

A Longitudinal Corpus-Based Study of Hesitation Markers in Mexico City Spanish: *Este* and *Eh* Then and Now

Lamar A. Graham¹¹ University of North Carolina at Chapel Hill

Keywords: discourse markers, Mexico City, sociolinguistics, variation, hesitation markers

Hispanic Studies Review

Vol. 7, Issue 1, 2023

The discourse markers *este* and *eh*, very common in spoken Spanish in Mexico City, frequently function as filled pauses and reformulative markers. Despite their prevalence in Mexico City and other Latin American countries, these markers are critically understudied in linguistic scholarship. Some previous studies discuss the marker *este* in particular with a focus on its social distribution, paying special attention to age, educational attainment, and sex as possible factors conditioning its use.

The current study continues the work of these previous studies, comparing the data between two corpora of spoken Mexico City Spanish. The distribution of *este* and *eh* within the *Habla culta* and *Habla popular* corpora from the 1970s is compared to that of the *Corpus sociolingüístico de la ciudad de México* from the 2000s. Through a variationist analysis of the same factors of sex, age, and educational attainment in both corpora, it is found that the effects of these factors shift in strength between the past and the present.

1. Introduction

In Spanish, and seemingly universally in human language, there exist certain particles that fulfill various functions of guiding discourse that are, accordingly, called *discourse markers* (Martín Zorraquino & Portolés, 1999; Portolés, 2001). These markers carry out various roles such as reformulation (explication and correction), mitigation, and augmentation of the discourse, among many others. Examples (1-4) give instances of these uses.

(1) *Reformulation – explication*

Ella es obstetra, **es decir**, médica para mujeres embarazadas.

‘She is an obstetrician, **es decir**, a doctor for pregnant women.’

(2) *Reformulation – correction*

No comí el chocolate... **más bien**, tomé dos o tres pedazos...

‘I did not eat the chocolate... **más bien**, I took two or three pieces...’

(3) *Mitigation*

A: Cómo te gustó la película?

B: **Pues, no sé**, no fue mi favorita.

‘A: How did you like the movie?

B: **Pues, no sé**, it wasn’t my favorite.’

(4) *Augmentation*

Limpié mi dormitorio. Además, saqué la basura.

'I cleaned my room. *Además*, I took out the trash.'

Both (1) and (2) are reformulative markers, but with different functions. In explicative contexts as in (1), the information before the marker is equivalent to what follows it, simply restated in other turns. This contrasts with the corrective function in (2) in which the information following the marker contradicts what precedes it. Markers with a mitigating function, such as those in (3), reflect the desire of the speaker to reduce the impact of the statement by not committing to it completely. Finally, augmentative markers such as in (4) reflect the speaker's desire to add to the previously expressed information.

Even though most discourse markers (alternatively DMs) are commonly composed of words that exist elsewhere in the language, they do not carry the same lexical-semantic weight in these contexts. That is to say, for example, that the word *bueno* used as a discursive marker does not convey the adjectival meaning of 'good', e.g. *el jamón está bueno* 'the ham is good'. Rather, a speaker would use *bueno* to indicate agreement or acknowledgment of the statement made by their interlocutor – in this case, *bueno* is understood as a communicative device and not an adjective, and the pragmatic weight in the discourse is the acknowledgment of the interlocutor by the speaker. Such is the case with discourse markers evolved from lexical words: their semantic meaning is largely bleached, but their communicative function – pragmatic weight, in other words – is still readily understandable.

Less commonly studied are *hesitation markers*, a subset of discourse markers whose primary role is to affect the structure of the conversation instead of the content. Canonically, the speaker that employs a hesitation marker (alternatively HM, in this paper) during their turn communicates intent to pause, hesitate, or reconsider for any reason. Consider the following examples (5-6) in which speakers use two different markers to carry out the same discursive function:

(5) I: también participé y/ recibimos golpes / *eh*/ hemos recibido de todo/ humillaciones/ toletazos [...]

'I: I also participated and/ we received strikes/ *eh*/ we received all types of/ humiliation/ strikes with clubs [...]

(CSCM,¹ interview 72)

¹ "CSCM" = *Corpus sociolingüístico de la ciudad de México*. Hereafter, "HC" = *Habla Culta* y "HP" = *Habla Popular*. These three corpora will be introduced further in Section 3.

(6) Inf. B.- Pues sí. ¿Ps para qué vamos a pedir más de lo que no tenemos... de querer... **este...** de lo que no hay?
 ‘Inf. B. – Well, yes. Well, for what reason are we going to ask for what we don’t have... to want... **este...** what isn’t there?’
 (HP, interview 20)

In example (5), we see that the particle *eh* carries no semantic information; its function is purely pragmatic. The speaker uses *eh* to indicate that, even though there is a pause or break for whatever reason, her turn is not over and that she wishes to continue. The hearer perceives *eh* within the speaker’s turn and intuitively knows to wait. The marker *este* in example (6) serves the same purpose as *eh* in (5). There are other discourse particles that have this same function such as *mm* and *ah*; moreover, other markers that have different main functions such as *pues* ‘well’, *bueno* ‘good’, and *o sea* ‘that is’, are also employed as HMs in order to fill pauses and maintain turns. These functions of pausing and mediating turns are essential to the structure and coherence of discourse (Blas Arroyo, 1995; Briz Gómez, 1993; Swerts, 1998).

Hesitation markers tend to be stigmatized as signals of poor speaking skills. Through the years, these markers have been labeled as *muletillas* ‘crutches’ (Christl, 1996), *bordones/latiguillos* ‘clichés’ (Stoesslein, 2014), or *disfluencias* ‘disfluencies’, each one of which carries a negative connotation. This negative impression of said markers, of being empty or unnecessary elements, arises from schoolteachers who are of the opinion that their use is a manifestation of poor management or command of the language (Soler Arechalde, 2006).

(7) E: (risa)/ pero no usas muchas muletillas fíjate
 I: ¿no?/ ¿qué muletillas?
 P: bueno como ‘y y y’
 E: o el [‘tons’]
 I: [este]
 E: este
 I: este
 E: [bueno]
 I: [no uso] el ‘este’/ [ese sí]
 E: [¿no verdad?]
 I: te lo corriges dando clase/ o sea no puedes
 E: ándale
 I: estar con ‘este este’/ eso sí lo corriges
 ‘E: (laughter)/ but you don’t use many crutches look
 I: no?/ what crutches?
 P: well, like ‘y y y’
 E: or [‘tons’]
 I: [este]
 E: **este**
 I: **este**
 E: [bueno]

I: [I don't say] '*este*'/ [that yes]
 E: [no, really?]
 I: you correct that out when teaching/ that is, you can't
 E: *ándale*
 I: be there with '*este este*'/ yes you correct that out'
 (CSCM, entrevista 35)

In example (7) here, we see a teacher who admits to avoiding *este*, for example, while she teaches. It is understood from her tone that these types of markers are inappropriate in her role as an educator. This stigma carried by HMs has caused them to be understudied in linguistic research in favor of other markers such as the aforementioned *pues* and *o sea*, as well as *claro* 'of course (lit. 'clear')', *entonces* 'so, then', etc., whose semantic import to the discourse is more readily apparent. But the truth about HMs is that, even though they communicate no semantic information whatsoever, their pragmatic and discursive importance is immeasurable, especially in terms of discursive cooperation.

