Variation and change in the use of hesitation markers in Germanic languages

Martijn Wieling[1*], Jack Grieve[2], Gosse Bouma[1], Josef Fruehwald[3], John Coleman[4], and Mark Liberman[5]

[1]University of Groningen, [2]Aston University, [3]University of Edinburgh, [4]University of Oxford, [5]University of Pennsylvania

In this study, we investigate cross-linguistic patterns in the alternation between UM, a hesitation marker consisting of a neutral vowel followed by a final labial nasal, and UH, a hesitation marker consisting of a neutral vowel in an open syllable. Based on a quantitative analysis of a range of spoken and written corpora, we identify clear and consistent patterns of change in the use of these forms in various Germanic languages (English, Dutch, German, Norwegian, Danish, Faroese) and dialects (American English, British English), with the use of UM increasing over time relative to the use of UH. We also find that this pattern of change is generally led by women and more educated speakers. Finally, we propose a series of possible explanations for this surprising change in hesitation marker usage that is currently taking place across Germanic languages.

1. Introduction

Two basic *hesitation markers* (also referred to as *fillers* or *filled pauses*) are common in modern Germanic languages: the UM form, which consists of a neutral vowel followed by a final labial nasal, and the UH form, which consists of a neutral vowel in an open syllable. For example, in the English language these forms are generally written as *um* and *uh* in American English and as *erm* and *er* in British English. Similarly, in German a distinction is made between in *ähm* or *öhm* and *äh* or *öh*, whereas in Dutch a distinction is made between *ehm* or *uhm* and *eh* or *uh*. Similar forms appear to exist in all other Germanic languages.

Hesitation markers, including UM and UH, have long been studied in linguistics, primarily because their use has been seen as being directly related to the cognitive processes responsible for the production of speech, specifically marking disfluencies (e.g., Maclay and Osgood, 1959; Goldman-Eisler, 1968; Rochester, 1973; Crystal 1982; Levelt, 1983; Levelt & Cutler, 1983; Schachter et al., 1991). For example, Schachter et al. (1991) found that lecturers in the humanities used more hesitation markers than lecturers in the natural sciences when teaching, but not when being interviewed. They argued that this difference is due to the larger number of words from which a lecturer in the humanities must choose compared to a lecturer in the natural sciences, where technical vocabulary is more strictly defined. Because humanities lecturers have to make more decisions during speech production, they tend to use more hesitation markers. In other words, hesitation markers are seen as marking *disfluency* during language production. This general explanation for the use of hesitation markers has been referred to as the *symptom hypothesis* (De Leeuw, 2007).

Although disfluencies during language production would appear to explain many occurrences of hesitation markers in spoken language, other explanations for the use of UM and UH have been identified. For example, in a series of reaction time experiments, Brennan and Schober (2001) found that hesitation markers were beneficial to comprehension, as listeners were faster to select a target object after a filler was used in the stimulus sentence. Indeed, listeners appear to show a disfluency bias when encountering a hesitation marker. For example, Arnold et al. (2013) found that when encountering disfluent speech, listeners were more likely to expect a discourse-new referent. In line with this, Bosker et al. (2014) showed that listeners were more likely to expect a low frequency as opposed to a high frequency word after a disfluency marker, though listeners adapted this expectation on the basis of the speaker (i.e. there was no higher expectation for a low-frequency word when listening to non-native speakers). Similarly, Fox Tree (2001) showed that UH (but not UM) facilitated the speed with which listeners were able to recognize upcoming words. Fraundorf and Watson (2011) showed that hesitation markers improve recall whether or not they predict upcoming discourse boundaries and that no such effect results from coughs of equal duration, ruling out a processing time effect. In contrast to the symptom hypothesis, this type of explanation for the usage of hesitation markers has been referred to as the *signal hypothesis* (De Leeuw, 2007). Still other researchers have pointed out that UM and UH can be used to fulfill various discursive functions (e.g. Swerts, 1998; Rendle-Short, 2004; Tottie, 2014). For example, Swerts (1998) showed that hesitation markers can be used as markers of discourse structure, with hesitation markers occurring more often with stronger discourse breaks than with weaker discourse breaks. Similarly, Tottie (2014) argued that UM and UH can be used as discourse markers, with a similar meaning as the discourse markers *well* and *you know*.

Linguists have also directly compared the usage of UM and UH. For example, as noted above, Fox Tree (2001) found that UH but not UM facilitated word recognition by listeners. Alternatively, Shriberg (1994) reported that UM was more frequently found in sentence-initial position than UH in American English, a result that Swerts (1998) replicated based on the analysis of Dutch data. Similarly both Swerts (1998) and Clark and Fox Tree (2002) found that UH tends to be used by speakers to mark minor delays, whereas UM tended to be used to mark major delays. Such findings, however, have not been replicated by all researchers. For example, O'Connell and Kowal (2005) argued that there are no functional differences in the usage of UM and UH based on their analysis of six media interviews of Hillary Clinton. Furthermore, based on a review of previous research, Corley and Stewart (2008) concluded that there is no evidence that speakers have intentional control over the production of UM or UH (see also Finlayson and Corley, 2012). Differences have also been found in the use of UM and UH across Germanic languages. For example, De Leeuw (2007) reported that whereas English and German speakers had a higher frequency of use of UM, Dutch speakers generally had a higher frequency of use of UH.

The aforementioned studies have all focused on the different functions of UH and UM from a structural perspective. However, researchers have also analyzed the effect of various social factors on the choice between these two forms. For example, Rayson et al. (1997) showed on the basis of a corpus analysis of the British National Corpus (BNC) that *er* (i.e. UH) was the second-most characteristic word for male speech and the fourth-most characteristic word for the speech of older (35+) speakers, whereas *erm* (i.e. UM) was the ninth-most characteristic word for people from the upper social class, although they did not directly contrast social patterns in the use of UM and UH. Liberman (2005), however, found clear gender- and age-related patterns in the use of UH versus UM in corpora of transcribed English-language telephone conversations (i.e. the Switchboard, Fisher Part 1 and Fisher Part 2 collections; Godfrey & Holliman, 1993; Cieri et al., 2004; Cieri et al., 2005). He observed that the use of UH was higher for men than for women and for older speakers than for younger speakers, whereas the use of UM was higher for women and younger speakers. In other words, the frequency of UM relative to UH (i.e. the UM/UH ratio) was greater for younger speakers and women.

More recently, various other corpus-based studies have analyzed the use of hesitation markers in English and have obtained similar results (see Tottie, 2011 for an overview). For example, on the basis of two sub-corpora of the BNC (i.e. BNC-Demographic and BNC-Context Governed), Tottie (2011) showed that women, younger people, and people from higher socio-economic classes had a higher UM/UH ratio than men, older people and people from lower socio-economic classes—a result that once again suggests that UM usage is rising over time, led by women and speakers from higher classes. Similarly, Acton (2011) analyzed the UM/UH ratio in American English based on the relatively recent Speed Dating Corpus (SDC; Jurafsky et al., 2009) and the older Switchboard corpus (Godfrey and Holliman, 1993) and obtained similar results, with women showing a greater UM/UH ratio than men in both corpora. Based on the Switchboard corpus, Acton (2011) also showed that this pattern persisted at the dialect-region level and when the gender of the hearer was taken into account (i.e. same-gender dyads appeared to show a greater UM/UH ratio than different-gender dyads). He also found that younger speakers had a greater UM/UH ratio than older speakers and that the UM/UH ratio was greater for the more recent SDC than the Switchboard corpus and therefore suggested that these results (together with the gender difference) might indicate that a linguistic change is in progress. Similarly, Laserna et al. (2014) analyzed transcripts of conversations collected by 263 American participants from five different studies (Mehl & Pennebaker, 2003a, 2003b; Mehl, Gosling & Pennebaker, 2006; Fellows, 2009; Baddeley, Pennebaker & Beevers, 2013), which were collected via electronically activated recorders carried by the participants for two to three days, allowing for truly spontaneous conversations to be obtained. Laserna et al. (2014) did not explicitly contrast the use of UM and UH in their study, but they reported a significant correlation between gender (male: 1, female: 2) of $r = -.15$ ($p < .05$) for UH, and $r = -.09$ ($p > .05$) for UM. Consequently, they concluded that women showed a lower frequency of use for both UH and UM than men (since the correlation coefficients are negative) (see also Bortfeld et al., 2001). However, as the reduction appears to be greater for UH than UM, this result suggests that women in this study are characterized by a greater UM/UH ratio than men. In addition, Laserna et al. (2014) reported a negative correlation between age and UM use ($r = -.21$, $p < .001$), but not between age and UH use ($r = -.01$, $p > .05$). As the use of UM (but not UH) decreases for older people, this implies that the UM/UH ratio also decreases for older people, which once again implies that a change in English hesitation marker usage is currently underway.

