

# A multifactorial analysis of (un)attended *this* in academic writing

Stefanie Wulff, Ph.D.



This is part of a larger joint research project in progress  
with John Swales and Ute Römer (University of Michigan).

# Introduction

The English demonstrative form *this* can function either as a free-standing pronoun as illustrated in (1) or as a determiner attending a head noun phrase as in (2).

- (1) **This** is an example.
- (2) **This sentence** is an example.



**What determines writers' choices of (un)attended this?**

# Previous Research

- (un)attended *this* has largely been seen not as a grammatical phenomenon, but a stylistic, rhetorical, or information-processing-related matter (Geisler 1985, Finn 1995)
- textbooks and style guides mostly adopt a rather prescriptive view on (un)attended *this*, equating unattended *this* with “vague reference” that is to be avoided (Ede 2004, Faigley 2007, Axelrod & Cooper 2008); no discussion of where unattended *this* might be warranted

# Previous Research

- Swales (2005): use of sentence-initial (un)attended *this* in published academic writing (Hyland Corpus)
  - found much higher occurrences of unattended *this* than occasional usage would suggest, specifically in certain disciplines (25% in dentistry, 56% in philosophy articles)
  - tentatively concluded that the absence of a noun phrase is associated with main verbs that are “syntactically and semantically simple” (Swales 2005:13)

# Previous Research

- Römer & Wulff (to appear): case study on (un)attended *this* on a pre-release version of the *Michigan Corpus of Upper Level Student Papers* (MICUSP)
  - *this* is the 11<sup>th</sup> most frequent word
  - average percentage of attended *this* (across sentential positions) is higher (73%) compared to Swales' analysis of sentence-initial *this*
  - with the exception of philosophy, disciplinary variation of (un)attended *this* is rather muted
  - shares of attended *this* steadily increase from final year undergraduate to third year graduate student writing
  - nouns attending *this* are mostly meta-discoursal or methodology-related (study, paper, argument, ...)

# Hypotheses and Goals

## Hypotheses

- sentence-initial (un)attended *this* is (partially) driven by local patterns emerging from the interplay of (un)attended *this* and the accompanying main verb
- the distribution of these local lexical patterns is a function of
  - academic discipline
  - level of academic proficiency
  - native speaker status

## Goals

- large-scale quantitative analysis of sentence-initial *this* to identify probabilistic trends that are difficult to see from a qualitative/monofactorial perspective
- (qualitative, in-depth pattern analysis of local patterns, guided by results of quantitative analyses)

# Data

- pre-release of MICUSP (MICUSP\_Jun09)
    - 2.3 million words (810 A-graded student papers)
    - from 16 disciplines/4 academic divisions
    - written by native and non-native writers
    - ranging from final year undergraduate to third year graduate students
  - retrieval of all instances of *this* with MonoConc Pro 2.2: 15,711 hits
  - manual cleaning of hits: 5 instances in which *this* is not a demonstrative determiner or pronoun
  - (semi-)manual coding of data according to sentential position of *this*
- ⇒ 5,827 hits of sentence-initial *this*: 2,499 (43%) unattended, 3,328 (57%) attended

# Analysis I: Binary Logistic Regression\*

- a binary logistic regression is used to **determine the probability of attended vs. unattended *this*** on the basis of several predictor variables:
  - **LOGFREQVERB**: the lemma frequency of the main verbs, obtained from more balanced reference corpus (ICE-GB)
  - **DIVISION**: Biology & Health Sciences, Humanities & Arts, Physical Sciences, or Social Sciences
  - **LEVEL**: final-year undergraduate, first-year graduate, second-year graduate, or third-year graduate
  - **GENDER**: female or male
  - **NATIVENESS**: native speaker or non-native speaker

