

“I’ll have to steal that... um borrow it”: Investigating *uh* and *um* in the instant messages of teens and twentysomethings

Tim Gadanidis

CVC10

May 4, 2018



Introduction

The variable

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

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

¹/əhʌm/

²Also spelled *uhm*.

Previous views on (UHM)

- Levelt (1989): (UHM) is a **symptom** of processing delay
- Maclay and Osgood (1959): (UHM) is used for **floor/turn management**
- Clark and Fox Tree (2002): (UHM) **signals an incoming delay** in speech; *uh* signals a short delay while *um* signals a long one

(UHM) as a discourse-pragmatic marker

- Most recently, Tottie (2016) argues that in speech, (UHM) is a pragmatic marker indicating **planning**
- (UHM) used more frequently in **word-search, long turns** and **responses to questions**

(UHM) in writing

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

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

- (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

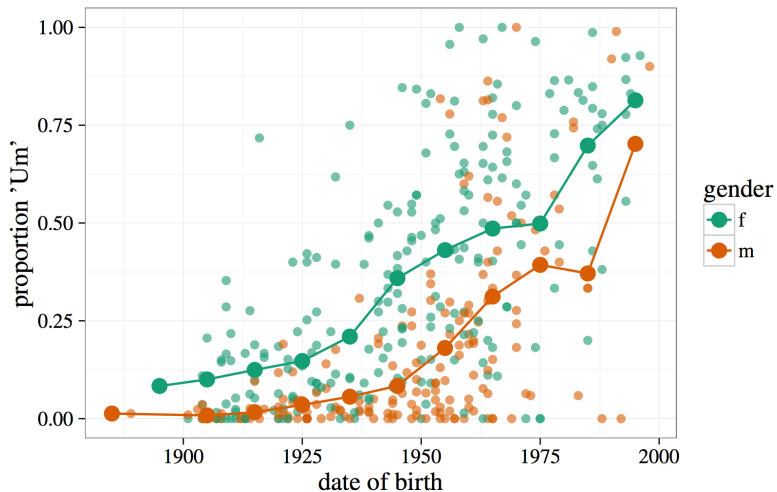


Figure 1: Proportion um by date of birth (Fruehwald, 2016)

Change in progress

- Fruehwald (2016), Wieling et al. (2016): **Women and young people** more likely to use *um* than *uh*⁴
- **Real-time data** from these papers and from Denis and Gadanidis (2018) also indicates a **change in progress**
- Wieling et al. (2016) suggest that *um* may have **taken on a new function**, leading to its rise, but are unable to identify a functional difference in speech

⁴In English, Dutch, German, Norwegian, Danish and Faroese

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 function** and reveal functional **differentiation**, if it exists

Data and method

- 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'm building from 9 Toronto-area students in my own social network, 2014–2017, birth years 1993–1997

Extraction

- (UHM) can appear **almost anywhere** in a sentence—no well-defined variable context
- Searched for and extracted every instance of <uh>, and <um>/<uhm>, allowing for reduplication of any letter e.g. *uhhhhhhhhhh* or *uuuummm*
- Coded for a number of predictors including:
 - social factors—year of birth, gender
 - position in message
 - sentence type (question, response, &c.)
 - polarity
 - turn-taking

Discourse-pragmatic function

Some metalinguistic commentary...

Some metalinguistic commentary...



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



Some metalinguistic commentary...



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 

Some metalinguistic commentary...



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 

As in writing (Tottie, 2017), (UHM) mainly marks **stance** in IM, but the stances marked by each variant have different connotations

Examples

- (3) *Um*: Apologetic, polite, concerned
- Uhm** // I accidentally did something really terrible T.T (F/1993)
 - Um** // You might wanna ease up on the ol liver there (M/1995)
 - uhm** this is a bit random but would you mind having [the party] at your place? (F/1995)
- (4) *Uh*: Disagreeing, disapproving, mocking
- uh** why not [make them pay for tickets]? // they're grown-ass adults (M/1995)
 - UHHHHHHHHH** // LITERAL WORDS TO COME OUT OF HIS MOUTH (M/1995)
 - she was like “**uhh...uhhh**...well that's why god created diseases and such” (M/1985)