Previous research on social variation in the use of UM and UH in British and American English has thus repeatedly identified the same basic patterns: younger speakers and women use relatively more UM than UH compared to older speakers and men (irrespective of the potential categorical functional differences between the two alternatives). This type of pattern is commonly identified in apparent-time sociolinguistic research and is seen as being indicative of a linguistic change in progress (Labov, 1994) with the use of UM relative to UH increasing over time. The apparent-time hypothesis assumes that most language is acquired during childhood and remains relatively stable afterwards. Correspondingly, the speech of older people is assumed to reflect the linguistic situation when these speakers were young. Furthermore, variationist sociolinguistics studies have repeatedly found that language change is led by women (e.g., see Labov, 2001). The first goal of this paper is therefore to assess whether a change in hesitation markers usage is truly underway in the English language based on detailed quantitative analyses of both longitudinal and apparent-time data. Furthermore, because other Germanic languages have comparable hesitation markers, the second goal of this paper is to investigate whether similar patterns of variation and change in the use of UM and UH can be found in other Germanic languages, including Dutch, German, Norwegian, Danish and Faroese.[1]

2. Data: Spoken language corpora

To compare patterns of linguistic variation and change in the use of the hesitation markers UM and UH in Germanic languages, we analyzed a range of spoken language corpora representing the English, Dutch, German, Norwegian, Danish and Faroese languages. For

[1] While we focus on Germanic languages in this study, note that a similar gender-related pattern has been recently observed in Mandarin speech (Yuan et al., submitted).

each of these corpora we generated a primary data set by extracting information about the usage of UM and UH[2] in the corpus as well as a range of social information about each speaker.[3] Most notably, we included gender and age. The age of the speakers may be used as a way to assess linguistic change. This type of *apparent time* analysis is a common technique in sociolinguistic research (see Labov, 1994) and is based on the assumption that if a change in progress is taking place, younger speakers will tend to use the more modern form, whereas older speakers tend to use the original form.

2.1. English

For the English language, we analyzed five spoken language corpora, including three corpora of American English, one corpus covering a wide range of British English dialects, and one corpus of Scottish English.

First, we analyzed the *Switchboard Corpus* of American English (Godfrey & Holliman, 1993), which contains data from approximately 2,400 two-sided telephone conversations collected in 1990. We extracted all 91,001 tokens of UM (i.e. *um*) and UH (i.e. *uh*) from the corpus, which were produced by a total of 520 different speakers. In addition, we recorded the position (counted from the start of the utterance) and duration of the hesitation marker and the duration of preceding and following pauses, as well as the age and gender of each speaker (education level was not included), and the total number of words that they contributed to the corpus.

Second, we analyzed the *Fisher Corpus* of American English (Part 1 and Part 2) (Cieri et al., 2004; Cieri et al., 2005), which contains transcripts of almost 12,000 telephone conversations collected from 2002 to 2003. We extracted all 19,753 tokens of UM (i.e. *um*) and UH (i.e. *uh*) from the corpus, which were produced by a total of 10,313 different speakers. In addition, we obtained the age, gender and amount of education (in years) of each speaker, and the total number of words that they contributed to the corpus.

Third, we analyzed the *Philadelphia Neighborhood Corpus* (PNC; Labov et al., 2013), which contains transcripts of interviews with speakers from the Philadelphia area conducted from 1973 to 2013. We extracted all 25,514 tokens of UM (i.e. *um*) and UH (i.e. *uh*) from the corpus, which were produced by a total of 395 different speakers. In addition, we recorded the duration of the hesitation marker, whether a pause occurred before or after the hesitation marker, the year of recording, the age, gender and number of years of schooling of each speaker, and the total number of words that they contributed to the corpus.

Fourth, we analyzed the spoken component of the *British National Corpus* (BNC; Coleman et al., 2012), which contains approximately seven million words recorded in 1993. We extracted all 25,498 tokens of UM (i.e *erm*) and UH (i.e. *er*) from the corpus, which were produced by a total of 960 different speakers. In addition, we recorded the duration of the hesitation marker and the duration of the pause following the hesitation marker, as well as the age and gender of each speaker, and the total number of words that they contributed to the corpus.

Fifth, we analyzed the *HCRC Map Task Corpus* of Scottish English (HCRC Map Task Corpus, 1993), which contains transcribed speech collected from undergraduates at the University of Glasgow in 1990, who were participating in a map task in which a guide had to explain a route drawn on a paper map to a follower who only had a map without the route. We

[2] Of course, transcribers may have made errors in assigning the label of the hesitation marker. However, it is unlikely that these errors are specific to the gender and age of the speakers.

[3] The data, methods and results associated with this analysis are available for download as supplementary materials at the first author's website (http://www.martijnwieling.nl) and at the Mind Research Repository (http://openscience.uni-leipzig.de).

extracted all 1,987 tokens of UM (i.e. *ehm, erm, mm*[4], *um*) and UH (i.e. *eh, er, uh*), which were produced by a total of 64 different speakers (of which 61 subjects were Scottish). In addition, we recorded the position of the hesitation marker in each utterance, as well as the age, gender, and role (i.e. follower or guide) of each speaker, and the total number of words that they contributed to the corpus.

2.2. Dutch

For the Dutch language, we analyzed the *Corpus Gesproken Nederlands* (version 2.0) (CGN, 2006), which contains spoken transcribed speech (about 9 million words) from various sources (e.g., spontaneous conversations, interviews, telephone dialogues) recorded from 1998 to 2004. We extracted all 228,619 tokens of UM (i.e. *ehm*, *uhm*) and UH (i.e. *eh*, *uh*) from the corpus, which were produced by a total of 3,433 different speakers. In addition, we recorded the position and duration of the hesitation marker, the duration of preceding and following pauses, the preceding and following word, the part-of-speech tag of the preceding and following word, as well as the age, gender, education level, nationality (Dutch, Belgian), and level of preparedness (i.e. low for spontaneous speech, high for a televised speech) of each speaker. Furthermore, we also extracted the total number of words that each speaker contributed to the corpus.

2.3. German

For the German language, we analyzed the *Forschungs- und Lehrkorpus Gesprochenes Deutsch* (FLGD; Depperman, 2014), which contains about 100 hours of recorded speech (about 1 million words) collected from 2005 to 2014. We extracted all 16,221 tokens of UM (i.e. *ähm, öhm*) and UH (i.e. *äh, öh*), which were produced by a total of 238 different speakers. In addition, we recorded the age and gender of each speaker.

2.4. Norwegian

For the Norwegian language, we analyzed the *Nordic Dialect Corpus and Syntax Database* (NDCSD; Johannessen et al., 2009), which contains approximately 2.8 million words from conversations and interviews collected between 1951 and 2012. We extracted all 47,604 tokens of UM (i.e. *em, EM, m, M, m-m, m_m*) and UH (i.e. *e, E, h-e*) from the corpus that were tagged as hesitation markers, which were produced by a total of 554 different speakers. In addition, we recorded the year of recording, the age group (old: aged 50+, young: aged between 18 and 30) and gender of each speaker, and the total number of words that they contributed to the corpus.

2.5. Danish and Faroese

Finally, for the Danish and Faroese languages, we analyzed the *Faroese Danish Corpus Hamburg* (FADAC; Braunmüller, 2011), which contains 440,000 words collected on the Faroe Islands from 2005 to 2009. We extracted all 4,504 tokens of UM (i.e. *ehm, ehhm, eehm, æhm, ææhm, øøhm,* etc.) and UH (i.e. *eh, ehh, eeh, æh, ææh, øøh,* etc.) from the corpus, which were produced by a total of 57 different speakers. In addition, we recorded the language in which the interview was conducted (Danish, Faroese), the age and gender of each speaker, and the total number of words that they contributed to the corpus.

---

[4] We included *mm*, which accounts for 7.7% of all hesitation markers in this dataset, as it generally appears to be used to mark hesitations, rather than for indicating assent (as opposed to *mhm*).

3. Data: Twitter corpora

In addition to analyzing various spoken language corpora, we also analyzed the use of UM and UH in both American and Dutch Tweets, a written register that is especially informal and shares several features with spontaneous speech. Notably, in instant message conversation, a similar register of computer-mediated communication, Tagliamonte & Denis (2008) found that the usage rates of discourse-pragmatic variables were broadly comparable to spoken language corpora. Of course, the function of UH and UM will often be different in writing than in speech, in large part because the use of UM or UH in written language is generally a conscious process (i.e. it has to be typed) and Twitter is less interactional than spoken language. This results in these forms being used primarily as discourse markers as opposed to hesitation markers. For example, the following Twitter conversation shows that UM in Twitter can be used to indicate irony, which appears to be far less common in spoken language:

A: “Make fun of Jeb Bush's brother all you want, but he would've been dropping bombs months ago.”
B: “um that's why everyone hates him”

Nevertheless, it is informative to test if patterns in the use of UM and UH can also be observed in written language.