2. Hesitation markers: overview and previous studies

Even as HMs are now beginning to be more frequently studied for their importance to the fields of sociolinguistics and pragmatics, there has traditionally been a lack of attention paid to these particles. For example, Blas Arroyo (1995) and Soler Arechalde (2006) emphasize the lack of research attention paid to the marker, noting that it has traditionally been considered empty language and "evidence of a poor and inadequate grasp of the oral language" (Soler Arechalde, 2006). Furthermore, she notes that she is not completely in agreement with those accounts, among them Martín Zorraquino and Portolés (1999), that claim that the only use of HMs is of hesitation or turn-keeping. In this section, I detail some of the most relevant previous studies to HM use, function, and behavior.

2.1. Hesitation markers *este* and *eh*: uses and behaviors

Even though the discursive function of hesitation is universal, the world's languages have their own distinctly codified markers to serve this function (Blondet, 2001). Speakers of American English most frequently use *um* [ʊ:m] and *uh* [ə: ~ u:] (see Clark & Fox Tree, 2002 for their analysis of these markers). French speakers tend to use the vowels [ə:] and [œ:] to indicate hesitation. More pertinently, Spanish speakers prefer sounds such as *eh* [e:], *em* [e:m], and *este* [este: ~ ehte:] as manifestations of discursive hesitancy (Erker & Bruso, 2017). The interesting aspect of hesitation is the frequent use of neutral vowels in these markers and, additionally, the tendency to elongate the marker with level or falling intonation (Martínez et al., 2004), which communicates to the hearer that the speaker intends to maintain his or her turn after the pause.

2.1.1. *Eh as a marker*

Blas Arroyo (1995) shows the uses of *eh* as a functional marker which is integral to inter-speaker communication. In his chapter he focuses on two distinct functions of *eh* as a marker: (1) as an interjection at the end of an utterance in the form of a “tag question”, and (2) as a facilitator of cohesion within the turn as a signal of reinforcement, of thematic contrast, or as a guide to the hearer’s understanding. Blas Arroyo observes that it is mistaken to claim that the use of *eh* is only associated with speakers of low education, given that the diversity in use of the marker extends to a variety of communicative contexts, as well as social situations. He proposes at the end that the frequency of use of *eh* as a marker varies inversely as the familiarity of the interlocutors; that is, that usage of *eh* decreases as speaker familiarity increases.

Continuing with the use of *eh* as a marker, Roggia (2012) investigates the frequency and functions of this marker in the speech of the Dominican Republic. He describes *eh* as invariable, extrasyntactic, and conversational, as is the case of other DMs. Roggia points out that *eh* is a polyfunctional marker: “ [...] as prototypically having a hesitation function and peripherally having a reformulation function, functions shared by other information structuring and reformulating DMs” (2012, p. 1796). While in agreement with Blas Arroyo concerning the discursive utility of *eh*, Roggia also finds that the frequency of *eh* in Dominican Spanish is sensitive to social factors such as sex, social class, and age.

2.1.2. *Este as a marker*

Also relevant to the current study is the HM *este* which has also gone understudied in traditional scholarship. Reyes Trigos (2002) describes the presence and the sociopragmatic uses of *este* in narratives. The marker lacks semantic import; ostensibly it was discursivized² from the demonstrative *este* ‘this’³ and now serves no referential purpose whatsoever. Nevertheless, it figures prominently in the development and structuring of discourse. Reyes Trigos mentions the recursivity of the marker, meaning that it can appear both at the onset and in the development of the narrative. Where *este* occurs in a story, it signals the beginning of the next stage of the narrative, such as a new discursive frame or theme, the building of said frame, and the beginning of the end of the story. The most interesting aspect of Reyes Trigos’ work is that she sets aside the metadiscursive hesitation function of *este* in favor of its discourse-structuring functions.

² Ocampo (2006) defines *discursivization* as, essentially, a process paralleling grammaticalization in which a form evolves from more lexical to more discursive/pragmatic.

³ This is debatable given the lack of contextual fit (as remarked by Martínez et al., 2004). The true origins of *este* as a HM are reserved for future research.

In a comparative study of discourse markers in the Spanish of Caracas, Galué (2002) compares the relative frequencies of various markers such as *claro*, *la verdad* ‘the truth (is)’, *mira* ‘look’, *fíjate* ‘check it out’, *okey*, etc. alongside *este*. She classifies *este* as a conversational metadiscursive marker, indicating the taking or keeping of the turn by the speaker. During each turn-keeping instance, the speaker processes information before speaking; this use concurs with Roggia’s (2012) account of the pragmatic functions of *eh*, indicating that these markers are capable of sharing the same discursive space.

2.2. Structural and social factors conditioning HM production

The primary studies of discursive and sociopragmatic variation in HM production upon which the current study is based are those of Soler Arechalde (2006, 2012), Soler Arechalde and Serrano (2010), and Graham (2013, 2018). Toward the question of sociopragmatic variation between specific HMs, Graham (2013) conducts a preliminary study of the use of such markers in the speech of San Juan (Puerto Rico). More recently, Graham (2018) conducted a study of the differences of use of *este* and *eh* within four metropolitan areas of Latin America, including Mexico City.⁴ The commonality among the studies by these authors is the focus paid specifically to Mexico City based on the social and discursive factors listed here.

2.2.1. Position of the marker within the turn

Soler Arechalde (2006, 2012) introduces other functions of *este* – which in Graham (2013) are almost equally applicable to both *este* and *eh* – that are recognizable depending on the position of the marker within the turn. The formation function is ascribed to the marker in turn-initial position, reformulation in turn-internal position, and topic-closing in turn-final position. Focusing on the prototypical function of *eh*, Roggia finds the same: that the nature of the hesitation changes depending on the position of the marker within the turn or utterance. If it occurs at the beginning, it indicates cognitive organization of the discourse on the part of the speaker. Within the turn, it signals to the hearer that the speaker wishes to maintain his or her turn while he or she is organizing thoughts, a characteristic shared with other markers such as *pues* or *o sea* (Cortés Rodríguez, 1991; Félix-Brasdefer, 2006; Schwenter, 1996). When comparing *eh* and *este* production based on utterance position, Graham (2018) finds that speakers across the study tend to prefer using *este* over *eh* in turn-medial position.

2.2.2. Sex

In her more recent study of contextual uses of *este*, Soler Arechalde (2012) contrasts the relative frequencies of *este* according to various social and discursive variables. Working solely with a small set of interviews from

⁴ Alongside Mexico City, the other areas under study were San Juan (PR), Montevideo (UY), and Medellín (CO).

the HC corpus,⁵ Soler Arechalde researches the difference in prevalence of *este* based on the communicative situation, the sex of the participants, the relationship between the speaker and other participants, the topic of discussion, and the time at which the discourse occurs. Regarding personal differences, she finds that women tend to avoid *este* in contexts of greater formality, showing consistency with her earlier study (Soler Arechalde, 2006). It was found in Graham (2013, 2018) that women tended toward *este* more frequently than men, irrespective of other factors.

2.2.3. Register, formality, or educational attainment

In searching for details of social stratification in the use of the marker, Soler Arechalde finds that the vast majority of instances of *este* occur in popular speech versus in higher registers. She observes that discursive contexts – i.e. the type of discourse being examined, such as spontaneous conversation versus guided interviews – condition the occurrence of *este* more frequently than social differences between participants.⁶ Notable about Soler Arechalde's results is the intersection of sex and register: men tend to predominate in the frequency of *este* in higher registers, while women favor it more in popular speech. This reflects an observation by Silva Corvalán (2001) that women tend to be more conscientious of their language in formal situations, preferring to avoid less prestigious forms such as HMs (and *este* in particular). This intersection was not analyzed in Graham (2013, 2018); however, the latter study did concur with Soler Arechalde in that *este* was strongly associated with less education on the part of the speaker.