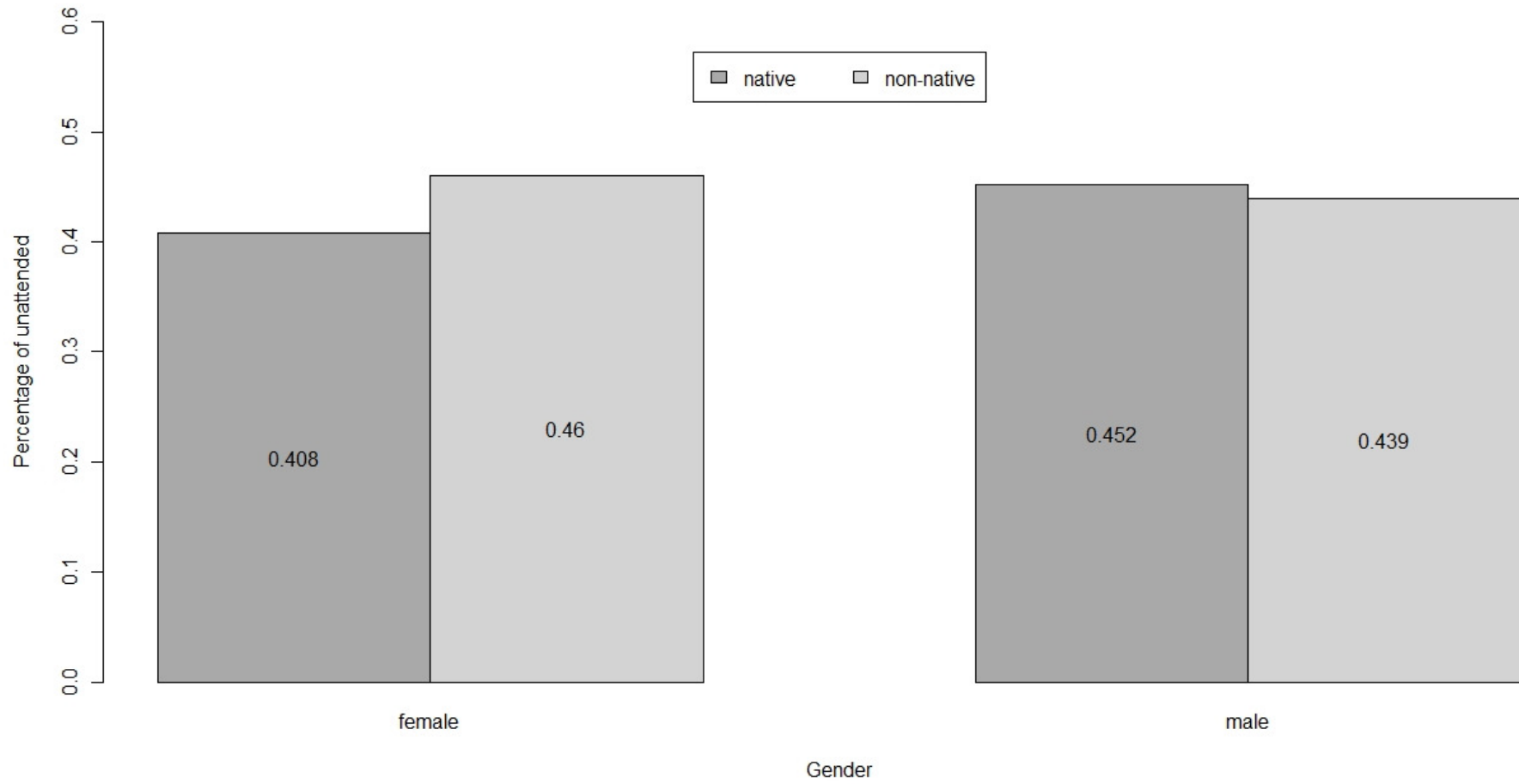
\*Thanks to Stefan Th. Gries for his advice on the statistics 😊



PREDICTOR/INTERACTION	Coeff.	p
LOGFREQVERB	-2.302	.000
DIVISION=artshumanities*LEVEL=firstyeargrad	.922	.000
DIVISION=physicalsciences*LEVEL=thirdyeargrad	-.793	.007
DIVISION=biohealthsciences*LEVEL=firstyeargrad	.584	.002
DIVISION=artshumanities*LEVEL=thirdyeargrad	-.503	.045
LEVEL=firstyeargrad	-.492	.000
DIVISION=biohealthsciences*GENDER=male	-.430	.011
GENDER=male*NATIVENESS=nonnative	-.370	.014
GENDER=male	.340	.001
LEVEL=secyeargrad	-.269	.043
NATIVENESS=nonnative	.256	.007

[log-likelihood ratio  $\chi^2 = 593.39$ ;  $df = 22$ ;  $p = 0$ ; Nagelkerke's  $R^2 = .131$ ;  $C = .684$ ,  $D_{xy} = .368$ ; predictive classification accuracy: 65%]

Interaction Gender x Nativeness



# Interim Summary

- strongest predictor by far: LOGFREQVERB
    - high lemma frequency increases likelihood of attended *this*
  - as academic proficiency increases, so does bias towards attended *this*
    - strongest shift from unattended to attended *this* in Humanities and Biology & Health Sciences
  - non-native speakers generally prefer unattended *this*
  - male students generally prefer unattended *this*
  - male non-native students, particularly in the Biology & Health Sciences, prefer attended *this*
- ⇒ first evidence for a strong lexical drive that is moderated to some extent by disciplinary demands, gender, and native speaker status

# Analysis II: Distinctive Collexeme Analysis

- a distinctive collexeme analysis (DCA; Stefanowitsch and Gries 2003; Gries and Stefanowitsch 2004) helps us to identify the **specific verbs** responsible for the strong LOGFREQVERB effect
- DCA measures the **distinctive association** between a given verb and attended and unattended *this*, respectively

## Top distinctive collexemes of unattended sentence-initial *this*

Verb	FYE <sub>log</sub>
be	103.631
mean	23.027
do	6.055
lead	5.928
result	3.495
happen	2.412
attribute	2.289
leave	1.944
imply	1.835
seem	1.694
accomplish	1.599
fall	1.562
measure	1.472