3.1. English

For English Twitter, we analyzed a corpus of 6 billion words of American Tweets collected by Diansheng Guo of the University of South Carolina in 2013, which only contains tweets where the longitude and latitude of the user at the time of posting is known, as it was designed for the analysis of geolinguistic variation. We extracted the 69,075 tokens of UM (i.e. *um*) and UH (i.e. *uh*) from the corpus that were produced by the 25,852 users who contributed at least 1,000 total words to the corpus and whose username contained an unambiguous male or female name (e.g. *John2002* was designated as male, whereas *Kate_1234* was designated as female). Although this approach to identifying gender is not perfect, as some names will be misclassified, we assume that the chances of misclassifications are relatively modest. In addition, we recorded the gender of each user and the total number of words that they contributed to the corpus. It would have also have been possible to use the username to determine the year of birth, but very few users had a username containing a potential year of birth (i.e. less than 1% of the 25,852 users).

3.2. Dutch

For Dutch Twitter, we analyzed a corpus of 28.9 billion words of Dutch Tweets collected by the Department of Information Science at the University of Groningen between 2011 and 2014. We extracted the 68,089 tokens of UM (i.e. *uhm, um, euhm, ehm*, etc.) and UH (i.e. *uh, uuh, eh, eeh, euh*, etc.) from the corpus that were produced by the 38,651 users who contributed at least 1,000 total words to the corpus and whose username contained an unambiguous male or female name (as described above) and/or a four digit number ranging between 1930 and 2009, which we used to estimate that user’s year of birth. In contrast to the English dataset, the (much larger) Dutch Twitter dataset contained this four digit number frequently in the usernames of Dutch Twitter users. This approach to identifying age also is not perfect, as some names will be misclassified, but we assume that the chances of misclassifications are relatively modest.

4. Analysis[5]

Because the dependent variable for each of the primary data sets is binary (i.e. the use of UM versus UH or the number of tokens of UM versus the number of tokens of UH), we assessed the effect of each of our predictor variables (e.g., age, gender, hesitation marker duration) on the use of UM and UH using mixed-effects logistic regression (Agresti, 2007). By using mixed-effects regression we are taking the structural variability associated with speakers into account (see Baayen, 2008). This is important because some speakers may be more likely to use UM (relative to UH) than others (i.e. modeled via a random intercept for speaker). Similarly, the effect of each predictor may vary across speakers. For example, for some speakers a longer duration of the pause following a hesitation marker may be more predictive of the usage of UM than for other speakers. This would be modeled with a by-speaker random slope for the duration of a following pause. Since we are using logistic regression, the estimates need to be interpreted with respect to the logit scale (i.e. the logarithm of the odds of observing UM rather than UH). Positive estimates indicate an increased probability of observing UM together with increasing values of the predictor, whereas negative estimates signal the opposite. An estimate of zero indicates that it has no effect on the probability of observing UM.

For all of the primary data sets except one, we obtained the best-fitting model including only significant predictors and supported random intercepts and random slopes. Predictors and random intercepts and slopes were included if they reduced the Akaike Information Criterion (AIC; Akaike, 1974) by at least 2, compared to the model without the random intercept or slope (see also Wieling et al., 2014 for a similar approach). A reduced AIC indicates that the additional complexity of the model is warranted given the increase in goodness of fit. Due to the large number of predictors in the Dutch data set, however, we did not fit the best model but rather fitted a random-intercepts-only model and assessed if the inclusion of individual random slopes affected the significance of the predictors. We only included predictors that remained significant in all cases in the final model. Given the large number of predictors in this model, we also did not evaluate all possible interactions.

We assessed the goodness of fit of these models (including the random-effects structure) by calculating the index of concordance *C*, which is known as the receiver operating characteristic curve area ‘C’ (Harrell, 2001). Values of *C* greater than 0.8 indicate a successful classifier, whereas a value of 0.5 indicates the classifier has no predictive power at all. All models had C values close to or over 0.8 (see supplementary materials for exact values).

5. Results: Spoken language

Table 1 presents the effects (including associated estimations of effect size: the increase in logits of the dependent variable for the categorical predictors, or per 1 standard deviation increase of the numerical predictors) of the speaker-related predictors that were present in at least two data sets (i.e. gender, age, education level, and year of recording) on the use of UM over UH. Table 1 clearly shows that women are more likely than men to use UM as opposed to UH across all data sets. Similarly, Table 1 shows that younger speakers are generally more likely than older speakers to use UM as opposed to UH; only in the case of the relatively small HCRC Corpus, does the effect of age not reach significance ($p = 0.07$). Table 1 also

[5] Given that we analyzed nine independent data sets, we provide a simplified summary of the results for all models together in this section, rather than reporting each individual model. The full details for each model can be found in the supplementary materials (available at the Mind Research Repository: http://openscience.uni-leipzig.de), which contains all data, all *R* commands used to generate the models, and all results for each individual model, as well as detailed instructions on how to conduct the analysis.

shows that more or longer educated people are more likely to use UM as opposed to UH in the Fisher corpus and the Dutch corpus, but that the effect of education in the PNC was non-significant. In addition, the effect of education is much smaller than that of age. Finally, Table 1 shows that the use of UM over UH has increased over real-time in the PNC, the Norwegian Corpus, and in the Dutch corpus. Figure 1 visualizes this result for the three data sets. For each data set, the graph shows the proportion of UM over UH (i.e. UM/[UM+UH]) by year of recording (divided into four groups containing roughly the same number of speakers) and gender. The error bars indicate the 95% confidence interval (i.e. 1.96 standard errors below and above the mean). It should be noted, however, that whereas the PNC (1973-2013) and the Norwegian corpus (1951-2012) each span at least 40 years, the Dutch Corpus only spans 13 years and 90% of the data was recorded between 1999 and 2003. The effect of year of recording is significant even while controlling for the age of the speaker (i.e. it is not an effect of age grading; see Table 1). When year of recording is excluded from the analysis for the PNC and instead only year of birth and age are taken into account, the most important predictor clearly is year of birth; the effect of increasing age (i.e. older people are still more likely to use UH) is only minimal (p = .04).

Significant interactions (e.g., between age and gender) were identified in some models; however, because these interactions did not change the direction of the general effect (e.g., the age effect was negative for both men and women, but less so for men than for women), we did not explicitly include these interactions in Table 1 (see, however, supplemental materials for the precise model specifications). Most important, these effects were found to be significant, while controlling for the effect of other potential important predictors, such as the duration of the pause before and after the hesitation marker (see Table 3, discussed below). Also note that for the PNC (and for the Switchboard corpus, but not for the HCRC, nor the BNC), the predictive value of the duration of the pause after the hesitation marker has diminished for people born in more recent years (i.e. a longer pause is more likely to predict the occurrence of UM over UH for older people than for younger people; see supplementary material). This suggests, for these datasets, that younger people are using UM more across the board, and are not simply more frequently signaling longer pauses.

Table 1. Effects of subject-related predictors on the choice of UM over UH for all data sets

	Gender: Male vs. Female		Age: Old vs. Young		Education: High/More vs. Low/Less		Year of Recording: Increase vs. Decrease	
Switchboard	F	(1.03)	Y	(0.6_z - 0.7_z)				
Fisher	F	(1.37)	Y	(0.39_z)	More	(0.11_z)		
PNC	F	(1.31)	Y	(1.2_z -1.7_z)	(More)	(0.03_z)	Increase	(0.54_z)
BNC	F	(0.45)	Y	(0.45_z)				
HCRC	F	(2.30)	(Y)	(0.35_z)				
German	F	(0.43)	Y	(0.94_z)				
Norwegian	F	(0.23)	Y	(0.65)			Increase	(0.35_z)
Danish/Faroese	F	(0.59)	Y	(0.4_z - 0.6_z)				
Dutch	F	(0.5 - 0.9)	Y	(0.3_z - 0.6_z)	High	(0.15_z)	Increase	(0.09_z)

Significant ($p < 0.05$) and non-significant (category name put between parentheses) effects are listed; an empty cell indicates the absence of that predictor in that data set. The values between parentheses indicate the effect size (in terms of logits: the increase in probability of observing UM rather than UH) when the category changes to the one indicated or (when a subscripted z is shown) when the value of the numerical predictor increases with 1 standard deviation. A range of values indicates the predictor is involved in an interaction. In other words, the effect of age in the Switchboard corpus varies based on the hesitation marker being phrase final (smaller effect) or not (larger effect), while the effect of gender and age varies per country for the Dutch data set (larger for Belgium than for the Netherlands), and the effect of age varies per language in the Danish/Faroese data set (larger for Faroese), and for gender in the PNC (larger for men).

