic features, as in Biber et al. (2004) and Biber (2006), designed to study university language and textual variations for general purposes (see discussions in Friginal, 2013).

- Feature selection process (Figure 2)

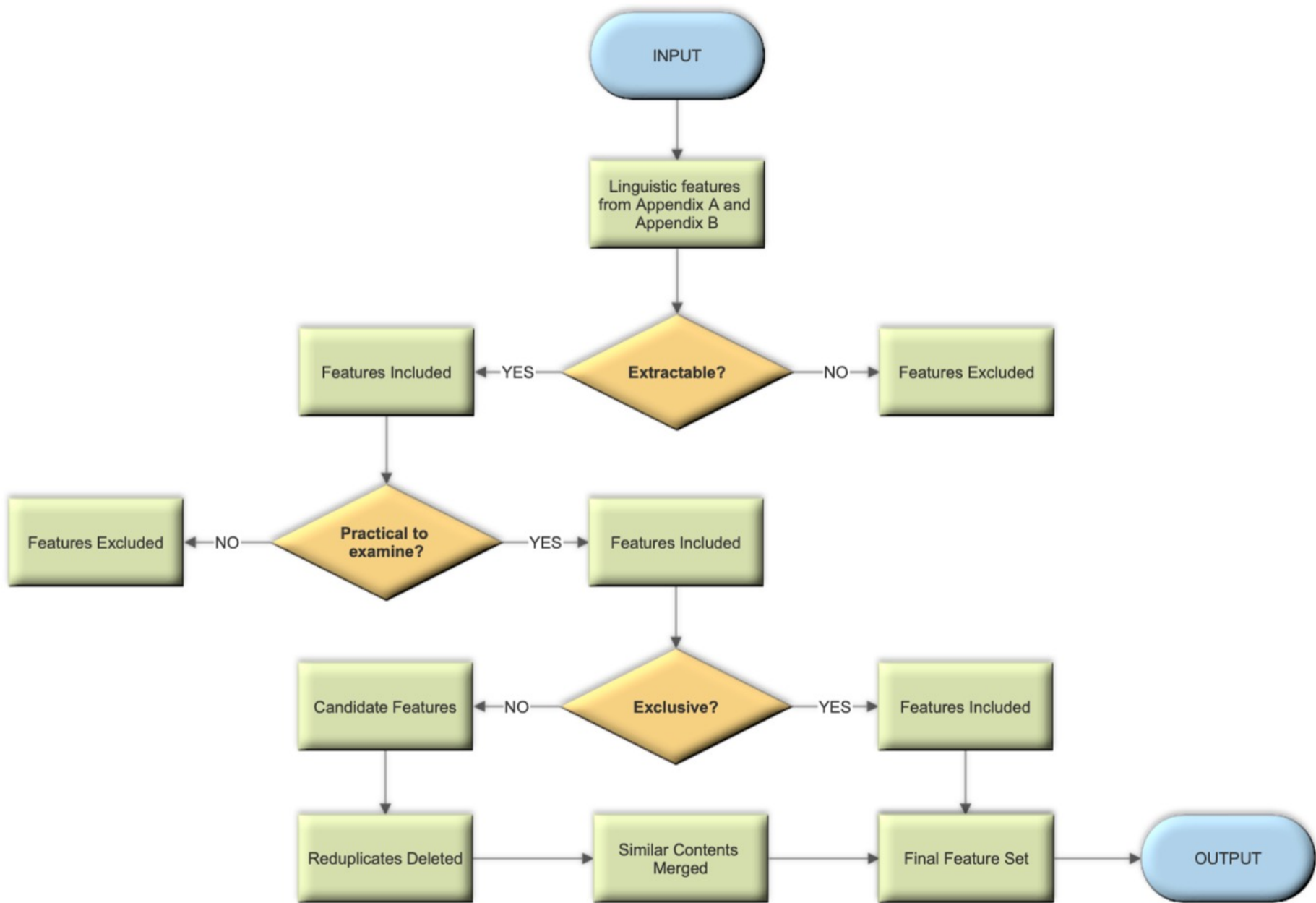


Figure 2. Process of feature selection

Linguistic features under analysis

Table 2
Linguistic features under analysis

| Structural Features | |
|--|--|
| words per utterance (WPU); utterances (UTT); 6-letter words and above (SLW); mean word length (WML)*; type-token ratio (TTR)* | |
| Conversational Features | |
| overlap (OLP); contraction (CONT)* | |
| Sentential Features | |
| Passive voice | agentless passive (AGPA)*; by passive (BYPA)* |
| Tense | past tense (PAST)*; perfect aspect (PEAS)*; non-past tense (NONP); progressive tense (VING) ^a |
| Split structure | split auxiliary (SPAU)* |
| Coordination | phrasal coordination (PHCO)*; independent clause coordination (CLCO)* |
| WH structure | WH question (WHQ)*; WH clause (WHC)* |
| Nominal postmodifying clause | That relative (THRA); WH relative on subject position (WHRS)*; WH relative on object position (WHRO)*; WH relative with fronted preposition (WHRFP)*; past participial postnominal clause (PPPCL)* |
| “To” clause preceded by | speech act verb (TSAV); cognition verb (TCOV); desire/intent/decision verb (TDIDV); modality/cause/effort verb (TMCEV); probability/simple fact verb (TPSFV); certainty adjective (TCAJ); ability/willingness adjective (TAWAJ); personal affect adjective (TPAAJ); ease/difficulty adjective (TEDAJ); evaluative adjective (TEVAJ); control noun (TCNO) |
| “That” clause preceded by | nonfactive noun (THNFN); attitudinal noun (THATN); factive noun (THFAN); likelihood noun (THLKN); nonfactive verb (THNFV); attitudinal verb (THATV); factive verb (THFV); likelihood verb (THLKV); likelihood adjective (THLKAJ); attitudinal adjective (THATAJ) |

Linguistic features under analysis

| Lexical Features | |
|----------------------------|--|
| Part of speech | nouns (NOUN); verb (VERB); noun modifier (NMOD); article (ART); modal (MD); negator (NEG)*; preposition (PREP)* |
