Revista Asiática de Estudios Iberoamericanos 32.3 (2021): 1-25.
https://doi.org/10.22927/snuibe.2021.32.3.1

A Variationist Approach to Tag Questions in Madrid Spanish – The Case of *¿no? and ¿eh?*

Jihee Hwang
Seoul National University

Hwang, Jihee(2021), "A Variationist Approach to Tag Questions in Madrid Spanish – The Case of *¿no? and ¿eh?*", *Revista Asiática de Estudios Iberoamericanos*, 32(3), 1-25.

Abstract The two frequently used tag question forms in Spanish, *¿no? and ¿eh?*, are considered linguistic variables since they are interchangeable in a discourse without causing any differences in the meaning. However, there has been a lack of analyses on these two forms from the variationist perspective, to my knowledge. The current study attempts to fill this gap by conducting a quantitative analysis on tag question usage in Madrid Spanish with corpus data within the framework of variationist sociolinguistics. The results demonstrated that the linguistic factors analyzed in the present study show similar patterns of usage with those of previous literature, in general: (i) most of the tags are used with declarative anchors, (ii) turn-medial tags are favored, and (iii) tag questions are more likely to be used in order to keep contact with the interlocutor or to get some responses from the interlocutor. On the other hand, the results of extralinguistic factors did not align with the patterns reported previously in the literature: (i) males use more tag questions than females, (ii) younger generations tend to use tag questions more frequently, and (iii) more tag questions are used by speakers with a middle level of education.

Key words Tag question; Variationist approach; Quantitative analysis; Corpus analysis; Madrid Spanish

I. Introduction

Tag questions are short "utterances with an interrogative tag" (Kimps 2018, 1) which are "not truth-conditional and have little or no propositional meaning" (Gómez González 2014, 95). They consist of two parts: an anchor and a tag (cf. Tottie and Hoffmann 2006). 'Anchor' refers to the main clause of the utterance and 'tag' refers to the question attached to the anchor. In general, tags are added to anchors in a form of conversations or written representations of speech (cf. Biber et al. 1999, cited in Gómez González 2014, 95). In the literature, studies on tag questions were mostly carried out qualitatively focusing on their pragmatic properties, and there have not been as many quantitative analyses from the sociolinguistic perspective so far. The present paper aims to fill this gap by comparing quantitatively the two frequently-used tag question forms in Madrid Spanish.

In Spanish, according to Gómez González (2014), *¿verdad?* and *¿no?* are considered to be 'canonical tag questions', which are counterparts of reduced interrogative 'Auxiliary+Subject' pattern in English (e.g. *is/isn't it?*). However, numerous previous studies provide a comparison between *¿no?* and a different discourse marker, *¿eh?* (cf. García Vizcaíno 2005; Rodríguez Muñoz 2009) or focus on analyzing either one of them, paying little or no attention to *¿verdad?* (cf. Blas Arroyo 1995; Ramírez Gelbes 2003; Montañez Mesas 2007). Also, the linguistic data analyzed in the present study revealed low frequency in the use of *¿verdad?* which led me to exclude this tag from the analysis. Therefore, out of the three tag question forms discussed in previous literature (i.e. *¿verdad?*; *¿no?*; *¿eh?*), the two forms which showed higher frequency (i.e. *¿no?*; *¿eh?*) will be considered representative examples in the current study, as in (1), for a comparative analysis between them:

(1) a. es la sensación que tú tienes de siempre, ***¿no?***
(MADR_H13_013, PRESEEA)
b. pues todos fumamos, me incluyo, ***¿eh?***
(MADR_H11_002, PRESEEA)

These two forms are analyzed as linguistic variables in the sense that they are "alternatives (i.e. options) within the same grammatical system which have the same referential value (meaning) in running discourse" (Sankoff 1988, cited in Tagliamonte 2012, 4). For example, in (1b), using the other form (e.g. *pues todos fumamos, me incluyo, ¿no?*) does not make a difference in the interpretation of the utterance.

Moreover, albeit a confusion of terms regarding this linguistic system in Spanish (e.g. *muletillas, apéndices modalizadores, conectores pragmáticos/metadiscursivos, marcadores (interactivos/interaccionales) del discurso, partículas discursivas, marcadores conversacionales, interjección, microunidades lingüísticas polivalentes en el discurso* (cf. Blas Arroyo 1995; Ramírez Gelbes 2003; García Vizcaíno 2005; Montañez Mesas 2007; Rodríguez Muñoz 2009; Gómez González 2014)), the current study employs the term 'tag question' to minimize confusion.