Figure 2 presents four graphs for the American English Switchboard data set, which visualize the relationship between age, gender and the use of UM and UH. The first graph (top-left) plots the proportion of UM over UH (i.e. UM/[UM+UH]) by age (divided into four age groups containing roughly the same number of speakers) and gender. Similarly as before, the error bars indicate the 95% confidence interval (i.e. 1.96 standard errors below and above the mean). This graph shows a clear increase in the proportion of UM over UH across age groups for both men and women, with women consistently showing a higher rate of UM usage than men. Note that all speakers in this corpus mostly use UH, with only women in the two youngest age groups approaching 50% UM usage. The second graph (top-right) plots the relative frequency of UM and UH taken together (i.e. total hesitation marker frequency relative to all words in the corpus) by age and gender. This graph shows a clear decline in hesitation marker usage across age groups for both men and women, with men consistently using more hesitation markers than women (but note that this pattern is not observed in the smaller HCRC and Danish/Faroese datasets, likely due to the large individual differences in hesitation marker frequency; also Bell et al. (2000) found no gender differences in hesitation marker usage for Swedish speakers). The third graph (bottom-left) charts the frequency of UM relative to all words in the corpus by age and gender. This graph shows a clear increase in UM use over age groups with women consistently using UM more frequently than men, even though this gap appears to be closing in the youngest age group. Finally, the fourth graph (bottom-right) plots the frequency of UH relative to all words in the corpus by age and gender. This graph shows a clear decrease in UH usage across age groups with men consistently using UH more frequently than women.

Figure 3 presents the same four graphs for the Dutch data set. Overall, the Dutch results are similar to the American English results presented in Fig. 2. In particular, the first graph (top-left) also shows a clear increase in the usage of UM over UH across age groups with women showing a higher proportion of UM over UH than men, while the third graph (bottom-left) shows a clear increase in the relative frequency of UM across age groups with women using UM more often than men. Finally, the fourth graph (bottom-right) shows a decrease in the relative frequency of UH across age groups, especially for women. Despite these similarities, differences between the American English Switchboard data and the Dutch data are apparent. Whereas hesitation markers in English have been showing a clear decrease in frequency across age groups, the second graph (top-right) shows that there is no clear trend in the overall usage of hesitation markers in Dutch (though the distinction between men and women is similar).

The visualizations for the other data sets, which can be found in the supplemental material, all show relatively similar patterns. Most important, all data sets show an increase across age groups in the use of UM over UH (with women having the highest proportion of UM use) and an increase across age groups in the relative frequency of UM. In addition, most data sets show a decrease across age groups in the use of UH. There are, however, differences between the nine data sets. In particular, the relative frequency of hesitation markers across age groups (i.e. the second graph in Figs 2 and 3) varies considerably across the nine data sets.

Despite generally following the same basic trends, there are also considerable differences in the average overall proportions of UM over UH and the relative frequencies of UM and UH across the nine data sets. These results are summarized in Table 2. For example, the average proportion of UM over UH ranges from 27% to 64% for the five English corpora, compared to 50% in the German corpus, 17% in the Danish corpus, 13% in the Norwegian corpus, and 11% in the Dutch corpus.

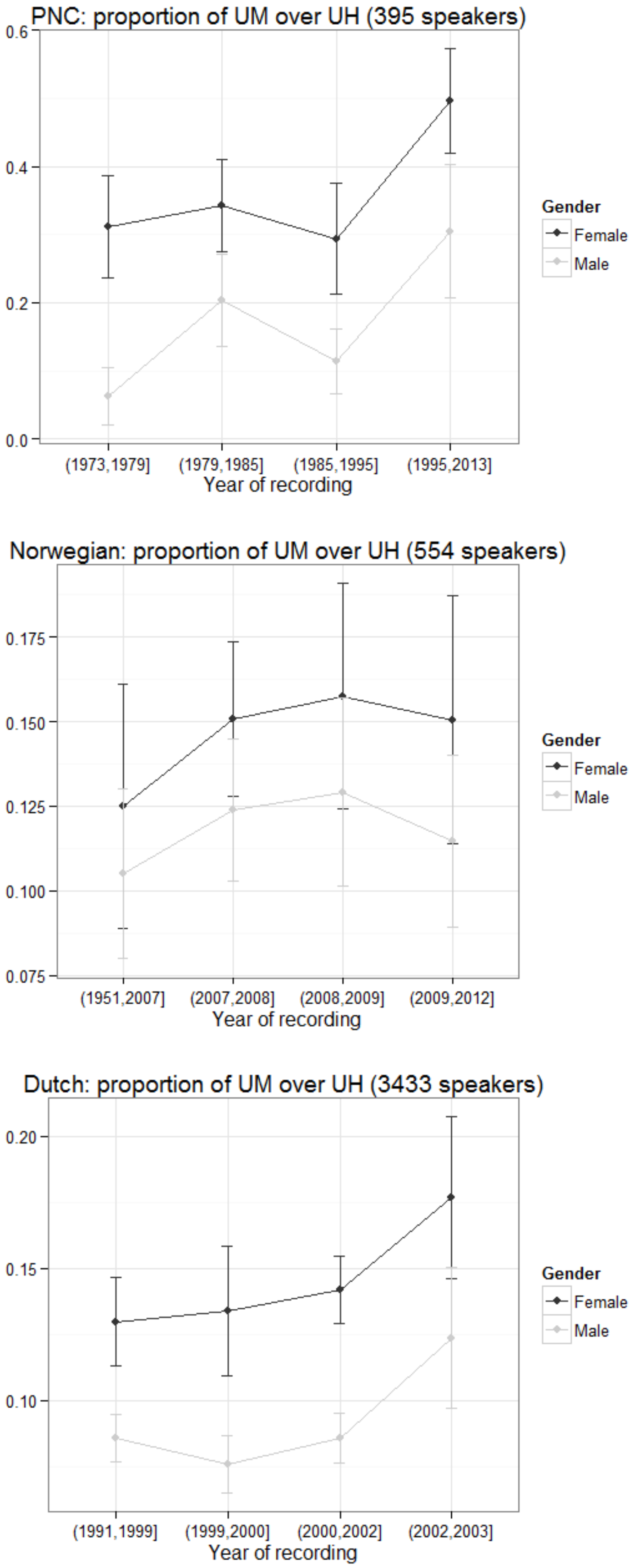

Figure 1. Proportion of UM over UH for three data sets: PNC (top), Norwegian (middle) and Dutch (bottom) by year of recording and gender.

Finally, Table 3 presents the effects (again including estimations of effect size) of the hesitation marker-related predictors that were present in at least two data sets (i.e. the duration of the hesitation marker, the duration or presence (for the PNC) of a pause before the hesitation marker, the duration or presence (for the PNC) of a pause after the hesitation marker, the presence of the hesitation marker at the start of the utterance, and the presence of the hesitation marker at the end of the utterance) on the use of UM over UH. Table 3 only presents results for the five data sets for which we were able to include information about the duration and position of hesitation markers and pauses.