## Top distinctive collexemes of attended sentence-initial *this*

Verb	FYE <sub>log</sub>
use	8.697
examine	8.275
focus	7.321
find	6.586
explore	4.631
base	4.386
seek	3.898
provide	3.446
contain	3.410
investigate	3.410
have	3.390
discuss	3.218
aim	3.166

# Interim Summary

- DCA confirms strong verb-specific associations between specific verbs and (un)attended *this*
- unattended *this*:
  - high-frequency, semantically bleached verbs (*be, mean, do*) working against the general trend of high frequency verbs to favor attended *this*
  - distinctive collexemes signal upcoming commentary on or discussion of some previously described process or result
- attended *this*:
  - attracts a wider range of verb types, but not as significantly as the “heavy hitters” in unattended *this*
  - distinctive collexemes initiate the description or the structural outline of a paper, study, or methods used

# Outlook: Analysis III Pattern Analysis

- John and Ute have been looking at **discipline-, level, and text-position-specific patterns** around the most distinctive *this*+verb combinations identified earlier
- some highlights:
  - **specific patterns exhibit discipline-specific biases:** e.g. *this is* is particularly prominent in Philosophy essays in contexts of deeper explanations or reasoning
  - **unattended *this*-patterns prefer text- and paragraph-initial positions, while attended *this*-patterns occur more in text- and paragraph-final positions** (accords with the semantic distinctions observable in the DCA)



# Outlook: Future Research Agenda

- expanded data sample of (un)attended *this* across
  - different sentential positions
  - all demonstrative forms (*this, that, these, those*)
- impact of morphology as predictor (we observed some biases towards tense and passivization)

# Conclusions

- **moderate influence of academic discipline and academic proficiency**, most obvious with **specific patterns** (rather than (un)attended *this* distributions in general)
  - strong evidence for **local lexical patterns** of unattended *this*:  
*this is, this means, this implies*
    - override the general bias towards attended *this* in academic writing
    - most strongly associated with contexts of explanation, interpretation, and reasoning
    - prefer text- and paragraph-final positions
- ⇒ indicative of **delexicalization** of *this is* into a “interpretation-marker” (not just “vague reference”!)

# Thank You!

If you're interested in the slides of this presentation, or would like to read the pre-final draft of our joint research paper, please be in touch:

**Stefanie.Wulff@unt.edu**

# References

- Axelrod, R. & Cooper C. (2008). *The St Martin's guide to writing* (8<sup>th</sup> ed.). Boston: Bedford/St Martin's.
- Ede, L. (2004). *Work in Progress* (6<sup>th</sup> ed.). Boston: Bedford/St. Martins.
- Faigley, L. (2007). *Writing: A guide for college and beyond*. New York: Pearson/Longman.
- Finn, S. (1995). Measuring effective writing: Cloze procedure and anaphoric “this”. *Written Communication*, 12: 240-266.
- Geisler, C., Kaufer, D. S. & Steinberg, E. R. (1985). The unattended anaphoric “this”: When should writers use it? *Written Communication*, 2: 129-155.
- Römer, U. & Wulff, S. (to appear). Applying corpus methods to writing research: Explorations of MICUSP. *Journal of Writing Research*.
- Swales, J. M. (2005). Attended and unattended “this” in academic writing: A long and unfinished story. *ESP Malaysia*, 11: 1-15.



# Appendix

Discipline	UA <i>this</i>	%	A <i>this</i>	%
English	174	35	329	65
History & Classical Studies	134	37	229	63
Linguistics	144	51	138	49
Philosophy	175	57	133	43
Economics	90	43	118	57
Education	135	41	195	59
Political Science	176	37	296	63
Psychology	342	42	480	58
Sociology	182	40	271	60
Biology	129	37	224	63
Natural Resources	207	51	200	49
Nursing	181	49	190	51
Civil & Environmental Engineering	83	32	178	68
Industrial & Operations Engineering	160	47	181	53
Mechanical Engineering	105	52	98	48
Physics	82	55	68	45
Total	2,499	43	3,328	57

<b>Proficiency level</b>	<b>UA <i>this</i></b>	<b>%</b>	<b>A <i>this</i></b>	<b>%</b>
Final year undergraduate	1,144	45	1,416	55
First year graduate	727	43	960	57
Second year graduate	361	39	575	61
Third year graduate	267	41	377	59
<b>Gender</b>	<b>UA <i>this</i></b>	<b>%</b>	<b>A <i>this</i></b>	<b>%</b>
Female	1,396	42	1,954	58
Male	1,103	45	1,374	55
<b>Native speaker status</b>	<b>UA <i>this</i></b>	<b>%</b>	<b>A <i>this</i></b>	<b>%</b>
Non-native speaker	499	45	607	55
Native speaker	2,000	42	2,721	58
<b>Total</b>	<b>2,499</b>	<b>43</b>	<b>3,328</b>	<b>57</b>