2.3. The current study: hypotheses

As a continuation of the study of social effects on the choice of HMs by Mexico City speakers, the current study takes advantage of the availability of sociolinguistic corpora from the past and the present. The goal of this type of study is to identify trends in discourse marker usage in two distinct time periods, and, furthermore, to discover how much the use of said markers has changed across eras. This study continues with the factors under consideration in Graham (2013, 2018) of age, sex, and educational attainment of the speakers; a new factor of time period (past versus present) is introduced, within which the social factors will also be analyzed.

As such, the hypotheses of the current study are as follow:

1. *Este* will be the most frequently produced HM across the study.

⁵ Due to the lack of variability in the types of discourse from the HP informants, Soler Arechalde does not quantitatively analyze this corpus.

⁶ Notable in these results from Soler Arechalde (2012) is the fact that, in a previous study by Soler Arechalde and Serrano (2010), neither sex nor the type of interview had any effect on the frequency of *este*. The authors attribute this to a low number of occurrences.

2. Marker production will differ proportionally across the two time periods under study, causing time period to exert a significant influence, though passive, on marker choice.
3. Sex will be a significant factor in HM choice. Following the studies by Soler Arechalde (2006, 2012), Soler Arechalde and Serrano (2010), and Graham (2018), women will show the greater preference for *este* than men. These proportions will differ in the two time periods under study.
4. Educational attainment will be significant as a factor in HM production. Speakers of lower formal education will produce more such markers than more highly-educated speakers, and, furthermore, they will show a stronger tendency toward *este* as a marker. These tendencies will not differ in strength between the past and the present.
5. Age will be a significant factor in HM choice. Younger speakers will be the most likely population to favor *este* as a marker. This tendency will not differ between time periods.

As a departure from Graham (2013, 2018), this study focuses solely on *societal/social* factors determining variation in marker selection, keeping the tradition of the previous work on *este* by Soler Arechalde (2006, 2012). These previous works, specifically the 2006 study, found that *este* was overwhelmingly encountered within a speaker's turn, much more frequently than in turn-initial or turn-final position. This distribution coincides with the finding for Mexico City speakers in Graham (2018), despite the time elapsed between the publishing of each corpus. At this juncture, it is not the expectation that the discursive/pragmatic functions of *este* would have evolved over time in a manner observable in the corpus data, and therefore the factor of position within turn will not be considered. The analysis of this factor may be considered in future studies.

3. Methodology of the current study

3.1. Corpora

In the current study, the data have been compiled from two different corpora. To analyze the speech of the past, the corpus *El habla de la ciudad de México* (Lope Blanch, 1971, 1976; Serrano Morales, 2014) is consulted. These three sections of the corpus consist of sociolinguistic interviews conducted between 1967 and 1975. Each interview was tagged according to the speaker's characteristics, specifically sex, age, and educational level. With the division between the learned norm (*Habla culta*; Lope Blanch, 1971; Serrano Morales, 2014) and popular speech (Lope Blanch, 1976), this facilitates the distinction between highly- and lower-educated speakers. All the speaker data available from this corpus are utilized, which amount to 37

interviews from *Habla popular* (HP) and 56 from the two sections of *Habla culta* (HC). Because of this division between HC and HP, we can distinguish between the sociolect of the higher-educated speakers and that of the lower-educated population.

The analysis of more modern speech incorporates a section of the *Corpus sociolingüístico de la Ciudad de México* (CSCM, Martín Butragueño & Lastra, 2011–2015), which was performed as a component of the *Proyecto para el estudio sociolingüístico del español de España y de América* (PRESEEA). This corpus consists of interviews conducted between 1997 and 2007. The section of CSCM as a PRESEEA component contains 108 transcriptions, and each interview is tagged according to sex, generation (age), and educational level.

The principal benefit of these corpora is that, as they contain decidedly sociolinguistic interviews, each and every phonetic/phonological production by the participants is registered in the transcriptions. That is, instead of ignoring sounds that are not purely lexical, such as for example metadiscursive markers such as *mm*, *ah*, *eh*, etc., these productions are faithfully transcribed. In the CSCM, this information is shown even more explicitly by the inclusion of phonetic variations such as lengthened vowels. This inclusion makes sociopragmatic study of HMs possible, contrasted with literary works within which these markers never appear.

3.2. Token mining and the dependent variable

Because the transcribed interviews are in electronic formats (either plain text files or PDFs), it would be more efficient to use computer software to carry out each phase of the project. The token search is conducted using *AntConc* (Anthony, 2015), which contains tools to delimit searches in order to return all the desired occurrences. Since this study aims to compare the effects of social factors on HM production, it makes sense that the dependent variable should be the speaker's choice of *este* or *eh*.

In order to get every possible occurrence of the markers in questions, including phonetic productions with elongated vowels (such as *esteeee*) and intrusive final consonants (such as *ehm*), it is necessary to use specific search strings in AntConc. The strings are composed accordingly:

- *este** (for all predictable instances of *este*)
- **eh** (for all predictable instances of *eh*)

The asterisks function as wildcards to capture any variant of the marker as transcribed in the interviews, including lengthening (*eeh*, *ehhh*, *esteee*, etc.). Obviously, the wildcard causes other words to appear in the results. Therefore, before the next step in the data analysis, it is necessary to exclude various sequences:

- Demonstrative *este* (*este país* 'this country')

- Tag question *eh* (¿... *eh*?)
- Repetitions of the same marker (*eh eh eh* or *este este este*) after the first in the sequence⁷
- Markers produced by non-participants, e.g. by the interviewers or other bystanders

After removing invalid hits, there were 13,151 usable instances of *este* and *eh* to be analyzed.

In the process of coding tokens for variable rules analysis (VARBRUL), it is necessary to mark the dependent variable and each independent factor with a lone character associated with a distinct category. The syntax of the program used to conduct the VARBRUL recognizes each place in the token string as a factor and each character as a distinct factor level, and the strings are by default case-sensitive. Each instance of *este* was coded as ‘E’ and each instance of *eh* was coded as ‘e’.

3.3. Independent factors and their codes

In the following subsections, the distinct independent factors and their values are detailed. The significance level of each factor will be measured, both on an overall basis and within each particular time period. It bears mentioning that each interview from CSCM involves only one participant (besides the interviewer), and so the coding is straightforward. Many of the HC/HP transcriptions also involve a single informant, though in contrast with CSCM, there are a number of interviews involving two informants. In order to correctly code each marker occurrence in HC/HP, it was necessary to pay closer attention to who was speaking at each moment, particularly in the interviews that involved speakers of either different ages, sexes, or educational levels.

3.3.1. Time period

In general, the most important independent factor in this study is whether the marker in question was produced in the older interviews – from HC/HP – or in the more recent ones – from CSCM.

(8) I: hace mucho que no voy a la facultad/ **este**/ te recargabas ahí/ te/ te acostabas en el pasto/
 ‘I: it’s been a long time since I’ve gone to the building/ **este**/ you could recharge there/ you/ you could lie back on the grass/’
 (CSCM, interview 24, R)

⁷ If a speaker produces a cluster of both markers under study, such as *este eh* or *eh este*, both are included.

