

# Um, about that, uh, variable: UH and UM in teen instant messaging

---

Tim Gadanidis

University of Toronto

DiPVaC4

May 28, 2018



# Introduction

---

# The variable

- The “filled pauses”/“hesitation markers”/“disfluencies”/...  
**UH and UM**, hereafter (UHM)<sup>1</sup> in **instant messaging** (IM)
- Variants: **uh** or **um**<sup>2</sup>

- (1) a. **um**, hostile much? (F, 1986)  
b. **uh** dude, They're having the meeting NOW (M, 1995)

---

<sup>1</sup>/əhʌm/

<sup>2</sup>Also spelled uhm.

## (UHM) as a pragmatic marker

- Views on (UHM) vary; I follow Tottie (2016) who argues that in speech, (UHM) is a pragmatic marker indicating **planning**
  - (UHM) is used more frequently in **word-search**, **long turns** and **responses to questions**

## (UHM) in writing

- Tottie (2017): in writing, (UHM) are “**stance adverbs**”<sup>3</sup>
  - Initial position: convey attitude towards proposition (**attitude** adverbs)
  - Medial position: comment on the manner of speaking (**style** adverbs)

---

<sup>3</sup>Term drawn from Biber, Johansson, Leech, Conrad, and Finegan (1999: 853).

## Stance adverbs

- (2) Tottie (2017: 5):
- a. **Um, senator**, the market already views those firms as having implicit government backing, because they do ... (Paul Krugman, *NYT*, 2010)
  - b. Obama is more, **um**, seasoned. Barack Obama's ... closely shorn hair appears to be increasingly gray. (*Washington Post*, 2010)

## Planning as a source for written (UHM)

### Sentence-initially:

... whereas *speakers hesitate* to produce answers to questions because they are uncertain of what to say or how to say it, *writers merely pretend to hesitate*, out of reluctance to say something tactless or hurtful.

(Tottie, 2017: 21)

### Sentence-medially:

*The writer pretends to be searching for a word and pretends to hesitate* before making an ironic, funny, somewhat derogatory or naughty choice.

(Tottie, 2017: 20)

- Tottie finds **no positional difference** between UH and UM
- Classifies both variants as **one lexeme**



Change in progress

---

## Change in progress

- Both real- and apparent-time data indicate that UM is rising relative to UH (Fruehwald, 2016; Wieling et al., 2016)
- Fruehwald (2016), Wieling et al. (2016) suggest that UM may have **taken on a new function**, leading to its rise, but are unable to identify such a functional difference

## The present study

---

# Motivation

- Tottie (2017) says that (UHM) is on a **lexical cline**:
  - *and-uh, but-uh* clitics in speech on the least wordlike end
  - “stance adverbs” in writing on the most wordlike end
- IM is a hybrid register (Tagliamonte, 2016; Tagliamonte & Denis, 2008)—it’s **conversational** and **interactive**, like speech, but in a **written medium**
- Thus investigating (UHM) in IM can give us clues to its **discourse/pragmatic function** and reveal functional **differentiation**, if it exists

## Data and method

---

# Corpora

- TTT: Data from 11 **17–20-year-olds** in one social network, 2004–2005, birth years **1985–1987** (Tagliamonte & Denis, 2008)
- TEEN: Data from 17 **teenagers in Toronto schools**, 2004–2006, birth years **1987–1990** (Tagliamonte & Denis, 2008)
- FBC: A corpus I built from 9 Toronto-area **students in my own community of practice** (Meyerhoff, 2002), 2014–2017, birth years **1993–1997**
  - Members of a Japanese martial arts club at the University of Toronto; members of its organizational committee

- Coded for a number of predictors including:
  - social factors—year of birth, gender
  - position in message
  - sentence type (question, response, &c.)
  - polarity
  - turn-taking

## Qualitative findings

---



## Some metalinguistic commentary...<sup>4</sup>



**John Scalzi**  @scalzi · 11 Jan 2013

Thesis: Online, when people put "**um**" before correcting you, "**um**" is passive-aggressive for "hey, you ignorant asshole." Discuss.

 120  91  37 



**John Scalzi**  @scalzi · Feb 28

Folks, unless you are actually 14, may I suggest that prefixing "**um**," to whatever argument you are about to make is not the effective rhetorical tool you think it is. I usually don't read past "**um**" in tweets because my eyes are already rolling in my head.