| Pronoun | first person pronoun singular (FPPS) ^b ; first person pronoun plural (FPPP)*; second person pronoun (SPP)*; third person pronoun (TPP)*; pronoun it (PIT)*; demonstrative pronoun (DEMP)*; indefinite pronoun (INDP)* |
| Noun sub-categories | nominalization (NOMZ)*; animate noun (ANMN); cognitive noun (COGN); concrete noun (CONN); technical noun (TCHN); quantity noun (QUAN); place noun (PLAN); group/institution noun (GIN); abstract/process noun (APN) |
| Verb sub-categories | “be” as main verb (BEMV)*; pro-verb <u>do</u> (PROD)*; activity verb (ACTV); communication verb (COMV); mental verb (MENV); causative verb (CAUV); occurrence verb (OCCV); existence verb (EXV); aspectual verb (ASPV) |
| Adjective sub-categories | attributive adjective (ATTAJ)*; predictive adjective (PREAJ)* |
| Adverb sub-categories | place adverb (PLAAD)*; time adverb (TMAD)*; nonfactive adverb (NFAD); attitudinal adverb (ATAD); factive adverb (FAD); likelihood adverb (LKAD) |
| Conjunction subcategories | causative adverbial subordinator (CAUSA)*; conditional adverbial subordinator (CONDA)*; contrastive adverbial subordinator (CONCA) ^c ; <u>other</u> adverbial subordinator (OTHA)* |
| Modal subcategories | possibility modal (POSMD)*; necessity modal (NECMD)*; predictive modal (PRMD)* |
| Stance-related expressions | conjunct (CNJT)*; downtoner (DTN)*; amplifier (AMP)*; hedge (HEDG)*; emphatic (EMP)*; polite expression (PLEP); general evidential expression (EVIEP) |

Linguistic features under analysis

- 98 general linguistic features informed by two sources:
 - a) 73 linguistic features summarized from 35 independent quantitative studies in language & gender;
 - b) 129 linguistic features, as in Biber et al. (2004) and Biber (2006), designed to study university language and textual variations for general purposes (see discussions in Friginal, 2013).

- Feature selection process (Figure 2)

- Linguistic Feature Extractor (LFE): a Python program created to automate the feature extraction. See: https://github.com/jaaack-wang/ling_feature_extractor/tree/Thesis_Project_Version.