(9) Inf. A. -y ¿qué es, precisamente, la actual... **este...** juventud?
Es una rebeldía hacia esa situación de... de guerras nada más.
Realmente eso es.
'Inf. A. -and what is, precisely, the current... **este...** youth? It is
a rebellion toward this situation of... of wars, that's all. Really
that's what it is.'
(HC, interview 21, P)

Instances from the past are coded as 'P', while instances from the present are coded as 'R'.

3.3.2. Age/generation

Following the pattern of previous studies (Graham, 2013, 2018; Soler Arechalde, 2006), the participant's age is one of the factors under consideration. The interviews from CSCM appear with age already tagged in the title of each transcription. Following the PRESEEA coding system, participants are identified according to generation instead of exact age: speakers from 18-34 are of the younger generation '1', those aged 35-54 are of the middle-age or "adult" generation '2', and those 55 and older are of the "elder" generation '3'.

(10) Enc.- ¿Dónde es la Liga?
Inf. B.- La Liga está en Congreso número treinta, aquí, en Tlalpan.
Enc.- Aquí en Tlalpan.
Inf B.- Sí. Y **este...** los martes es la junta. La junta consiste en... en recordar lo... **este...** acontecido en la... en la junta anterior.
'Int.- Where is the *Liga*?
Inf. B.- The *Liga* is on Congress number 30, here in Tlalpan.
Int.- Here in Tlalpan.
Inf. B.- Yes. And **este...** the meeting is on Tuesdays. The meeting consists of... of recording the... **este...** business of the... of the previous meeting.'
(HP, interview 16 – younger generation 1)

(11) Enc.- Pero se amparaban con la pluma.
Inf.- Con la pluma, sí; escribían... preciosamente los dos. Fueron grandes. Del padre Méndez... **este...** ¿cómo se llama?... Gabriel, yo recuerdo el elogio fúnebre que le hizo monseñor Octaviano Valdés... **este...** en el seminario conciliar de México, en sus honras fúnebres.
'Int.- But they exercised their rights with the pen.
Inf.- With the pen, yes; they wrote... precious, the two of them. They were great. From Father Méndez... **este...** what's his name?... Gabriel, I remember the somber eulogy that

Monsignor Octaviano Valdés delivered of him... *este...* in the council seminar of Mexico, in his funerary honors.’
(HC, interview 4 – middle-age generation 2)

(12) I: porque íbamos a llevar/ *eh*/ al santísimo que son *este*/ las hostias consagradas

E: mm

I: y bueno pues para nosotros es muy importante y/ mm// muy importante/ ¿no?/ *este*/ lo del santísimo/ y la gente bien contenta también/ participaron/ *ehm*/ pues eso fue lo de/ Michoacán/ y ahorita estoy esperando el otro paseo (risa)

‘I: because we were going to take/ *eh*/ the most holy which are *este*/ the holy hosts

E: mm

I: and yes well for us it is very important and/ mm// very important/ right?/ *este*/ the most holy/ and the very content people/ participated/ *ehm*/ well that was what it was in Michoacán/ and right now I’m waiting for the next outing (laughter)’

(CSCM, interview 70 – elder generation 3)

The interviews from HC/HP are not explicitly tagged in this manner. However, thanks to the biographical information available at the beginning of each file, which includes the ages of all the participants, it is possible to categorize the speakers into the three generations to which they belong and maintain consistency with the PRESEEA norms.

3.3.3. Educational attainment

For this study, participants are divided along this line: those with college/university degrees are coded as ‘h’ (for *higher education*), and those who did not surpass high school are coded as ‘d’ (for *diploma*).

(13) I: había ahí/ toda clase de verduras/ *eh* ganado no/ toda clase de verduras vendían ahí

‘I: there there were/ all types of greens/ *eh* not cattle/ they sold all types of greens there’

(CSCM, interview 108 – high school or less d)

(14) I: el pensar que la mujer había nacido para para/ ser ama de casa/ para ser madre de familia/ buena esposa/ madre abnegada/ y *este*/ esposa la compañera de su esposo/ en en/ en el hogar y en *este*/ digamos una cosa muy idealizada [...]

‘I: the thought that women had been born to to/ be a housewife/ to be a mother to a family/ good wife/ a self-sacrificing mother/ and *este*/ wife and partner to her husband/

in in/ in the house and in *este*/ let's say it's a very idealized thing
[...]/
(HC, interview 6 – higher education h)

Again, the CSCM transcriptions are tagged according to the educational level of each participant. The set of transcriptions is divided into sections based on the education attained by the participant: without a diploma, with no more than a diploma, and, at the highest level, with a university degree. The HC/HP corpora do not carry these tags within the transcriptions; instead, they tell what the career of each participant is. However, the division between HC and HP leads to the assumption that the HC speakers have a higher educational attainment than those of HP. Therefore, the fact that there are three levels of education specified in CSCM but only two between HC/HP appears to be a complication. There is a clear demarcation, though, between those who have university degrees and those who do not.

3.3.4. Sex

In both corpora – CSCM and HC/HP – interviews are marked according to the sex of each participant. In this study, 'M' indicates *male* or man, while 'F' indicates *female* or woman.

(15) X: ay no sí es bien [bonito]
I: [no] no es cierto pero/ [...] los que viven cerca del desierto
X: mh
I: ellos/ *este*/ tienen menos posibilidad de/ de construir una mejor casa/ y tienen todo muy árido
'X: ay no yes it's very [pretty]
I: [no] it's not true but/ but/ [...] those who live near the desert
X: mh
I: they/ *este*/ have less possibility of/ of building a better house/ and they have everything very dry'
(HC, interview 9 – women⁸ F)

(16) I: [hasta] que llegan a pesar hasta siete/ [siete kilos seis kilos]
F: [¡ihh/ siete kilos!]
I: ¡no! / pues unas gallinotas/ pero gallinotas // *eh*/ llegó a criar unos gallos/ pero así gallotes
'I: [until] they grow to weigh up to seven/ [seven kilos six kilos]
F: [ihh!/ seven kilos!]
I: no! / well some big hens/ but hens // *eh*/ came to raise some roosters/ but big roosters'
(CSCM, interview 37 – man M)

⁸ Interview 9 from HC was one of those with two participants I and X. Both of these participants, though, were women, thus not complicating the coding.

The CSCM interviews correspond with a single participant, and so tagging according to sex is relatively easy. The HC/HP files are not tagged in this manner, but again the biographical information at the beginning of each file gives the necessary details for appropriate coding.

All told, the distribution of participants is as follows, in tables 1 and 2. We begin with the HC/HP informants:

Table 1. Distribution of participants from HC/HP

Education →		<i>Habla Culta</i>		<i>Habla Popular</i>	
↓ Age	Sex →	Women	Men	Women	Men
Elder speakers		15	6	5	7
Middle-age speakers		16	12	7	10
Younger speakers		19	20	9	6
TOTAL = 132		50	38	21	23

The number of participants differs from the number of transcribed interviews due to the fact that, in some of the interviews, there were two informants. This is not the case of the CSCM transcriptions, the participant distribution of which appears below:

Table 2. Distribution of CSCM participants

Education →		<i>Higher education</i>		<i>High school education</i>		<i>Lesser education</i>	
↓ Age	Sex →	Women	Men	Women	Men	Women	Men
Elder speakers		6	6	6	6	6	6
Middle-age speakers		6	6	6	6	6	6
Younger speakers		6	6	6	6	6	6
TOTAL = 108		18	18	18	18	18	18

All told, the characters representing the dependent variable of marker choice and the four independent factors are concatenated, following the symbol ‘(’, in a token recognizable by the computer program GoldVarb (2005), which will be used to conduct the VARBRUL analysis. For example, a sequence such as (EP2dF represents a production of *este* by a middle-aged woman with no education past high school from the past, as in the following example:

(17) Inf. -La niña grande ya tiene diez años.