Results

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

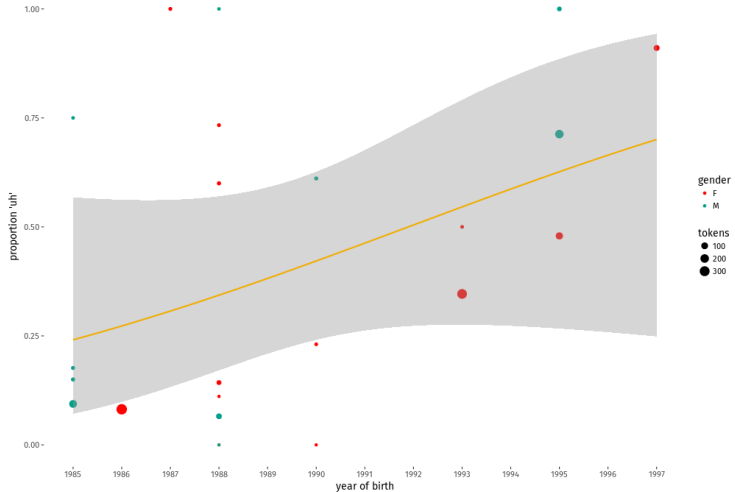


Figure 2: Proportion of *uh* by date of birth

Individual variation

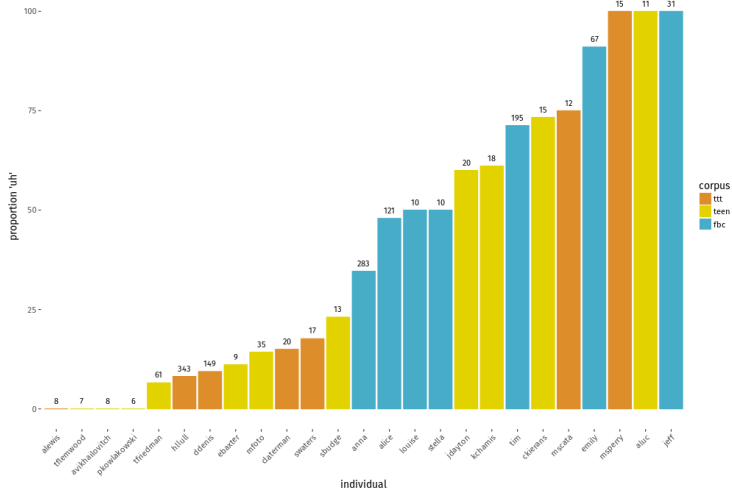


Figure 3: Individuals' rate of *uh*, sorted

Gender

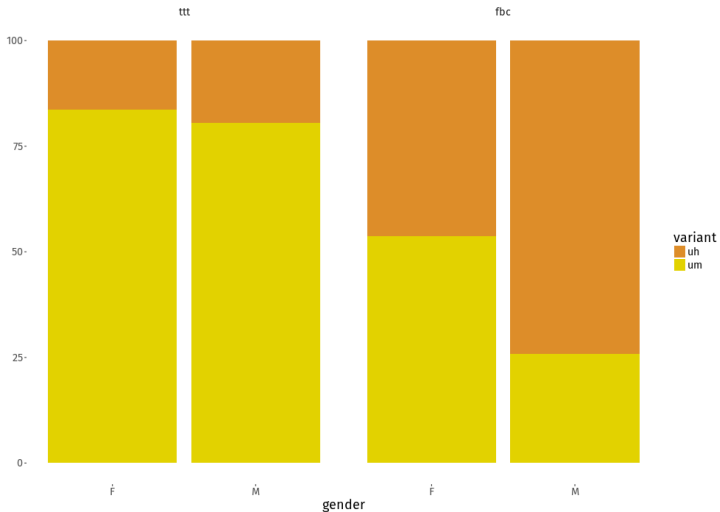


Figure 4: *Uh* vs. *um* by gender in each corpus

Message position

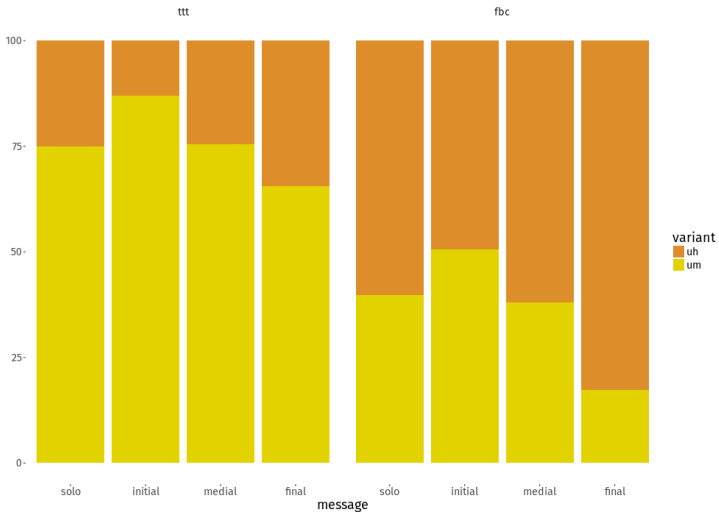


Figure 5: *Uh* vs. *um* by message position in each corpus

Message position

- (5) Initial and solo: *um*
- a. **Ummmm** I don't rllly wanna risk it (F, 1997)
 - b. **uhmmmmmmm** // x.x (F, 1993)
- (6) Medial and final: *uh*
- a. ...gettin a bit, **uh**, scary there. (M, 1995)
 - b. I **uh**... // I just woke up LOL (F, 1995)

Questions and answers

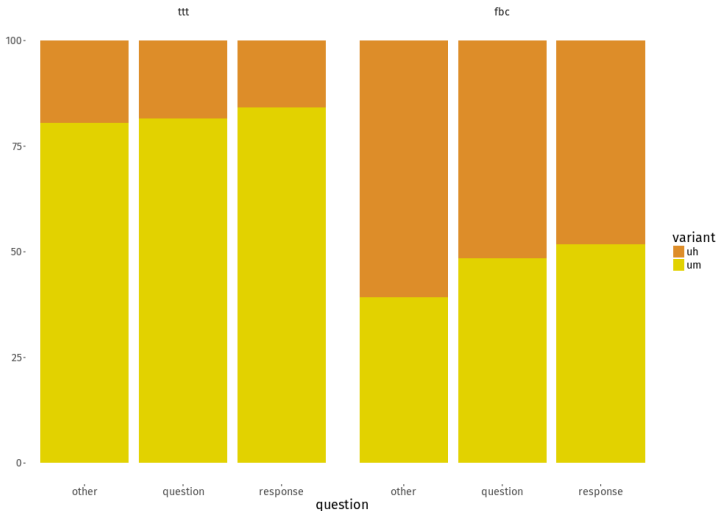


Figure 6: *Uh* vs. *um* by sentence type in each corpus

Questions and answers

(7) Questions:

- a. Hey, **uhm**, are you still gonna make tea? (F, 1993)
- b. **uh**, what did they talk about then (M, 1995)

(8) Responses:

- a. **uhmmmm** // you go to a textile store and buy dem fabrics ;0 (F, 1995)
- b. **um...** // nope, I've been trying all day (F, 1988)

Statistical modelling

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	
position = medial	-0.71352	0.29875	-2.388	0.016923	*
position = final	-1.65438	0.42993	-3.848	0.000119	***
type = other	<i>reference</i>				
type = question	-0.07511	0.23997	-0.313	0.754297	
type = response	0.41239	0.15411	2.676	0.007451	**
year of birth (asc.)	-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.

- *Um* is favoured *message-initially* and, in FBC, in *answers*
- Younger speakers favour *uh* relative to *um*

Discussion

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)

- 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**

A possible trajectory

- Early state: *uh* dominant, *um* at 11%⁵
- *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

⁵Based on data from Ontario farmers, year of birth 1890–1919 (Denis & Gadanidis, 2018)

Summary and takeaways

The nature of (UHM)

- At least in IM, both variants fill various and different discourse-pragmatic functions—not just undesirable noise as they are often construed
- (UHM) provides the interlocutor with crucial info about emotion, stance, &c.

The emergence of a convention

- One part of a **developing register** of online English characterized by extensive use of **discourse markers** (e.g., *uh/um, hmm, lol*) (Tagliamonte & Denis, 2008), **emoticons** (Tagliamonte & Denis, 2008), &c.
- 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

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
- 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

- 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*.
- Clark, H. H. & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111.
- Denis, D. & Gadanidis, T. (2018). Before the rise of *um*. To be presented at DiPVaC4, Helsinki, Finland.

- 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).
- Levelt, W. J. (1989). *Speaking: from intention to articulation*. MIT Press Series in Natural-Language Processing. MIT Press, Cambridge, Massachusetts.
- Maclay, H. & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15(1), 19–44.

- 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/>
- 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.

- 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
- 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