 101  43  557 



**John Scalzi**  @scalzi · 12 Dec 2015

The Five Point Exploding Heart Technique, but for stopping "**Um**, ACTUALLY" statements instead of murdering people.

 11  36  100 



---

<sup>4</sup>Thanks to Marisa Brook for pointing me in the direction of these tweets!

## Discourse-pragmatic function

- As in writing (Tottie, 2017), (UHM) mainly marks **stance** in IM, but the stances marked by each variant are different
- Core indexicality: **hesitation**
  - But when embedded in different contexts, different indexicalities emerge—indexical moves (Eckert, 2008; Silverstein, 2003)
- I will show examples from FBC, a corpus from a community of practice (Meyerhoff, 2002) that I participate in

## Both variants: epistemic stance

Uncertainty:

1 A: [*sends a picture of groceries*]

2 A: [*buy*] Something like that

3 B: What aisle lol

4 A: **Uhhhh**

5 A: **Uhm**

6 A: Idk

7 A: LOL

## Both variants: epistemic stance

Unexpectedness:

- 1 A: at [*bar*] at [*club social event*] he like fed her  
a peice [*sic*] of chicken
- 2 A: by just like
- 3 B: **Um**
- 4 A: shoving it in her mouth
- 5 A: lmao
- 6 C: :/
- 7 B: **Uh**

## um: politeness and face-work

Mitigating criticism:

- 1 A: **Uhm**, the rice cooker is super hot cuz it was still in keep warm mode o-o
- 2 B: Holy fuck sorry
- 3 A: It's okay, let's just be careful next time o.o

## UM: politeness and face-work

Mitigating a request:

- 1 A: Hi [B] :D **uhm** this is a bit random but if we were to have takoyaki party, would you mind having it at your place? Since you have the mobile gas burner (convenient) and [C] will know the ppl attending ([club] peeps) , I don't want my third roommate to feel uncomfortable by doing it at my place .-
- 2 B: Of course I don't mind
- 3 A: Oh wow thanks !! :DDDDDDDD

## um: politeness and face-work

Mitigating criticism:

- 1 A: I've been drinking
- 2 A: Since Friday
- 3 A: Last week
- 4 A: I drank from Wednesday to Sunday
- 5 B: **Um**
- 6 B: You might wanna
- 7 B: ease up on the ol liver there

## uh: challenging and disapproving

Disapproving:

- 1 A: he said they spent \$7000 on [*event*]
- 2 B: 1/7th on the venue change :)
- 3 A: lol yup
- 4 B: :/
- 5 A: apparently it was \$5000 for [*new venue*]
- 6 B: **uh**
- 7 B: so
- 8 B: like
- 9 B: they just wasted money themselves



## UH: challenging and disapproving

Disapproving, disaligning:

- 1 A: I remember he used to say shit like he didn't
- 2 A: Respect any of his sensei
- 3 A: And I was like **uhh**
- 4 B: "Oh Melissa thinks she's hot shit just cause she's been practicing longer and started this club"
- 5 B: **UHHHHHHHHH**
- 6 B: LITERAL WORDS TO COME OUT OF HIS MOUTH

## UH: challenging and disapproving

Challenging:

- 1 A: how did i treat her like a thing
- 2 B: **uh** hello
- 3 A: you've been trying to change her mind
- 4 B: trick her into liking you back again

## UH in reported speech

- UH is used in reported speech (along with *duh* and *ah*) to suggest ignorance or disfluency, but this does not occur with UM

- (3)
- a. she was like “**uhh...uhhh**...well that’s why god created diseases and such” ..or something to that effect
  - b. “duhhhhh, does **uhh** God do all of that or ahhhh is it like the moon?”
  - c. My Japanese is garbage // Rip // Can’t do polite words // I was like “**Uhhh**” “ahhh”

# Discussion

- Both variants primarily used as **stance markers** (as in Tottie, 2017)
- The distinction between the two variants is blurry, and not always as clear as in the extracts I have highlighted
- However, **UM has a distinctively politer connotation than UH does**
  - In line with suggestion by Wieling et al. (2016) that “UM is arguably more polite than UH [...] given that UH leaves the mouth in an open position and that the UH sound is also a common reaction to physical pain, fatigue, sadness, and anger” (Wieling et al., 2016: 229)

## Quantitative findings

---

## Overall distribution

- 1513 tokens
- Across all data: 36% UH
- Corpus-by-corpus:
  - TTT: 13% UH (573 tokens)
  - TEEN: 30% UH (217 tokens)
  - FBC: 55% UH (723 tokens)