Enc.- Mm.

Inf.- Y es mala ella. Luego le digo: “Mira, *este...* le digo a m’hija, le digo- tú no te vayas a meter allá a su casa d’ella, ni cosa ninguna”. Porque ai, como viven todas allí donde vive m’hijo el otro... y yo vivo atrás de la vecindá donde ellos viven...

‘Inf. -The big girl is already ten years old.

Int.- Mm.

Inf.- And she is bad. So I’ll tell you: “Look, *este...* - I tell my

Table 3. Frequencies in the past with respect to age

Age		<i>este</i>	<i>eh</i>	Total N	%	Markers/ speaker
Younger speakers	N	1377	622	1999	58.1	37.02
	%	68.9	31.1			
Middle-age speakers	N	611	263	874	25.4	19.42
	%	69.9	30.1			
Elder speakers	N	473	92	565	16.4	17.12
	%	83.7	16.3			
Total	N	2461	977	3438		26.05
	%	71.6	28.4			

daughter, I tell her- don't you go into her house, not for a thing." Because there, how everyone lives over where my other son lives... and I live behind the neighborhood where they live...'

(HP, interview 30)

After entering the token file into GoldVarb, we received various details about the frequencies of marker production and significance of each factor. These results are analyzed in the next section.

4. Results

A total of 13,151 usable occurrences of hesitational *este* and *eh* were retrieved from the two corpora. Throughout this section, we will see the connection between the frequencies of use of each marker and how these frequencies, in association with the defined independent factors, differ between the two time periods under study. In the following subsections the factors will be analyzed for significance, first in the past and then in the present, culminating in an analysis of the significance of time period regarding marker usage.

4.1. Marker usage in the past

There are 3438 instances of the relevant HMs from the past corpora, 2461 of them being *este* (71.6%) and 977 of them being *eh* (28.4%). Let us examine how these frequencies differ with respect to the three independent factors under consideration.

4.1.1. Age

Looking at the percentages of *este* and *eh* associated with the generations in HC/HP, the difference in use among the age groups is evident.

First, all three groups produced *este* more frequently than *eh*. One can see the differences in marker production in general among the generations. Younger speakers produced these markers, without distinction, more frequently than the elder two generations combined.

Table 4. Frequencies in the past with respect to educational attainment

<i>Educational attainment</i>		<i>este</i>	<i>eh</i>	<i>Total N</i>	<i>%</i>	<i>Markers/speaker</i>
High school or less	N	1363	64	1427	41.5	32.43
	%	95.5	4.5			
Higher education	N	1098	913	2011	58.5	22.85
	%	54.6	45.4			
<i>Total</i>	N	2461	977	3438		26.05
	%	71.6	28.4			

Regarding marker choice, the proportion of *este* to *eh* is nearly equal between younger and middle-age speakers, with a frequency of just under 70%. Interestingly and unexpectedly, the *elder* speakers produced *este* at the highest rate, with a frequency of 83.7%.

4.1.2. Educational attainment

With this factor, we can see a large gap between the frequencies of use of the markers.

The participants with higher education produced the two markers fairly evenly, with *este* produced at a 54.6% rate. This outcome is very distinct from that of the lower-education group. Said group demonstrated an extremely strong preference for *este*, producing this marker at a rate of 95.5%. It is also noteworthy that the lower-education group produced the HMs in question at a rate higher than average, while the higher-education group did so at a rate lower than average.

4.1.3. Sex

The raw frequencies of marker usage between men and women in this study reveal distinct tendencies and appear to confirm the observations of Soler Arechalde (2006).

Women produced the higher number of HMs (56.5% of total). In accordance with the comparisons of previous studies by Soler Arechalde (2006) and Graham (2013, 2018), we see that women also demonstrate a relatively higher frequency of *este* in contrast to men.

4.1.4. Preliminary statistical analysis of independent factors

At this point we employ the VARBRUL analysis in order to measure the significance of each factor. The three factor groups – factors plus their levels – are compared to each other, taking into account the relative and absolute frequencies of each, in isolation and in conjunction with the other factors, in order to calculate relative factor weights and tendencies toward one marker or the other ([table 6](#) below) shows the factors and their associated factor weights as calculated by GoldVarb.

Factor weight values of >0.50 in this VARBRUL indicate a preference for *este*, while values <0.50 indicate a dispreference of *este* (and a preference for *eh*). The closer to the extremes of 1.00 and 0.00, the stronger the

Table 5. Frequencies in the past with respect to sex

Sex		<i>este</i>	<i>eh</i>	Total N	%	Markers/speaker
Men	N	984	511	1495	43.5	24.51
	%	65.8	34.2			
Women	N	1477	466	1943	56.5	27.37
	%	76	24			
Total	N	2461	977	3438		26.05
	%	71.6	28.4			

preference in either direction. The factor weight calculations show that, for now, there are two significant factor groups: educational attainment and sex. The calculation of range, which indicates the numerical distance between the highest-weighted factor level and the lowest, is a representation of strength of that factor. In this case, the factor of educational attainment is the strongest, with sex, though significant, being weaker.

The VARBRUL calculations determined that age was not a significant factor in this study. Such a determination normally indicates that the relative frequencies are not appreciably affected by differing factor levels. However, recall from [table 5](#) (and repeated in the “Apps./Total” column of [table 8](#)) that the elder group demonstrated a considerably higher frequency of *este* usage than the other two generations, yet the factor weights do not reflect this higher tendency. When a factor group is determined not to be significant despite such a discrepancy in outcomes, it is necessary to consider the possibility of *interaction* between factor groups. If factors interact, this means that they are not completely independent from one another. The most straightforward manner to determine which factors are interacting is to cross-tabulate two factors and observe the distributions of outcomes. In the interest of space, I will only show below the cross-tabulation of age and sex.

Table 6. Factor weights for past corpus

Group	Factor	Weight	App/Total
Education	High school or less	0.84	0.96
	Higher education	0.24	0.55
	Range = 60		
Sex	Women	0.55	0.76
	Men	0.44	0.66
	Range = 11		
Age	Elder speakers	[0.54]	0.84
	Middle-age speakers	[0.49]	0.70
	Younger speakers	[0.49]	0.69

Total Chi-square = 52.1850

Chi-square/cell = 4.3488

Log likelihood = -1631.388

Table 7. Age and sex cross-tabulated in the past corpus

↓ Sex	Age →	<i>Elder</i>	%	<i>Middle-age</i>	%	<i>Younger</i>	%	Total	%
Men	<i>este</i>	158	92	257	68	569	60	984	66
	<i>eh</i>	13	8	119	32	379	40	511	34
	Σ	171		376		948		1495	
Women	<i>este</i>	315	80	354	71	808	77	1477	76
	<i>eh</i>	79	20	144	29	243	23	466	24
	Σ	394		498		1051		1943	
Σ	<i>este</i>	473	84	611	70	1377	69	2461	72
	<i>eh</i>	92	16	263	30	622	31	977	28
Total		565		874		1999		3438	

Table 8. Factor weights with interaction considered

Group	Factor	Weight	App/Total
Education	High school or less	0.84	0.96
	Higher education	0.24	0.55
	<i>Range = 60</i>		
Age + sex	Older men	0.74	0.92
	Younger women	0.58	0.77
	Middle-aged women	0.50	0.71
	Older women	0.49	0.8
	Middle-aged men	0.47	0.68
	Younger men	0.39	0.6
	<i>Range = 35</i>		

Total Chi-square = 18.5825

Chi-square/cell = 1.5485

Log likelihood = -1613.687

The cross-tabulation of age and sex (as shown in [table 7](#)) makes the picture clear. In terms of sex, we see that elder men far outpace the overall frequency for men (92% *este* versus 66% overall average), while younger men (60%) are slightly below the overall frequency. Regarding women, there is not as strong an effect; elder women slightly outpace the women's average (80% *este* versus 76% overall), and middle-age women are slightly behind.