Since most of the studies of Spanish tag questions, particularly the use of *¿no?* and *¿eh?*, have analyzed them as members of a wider set of elements that behave in a similar way (cf. García Vizcaíno 2005) from a discourse-pragmatic perspective, the goal of this study is to approach them from the variationist sociolinguistic perspective by employing quantitative methods in order to draw a different interpretation of the two forms in question. Tag questions are known "to adopt very different forms from one language to another and even within a language" (Cuenca 1997, 3), which makes it ideal for conducting a variationist analysis. Specifically, the present study measures the use of each form (i.e. *¿no?* and *¿eh?*) in the capital city of Spain, Madrid, with data provided by PRESEEA (*Proyecto*

para el estudio sociolingüístico del español de España y de América) corpus.

Various linguistic and extralinguistic factors related to the use of the two forms in question will be analyzed via Rbrul (Johnson 2009) for mixed-effects variable rule analysis. With the statistical findings of the study, the present paper aims to provide interpretations for the possible similarities and discrepancies between the two tag questions frequently used in Madrid.

II. Literature review

1. Qualitative analysis on *¿no?* and *¿eh?*

Rodríguez Muñoz (2009, henceforth RM) approached in a descriptive manner the pragmatic functions that *¿no?* and *¿eh?* fulfill in the discourse, through observations of spontaneous speech samples extracted from the CREA (*Corpus de Referencia del Español Actual*) corpus. As for *¿no?*, its main usage in the discourse is to corroborate or ratify the information or opinion. This 'confirmative' function implies the presence of another interlocutor who, in a certain way, is signaled by the speaker in order to obtain a verbal or non-verbal response, as in (2). It is more usual for the confirmative tag *¿no?* to appear in the turn-final position followed by a response from the interlocutor since turn-final tags "encourage interlocutors to verbally engage in the interaction" (Moore and Podesva 2009, 455) by facilitating their turn transition. However, it is also possible to be used in the turn-medial position, as illustrated in the following example:

(2) SM: *eh, me sale esto aquí en las cartas, ustedes tienen un grupo de amistades en común,* ***¿no?*** (looks up from the table and looks at the camera) / *o, o ¿se juntan con un grupo de personas?*
P: *mmm, no:* (Rodríguez Muñoz 2009, 89)

Different from the confirmative function mentioned above that is used upon requesting verification of information or opinion, another function associated with *¿no?* is characterized as 'phatic'. This function is used to keep contact with the interlocutor in a discourse. Such use of *¿no?* does not require a response from the part of the interlocutor, although sometimes it can be expressed through verbal or non-verbal elements, as in (3):

(3) A: *Porque resulta, ¿sabes?, que ha habido problemas,* ***¿no?****, y allí y con eso, y todas esas historias, está [un poco]*
B: *[¡Sí:], sí ese es el rollo! Está un poco ¿me entiendes?, ¡nervioso!,* ***¿no?*** *¡Ese es el rollo! El que ha habido problemas…*
(Rodríguez Muñoz 2009, 91)

As can be seen in (3), phatic use of *¿no?* tends to occur frequently in the turn-medial position since it does not necessarily require a response from the interlocutor. However, there are cases in which it is placed in the turn-final position.

On the other hand, RM summarizes the discourse-pragmatic functions of *¿eh?* as confirmative, phatic, intensifier, and attenuator. The first two functions are shared with *¿no?*, whereas the other two are the characteristics that differentiate the two forms. First, 'intensifier' refers to the function that reinforces the propositional content of the statement. In most cases, this intensifier *¿eh?* is followed by a paraphrase of the idea conveyed in the anchor in order to clarify its propositional content. Therefore, it is easily followed by an additional explanatory sentence preceded with the marker *o sea* or *es decir*, as in (4):

(4) *Sobre el tema de la vivienda, cuando una familia vive en situación de indigencia […], esas bolsas de deterioro urbano que todos conocemos, […] a esas familias se les da vivienda gratuita,* ***¿eh?*** *[o sea, es decir] del derecho, tenemos desde hace años un plan de realojo de la población marginal […]* (Rodríguez Muñoz 2009, 97)

Second, the attenuator function of *¿eh?* mitigates the propositional content of the utterance. In other words, it attenuates the illocutionary force of the anchor, which are usually imperative, exhortative, or other face-threatening acts:

(5) SM: … *Piscis, me tienes que bajar la tele, ni se te ocurra subirla,* ***¿eh?*** (laughing)
P: *sí, sí* // (Rodríguez Muñoz 2009, 98)

2. Quantitative analysis on *¿verdad?* and *¿no?*

Gómez González (2014, henceforth GG) provides comparative statistics of 'canonical' tag questions within three different languages: British English (718 tokens), Peninsular Spanish (735 tokens), and European Portuguese (1,020 tokens). The independent factors analyzed in her study include (i) frequency, (ii) mood of the anchor, (iii) position of the tag, (iv) polarity of the anchor, (v) distribution across genres, and (vi) functional characteristics. A brief summary of the findings of each factor will be discussed here except for 'distribution across genres', which is irrelevant to the present study.

