Before the rise of um

Derek Denis and Timothy Gadanidis

August 7, 2019

1 Introduction

One of the most dramatic discourse-pragmatic changes in twentieth-century English has progressed under the radar of laypeople and (until recently) linguists: the rise of *um* as the predominant variant of the 'filled pause' variable (UHM) at the expense of *uh* (Fruehwald, 2016; Tottie, 2011; Wieling et al., 2016). Fruehwald (2016: 43) documents this "textbook" change over 100+ years of apparent time: *um* increases incrementally between generations and the rise is led by women. In this chapter, we investigate (UHM) at an early stage of change to determine what triggered the rise of *um*.

2 Um and uh

[əː] [əːm]

3 Change in progress

The rise of *um* has now been described extensively in the variationist and corpus-linguistic literature, across a number of corpora and speech communities.

In the British National Corpus, Tottie (2011) observed that *um* was used more frequently than *uh* by women, younger speakers, and more educated speakers; men, older speakers and educated speakers used (UHM) more often overall. Fruehwald (2016)

While these accounts demonstrate definitively that a change is underway, an explanation for the change remains elusive. What was the trigger for this "textbook" change?

In this chapter, we investigate data from before the rise of *um* with the goal of evaluating the functional expansion hypothesis.

4 Data

The data for this study are from the *Farm Work and Farm Life Since 1890* oral history collection (Denis, 2016). The corpus consists of oral history interviews with 155 elderly farmers, recorded in 1984. The corpus covers five regions of Ontario, Canada: Temiskaming, Essex, Dufferin, Niagara Region, and Eastern Ontario; for this study, speakers from the latter two regions were considered. Speaker birth years range from 1891 to 1919, just before *um* began to take off per Fruehwald (2016).

We extracted each instance of *uh* and *um* from the transcripts, excluding unrelated instances such as *uh-oh*. Tokens from the two much-younger interviewers was also extracted, and analyzed separately. The transcription protocol emphasized faithful reproduction of *uh* and *um*.

5 Coding

We coded for the following social factors: year of birth, gender, and region (Niagara or Eastern Ontario).

To operationalize the functional expansion hypothesis, we coded for utterance position (initial or non-initial).

6 Results

Table 1 shows how our data compare with previous communities analyzed. The first block summarizes our data from Niagara and Eastern Ontario, as well as F-INT and M-INT, the two younger interviewers. The second block summarizes results from previous work on the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992), the Fisher corpus (Cieri, Miller, & Walker, 2004), the Philadelphia Neighborhood Corpus (PNC) (Labov & Rosenfelder, 2011), and the British National Corpus (BNC) (2007). The numbers for all of these other corpora are drawn from Wieling et al. (2016).

Community	Raw N *uh*	Raw N *um*	% *um*	Mean *uh* /1000	Mean *um*/1000	Mean UHM/1000
Niagara	1864	357	16.1	21.3	4.1	25.4
E. Ont.	1563	168	9.7	22.6	2.4	25.0
F-INT	321	318	49.8	12.4	12.3	24.7
M-INT	255	51	16.7	13.2	2.6	15.8
Switchboard	—	—	28.3	22.1	7.5	29.6
Fisher	—	—	64.1	6.8	9.9	16.7
PNC	—	—	27.6	13.2	4.5	17.7
BNC	—	—	46.1	4.5	4.3	8.8

Table 1: Cross-community comparison

As can be seen in the table, *um* is less frequent in our data compared to the more recent corpora; the female interviewer uses it around half the time, while the male interviewer's rate is comparable to the farmers'. Relative frequency of (UHM) taken as a whole is on par with other corpora, but we are cautious about making such a comparison because each corpus was collected and transcribed differently (for related discussion, see Pichler, 2010).

References

Cieri, C., Miller, D., & Walker, K. (2004). The fisher corpus: a resource for the next generations of speech-to-text. In *Lrec* (Vol. 4, pp. 69–71).

Denis, D. (2016). Oral histories as a window to sociolinguistic history and language history: Exploring earlier Ontario English with the Farm Work and Farm Life Since 1890 oral history collection. *American Speech*, *91*(4), 513–516.

Fruehwald, J. (2016). Filled pause choice as a sociolinguistic variable. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 6.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). Switchboard: telephone speech corpus for research and development. In *[proceedings] icassp-92: 1992 ieee international conference on acoustics, speech, and signal processing* (Vol. 1, pp. 517–520). IEEE.

Labov, W. & Rosenfelder, I. (2011). The Philadelphia Neighborhood Corpus.

Pichler, H. (2010). Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics*, *14*(5), 581–608.

The British National Corpus. (2007). Version 3. Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium.

Tottie, G. (2011). Uh and um as sociolinguistic markers in British English. *International Journal of Corpus Linguistics*, *16*(2), 173–197.

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.