# Change over time

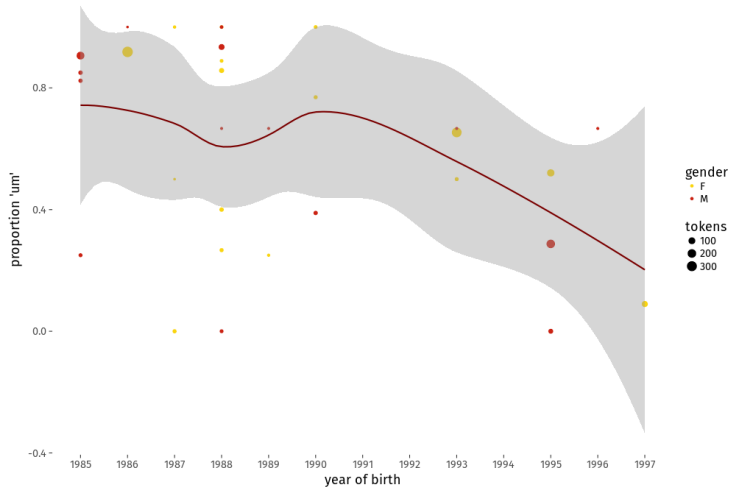


Figure 1: Proportion of UM by year of birth

# Individual variation

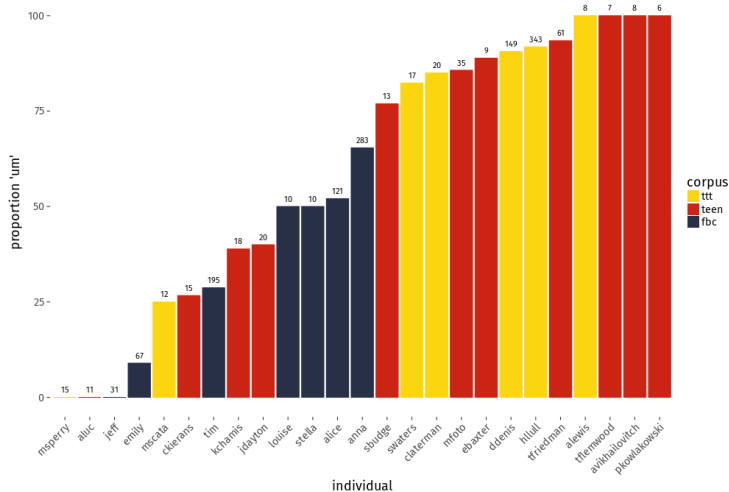


Figure 2: Individuals' rate of UM, sorted



# Message position

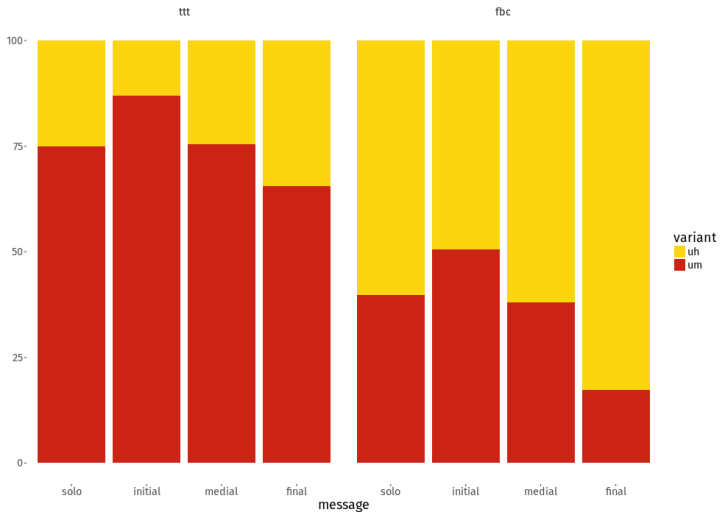


Figure 3: UH vs. UM by message position in each corpus

## Message position

- **But** the vast majority of tokens are initial or solo—only 12% overall are medial or final, and this declines from 20% in the oldest corpus to 6% in the newest corpus
- Examination of these tokens reveals they are used mainly to explicitly indicate lexical/memory access or planning, much like in speech (4)
- This usage is rare for UM and has a polite/sensitive connotation (5)

- (4) a. ok, i am trying to play that game.. **uh** Hearts... right  
b. Not until **uhh** // Let me check
- (5) a. ... did his family member, **um**... pass away... ?  
b. do you **uum** still want [to see] me today?