As for the frequency, Spanish showed 22.04 tag question uses per 10,000 words, twice more frequent than English due to the fact that Spanish tag questions are used in wider contexts and serve more diverse functions than their English counterparts. Also, it was found that *¿no?* is much more frequent than *¿verdad?* with the rate of 94.6% and 5.4%, respectively. This finding led me to question to what extent this pattern of frequency will be similar or different when the use of *¿no?* is compared to *¿eh?* rather than to *¿verdad?*.

Also, it was demonstrated that the vast majority of *¿no?* and *¿verdad?* was attached to anchors with declarative mood (96%), followed by exclamative (2.9%), interrogative (1%), and imperative (0.1%). Although both of the

tag questions mostly combine with declarative, *¿verdad?* didn't show any tokens used with imperatives, whereas *¿no?* was combined with all possible mood choices.

In terms of the position of the tags, the turn-final position (89.4%) was highly frequent while the turn-medial position (10.6%) was not. Tags in the turn-final position show a full range of combinations with the moods of the anchor, but the turn-medial tags are almost exclusively used with declarative anchors. GG points out that "these results seem to indicate that speakers prefer to avoid double marking in both the anchor and tag moves of the construction; in other words, if the tag is marked position-wise, then the anchor tends not to be marked in terms of mood choices" (110).

Regarding the polarity of the anchor, Carvalho and Kern (2019, 468) state that tags showing reverse polarity are more common while "tags that preserve the same polarity as the anchor are also possible". Since all possibilities of polarity combinations between anchors and tags are available —reversed polarity type (e.g. affirmative+negative, negative+affirmative) and constant polarity type (e.g. affirmative+affirmative, negative+negative)—, it is worth analyzing which type is preferred by speakers in authentic conversations. The findings of GG showed that the reversed type (86.9%) is more frequent than the constant type (13.1%), with a detailed hierarchy of frequency as follows: affirmative+negative (86.3%) > negative+negative (8.7%) > affirmative+affirmative (4.4%) > negative+affirmative (0.6%). The author attributes this overwhelming preference for the negative tag (i.e. *¿no?*) to its overall greater frequency of use: *¿no?* (94.6%) vs. *¿verdad?* (5.4%). Along with this account, it is also stated that the comparative phonotactic ease of the articulation or production of the word *¿no?* over *¿verdad?* might be another factor that makes it be preferred by the speakers.

Finally, in GG, the percentage of occurrences based on the functional characteristics was observed in the following hierarchy: Informational (45.6%) > Attitudinal (23.4%) > Regulatory-delaying (15.5%) > Focusing (8.8%) > Others (6.7%). These terms used in GG are different from those of RM, summarized earlier in II.1, however, what they refer to is similar. In fact, given the definitions and examples provided in GG and RM, it is plausible to consider 'Informational (GG) - Confirmative (RM)', 'Attitudinal (GG) - Intensifier (RM)', and 'Focusing (GG) - Phatic (RM)' as equivalents, and the present study will use the latter terminology of RM. The 'Regulatory-delaying' function which refers to the "use by the speaker to organize or to delay the processing of information" (GG, 119) was not studied in RM; thus, this function will be added to the present analysis and dubbed 'Delaying', following GG's term.

Although GG provided a thorough analysis of various factors that affect the use of the two tag question forms quantitatively, it only measured linguistic factors, not considering extralinguistic ones. Additionally, if a mixed-effects model which includes random factor(s) was analyzed statistically, it might have improved our understanding of *¿no?* and *¿verdad?*.

In sum, previous studies provided analyses on discourse-pragmatic functions and comparative statistics based on linguistic factors of Spanish tag questions. From the literature review, it was possible to find some gaps that are needed to be filled, which serve as research questions of the present analysis:

RQ 1: How do extralinguistic factors condition the use of Spanish tag questions?

RQ 2: How will the statistics differ from those of GG if *¿eh?* (instead of *¿verdad?*) is considered to be a variable along with *¿no?*?

RQ 3: If 'Speaker' is included as a random factor in the model, will the statistical results be different?

III. Data and Methods

1. The PRESEEA Corpus

As mentioned earlier, the present study compares the use of two representative forms of Spanish tag questions, using a synchronic sociolinguistic corpus, the PRESEEA. The data represent the spoken language of four different cities in Spain: Alcalá de Henares, Granada, Madrid, and Valencia. Among them, this study particularly analyzes the Madrid corpus since it shows the largest number of tokens of tag questions: 442 tokens in Alcalá de Henares, 392 tokens in Granada, 1,283 tokens in Madrid, and 362 tokens in Valencia.