- Except words per utterance, mean word length and type-token ratio, the remaining features were normalized at 100 words

Data analysis

- **Data preprocessing:** 87 linguistic features retained
 - Features of extremely low frequency (defined as median = 0) were removed;
 - Features of high correlation (Pearson's $r > 0.8$) were reduced

- **The nature of the dataset** (Shapiro-Wilk tests): largely non-normal distributions

- **Measurements of gender differences:**
 - Nonparametric hypothesis tests: Mann-Whitney U test; Kruskal Wallis H test
 - Multiple testing \rightarrow False Discovery Rate (FDR): $\alpha < 0.05$
 - Effect size: Cohen's d (thresholds in absolute value: 0.2, small; 0.5, medium; 0.8 large)

- **Statistical procedures:** Three rounds of statistical analyses
 - a) Gender differences in the entire corpus
 - b) Overall effects of academic discipline
 - c) Gender differences within specific academic disciplines

Results

- **Overall gender effects:** only three features showing significant differences (Mann-Whitney U tests)

Table 3

| Linguistic Feature | Male $\mu \pm \sigma$ | Female $\mu \pm \sigma$ | Adjusted p-value | <i>d</i> | Effect Size |
|--------------------|-----------------------|-------------------------|------------------|----------|-------------|
| WHRO | 0.12 \pm 0.07 | 0.07 \pm 0.07 | 0.028 | 0.583 | Medium, M+ |
| TDIDV | 0.21 \pm 0.12 | 0.30 \pm 0.13 | 0.028 | -0.714 | Medium, F+ |
| MENV | 2.18 \pm 0.84 | 2.62 \pm 0.65 | 0.031 | -0.583 | |

- **Overall discipline effects** (Kruskal Wallis H tests): 34 significant comparisons (p values adjusted by FDR)
- **Specific gender effects within each specific academic discipline:** None

Interpretations: WHRO

- **WHRO:** WH relative clauses on object position, M+
- **Biber (1988):** explicit and elaborated identification of referents in a text
- **Example:**
 - I get minus-R E-to-the-minus-R-T here and then the S **which** I'm holding constant right just from the product rule... (from pslct015, by nm0765)
- **Observations:**
 - 1) Male instructors consistently used more WHRO than female instructors across disciplines, but the practical difference is very small (Figure 3)
 - 2) Male instructors and female instructors used comparable amounts of relative clauses (Figure 4)