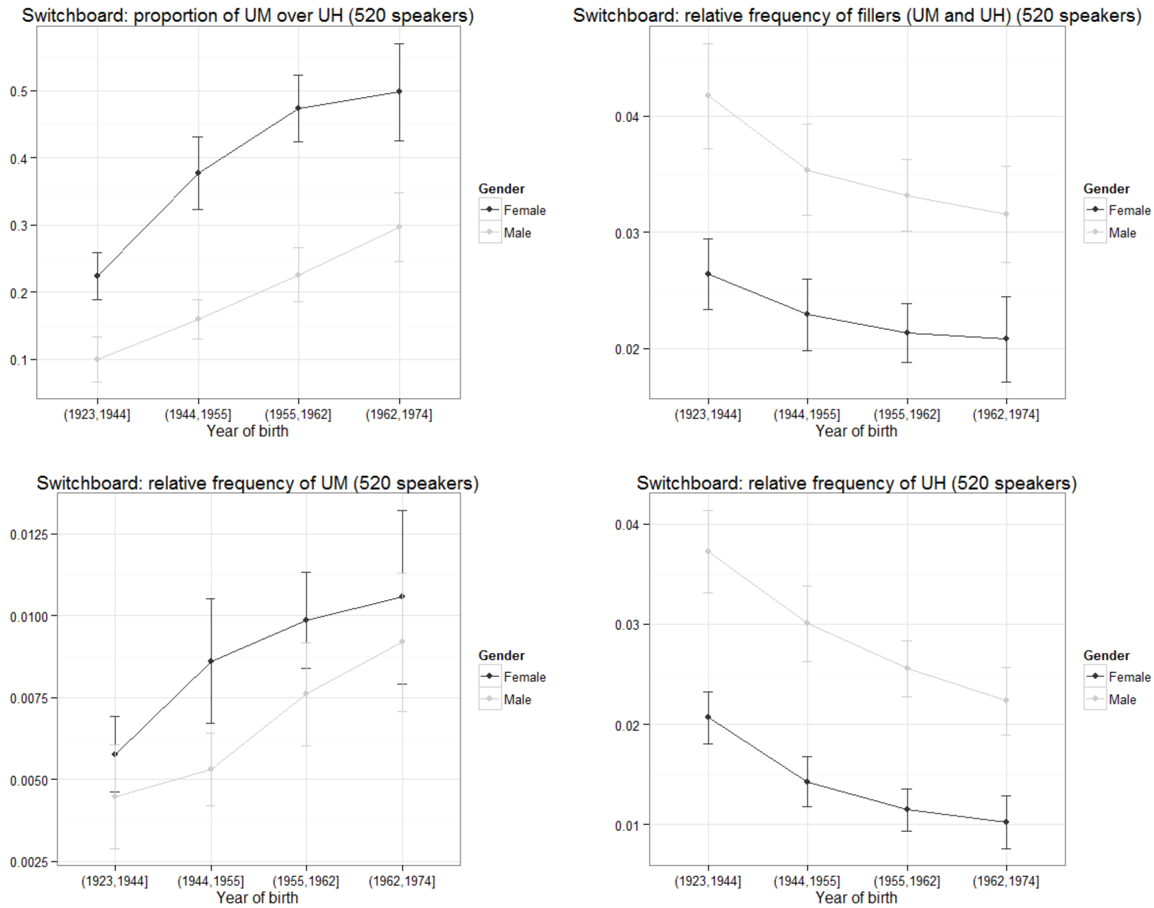

Figure 2. American English Switchboard data: proportion of UM over UH (top-left), relative frequency of hesitation markers (top-right), relative frequency of UM (bottom-left), and relative frequency of UH (bottom-right) by age and gender.

Table 2. Proportion of UM over UH and relative frequency of UM and UH for all data sets

	UM Proportion	UM Relative Frequency	UH Relative Frequency
Switchboard	0.2825	0.0075	0.0221
Fisher	0.6408	0.0099	0.0068
PNC	0.2765	0.0045	0.0132
BNC	0.4612	0.0043	0.0045
HCRC	0.5717	0.0081	0.0058
German	0.5017	(no word counts)	(no word counts)
Norwegian	0.1285	0.0026	0.0189
Danish/Faroese	0.1653	0.0020	0.0079
Dutch	0.1086	0.0037	0.0315

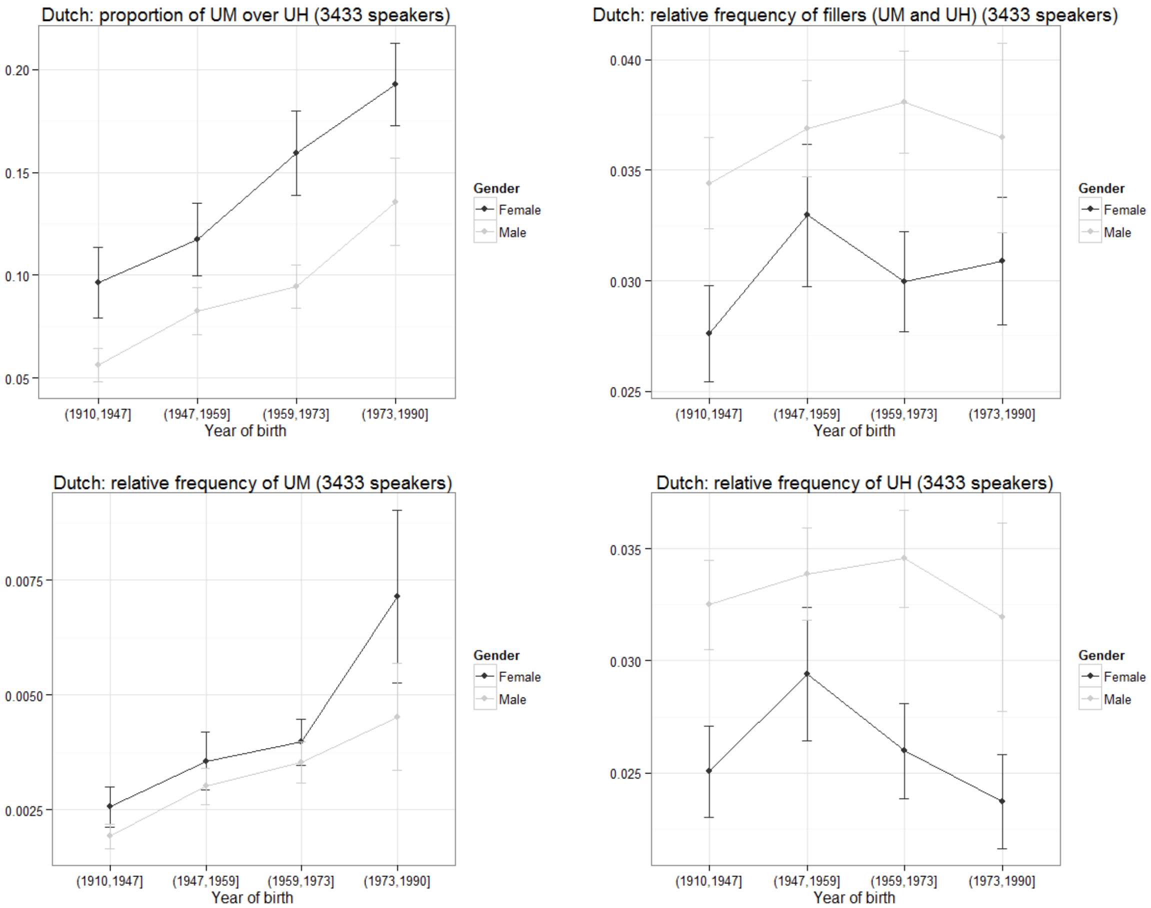

Figure 3. Dutch Spoken data: proportion of UM over UH (top-left), relative frequency of hesitation (top-right), relative frequency of UM (bottom-left), and relative frequency of UH (bottom-right) by age and gender.

Table 3. Effects of hesitation marker-related predictors on the choice of UM over UH

	Duration of Marker	Duration/Presence of pause before Marker	Duration/Presence of pause after Marker	Initial Position	Final Position
Switchboard	Longer (0.87_z)	Longer (0.12_z)	Longer (0.11_z)	Initial (0.67)	Final (1.06)
PNC	Longer (1.25_z)	(Absent) (-0.08)	Present / Longer (0.59) / (0.55_z)		
BNC	Longer (1.06_z)		Longer (0.44_z)		
HCRC				Initial (0.83)	Final (1.07)
Dutch	Longer (1.15_z)	Longer (0.17_z)	Longer (0.47_z)	Initial (0.51)	Final (0.96)

Significant ($p < 0.05$) and non-significant (category name put between parentheses) effects are listed; an empty cell indicates the absence of that predictor in that data set. The values between parentheses indicate the effect size (in terms of logits: the increase in probability of observing UM rather than UH) when the category changes to the one indicated or (when a subscripted *z* is shown) when the value of the numerical predictor increases with 1 standard deviation.

In general, all predictors showed positive estimates, indicating that higher values of the predictors are associated with a greater likelihood of observing UM as opposed to UH. Specifically, a longer duration (of the hesitation marker or the pause before or after the hesitation marker) is associated with a greater likelihood of the hesitation marker being UM rather than UH, while the occurrence of the hesitation marker in utterance-initial or utterance-final position is also associated with a greater likelihood of the hesitation marker being UM rather than UH. Note that in the case of the Philadelphia Neighborhood Corpus, the presence of a pause before or after the hesitation marker is similar to the hesitation marker being utterance initial or final as utterances were identified on the basis of the pauses (a pause of 200 ms. or more indicated the break between two utterances).

6. Results: Twitter

Figure 4 presents four graphs for the American English Twitter data set, which visualize the relationship between gender and the use of UM and UH. The first graph (top-left) plots the proportion of UM over UH and shows that women are more likely to use UM over UH than men. The logistic mixed-effects regression model indicates this effect was significant ($p < .001$). The second graph (top-right) plots the frequency of UM and UH taken together relative to all words in the corpus and shows that women are more likely to use hesitation markers than men. The third graph (bottom-left) plots the frequency of UM relative to all words in the corpus and shows that women are more likely to use UM overall than men. The fourth graph (bottom-right) plots the frequency of UH relative to all words in the corpus and shows that women are more likely to use UH overall than men. These results for the proportion of UM over UH and the relative frequency of UM agree with the results of the analysis of the American English spoken language data sets (e.g., see Fig. 2); however, unlike the results of the spoken analyses, women were found to have higher relative frequencies for UH and for hesitation markers in general, likely reflecting functional differences in the use of UM and UH in written language.