The PRESEEA corpus consists of recorded conversations conducted by researchers and their informants. The interviews were (semi-)directed with a series of conversation themes such as (i) greetings, (ii) weather, (iii) the place where the informant lives, (iv) family and friends, (v) customs, (vi) danger of death, (vii) important anecdotes in life, (viii) desire for economic improvement, and (ix) ending. Each theme has some sample questions that the interviewer-researchers could make use of during the interviews.

Although the interviews were (semi-)directed with aforementioned themes and questions, the order of each theme could vary following the circumstances of each interview, and the interviewer-researchers interrupted the informants as little as possible for the most naturalistic gathering of the data. Each interview lasted at least 45 minutes and was followed by a questionnaire to collect personal data from the informants.

The social factors provided by the corpus are gender, age group, and level of education of the participants. As for the age group, it is divided into three subgroups: Group 1 consisted of the speakers aged between 20 to 34; Group 2, 35 to 54; and Group 3, older than 55. On the other

hand, level of education is also divided into three subgroups: Low, speakers without education (illiterate) or those who only had primary education or approximately 5 years of schooling; Middle, with secondary education with approximately 10-12 years of schooling; and High, up to university level or higher with more or less 15 years of schooling.

2. Statistical analysis via Rbrul

Rbrul is a program for analyzing linguistic data from the variationist approach using a software program for statistical computing and graphics, R (R Core Team 2020). According to the developer, Daniel Ezra Johnson, Rbrul "is inspired by D. Sankoff's original variable rule program VARBRUL and its successor Goldvarb, as well as by Paolillo's R-Varb."[1] It is said that Rbrul can do everything that Goldvarb provides but better and faster. Also, it offers some useful types of analysis which Goldvarb fails to, such as including continuous predictors, including continuous responses, and fitting mixed models with random factors. The last function, i.e. making it possible to fit mixed-effects models, is the reason why Rbrul was employed in the current study.

Then, why are the mixed-effects models with random factors necessary? When there is a certain tendency observed in the data, it should result from the effects of controlled predictors, not due to the effects of random factors such as speakers. In other words, the observed tendency should not be drawn from interpersonal differences in order to show that the results of the data are generalizable. Concerning the present study, it is possible that some speakers produce much more tag questions than others and/or prefer one of the two forms to the other. To make sure that this possibility does not affect the statistical analysis, 'Speaker' will be included

1) Cited from *Rbrul* Manual (http://www.danielezrajohnson.com/Rbrul_manual.html).

as a random factor in the Rbrul analysis. Moreover, to answer the third and last research question (i.e. If 'Speaker' is included as a random factor in the model, will the statistical results be different?), two different models with and without the random 'Speaker' factor will be compared.

3. Independent factors and their predictions

First of all, all tokens of *¿no?* and *¿eh?* were extracted from the Madrid subcorpus of PRESEEA. A total of 1,192 tokens of the two forms out of 163,267 words were found but only 1,101 of them were analyzed excluding 91 non-tag type uses, as in (6) where the use of *¿no?* is an interrogative sentence without an anchor or main clause.

(6) I: … *nada / no hay nada de eso.*
E: ***¿no?*** *¿no hay nada?* (MADR_H22_026, PRESEEA)

Table 1 below summarizes the relative frequency of each tag form, and it shows a compatible result with GG where the counts of negative tag *¿no?* exceeded those of positive tag *¿verdad?* by a great degree (698 vs. 37, respectively)[2]:

Table 1. Relative Frequency of *¿no?* and *¿eh?*

	¿no?	*¿eh?*	TOTAL
Count	969	132	1,101
Percentage	88.01%	11.99%	100%

Each of the variables is then analyzed with linguistic factors that were studied in the literature: mood of the anchor, polarity of the anchor, position of the tag, and functional characteristics of the tag. In addition,

2) Although it is out of scope for the present study, the variable *¿verdad?* shows 27 counts in the Madrid subcorpus of PRESEEA. It was excluded from the analysis due to its low frequency compared to the other two variables.

three extra-linguistic factors provided from the corpus were also taken into consideration: gender, age group, and level of education.

Regarding linguistic factors, the first two of them are related to anchors whereas the other two are related to tags. Specifically, the mood of the anchor will be coded as declarative, interrogative, exclamative, or imperative and the polarity of the anchor will be coded as either affirmative or negative. With respect to the tags' factors, the position of the tag will be coded as turn-medial or turn-final and its function will be coded as confirmative, phatic, intensifier, or delaying as discussed in II.2. above.

Following previous investigations, it is expected that (i) the majority of the tags will be attached to declarative anchors, (ii) the combination of 'affirmative anchor+negative tag' will be the most frequent token, (iii) turn-final will be a major position where tags can be found, and (iv) confirmative and intensifier will be the main functions that the tags show.