# Questions and answers

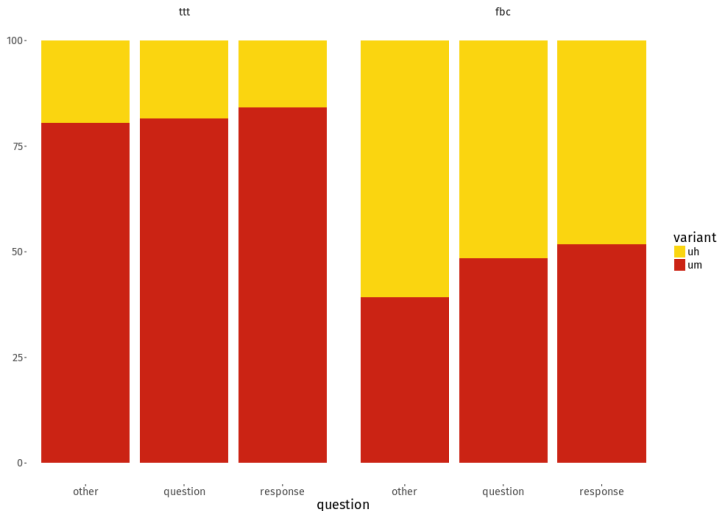


Figure 4: UH vs. UM by sentence type in each corpus

## Questions and answers

um in responses: indicates consideration of question, sensitivity to interlocutor face

- (6) *Do you want to be communications/PR lead?*  
**um** i guess out of all your suggestions that'd be my preferred position // i really would prefer equip manager tho...
- (7) *Do you want to get dinner?*  
**Uhm** // I ate already, sorry ><

## Questions and answers

When uh is used in responses, tends to indicate lack of knowledge or rejection of question entirely

- (8) *Do you know anything about the meeting today?*  
**Uh** I didn't hear anything... // And there was no post in the [facebook] group
- (9) *Why are they spending money on that and not the tournament subsidy?*  
**uhh** how would i know LOOOL

## Discussion

---

## Integration of qualitative and quantitative findings

The different indexicalities identified in the qualitative analysis are also reflected in the quantitative results—e.g., the more “basic” UH is also used in the most speech-imitative medial tokens; the polite/sensitive UM is used more often in responses to questions

## Reversal of the change in progress?

- The IM data is headed the opposite direction from the attested pattern—UH is rising
- A possible explanation: specialization (Kroch, 1994)



# Specialization

- Kroch (1994: 8): competition between members of a **doublet** will lead to one of two outcomes:
  1. one form **declines and disappears**
  2. the forms **differentiate in meaning** and stabilize

# Specialization

- While *uM* is rising in speech, *uH* is rising in IM
- Neither variant seems to be disappearing
- So we expect **specialization**—and that's what we find:
- Although they often overlap, the variants are used in **different contexts** and **message positions**, and they have **qualitatively different functions/indexicalities**

## A possible trajectory

- Early state: **UH dominant**, UM at 5–30% (Denis & Gadanidis, 2018)
- **UM rises** throughout 1900s and early 2000s, reaching up to 64% *um* (Wieling et al., 2016)
- **Competition** between incoming UM and preexisting UH may result in the **specialization** we see in IM

## Summary and takeaways

---

## The nature of (UHM) in IM

- At least in IM, both variants fill various and different discourse-pragmatic functions—clearly not just undesirable noise
- (UHM) provides crucial stance information (dependent on the variant and the context in which it is embedded) to the interlocutor
- Derived from a core indexicality of hesitation
- Further investigation is required to determine if specialization is underway in speech as well

## Next steps and future research

- Relative frequency analysis (e.g. Denis & Gadanidis, 2018; Fruehwald, 2016; Wieling et al., 2016) (underway!)
- Comparison to spoken data from the same informants (to be collected)
- Experiment to test social perceptions of UH vs. UM (e.g. Campbell-Kibler, 2007)
- Further investigation of apparently nonlexical discourse/pragmatic markers in IM, e.g. *hmm*, *ugh*