Figure 5 presents four graphs for the Dutch Twitter data set, which visualize the relationship between age, gender and the use of UM and UH. The first graph plots the proportion of UM over UH and shows that women and younger Twitter users are more likely to use UM than men and older Twitter users, although in this case the youngest users were found to reduce their use of UM compared to users from the second youngest group. The logistic mixed-effects regression model indicates that the age effect was significant ($p < .001$) but the gender effect was not ($p = .13$). However, note that a curvilinear pattern might be a better fit to the data, with a decrease in proportion of UM over UH for the youngest users. The second graph plots the frequency of UM and UH taken together relative to all words in the corpus and shows that women and younger Twitter users are more likely to use hesitation markers than men. Also in this case, the youngest users were found to reduce their use of hesitation markers compared to users from the second youngest group. The third graph plots the frequency of UM relative to all words in the corpus and shows that women and younger Twitter users are more likely to use UM than men, although once again the youngest users were found to reduce their use of UM compared to users from the second youngest group. The fourth graph plots the frequency of UH relative to all words in the corpus and shows that women and younger Twitter users are more likely to use UH than men, with a similar deviating pattern for the youngest users. In terms of gender, these results (though not significant) are in line with the results of the analysis of the American Twitter data.

Although the results of the analysis of both the American and Dutch Twitter data correspond reasonably well overall with the results of the analysis of the spoken language data sets, the relative frequency of the hesitation markers in the Twitter data is an order of magnitude lower than in the spoken language data, which likely reflects clear register differences between speech and writing. Table 4 lists these values, for comparison with the corresponding values for the spoken data sets presented in Table 2. Note that the proportion of UM versus UH for the Dutch Twitter data is much larger than for the Dutch spoken data. Again this is likely indicative of register differences between speech and writing.

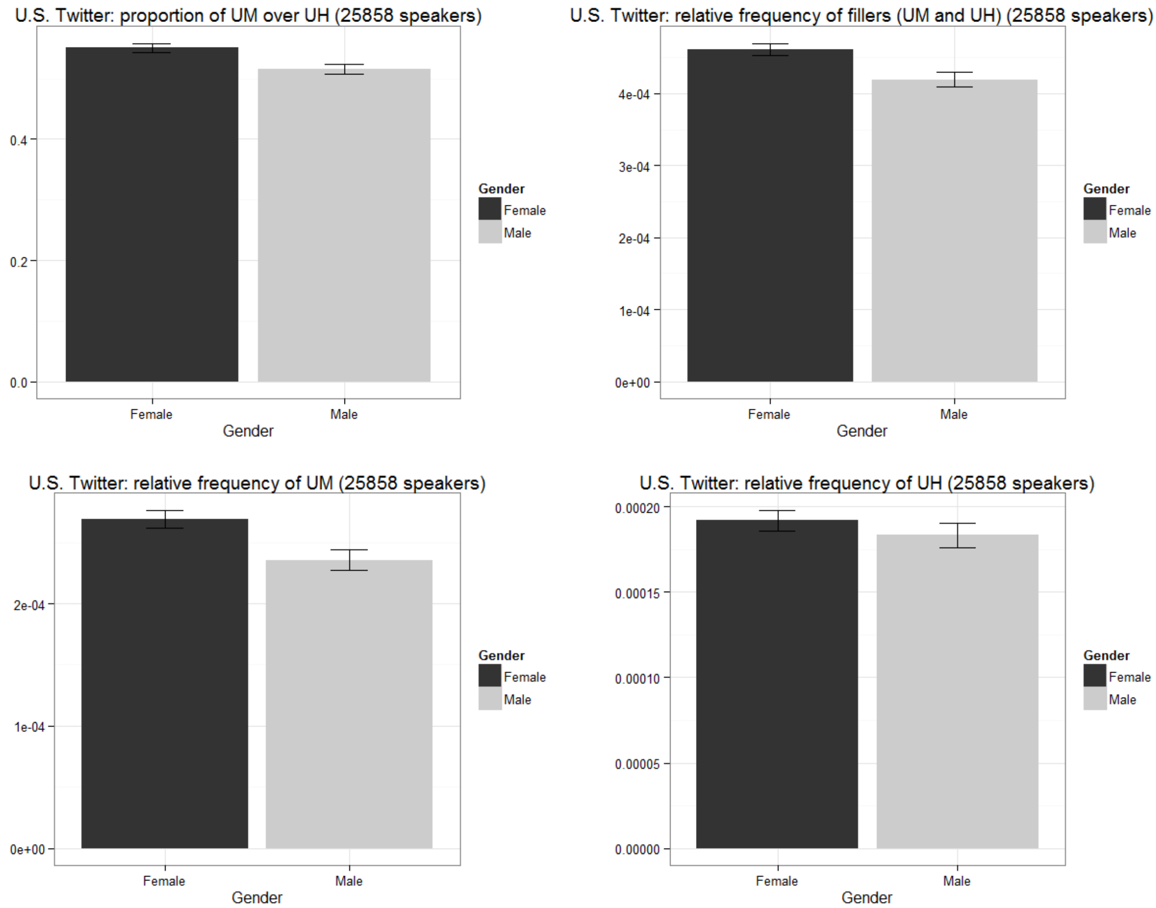

Figure 4. American Twitter data: proportion of UM over UH (top-left), relative frequency of UM and UH (top-right), relative frequency of UM (bottom-left), and relative frequency of UH (bottom-right) by gender.

Table 4. Relative proportion of UM vs. UH and versus all words for the Twitter data sets

	UM Proportion	UM Relative Frequency	UH Relative Frequency
American English	0.5334	0.00025	0.00019
Dutch	0.6518	0.00011	0.00006

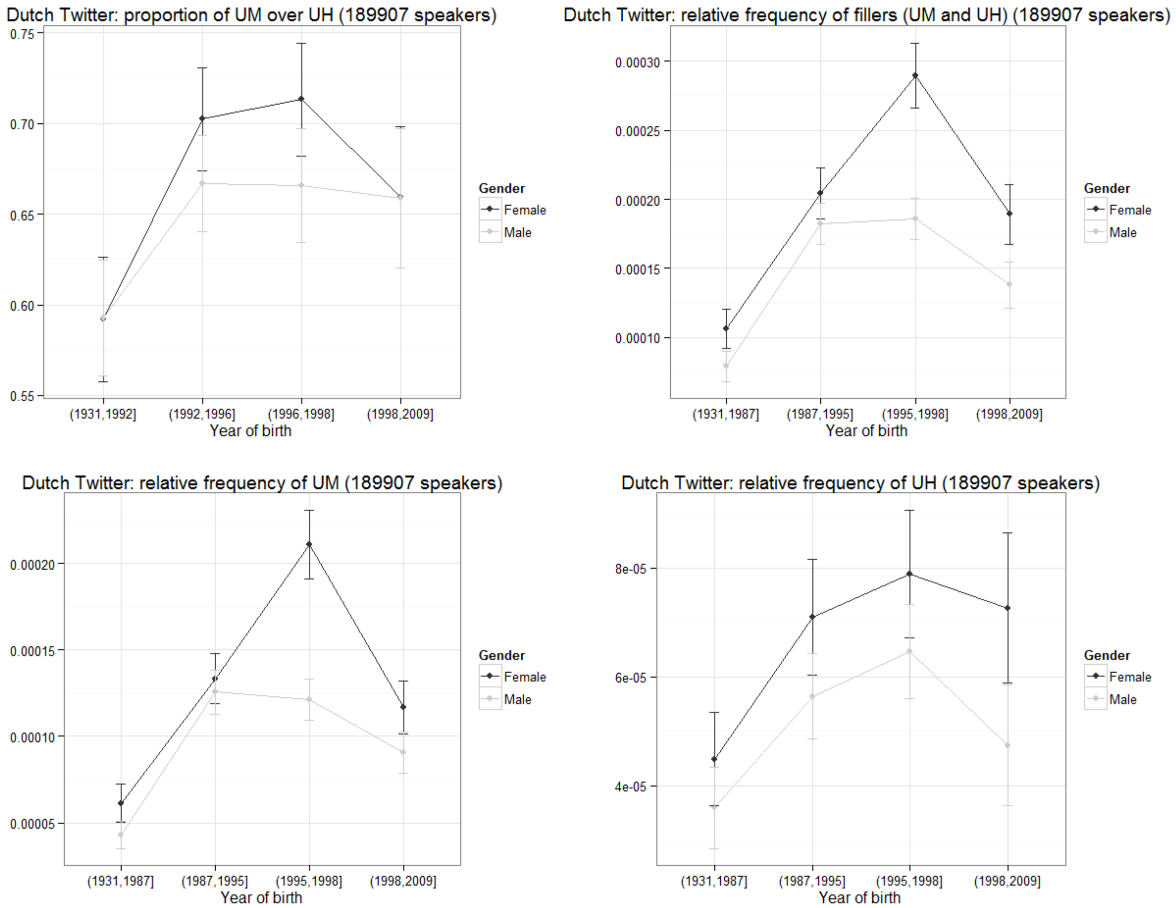

Figure 5. Dutch Twitter data: proportion of UM over UH (top-left), relative frequency of UM and UH (top-right), relative frequency of UM (bottom-left), and relative frequency of UH (bottom-right) by age and gender.