On the other hand, as for extralinguistic factors, previous literature on English tag questions can shed light on the predictions. According to Tottie and Hoffmann (2006), consideration of gender started with Lakoff (1973), who claimed that women show greater usage of tag questions as a sign of insecurity. Since then, various researchers have argued that it is power rather than gender which conditions the use of tag questions. However, there has not been a consensus drawn from those studies: some analyses show a higher frequency of use from powerless people's part, whereas others show the opposite result.

In the case of speaker age, it was found that "younger people use far fewer canonical tag questions than older people" (Tottie and Hoffmann 2006, 304). However, this result cannot be interpreted as older generations favoring tag questions since it is possible that younger generations prefer to use non-canonical forms which were excluded from Tottie and Hoffmann's study.

As it seems that there lacks a general agreement on the social factors conditioning the use of tag questions so far, it is of interest to measure which of the extralinguistic factors provided in the PRESEEA corpus (i.e. gender, age, and level of education) conditions the use of Spanish tag questions, if any. Note that since PRESEEA only contains data from the interview, not from oral speeches between interlocutors who know each other, the question of 'power' will be reserved for future study.

IV. Data analysis and discussions

As mentioned above, the raw frequency of *¿no?* is a lot higher than that of *¿eh?*: 969 tokens vs. 132 tokens, out of 1,101 occurrences in total. This result gives support to GG, where the author mentions that the negative semantics of *¿no?* makes it much more frequent than its positive counterpart, *¿verdad?*. Although it was *¿eh?* not *¿verdad?* which was contrasted in the present study, the negative tag *¿no?* still manifests its predominance when it comes to tag question usage in Madrid, Spain.

Table 2 below summarizes the results of the social factors, analyzed via Rbrul. In terms of factor weight, the factors with their factor weights

Table 2. Summary of the Results of Social Factors for *¿no?* and *¿eh?*

	¿no?	*¿eh?*	Total	Factor weight
Gender				
Male	650 (67.08%)	62 (46.97%)	712 (62.67%)	**0.583**
Female	319 (32.92%)	70 (53.03%)	389 (35.33%)	0.417
Age group				
1 (20 to 34)	291 (30.03%)	17 (12.88%)	308 (27.97%)	**0.698**
2 (35 to 54)	542 (55.93%)	45 (34.85%)	588 (53.41%)	**0.598**
3 (older than 55)	136 (14.04%)	69 (52.27%)	205 (18.62%)	0.225
Level of education				
Low	222 (22.91%)	51 (38.64%)	273 (24.80%)	0.427
Middle	526 (54.28%)	42 (31.82%)	568 (51.59%)	**0.634**
High	221 (22.81%)	39 (29.54%)	260 (23.61%)	0.436

above 0.5 will be interpreted as favoring the use of *¿no?* to *¿eh?*.

These results show that males use more tag questions than females (62.67% vs. 34.33%, respectively), contradicting Lakoff (1973). Again, it could be a matter of power rather than gender which is conditioning the usage, but the power factor can't be measured with the given corpus. Also, it is shown that males slightly favor the *¿no?* form with a factor weight of 0.583. As for the age factor, more than half of the tokens were from Age group 2, with the range between 35-54 years old. With factor weights of 0.698 and 0.598, respectively, the two younger generations favor *¿no?* whereas the oldest group disfavors it. Finally, with the level of education, the middle group shows more than half of the occurrence with a percentage of 51.59%. This group also favors the use of *¿no?* while the other two groups do not show this favoring effect. Below are some examples in which the factors that most strongly condition the use of *¿no?* over *¿eh?* from each factor group are included:

(7) a. … *prácticamente no ha nevado, ¿no?* (MADR_H12_007) [Male]
b. … *ahora casi todo el mundo es hijo único, ¿no?* (MADR_M13_018) [Age group 1]
c. […] *porque yo soy alumna, ¿no?* (MADR_M12_010) [Middle level of education]

To sum up, among the three extralinguistic factors, the factors conditioning the use of *¿no?* were age group and level of education, indicating that the youngest age group (20 to 34 years old) and the speakers with up to the middle level of education favored its use. On the other hand, the oldest generation was the only people who favored the use of *¿eh?*.