After identifying the interaction between age and sex, the next step is to recode and perform the VARBRUL analysis again. The factor group of education remains unchanged, and the intersection of age and sex becomes our new factor group, with six levels: older men, middle-age men, younger men, older women, middle-age women, and younger women. Below is the new factor weight table:

With the interaction accounted for, we see that both the educational attainment and the intersection of age and sex are significant. The factor weights associated with the levels of age + sex correspond with the frequencies in [table 7](#). With the intersection as a factor group, now we can correctly identify how age affects marker production: it is dependent upon sex and is

Table 9. Frequencies in the present with respect to age

Age		<i>este</i>	<i>eh</i>	Total N	%	Markers/ speaker
Younger speakers	N	2014	434	2448	25.2	68
	%	82.3	17.7			
Middle-age speakers	N	2712	992	3704	38.1	102.89
	%	73.2	26.8			
Elder speakers	N	2626	935	3561	36.7	98.92
	%	73.7	26.3			
Total	N	7352	2361	9713		89.94
	%	75.7	24.3			

thus not truly independent. Older men show a favoring effect toward *este*, while younger men, the group least disposed to *este*, are represented by a factor weight closer to *eh*. Note that the factor weights for education are identical to the previous analysis: 0.84 for high school or less, and 0.24 for higher education.

4.2. Marker usage in the present

There are 9713 instances of *este* and *eh* as HMs from the present-day corpus, 7352 of *este* (75.7%) and 2360 of *eh* (24.3%). In the same manner as the analysis of the past corpus, let us examine the frequencies according to age, education, and sex and evaluate each factor for significance.

4.2.1. Age

Upon examination of the frequencies of *este/eh* usage among the three generations, we can observe important differences among the groups.

Each generation produces *este* more frequently than *eh*; middle-age and elder speakers actually produce more HMs overall than younger speakers. The most notable aspect of this distribution of *este* versus *eh* is the relative frequency among younger speakers: 82.3% of markers produced by them are *este*, while the two groups of more advanced age produce the marker at a rate closer to that of the population as a whole.

4.2.2. Educational attainment

Again, both groups produce *este* more frequently than *eh*, with the lower-education group producing 67% more markers total than the higher-education group – though, on a per-speaker basis, the higher-education group produces more markers overall.

We can see the considerable divergence in percentages between both groups. The group with the university education produces *este* at a rate of nearly 9 percentage points lower than the average of the entire population. Moreover, the proportion of *este* by the lower-education group is higher than the overall frequency by 5.3 percentage points.

Table 10. Frequencies in the present with respect to educational attainment

<i>Educational attainment</i>		<i>este</i>	<i>eh</i>	<i>Total N</i>	%	<i>Markers/speaker</i>
Higher education	N	2434	1207	3641	37.5	101.14
	%	66.8	33.2			
High school or less	N	4918	1154	6072	62.5	84.33
	%	81	19			
<i>Total</i>	N	7352	2361	9713		89.94
	%	75.7	24.3			

Table 11. Frequencies in the present with respect to sex

<i>Sex</i>		<i>este</i>	<i>eh</i>	<i>Total N</i>	%	<i>Markers/speaker</i>
Men	N	2992	1146	4138	42.6	76.63
	%	72.3	27.7			
Women	N	4360	1215	5575	57.4	103.24
	%	78.2	21.8			
<i>Total</i>	N	7352	2361	9713		89.94
	%	75.7	24.3			

Table 12. Factor weights from the present

<i>Group</i>	<i>Factor</i>	<i>Weight</i>	<i>App/Total</i>
<i>Educational attainment</i>	High school or less	0.57	0.81
	Higher education	0.39	0.67
	<i>Range = 18</i>		
<i>Age</i>	Younger speakers	0.59	0.82
	Middle-age speakers	0.47	0.73
	Elder speakers	0.47	0.74
	<i>Range = 12</i>		
<i>Sex</i>	Women	0.54	0.78
	Men	0.45	0.72
	<i>Range = 9</i>		

4.2.3. Sex

As in the past, the women of the present use more HMs than men.

Women continue to produce HMs more frequently than men, comprising 57.4% of the total. We see that the proportions of *este-eh* between men and women are closer in the present than in the past, though women still predominantly produce *este* more frequently than women.

4.2.4. Statistical analysis of independent factors

As with the study of markers in the past, GoldVarb calculates the weights of each factor level and determines significance of each factor group. Below we see the factor weights as calculated.

Table 13. Factor weights of time period

Group	Factor	Weight	Apps./Total
Time period	Present	[0.50]	0.76
	Past	[0.49]	0.72

The factor group of educational attainment is the strongest according to the analysis, with a range of 18 between the two factor levels. We see that those with less education slightly favor *este*, while those with higher education tend not to prefer *este* as a marker (relatively), implicating a greater preference for *eh*. With respect to age, middle-age and elder speakers show a very slight preference toward *eh* as shown by the factor weights of just under 0.50, while younger speakers favor *este*. Finally, the weakest indicator is sex, with women only slightly favoring *este* and men only slightly favoring *eh*. The VARBRUL analysis determined all three factors to be significant, and, as a departure from the quantitative analysis of the speech of the past, there is no measurable interaction between any of the factors.

4.3. Significance of marker choice between the past and the present

Up to this point, we have discussed the significance of social factors – age, sex, education – with respect to their influences on the choice between *este* and *eh* in hesitation. Having seen the results, the question remains: is the period of time a significant factor in the preference for one marker over another? It is known that *este* is the most frequently produced HM during both the past and the present, and, furthermore, that the percentage of *este* usage in the present is slightly higher than that of the past (see [table 1](#)). Is this discrepancy statistically significant?

The factor weights of each level are nearly equal. Because of this, the VARBRUL analysis calculates, predictably, that the time period during which the interviews took place is not a significant indicator of marker choice or production.

5. Discussion of results

To summarize the results presented in the previous section, we revisit the stated hypotheses. Are the hypotheses of age, education, and sex as significant factors supported in HM choice of the speech of the past and of the present in Mexico City? And are these significances different between the two time periods?