7. Discussion

The results of our analyses have shown that there is a consistent pattern of sociolinguistic variation in the use of the hesitation markers UM and UH across many modern Germanic languages. In English, Dutch, German, Norwegian, Danish and Faroese, UM is relatively more common than UH in the language of women and younger speakers when compared to the language of men and older speakers. Although gender and age patterns in the use of UM and UH have been identified in previous research on the English language, this paper has shown that this pattern holds across a wide variety of Germanic languages, as well as several varieties of English (American, Scottish and other British dialects). Furthermore, because we analyzed a wide variety of different corpora, this paper has also shown that this pattern is even more pervasive, existing across a range of time periods and registers, including both speech and writing.

In addition to identifying a cross-linguistic pattern of language variation, the results of our study strongly suggest that what has actually been identified is a cross-linguistic pattern of language change. Because variation in the use of UM and UH shows a clear trend across age groups, with younger speakers using UM rather than UH more often than older speakers, it appears that there is a change in hesitation marker usage currently taking place across various Germanic languages, with the use of UM rising over time. This type of *apparent-time* evidence, which is common in sociolinguistic research (see Labov, 1994), is based on the assumption that if a change is taking place, then younger speakers will generally be more likely than older speakers to prefer the linguistic form that is on the rise. This interpretation of our age-based results is strongly supported by our longitudinal analyses of the Philadelphian English, the Norwegian, and (to a lesser extent) the Dutch corpora, which show that the use of UM is rising in real time. Finally, our finding that women consistently use UM more often than men is also consistent with this interpretation, as women have frequently been found to lead linguistic change (see Labov, 1990). This study has therefore uncovered clear evidence that a similar change is taking place in the use of the hesitation markers UM and UH (irrespective of whether one accepts a categorical distinction between the two variants or not) across a range of Germanic languages, with the use of UM as opposed to UH becoming more frequent over time.

This change in the use of hesitation markers is surprising because it is occurring simultaneously across a relatively large and mostly mutually unintelligible set of Germanic languages. Examples of cross-linguistic change are not well attested in the literature and it is unclear how this type of change could have developed or could be maintained. Perhaps the most basic question is whether this cross-linguistic change began in one language and then spread to other languages, or whether it developed in all languages simultaneously. This is a complex puzzle, one for which we cannot provide a definitive answer. In the remainder of this paper, we therefore present a number of possible explanations for this cross-linguistic change, discuss the strengths and weaknesses of each of these explanation, and consider how these competing theories could be tested in future research.

One possible explanation for this cross-linguistic change is that there are independent patterns of change in use of UM and UH occurring in all six of the Germanic languages for various different reasons, which are *coincidentally* all moving in the same direction. Although such an explanation is possible, completely independent changes progressing in unison across six different languages is highly improbable. It is therefore necessary to consider other hypotheses that directly explain why the same basic change is taking place across so many Germanic languages. There would appear to be two general types of *non-coincidental* explanations that could account for these results: the change may have spread through *contact* from one of the languages to the others or a true *parallel change* may be taking place caused by some factor that affects the use of these related hesitation markers across all the languages.

Language contact is one possible explanation for this cross-linguistic change in hesitation marker use. For example, lexical items in one language that refer to new concepts are often borrowed into other languages that do not have words to refer to those concepts, such as the English word ‘computer’. This word was borrowed into Dutch, German and Danish, although not into Norwegian (*datamaskin*) or Faroese (*telda*). English forms, in particular, would appear to be especially likely to spread through contact, because it is one of the primary languages of mass media and the Internet, as well as being commonly used as a second language by many speakers of other Germanic languages. Unfortunately, as all datasets show an increase in UM use for younger speakers, and the age range covered in each data set varies considerably, we are not able to identify when this potential development would have started.

Even though it is well known that linguistic forms can spread through language contact, which furthermore is often led by women (Van Ness, 1995), it is unclear if language contact could explain the type of cross-linguistic change in hesitation marker usage identified in this study. On the one hand, hesitation markers are relatively frequent in the English language, ensuring that they would be present in the language to which non-native English speakers are exposed. The proportion of UM (over UH) is also higher on average in the English language corpora compared to the corpora for other Germanic languages, which is what we would expect if the change originated in the English language. On the other hand, there is a considerable range in the average usage of UM over UH in the English corpora (see Table 2), which in some cases dips below the levels for German speakers in particular. The use of hesitation markers would also generally appear to be a highly subconscious process and the shift in usage of UM versus UH in the English language is a subtle change, only having been identified here through a careful statistical analysis of large amounts of language data. Furthermore, unlike the examples of language contact presented above, both forms involved in this change already existed in all the Germanic languages under analysis, so that it is not the specific form UM that would have spread but a pattern of change that affects a pre-existing alternation.

All of these factors presumably make it more difficult for variation in hesitation markers to spread through contact than, for example, a new word that refers to a new concept. However, perhaps that is what is happening here: UM might have taken on a new meaning or function in English, and it is this meaning or function has spread through contact to other Germanic languages, which already have a comparable form, To some extent we did control for functional differences in the use of UM and UH by including various linguistic predictors in our analyses. For example, UM tended to have a longer duration, was preceded and followed by longer pauses, and was more frequently found at the beginning or end of an utterance than UH. These results are in line with earlier studies (e.g., Clark and Fox Tree, 2002, Shriberg, 1994, Swerts, 1998), which found that UM is more likely to signal a major delay (but see O'Connell and Kowal, 2005). Of course, a longer duration of UM is not surprising, given that UM is essentially UH plus the labial nasal, but the gender and age-related patterns still hold when these potential linguistic differences between the two hesitation markers are controlled for. In addition, for most of the corpora analyzed here, the overall relative frequency of UM and UH combined was found either to be decreasing or have remained relatively steady over real or apparent time, which suggests that there has not been a substantial increase in the use of UM or UH as discourse markers over this period of time. We did not, however, analyze different *linguistic* functions of UM or UH. Most notably, as discussed in the introduction, it is clear that hesitation markers can be used as discourse markers, for instance to manage turn taking during a conversation, or to signal indecision, disagreement, focus, or confusion. If UM, for example, is becoming more common as a discourse marker over time compared to UH in English, then this change could explain the rise of UM in English and could have been passed on to other Germanic languages through contact. It should be noted, however, that relatively comparable age- and gender-related patterns were found in the Twitter data, where UM and UH generally have different functions than in spoken language.

In addition to language contact, a cross-linguistic change could also be the result of some linguistic or extra-linguistic process that causes each of the languages to change independently but in parallel. For example, parallel changes can be a result of general processes of sound change, such as elision, which involves the deletion of segments during speech to facilitate articulation. There does not appear, however, to be any phonological processes that would explain the rise in usage of UM compared to UH over time cross-linguistically, such as a tendency for open syllables to close. In fact, the opposite is true: open syllables are generally more common than closed syllables in languages of the world, and furthermore syllables consisting solely of a vowel, such as UH, tend to develop onsets as opposed to codas over time (Hyman, 2008). It also seems possible that UM could be reduced to UH through elision in natural speech so as to accelerate language production. General processes of phonological change therefore do not appear to explain the results of this study.

Alternatively, a general extra-linguistic force could be responsible for a parallel change in the usage of UM and UH across the six Germanic languages. For example, Biber et al. (2010) found that noun phrase modification in English newspaper writing has become syntactically more complex and compressed over time, and argue that this is due to the increasing amount of information incorporated into newspapers in modern times and the increasing use of word processing technology that has allowed reporters to devote more time to carefully preparing and editing their texts. Similar societal changes could be affecting the usage UM and UH cross-linguistically. For example, although there is general prescription against using both forms in the English language (Erard, 2007), UM is arguably more polite than UH (e.g., "polite yawning" is used to refer to yawning with the mouth closed; Hilgers et al., 2000), given that UH leaves the mouth in an open position and that the UH sound is also common reaction to physical pain, fatigue, sadness, and anger. Given the rise of living standards, education level, mass media, and the service economy in the Western World over the course of the 20th century, it is possible that people have become more self-conscious of their language use, resulting in the rise of UM over UH across Germanic languages. Unfortunately, the datasets analyzed in this study are not suitable for a more detailed diachronic analysis of a potential shift towards more self-conscious language use. Furthermore, to our knowledge there are no scientific studies describing a potential shift towards (or away from) more self-conscious speech during the 20th century.