Interpretations: WHRO

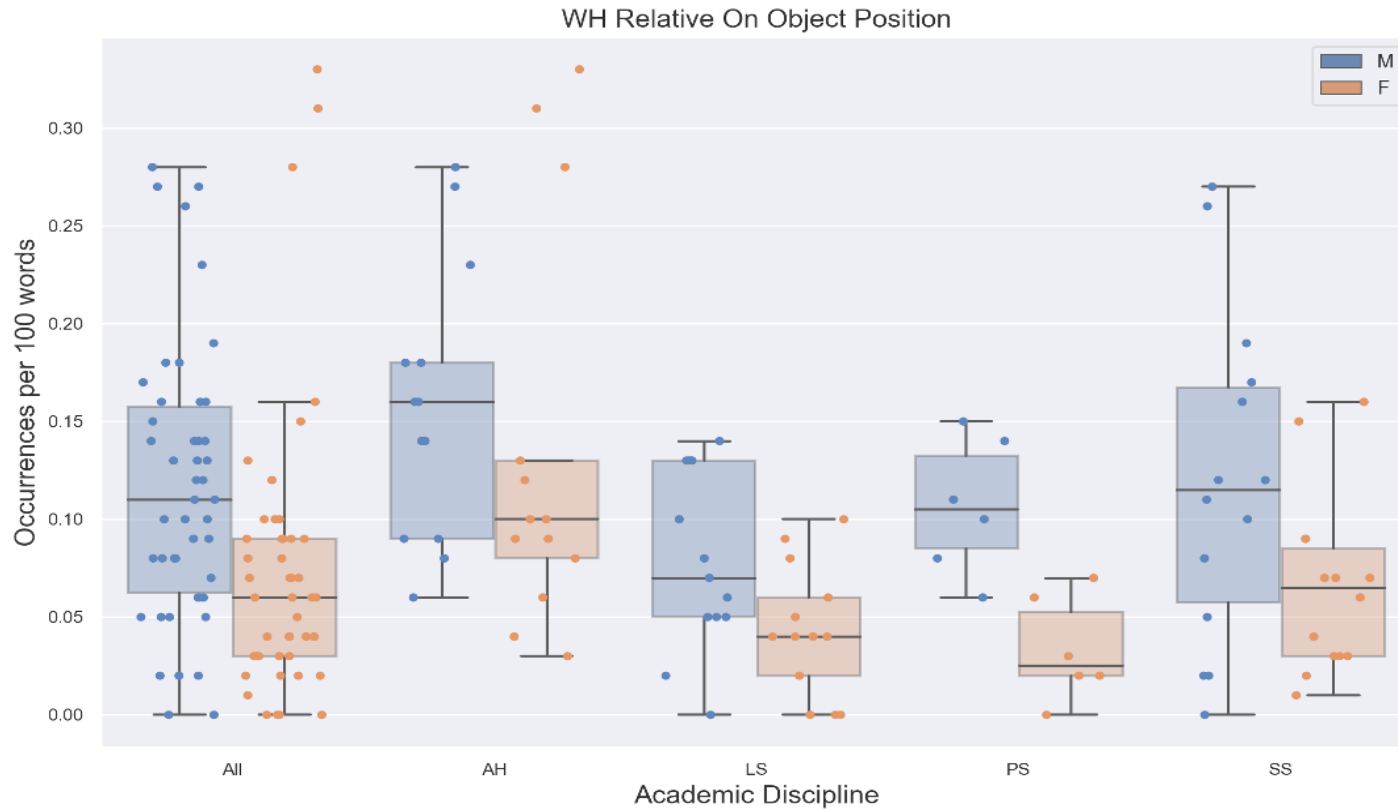


Figure 3. WHRO usage by gender and academic discipline

Interpretations: WHRO

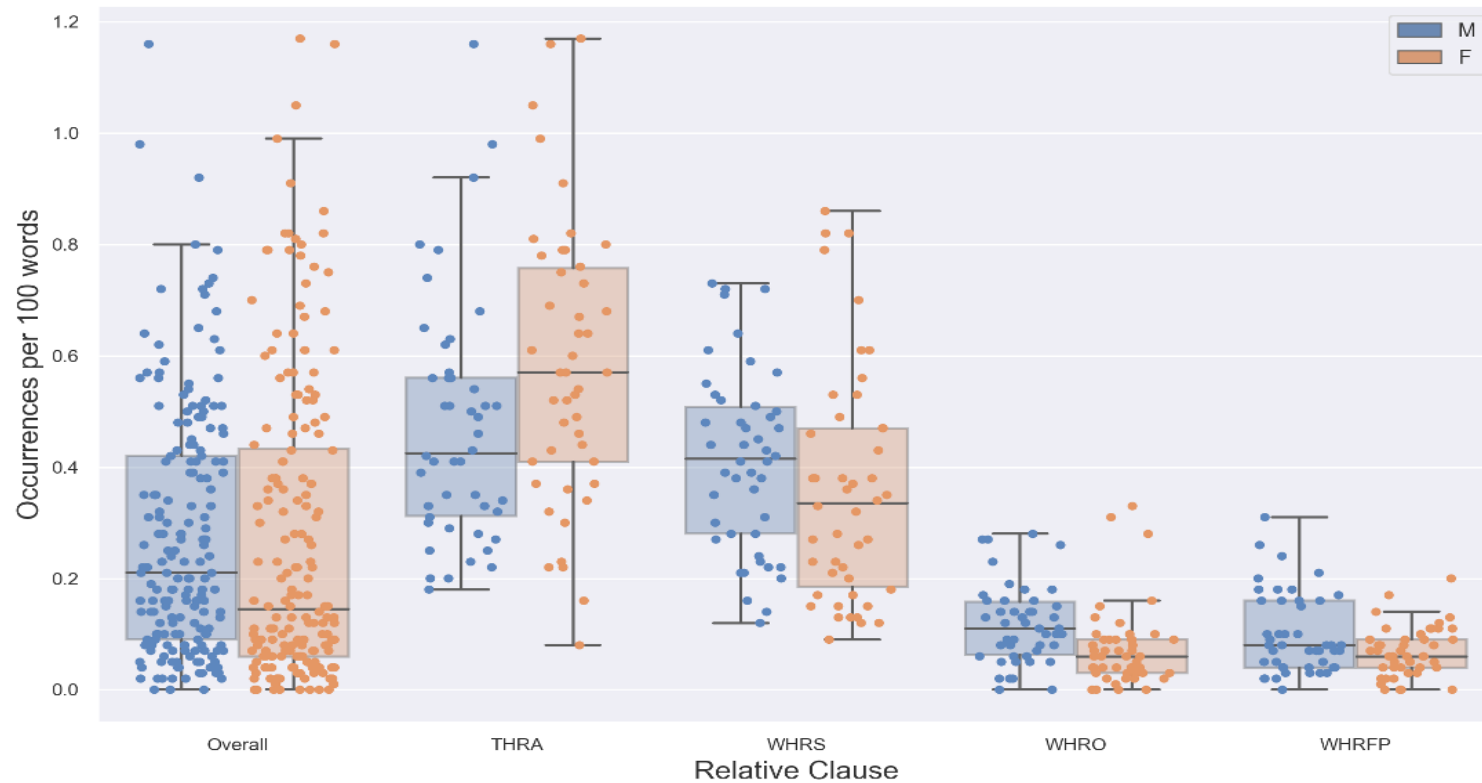


Figure 4. Usage of relative clauses by gender and academic discipline

Interpretations: WHRO

- **WHRO:** WH relative clauses on object position, M+
- **Biber (1988):** explicit and elaborated identification of referents in a text
- **Example:**
 - I get minus-R E-to-the-minus-R-T here and then the S **which** I'm holding constant right just from the product rule... (from pslct015, by nm0765)

- **Observations:**
 - 1) Male instructors consistently used more WHRO than female instructors across disciplines, but the practical difference is very small (Figure 3)
 - 2) Male instructors and female instructors used comparable amounts of relative clauses (Figure 4)

- **Interpretations:**
 - Low frequency of WHRO → formality (WHRFP, WH relative clauses with fronted preposition)
 - Small practical difference → overinterpretation should be discouraged

Interpretations: TDIDV

- **TDIDV:** desire/intent/decision verbs followed by a “to” clause, F+
- **Biber (1988):** makes a discourse more procedural (versus content-focused)
- **Examples:**
 - I get that's all **I want to** say about this sort of level of housekeeping... (from ahlct015, by nm0067)
 - if your solute concentration goes up **you will need to** pass more urine... (from lsct029, by nf440)
 - image now the question is that if **we want to** see both the irradiance... (from pslct034, by nf0934)
- **Observations:**
 - 1) Female instructors consistently used more TDIDV than male instructors across disciplines (Figure 5), especially inclusive pronouns *we* and *you* (Table 4)
 - 2) The basic pattern of TDIDV usage is similar between female and male instructors (Table 4)