5.1. Age – *este* as a generational marker, or trend in sociopragmatic change?

Regarding age as a factor, we find that it is a significant indicator in the present, with younger speakers producing significantly higher proportions of *este* compared to middle-age and elder speakers. This sharply contrasts with the past, during which the generation most likely to produce *este* was the elder

Table 14. Age across time

Age	Marker	Time period					
		Past	%	Present	%	Σ	%
Younger speakers	<i>este</i>	1377	69	2014	82	3391	76
	<i>eh</i>	622	31	434	18	1056	24
	Total	1999		2448		4447	
Middle-age speakers	<i>este</i>	611	70	2712	73	3323	73
	<i>eh</i>	263	30	992	27	1255	27
	Total	874		3704		4578	
Elder speakers	<i>este</i>	473	84	2626	74	3099	75
	<i>eh</i>	92	16	935	26	1027	25
	Total	565		3561		4126	

generation, once again an unexpected finding. Recall that, in the past corpora, age was not completely independent as a factor; we encounter an interaction with sex.

Another notable finding regarding age across time was the difference in rates of HM production overall between the past and the present.

The elder generation of the past produced a total of 565 markers, which was the fewest across all generations past and present. We see that the middle-age group produced 874 HMs, whereas the younger generation produced 1999 HMs, more than their elders combined. This contrasts considerably with the speakers from the present corpus in which the *younger* generation produces HMs less frequently than their predecessors.

In both time periods the younger generation appears to be an indicator of a new trend. The upshot of HM production by the younger generation in the HC/HP corpora portends an increase in popularity of, or a decrease in stigma surrounding, HMs. By contrast, the lower occurrence of HMs in general by the younger speakers in the CSCM corpus indicates a decrease in popularity of these markers. Interestingly, the rate of variation in HM production in the middle-age generation remains relatively flat from the past to the present.

5.2. Educational attainment – greater acceptance of *este* across time?

According to various studies of these markers in modern or current usage (studies which have already been mentioned in this paper), speakers with higher levels of education tend to produce *este* less frequently on a percentual basis, contrasted with lower-education speakers. This sociolinguistic characteristic is observed in both the past and the present, though the proportions differ:

Table 15. Education across time

Educational attainment	Marker	Time period					
		Past	%	Present	%	Σ	%
Higher education	<i>este</i>	1098	55	2434	67	3532	62
	<i>eh</i>	913	45	1207	33	2120	38

	Total	2011		3641		5652	
High school or less	<i>este</i>	1363	96	4918	81	6281	84
	<i>eh</i>	64	4	1154	19	1218	16
	Total	1427		6072		7499	

We can see that, for the lower-education group, the ratio of *este* to *eh* in the present is not as wide as it was in the past. Even though this group produces a lower proportion of *este* in the present than in the past, nevertheless, it is evident that *este* maintains a considerably strong association with the speech of the less educated. In the data from the past, the higher-education group was more balanced in their choice of HM, with only a slight tendency toward *este*. This tendency becomes stronger in the present, with *este* becoming twice as frequently produced as *eh*. It is possible to describe this proportional increase as a decrease in stigma, though small, around *este* by those with greater educational attainment.

5.3. Sex – marker of identity?

In the same manner as educational attainment, one can observe notable differences in the frequencies of HM production by men and women. There is always a greater frequency of *este* produced by either group, but do these proportions differ across the years?

The production patterns by women do not change much across the years; women produce *este* 76% of the time in the past compared to 78% in the present. We can see, though, a greater increase in use by men across the same time period: 66% *este* production in the past compared to 72% in the present. If there is a conclusion to be drawn by these statistics, it is that it appears that the relationship between sex and *este* usage has weakened slightly over the years.

Table 16. Sex across time

Sex	Marker	Time period					
		Past	%	Present	%	Total N	%
Women	<i>este</i>	1477	76	4360	78	5837	78
	<i>eh</i>	466	24	1215	22	1681	22
	Total	1943		5575		7518	
Men	<i>este</i>	984	66	2992	72	3976	71
	<i>eh</i>	511	34	1146	28	1657	29
	Total	1495		4138		5633	

5.4. Sex and education combined – implications for the value of formality between men and women?

An observation that appears to contradict the findings of Soler Arechalde (2006, 2012), perhaps for the fact that in the current study two HMs are compared as opposed to analyzing just one, is the rate of occurrences of *este* by both sexes. First, I present the data from HC/HP of the intersection of sex and education:

Table 17. Sex with respect to education in the past

Sex	Marker	Education (past)					
		High school or less	%	Higher education	%	Total N	%
Men	<i>este</i>	520	92	464	50	984	66
	<i>eh</i>	44	8	467	50	511	34
	Total	564		931		1495	
Women	<i>este</i>	843	98	634	59	1477	76
	<i>eh</i>	20	2	446	41	466	24
	Total	863		1080		1943	

Nearly three fifths of HMs produced by women with higher education were *este*, demonstrating their tendency toward that marker. Men with higher education, however, did not clearly display a preference for one marker or the other. These outcomes stand in stark contrast to the men and women with less education, both of which show a nearly categorical preference for *este*. In both groups, we find that the women's preference is still stronger toward *este* than that of men.

Now, let us consider the same intersection of data, but using CSCM:

Table 18. Sex with respect to education in the present

Sex	Marker	Education (present)					
		High school or less	%	Higher education	%	Total N	%
Men	<i>este</i>	2023	77	969	64	2992	72
	<i>eh</i>	605	23	541	36	1146	28
	Total	2628		1510		4138	
Women	<i>este</i>	2895	84	1465	69	4360	78
	<i>eh</i>	549	16	666	31	1215	22
	Total	3444		2131		5575	

We can see that, contrasted with the past, women and men with higher educational attainment do not differ as much as the same groups in the past in their use of *este* versus *eh*, with a difference of only 5 percentage points between the sexes. Furthermore, we also see that both men and women, compared to the past corpus, display a clear preference for *este* that either used to be weaker (in the group of women) or that did not exist before (in the

Table 19. Sex with respect to age in the present

↓ Sex	Age →	<i>Elder</i>	%	<i>Middle-age</i>	%	<i>Younger</i>	%	Total	%
Men	<i>este</i>	1089	69	1042	72	861	77	2992	72
	<i>eh</i>	480	31	407	28	259	23	1146	28
	Total	1569		1449		1120		4138	
Women	<i>este</i>	1537	77	1670	74	1153	87	4360	78
	<i>eh</i>	455	23	585	26	175	13	1215	22
	Total	1992		2255		1328		5575	
Total	<i>este</i>	2626	74	2712	73	2014	82	7352	76
	<i>eh</i>	935	26	992	27	434	18	2361	24
	Total	3561		3704		2448		9713	

group of men). Within each sex group, the differences in frequency of *este* use, based on education, are still considerable, but the discrepancy between the two educational levels is less pronounced in the present than in the past. Therefore, we can consider the possibility that, sociolinguistically, Mexico City speakers in the present do not tend to avoid lower-prestige forms as frequently in more formal contexts, be they men or women.

5.5. Sex and age – reduced interaction?

The variable rules analysis found that the factors of sex and age were not wholly independent in the HC/HP corpus. Rather, there was an interaction effect between these two factors that had to be analyzed. No such effect was found in the present corpus however. What does this outcome mean in context?

Recall from [table 7](#) (which I will not reproduce here for reasons of space) that the group which produced *este* at the highest proportion was the elder men, and that the younger men were the group least likely to use *este*. The other four factor levels (middle-aged men, and all three generations of women) did not exhibit a strong preference in either direction. This trend, though, is not retained in the modern corpus, as [table 19](#) shows.