In conclusion, this study has shown that there is a clear change taking place across modern Germanic languages, with UM rising in frequency relative to UH. Furthermore, we have considered some possible explanations for this surprising cross-linguistic change, with two hypotheses standing out as being most likely. The first explanation is that the change originated in English and spread through contact with other Germanic languages, which have similar forms, possibly reflecting semantic change in the use of UM, i.e. as a discourse marker. The second explanation is that a parallel change is underway due to general societal changes in communication in the Western World, for example with UM increasing in usage because it is more self-conscious than UH. To assess these hypotheses both individually and in conjunction, as well as potentially generating other explanations for the findings of this study, it is necessary to conduct more detailed functional and social analyses of UM and UH usage over time both within and across Germanic languages.

Acknowledgements

We thank Deborah Cameron, Michael Cysouw, Mark Dingemanse, Diansheng Guo, Daniel Ezra Johnson, Alice Kasakoff, John Nerbonne, Andrea Nini, Emily Waibel, and three anonymous reviewers whose comments have helped to improve this manuscript. Furthermore, we thank Thomas Schmidt in helping us to extract the relevant information from the *Forschungs- und Lehrkorpus Gesprochenes Deutsch.*

References

Acton, E. K. 2011. On gender differences in the distribution of um and uh. *University of Pennsylvania Working Papers in Linguistics* 17(2), article 2.

Akaike, H. 1974. A new look at the statistical identification model. *IEEE Transactions on Automatic Control* 19(6): 716-723.

Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Hoboken, NJ, 2nd edition.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. 2003. *Disfluencies signal theee, um, new information*. Journal of Psycholinguistic Research 32: 25-36.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.

Baddeley, J. L., Pennebaker, J. W., & Beevers, C. G. 2013. Everyday social behavior during a major depressive episode. *Social Psychological and Personality Science,* 4: 445-452.
Bamman, D., Eisenstein, J., & Schnoebelen, T. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2): 135-160.
Biber D., Grieve J., & Iberri-Shea H. 2010. Noun phrase modification. In Rohdenburg G. & Schlüter J. (eds.) *One Language, Two Grammars? Differences between British and American English,* 182-193. Cambridge University Press.
Bell L., Eklund R., & Gustafson J. 2000. A Comparison of Disfluency Distribution in a Unimodal and a Multimodal Human–Machine Interface. *Proceedings of the International Conference on Spoken Language Processing (ICSLP) 2000*, 16–20 October 2000, Beijing, China, vol. 3, 626-629.
Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F., & Brennan, S.E. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44: 123-147.
Bosker, H. R., Quené, H., Sanders, T. J. M., & de Jong, N. H. 2014. Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, 75: 104-116.
Braunmüller, Kurt 2011. Faroese Danish Corpus Hamburg (FADAC Hamburg), http://corpora.exmaralda.org/sfb_k8.html.
Brennan, S. E., & Schober, M. F. 2001. How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44(2): 274-296.
Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. 2004. Fisher English Training Speech Part 1 Transcripts LDC2004T19. Web Download. Philadelphia: Linguistic Data Consortium.
Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. 2005. Fisher English Training Part 2, Transcripts LDC2005T19. Web Download. Philadelphia: Linguistic Data Consortium.
Clark, H. H., & Fox Tree, J.E. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84: 73-111.
Coleman, J., Baghai-Ravary, L., Pybus, J., & Grau, S. 2012 Audio BNC: the audio edition of the Spoken British National Corpus. Phonetics Laboratory, University of Oxford, http://www.phon.ox.ac.uk/AudioBNC.
Corley, M., & Stewart, O. W. 2008. Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass*, 2(4): 589-602.
CGN. 2006. Corpus Gesproken Nederlands. Version 2.0. http://tst-centrale.org/nl/producten/corpora/corpus-gesproken-nederlands/6-17
Crystal, D. 1982. Profiling linguistic disability. London: Arnold.
De Leeuw, E. 2007. Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19(2): 85-114.
Deppermann, A., & Schmidt, T. 2014. Gesprächsdatenbanken als methodisches Instrument der Interaktionalen Linguistik - Eine exemplarische Untersuchung auf Basis des Korpus FOLK in der Datenbank für Gesprochenes Deutsch (DGD2). In: Domke, Christine & Gansel, Christa (eds.) *Korpora in der Linguistik - Perspektiven und Positionen zu Daten und Datenerhebung* [=Mitteilungen des Deutschen Germanistenverbandes 1/2014], 4-17.
Erard, M. 2007. *Um...Slips, Stumbles, and Verbal Blunders, and What They Mean.* New York: Pantheon Books.
Fellows, M. D. 2009. *An exploration of emotion language use by preschool-aged children and their parents: Naturalistic and lab settings*. PhD thesis. The University of Texas at Austin
Finlayson, I. R., & Corley, M. 2012. Disfluency in dialogue: an intentional signal from the speaker? *Psychonomic bulletin & review* 19(5): 921-928.

Fox Tree, J. E. 2001. Listener’s uses of um and uh in speech comprehension. *Memory and Cognition* 29: 320-326.
Fraundorf, S.H., & Watson, D.G. 2011. The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language* 65(2): 161-175.
Godfrey, J., & Holliman, E. 1993. Switchboard-1 Release 2 LDC97S62. DVD. Philadelphia: Linguistic Data Consortium.
Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.
HCRC Map Task Corpus 1993. LDC93S12. Web Download. Philadelphia: Linguistic Data Consortium.
Hilgers F. J., van Dam F. S., Keyzers S., Koster M. N., van As C. J., & Muller M. J. 2000. Rehabilitation of olfaction after laryngectomy by means of a nasal airflow-inducing maneuver: the polite yawning technique. *Archives of Otolaryngology–Head & Neck Surgery* 126(6): 726-732.
Hyman, L. M. 2008. Universals in phonology. *The linguistic review 25*(1-2): 83-137.
Johannessen, J.B., Priestley, J., Hagen, K., Åfarli, T. A., & Vangsnes, Ø. A. 2009. The Nordic Dialect Corpus - an advanced research tool. In: Jokinen, Kristiina and Eckhard Bick (eds.) *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA*. NEALT Proceedings Series Volume 4, 73-80.
Jurafsky, D., Ranganath, R., & McFarland, D. 2009. Extracting social meaning: Identifying interactional style in spoken conversation. In: *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics.* ACL, 638–646.
Labov, W. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2(2): 205-254.
Labov, W. 1994. *Principles of Language Change. Volume 1: Internal Factors.* Oxford: Wiley-Blackwell.
Labov, W. 2001. *Principles of Language Change. Volume 2: Social Factors.* Oxford: Wiley-Blackwell.
Labov, W., Rosenfelder, I., & Fruehwald, J. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89(1): 30-65.
Laserna, C. M., Seih, Y. T., & Pennebaker, J. W. 2014. Um... who like says you know. Filler word use as a function of age, gender, and personality. *Journal of Language and Social Psychology* 33(3): 328-338.
Levelt, W. J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14: 41-104.
Levelt, W. J. M., & Cutler, A. 1983. Prosodic marking in speech repair. *Journal of Semantics* 2: 205-217.
Liberman, M.(2005. *Young men talk like old women*. http://itre.cis.upenn.edu/~myl/languagelog/archives/002629.html (first accessed 6 November 2005).
Maclay, H., & Osgood, C. E. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15: 19-44.
Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. 2006. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology* 90: 862-877.
Mehl, M. R., & Pennebaker, J. W. 2003a. The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science* 14: 579-585.
Mehl, M. R., & Pennebaker, J. W. 2003b. The sounds of social life: A psychometric analysis of students’ daily social environments and natural conversations. *Journal of Personality and Social Psychology* 84: 857-870.

O'Connell, D. C., & Kowal, S. 2005. Uh and um revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research* 34: 555-576.

Rayson, P., Leech, G., & Hodges, M. 1997. Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics* 2: 133-152.

Rendle-Short, J. 2004. Showing structure: Using um in the academic seminar. *Pragmatics*, 14(4), 479-498.

Rochester, S. R. 1973. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research* 2(1): 51-81.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. 1991. Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology* 60(3): 362-367.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. PhD dissertation, University of California, Berkeley.

Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of pragmatics* 30(4): 485-496.

Tagliamonte, S. A., & Denis, D. 2008. Linguistic Ruin? Lol! Instant Messaging and Teen Language. *American Speech* 83(1): 3-34. doi:10.1215/00031283-2008-001

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

Tottie, G. 2014. On the use of uh and um in American English. *Functions of Language* 21(1): 6-29.

Van Ness, S. 1995. Ohio Amish women in the vanguard of a language change: Pennsylvania German in Ohio. *American speech* 70(1): 69-80.

Yuan J., Xiaoying X., Lai W., & Liberman M. Submitted. Pauses and pause fillers in Mandarin monologue speech: The effects of sex and proficiency.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. 2014. Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language* 90(3): 669-692.