Interpretations: TDIDV

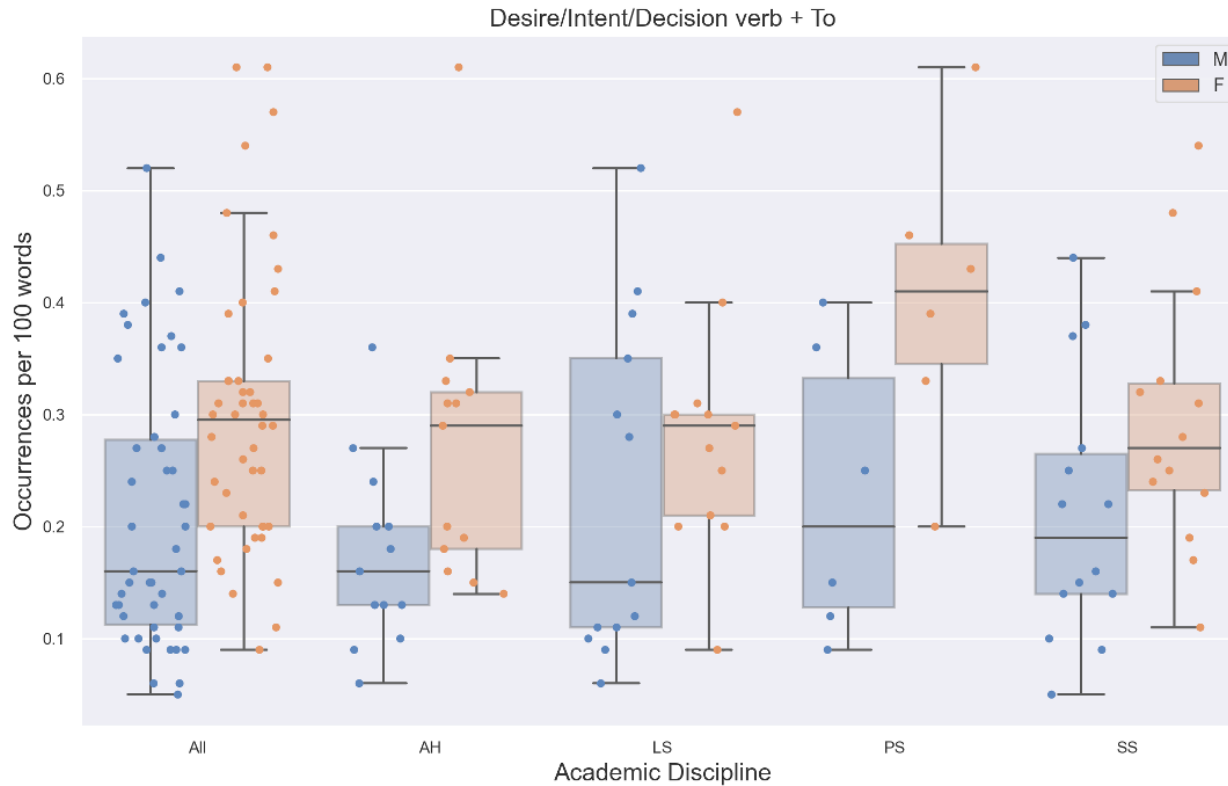


Figure 5. TDIDV usage by gender and academic discipline

Interpretations: TDIDV

Table 4

Top 10 lexical contexts (lemmatized) for TDIDV between male and female instructor discourses.

| Male instructor | | | | | | |
|--------------------------|-----------------------------|------------|----------------|------------|------------------------------------|------------|
| Index | Unigram: left1 ^a | Occurrence | Unigram: left2 | Occurrence | Bigram: Left1 + TDIDV ^b | Occurrence |
| 1 | you | 110 | I | 146 | I want to | 84 |
| 2 | I | 98 | you | 145 | you want to | 60 |
| 3 | not | 65 | do | 76 | not want to | 37 |
| 4 | we | 59 | not | 73 | you need to | 36 |
| 5 | would | 45 | we | 68 | would like to | 34 |
| 6 | they | 43 | that | 61 | we need to | 32 |
| 7 | be | 21 | would | 55 | we want to | 27 |
| 8 | just | 16 | if | 50 | they want to | 26 |
| 9 | to | 14 | they | 50 | just want to | 16 |
| 10 | who | 13 | what | 39 | not need to | 13 |
| Female instructor | | | | | | |
| Index | Unigram: left1 | Occurrence | Unigram: left2 | Occurrence | Bigram: Left1 + TDIDV | Occurrence |
| 1 | you | 150 | you | 212 | I want to | 93 |
| 2 | we | 113 | I | 168 | you want to | 85 |
| 3 | I | 109 | we | 136 | we need to | 58 |
| 4 | not | 60 | do | 95 | you need to | 54 |
| 5 | would | 53 | not | 67 | we want to | 47 |
| 6 | they | 51 | that | 61 | would like to | 44 |
| 7 | he | 28 | would | 61 | not want to | 39 |
| 8 | might | 22 | they | 59 | they want to | 37 |
| 9 | just | 22 | if | 46 | he want to ^c | 24 |
| 10 | be | 20 | what | 40 | just want to | 19 |

Interpretations: TDIDV

- **TDIDV:** desire/intent/decision verbs followed by a “to” clause, F+
- **Biber (1988):** makes a discourse more procedural (versus content-focused)
- **Examples:**
 - I get that's all **I want to** say about this sort of level of housekeeping... (from ahlct015, by nm0067)
 - if your solute concentration goes up **you will need to** pass more urine... (from lsct029, by nf440)
 - image now the question is that if **we want to** see both the irradiance... (from psct034, by nf0934)
- **Observations:**
 - 1) Female instructors consistently used more TDIDV than male instructors across disciplines (Figure 5), especially inclusive pronouns *we* and *you* (Table 4)
 - 2) The basic pattern of TDIDV usage is similar between female and male instructors (Table 4)
- **Interpretations:**
 - More inclusive pronouns → more engaging and interactive discourses
 - Comparable pattern → a common professionalized practice for university teaching

Interpretations: MENV

- **MENV:** Mental verbs, F+
- **Biber (1988) & Precht (2008)** : convey personal stance or uncertainty
- **Examples:**
 - know, think, see, mean, find, remember, learn and understand