Table 3 below summarizes the Rbrul results with the linguistic factors included in the analysis:

Table 3. Summary of the Results of Linguistic Factors for *¿no?* and *¿eh?*

	¿no?	*¿eh?*	Total	Factor weight
Mood				
Declarative	954 (98.35%)	129 (97.73%)	1,083 (98.37%)	**0.828**
Interrogative	10 (1.03%)	0 (0.00%)	10 (0.91%)	**>.999**
Exclamative	6 (0.62%)	2 (1.51%)	8 (0.73%)	**0.628**
Imperative	0 (0.00%)	1 (0.76%)	1 (0.09%)	<.001
Polarity				
Affirmative	863 (89.06%)	105 (79.55%)	968 (87.92%)	**0.593**
Negative	106 (10.94%)	27 (20.45%)	133 (12.09%)	0.407
Position				
Medial	642 (66.25%)	73 (55.30%)	715 (64.94%)	**0.647**
Final	327 (33.75%)	59 (44.70%)	386 (35.06%)	0.353
Function				
Confirmative	332 (34.26%)	63 (47.73%)	395 (35.88%)	0.173
Phatic	369 (38.08%)	54 (40.91%)	423 (38.42%)	0.461
Intensifier/Attenuator	102 (10.53%)	7 (5.30%)	109 (9.90%)	**0.653**
Delaying	166 (17.13%)	8 (6.06%)	174 (15.80%)	**0.749**

As can be seen from Table 3, nearly every token was found with a declarative anchor. Interrogatives only had *¿no?* as their tag, whereas the only one token of the imperative anchor had *¿eh?* as its tag. That the imperative anchor took the positive *¿eh?* form is not surprising given that "constant polarity is the unmarked case" (Kimps and Davidse 2008, 720) in imperative tag questions. Apart from the interrogative anchors that were only followed by the tag *¿no?*, which resulted in the largest factor weight, declarative anchors highly favor the use of *¿no?* along with exclamative anchors also favoring the negative form. Also, affirmative anchors tend more to be followed by tags, compared to negative anchors. In more detail, the combination of an affirmative anchor with a negative tag is slightly preferred. As opposed to GG, turn-medial tags outweighed turn-final tags, and it was found that the *¿no?* form was preferred in medial position. Concerning discourse-pragmatic functions that tags show, the order of frequency resulted in 'phatic > confirmative > delaying > intensifier/attenuator'. The negative tag *¿no?* was highly favored with delaying

function, followed by intensifier/attenuator. On the other hand, with the confirmative function, *¿eh?* was highly preferred. Again, some examples which include the factors that most strongly condition the use of *¿no?* over *¿eh?* from each factor group are provided:

(8) a. *¿y esa no la has encontrado todavía? ¿no?* (MADR_H12_007) [Interrogative anchor]
b. […] *estás lejos, ¿no?, es que ahora estoy* […] (MADR_M13_018) [Affirmative anchor]
c. …*más mayores que yo, ¿no?, o que, dan, te dan un* …(MADR_M12_010) [Turn-medial]
d. … *pero también es un barrio colindante justo, ¿no?, entonces el colegio estaba allí, ¿no?* (MADR_H23_033) [Delaying function]

In sum, the results show that (i) all the mood for the anchors but the imperative, (ii) turn-medial position, and (iii) delaying and intensifier/attenuator functions favor the use of *¿no?*. The factors which show preference of *¿eh?* were confirmative function and turn-final position. When *¿no?* and *¿eh?* are considered together, the raw frequency shows that tag questions are used more frequently with declarative and affirmative anchors with phatic and confirmative functions in turn-medial position.

Now, let us look into some interactions of the factors that were found to affect the variable use of the two tag forms in question. Firstly, Table 4 below shows the interaction of age group with the level of education. Since the third linguistic factor, gender, did not show a significant result, only the interaction of the aforementioned two factors was analyzed.

As for the use of tag questions in general, Age group 2 with a Middle level of education (23.80%) and Age group 1 with a Middle level of education (19.26%) showed the highest frequency of tag usage. This result can be interpreted as the younger generation, up to 54 years old, with secondary education use more tag questions than the oldest generation in

Table 4. Summary of the Results for "Age group : Level of education" Interaction

	¿no?	*¿eh?*	Total	Factor weight
Age group : Education				
1 : Low	46 (4.75%)	9 (6.82%)	55 (4.99%)	0.424
1 : Middle	208 (21.47%)	4 (3.03%)	212 (19.26%)	**0.690**
1 : High	37 (3.82%)	4 (3.03%)	41 (3.72%)	0.379
2 : Low	132 (13.62%)	25 (18.94%)	157 (14.26%)	0.318
2 : Middle	242 (24.97%)	20 (15.15%)	262 (23.80%)	0.241
2 : High	168 (17.34%)	1 (0.76%)	169 (15.35%)	**0.871**
3 : Low	44 (4.54%)	17 (12.88%)	61 (5.54%)	**0.745**
3 : Middle	76 (7.84%)	18 (13.64%)	94 (8.54%)	0.586
3 : High	16 (1.65%)	34 (25.76%)	50 (4.54%)	0.195

Madrid Spanish. It is opposed to what Tottie and Hoffmann (2006) found with English tag questions: older generations use more canonical tag questions than younger generations.