In contrast to the past corpus, elder men are the *least* likely to choose *este* over *eh*, whereas younger men have come to be the second-most likely group to make this same choice, alongside elder women. Thus, we see a return to this preference in the youngest speaker group. An interesting consideration is that the elder group in the CSCM corpus may largely corresponds to the *younger* group in the HC/HP corpus, age-wise. The youngest elder in CSCM – around 55 years old, according to convention – would have been 20 at the oldest when HC/HP was compiled. A possibility to consider is that *este* was simply not as popular among that group speakers of that generation compared to others. As a result, the speakers' speech patterns remained the same as they aged.

6. Conclusions and directions for future study

It has been demonstrated that the same social factors affecting the variation between hesitation markers *este* and *eh* in the speech of Mexico City in the past remain active in the speech of the present. Within each period of interest, the effects of certain social factors on the tendency toward each marker have varied, indicating changes in the social associations and attitudes toward each marker. It has also been demonstrated that the factors are not all independent of one another. There is a significant dependency of age upon sex in Mexico City Spanish in the past, one which does not persist in the present.

The discussion around HM usage in Section 1 of this paper revealed a social stigma associated to these forms. Blas Arroyo (1995) and Soler Arechalde (2006) mention that they are traditionally regarded as improper or imperfect speech, with no communicative value, and conversations within the corpus data, specifically those of CSCM, reflect the presence of this attitude among certain speakers in Mexico City. As alluded to in the subsections discussing educational attainment and its effects on marker choice, the results herein suggest that the association between *este* and poor education has weakened significantly in the time between the compilation of these corpora. It is quite possible that the attitudes surrounding *este*, and HMs in general, have become more relaxed in the modern day. A study of linguistic attitudes surrounding HMs would be a welcome extension of this discussion.

Another possible future direction along this line of research may be a more thorough investigation of the discursive or pragmatic variation between the past and the present, focusing on the myriad of uses of the markers and if these nuances are maintained across time. It is also the hope of the author that the discussion around HM variation in the Hispanophone world would extend to other countries and cities, to include those areas in which other markers are used to indicate hesitation.

Submitted: February 20, 2020 EDT. Accepted: August 31, 2022 EDT. Published: January 19, 2023 EDT.

References

- Anthony, L. (2015). *AntConc*. Waseda University (Japón).
- Blas Arroyo, J. L. (1995). La interjección como marcador discursivo: el caso de *eh*. *Anuario de Lingüística Hispánica*, 11, 81–117.
- Blondet, M. A. (2001). Las pausas llenas: marcas de duda e identidad lingüística. *Lingua Americana*, 8, 5–15.
- Briz Gómez, A. (1993). Los conectores pragmáticos en la conversación coloquial (II): su papel metadiscursivo. *Español Actual*, 59, 39–56.
- Christl, J. (1996). Muletillas en el español hablado. In T. Kotschi, W. Oesterreicher, & K. Zimmermann (Eds.), *El español hablado y la cultura oral en España e Hispanoamérica* (pp. 117–146). Iberoamericana.
- Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84, 73–111.
- Cortés Rodríguez, L. (1991). *Sobre conectores, expletivos y muletillas en el español hablado*. Editorial Librería Ágora.
- Erker, D., & Bruso, J. (2017). *Uh, bueno, em ...*: Filled pauses as a site of contact-induced change in Boston Spanish. *Language Variation and Change*, 29(2), 205–244. <https://doi.org/10.1017/S0954394517000102>
- Félix-Brasdefer, J. C. (2006). Pragmatic and textual functions of *o sea*: Evidence from Mexican Spanish. In T. L. Face & C. A. Klee (Eds.), *Selected proceedings of the 8th Hispanic Linguistics Symposium* (pp. 191–203). Cascadilla Proceedings Project.
- Galué, D. (2002). Marcadores conversacionales: un análisis pragmático. *Boletín de Lingüística (Universidad Central de Venezuela)*, 18, 27–48.
- Graham, L. A. (2013). Comparing hesitation markers in Sanjuanero Spanish. *Diálogo de la Lengua*, 5, 66–77.
- Graham, L. A. (2018). Variation in hesitation: the case of *este* vs. *eh* in Latin American Spanish. *Spanish in Context*, 15(1), 1–26. <https://doi.org/10.1075/sic.00001.gra>
- Lope Blanch, J. M. (Ed.). (1971). *El habla de la ciudad de México. Materiales para su estudio*. UNAM.
- Lope Blanch, J. M. (Ed.). (1976). *El habla popular de la ciudad de México. Materiales para su estudio*. UNAM.
- Martín Butragueño, P., & Lastra, Y. (Eds.). (2011–2015). *Corpus sociolingüístico de la Ciudad de México (CSCM)* (1st ed.). El Colegio de México.
- Martín Zorraquino, M. A., & Portolés, J. (1999). Los marcadores del discurso. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 3, pp. 4052–4213). Espasa.
- Martínez, H., Urdaneta, L., & Domínguez, C. L. (2004). Estudio fonético-sintáctico de algunos marcadores discursivos: propuestas de formalización para desambiguarlos automática o informáticamente. *Lengua y Habla*, 8, 59–79.
- Ocampo, F. (2006). Movement toward discourse is not grammaticalization: the evolution of *claro* from adjective to discourse marker in spoken Spanish. In N. Sagarra & A. J. Toribio (Eds.), *Selected proceedings of the 9th Hispanic Linguistic Symposium* (pp. 308–319). Cascadilla Proceedings Project.
- Portolés, J. (2001). *Marcadores de discurso*. Ariel.
- Reyes Trigos, C. (2002). *Este* y algunos otros marcadores pragmáticos en narraciones orales. *Signos Literarios y Lingüísticos*, 4(1), 1–75.

- Roggia, A. B. (2012). Eh as a polyfunctional discourse marker in Dominican Spanish. *Journal of Pragmatics*, 44(13), 1783–1798. <https://doi.org/10.1016/j.pragma.2012.08.010>
- Schwenter, S. A. (1996). Some reflections on o sea: A discourse marker in Spanish. *Journal of Pragmatics*, 25(6), 855–874. [https://doi.org/10.1016/0378-2166\(95\)00023-2](https://doi.org/10.1016/0378-2166(95)00023-2)
- Serrano Morales, J. C. (2014). *Norma lingüística culta (1964-1970) – nuevas transcripciones 2013*. UNAM. <http://www.iifilologicas.unam.mx/elhablamexico/>
- Silva Corvalán, C. (2001). *Sociolingüística y pragmática del español*. University Press.
- Soler Arechalde, M. Á. (2006). El uso de este... en el habla de la Ciudad de México. In A. V. Espinosa (Ed.), *Actas del XIV Congreso Internacional de ALFAL* (formato CD). Santiago de Chile.
- Soler Arechalde, M. Á. (2012). Algunos factores determinantes y contextos de uso para el marcador discursivo este... en el habla de la ciudad de México. *Anuario de Letras: Lingüística y Filología*, 46, 155–168.
- Soler Arechalde, M. Á., & Serrano, J. C. (2010). El marcador discursivo este... Algunos aspectos prosódicos y estadísticos. In *Actas del X Congreso Nacional de Lingüística, Asociación Mexicana de Lingüística Aplicada y Universidad Autónoma del Estado de México* (formato CD). Toluca.
- Stoesslein, H. E. (2014). Los marcadores del discurso más frecuentes en el español argentino. *Cuadernos Iberoamericanos*, 2(4), 138–146.
- Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30(4), 485–496. [https://doi.org/10.1016/s0378-2166\(98\)00014-9](https://doi.org/10.1016/s0378-2166(98)00014-9)