- **Observations:**
 - 1) Female instructors consistently used more MENV than male instructors across disciplines (Figure 6), especially inclusive pronouns *we* and *you* (Table 5)
 - 2) The basic pattern of MENV usage is similar between female and male instructors (Table 5)

Interpretations: MENV

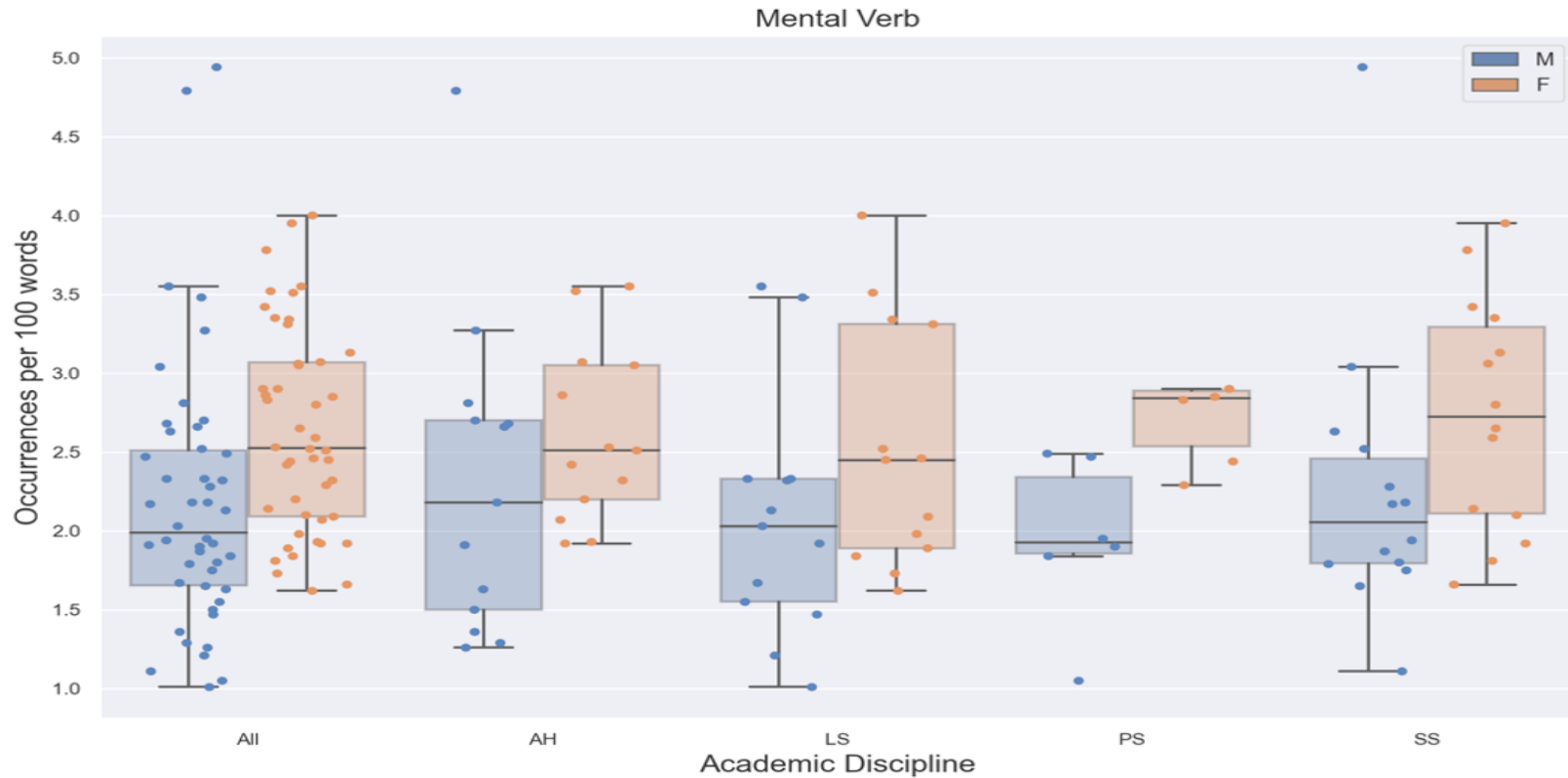


Figure 6. MENV usage by gender and academic discipline

Interpretations: MENV

Table 5

Top 10 lexical contexts (lemmatized) for MENV between male and female instructor discourses.