To contrast the preference for each tag form, factor weights are to be discussed. With the highest factor weight of 0.871, Agre group 2 with a High level of education highly favor the use of *¿no?*, followed by Age group 3 with a Low level of education (0.745), and Age group 1 with a Middle level of education (0.69). In contrast, Age group 3 with a High level of education highly prefer the use of *¿eh?*, followed by Age group 2 with a Middle level of education.

For the analysis of the interaction between the linguistic factors, Rbrul showed that there are not any statistically significant interactions with mood as a factor. This result is not surprising given that more than 90% of the tags were attached to declarative mood. With the other three factors (i.e. Polarity, Position, Function), the result is summarized in the table below:

Table 5. Summary of the Results for "Age group : Level of education" Interaction

	¿no?	*¿eh?*	Total	Factor weight
Polarity : Position				
Affirmative : Medial	573 (59.13%)	57 (43.18%)	630 (57.22%)	0.516
Affirmative : Final	290 (29.93%)	48 (36.36%)	338 (30.70%)	0.484
Negative : Medial	69 (7.12%)	16 (12.12%)	85 (7.72%)	0.516
Negative : Final	37 (3.82%)	11 (8.33%)	48 (4.36%)	0.484
Polarity : Function				
Affirmative : Confirmative	294 (30.34%)	52 (39.39%)	346 (31.43%)	0.431
Affirmative : Phatic	327 (33.75%)	43 (32.58%)	370 (33.61%)	0.455
Affirmative : Intensifier	88 (9.08%)	4 (3.03%)	92 (8.36%)	0.563
Affirmative : Delaying	154 (15.89%)	5 (4.55%)	160 (14.53%)	0.551
Negative : Confirmative	38 (3.92%)	11 (8.33%)	49 (4.45%)	0.569
Negative : Phatic	42 (4.33%)	11 (8.33%)	53 (4.81%)	0.545
Negative : Intensifier	14 (1.44%)	3 (2.27%)	17 (1.54%)	0.437
Negative : Delaying	12 (1.24%)	2 (1.52%)	14 (1.27%)	0.449

Although none of the interactions can be interpreted as statistically significant —all of them show factor weights in between 0.4 and 0.5—, some interesting findings could be worth mentioning. First of all, the combination of 'affirmative anchor : turn-medial tag' was used in more than half of the total occurrences. When it comes to the interaction of polarity with function, the order of 'affirmative anchor : phatic tag (33.61%) > affirmative anchor : confirmative tag (31.43%) > affirmative anchor : delaying tag (14.53%)' showed higher frequency. It was also an expected result because phatic tags are mostly found in turn-medial positions due to their pragmatic meaning, i.e. to keep contact with the interlocutor, not necessarily expecting her/his answer.

Also, the type of the data needs to be taken into consideration for a better interpretation of the results. Since the corpus consisted of interviews between unacquainted people (i.e. a researcher and one or two speakers living in the city of Madrid), most of the tag questions were found in the answers of the participants. Given the situation, it is plausible that turn-medial tags with phatic function were the most frequent types:

when participants try to answer the question while maintaining their turns, they would use tag questions to keep contact with their interlocutors in between their answers.

Finally, in order to answer the last research question which asked whether fitting a mixed-effects model would show any difference in the statistical analysis, I included 'Speaker' as a random factor and examined the revised model in Rbrul. It was reported through the program that overall factor weights were almost identical between the two models. In other words, the Rbrul results with a mixed-effects model showed no difference in terms of factor weights shown in Tables 2-5. However, after including the random factor, 'Speaker', it was recommended to exclude some of the factor groups to get a model with a better fit. Specifically, in the mixed-effects model, the best model fit was obtained when 'Education', 'Function', 'Gender', and 'Mood' were excluded in the analysis. With the rest of the factor groups (i.e. 'Age group', 'Polarity', and 'Position'), I present below a mosaic plot which demonstrates in which condition(s) the two tag forms are favored and disfavored.

A darker blue box in the plot indicates that the conditions surrounding it highly favor one dependent variable over the other. In Figure 1, there are three darker blue boxes, all on the upper side of the plot which is for the *¿eh?* form. These three boxes demonstrate that the tag *¿eh?* is most likely to be used by the oldest generation (Age group 3) when it follows an affirmative anchor. Also, when the same group of speakers uses a tag form after a negative anchor, *¿eh?* will be located in the turn-final position with high probability.