# Acknowledgments

- **Social Sciences and Humanities Research Council (SSHRC)** for research grants 2001–present to **Sali Tagliamonte**, which enabled the collection of the **Tagliamonte Internet Archive**, data from which was used in this project
- University of Toronto Linguistics SSHRC Institutional Grant (SIG) and SGS Conference Grant for sponsoring travel
- All the members of the UofT Linguistics 2017–2018 **MA Forum**, as well as my supervisor **Derek Denis**, for support, feedback and comments
- **All of the participants** who generously volunteered their IM data, without whom this project would not be possible



Social Sciences and Humanities  
Research Council of Canada

Canada



UNIVERSITY OF TORONTO  
VARIATIONIST SOCIOLINGUISTICS  
LABORATORY

## References i

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. *Edinburgh: Pearson Education Limited*.
- Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American speech*, 82(1), 32–64.
- Denis, D. & Gadanidis, T. (2018). Before the rise of *um*. Paper presented at DiPVaC4, Helsinki, Finland.
- Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics*, 12(4), 453–476.



## References ii

- Fruehwald, J. (2016). Filled pause choice as a sociolinguistic variable. *University of Pennsylvania Working Papers in Linguistics*, 22(2), 6.
- Kroch, A. (1994). Morphosyntactic variation. In *Proceedings of the thirtieth annual meeting of the Chicago Linguistics Society* (Vol. 2, pp. 180–201).
- Meyerhoff, M. (2002). Communities of practice. In J. K. Chambers, P. Trugill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 526–548). Oxford, U.K.: Blackwell.

## References iii

- R Core Team. (2017). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & communication*, 23(3-4), 193–229.
- Tagliamonte, S. A. (2016). So sick or so cool? The language of youth on the internet. *Language in Society*, 45(1), 1–32.
- Tagliamonte, S. A. & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American speech*, 83(1), 3–34.

## References iv

- Tottie, G. (2016). Planning what to say: uh and um among the pragmatic markers. In G. Kaltenböck, E. Keizer, & A. Lohmann (Eds.), *Outside the Clause: Form and function of extra-clausal constituents* (pp. 97–122). John Benjamins Publishing Company.
- Tottie, G. (2017). From pause to word: *uh*, *um* and *er* in written American English. *English Language & Linguistics*, 1–26.
- Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. (2016). Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change*, 6(2), 199–234.

# Acknowledgments

- **Social Sciences and Humanities Research Council (SSHRC)** for research grants 2001–present to **Sali Tagliamonte**, which enabled the collection of the **Tagliamonte Internet Archive**, data from which was used in this project
- University of Toronto SIG Travel Grant - add full name and SGS Conference Grant for sponsoring travel
- All the members of the UofT Linguistics 2017–2018 **MA Forum**, as well as my supervisor **Derek Denis**, for support, feedback and comments
- **All of the participants** who generously volunteered their IM data, without whom this project would not be possible



Social Sciences and Humanities  
Research Council of Canada

Canada



UNIVERSITY OF TORONTO  
VARIATIONIST SOCIOLINGUISTICS  
LABORATORY

## Bonus slides

---

## The emergence of a convention

- One part of a **developing register** of online English which exploits apparently sublexical markers (UM/UH, *hmm*, &c. for stance marking)
- Tracking (UHM) from 2004 to 2017 illustrates the **development of a convention** for its use as it moves **from the spoken domain to the written one**
- It remains to be seen whether the patterns I identify here apply in speech as well

## Mixed-effects model

Predictor	Estimate	Std. Error	z-value	Pr(> z )	
(Intercept)	-1.28518	0.87175	-1.474	0.140415	
position = solo	<i>reference</i>				
position = initial	0.14833	0.16350	0.907	0.364283	
<b>position = medial</b>	-0.71352	0.29875	-2.388	0.016923	*
<b>position = final</b>	-1.65438	0.42993	-3.848	0.000119	***
type = other	<i>reference</i>				
type = question	-0.07511	0.23997	-0.313	0.754297	
<b>type = response</b>	0.41239	0.15411	2.676	0.007451	**
<b>year of birth (asc.)</b>	-0.23306	0.10143	-2.298	0.021569	*

*formula:* dep.var ~ position + question + polarity + turn + yob + gender + (1|indiv)

*not selected as significant:* gender, turn, polarity

**Table 1:** Mixed-effects model using **lme4** (Bates et al., 2015) in R (R Core Team, 2017). Individual as a random effect.

## The model in a nutshell

- UM is favoured message-initially and in answers
- Younger speakers favour UH relative to UM