| Male instructor | | | | | | |
|--------------------------|----------------|------------|----------------|------------|----------------------|------------|
| Index | Unigram: left1 | Occurrence | Unigram: left2 | Occurrence | Bigram: Left1 + MENV | Occurrence |
| 1 | you | 1402 | you | 1861 | you know | 710 |
| 2 | I | 1177 | I | 1313 | I think | 491 |
| 3 | to | 626 | to | 689 | I mean | 436 |
| 4 | not | 413 | be | 503 | you see | 144 |
| 5 | be | 318 | we | 496 | not know | 143 |
| 6 | we | 299 | not | 489 | can see | 142 |
| 7 | can | 230 | would | 477 | I want | 134 |
| 8 | will | 213 | that | 326 | you want | 126 |
| 9 | would | 170 | have | 294 | you think | 108 |
| 10 | have | 136 | can | 291 | to think | 108 |
| Female instructor | | | | | | |
| Index | Unigram: left1 | Occurrence | Unigram: left2 | Occurrence | Bigram: Left1 + MENV | Occurrence |
| 1 | you | 1616 | you | 2266 | you know | 841 |
| 2 | I | 1215 | I | 1393 | I think | 546 |
| 3 | to | 713 | to | 865 | I mean | 428 |
| 4 | not | 521 | we | 811 | can see | 231 |
| 5 | we | 521 | would | 556 | not know | 224 |
| 6 | be | 365 | not | 534 | I want | 174 |
| 7 | can | 266 | be | 513 | you want | 174 |
| 8 | have | 256 | and | 427 | you need | 168 |
| 9 | would | 224 | have | 408 | you think | 152 |
| 10 | will | 204 | can | 383 | to think | 149 |

Interpretations: MENV

- **MENV:** Mental verbs, F+
- **Biber (1988) & Precht (2008)** : convey personal stance or uncertainty
- **Examples:**
 - know, think, see, mean, find, remember, learn and understand

- **Observations:**
 - 1) Female instructors consistently used more MENV than male instructors across disciplines (Figure 6), especially inclusive pronouns *we* and *you* (Table 5)
 - 2) The basic pattern of MENV usage is similar between female and male instructors (Table 5)

- **Interpretations:**
 - More inclusive pronouns → more approachable and less assertive discourses
 - Comparable pattern → a common professionalized practice for university teaching

Conclusions

- **RQ1: Overall gender differences**
 - ❑ only 3 of 87 features showing significant gender-related differences (medium effect sizes)
 - ❑ the overall gender differences are small and limited

- **RQ2: Discipline-level gender differences**
 - ❑ no discipline-level gender difference identified in the corpus
 - ❑ academic discipline has more effects on instructor discourses than gender

- **Interpretations of gender differences:**

(M+: WH relative clauses on object position; F+: desire/intent/decision verbs followed by a “to” clause & mental verbs)

 - Female instructor discourses are slightly more engaging and less formal than their male counterparts
 - no discipline-level gender differences + comparable patterns of language use → instructor discourses as a highly professionalized practice

Selected references

- Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. John Benjamins Publishing. <http://site.ebrary.com/id/10126066>
- Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43(3), 325–356. <https://doi.org/10.1080/0305764X.2013.786024>
- Leaper, C., & Ayres, M. M. (2007). A Meta-Analytic Review of Gender Variations in Adults' Language Use: Talkativeness, Affiliative Speech, and Assertive Speech. *Personality and Social Psychology Review*, 11(4), 328–363. <https://doi.org/10.1177/1088868307302221>
- Motschenbacher, H. (2012). *An Interdisciplinary Bibliography on Language, Gender and Sexuality (2000-2011)*. John Benjamins Publishing.
- Newman, M., Groom, C., Handelman, L., & Pennebaker, J. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes - DISCOURSE PROCESS*, 45, 211–236. <https://doi.org/10.1080/01638530802073712>
- Precht, K. (2008). Sex similarities and differences in stance in informal American conversation1. *Journal of Sociolinguistics*, 12(1), 89–111. <https://doi.org/10.1111/j.1467-9841.2008.00354.x>
- Thompson, P., & Nesi, H. (2001). The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research*, 5, 263–264. <https://doi.org/10.1177/136216880100500305>

Thank you!