There are two blue boxes in the figure, which also indicates a preference for one of the forms but its degree is weaker than the aforementioned darker ones. The first blue box on the upper side can be interpreted that the oldest generation favors *¿eh?* when it is used in a turn-medial position

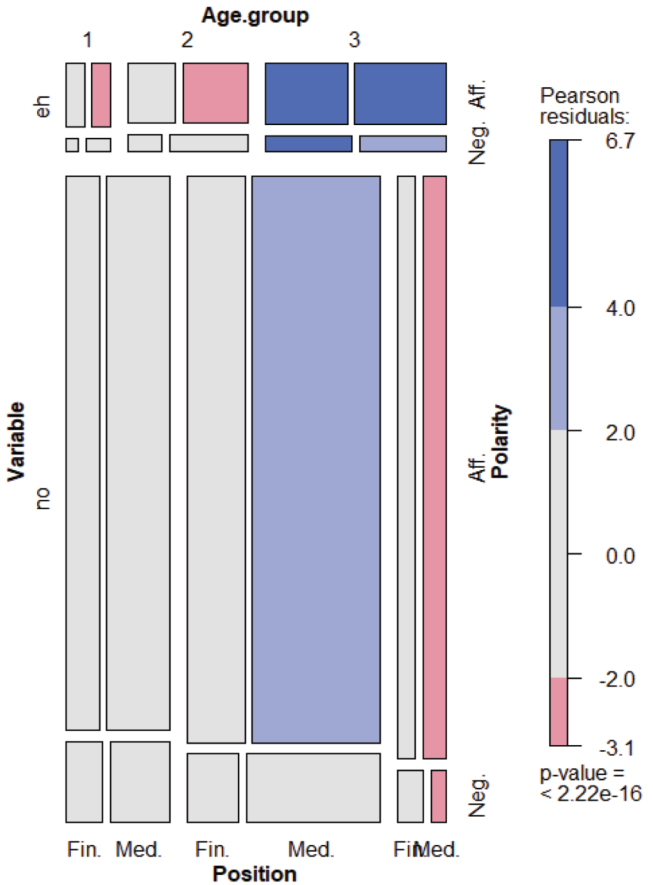

Figure 1. Significance of Relevance between the Dependent Variable, Age Group, Position, and Polarity

following a negative anchor. The other blue box indicates that the second generation (Age group 2) prefers the negative form *¿no?* in turn-medial position with an affirmative anchor.

In contrast, red boxes demonstrate disfavoring effect of the dependent variable in question. The first two red boxes on the upper side show that the two younger generations are likely to avoid using *¿eh?* in turn-medial positions with affirmative anchors. On the other hand, the other two red

boxes represent the probability that the oldest generation disfavor the *¿no?* tag form in turn-medial position with either polarity of the anchor.

V. Concluding remarks

Based on the Madrid subcorpus provided by PRESEEA, the present study aimed to analyze quantitatively the use of tag questions in Madrid Spanish. Although there has been numerous research on Spanish tag questions, especially from the discourse-pragmatic point of view, quantitative analysis with a variationist approach was previously lacking. To fill this gap, the present study intended to measure to what extent various linguistic and extralinguistic factors condition the use of the two frequent tag question forms in Spanish, *¿no?* and *¿eh?*.

The results of the corpus analysis can be explored in two folds: how is the pattern of usage of *¿no?* and *¿eh?* as a whole in regard to percentage, and which form among the two is favored under which conditions. First of all, it was found that (i) males use tag questions more frequently than females; (ii) the middle age group shows the highest frequency followed by the youngest, with the oldest generation using them less frequently; (iii) speakers with a middle level of education use more tag questions followed by speakers with a low and high level of education; (iv) almost every tag questions were attached to declarative anchors, and the only one imperative anchor found in the corpus was followed by *¿eh?*; (v) the majority of the tag questions were found in the combination of an affirmative anchor with the negative tag, *¿no?*; (vi) more tags were found in turn-medial position than in turn-final positions; and (vii) the most frequent function that the tags served was phatic, followed by confirmative, intensifier, and delaying.

In terms of the choice of forms, (i) males slightly favor *¿no?* whereas

females use more *¿eh?* form; (ii) the youngest two groups favored *¿no?* although the oldest generation favored *¿eh?*; (iii) speakers with a middle level of education used more *¿no?* form whereas the other two groups favored the use of *¿eh?*; (iv) all the other moods except for imperative favor the *¿no?* form; (v) affirmative anchors slightly favored *¿no?* whereas *¿eh?* was favored by negative anchors; (vi) *¿no?* was used more in turn-medial position while turn-final position favors *¿eh?*; and (vii) delaying and intensifier functions favor the use of *¿no?* whereas the confirmative function favors *¿eh?*, while the phatic function does not favor either form.

Concerning some interactions between different linguistic factors, it was found that (i) the combination of an affirmative anchor with turn-medial tag was used more than half of the total occurrences; and (ii) the affirmative anchor followed by a tag with phatic function showed the highest frequency.

Finally, when the effect of the random factor was checked, it was recommended to only include the following variables in the analysis to get the best model fit: 'Age group', 'Polarity', and 'Position'. This model predicts that the oldest generation is more likely to use *¿eh?* with an affirmative anchor, both in turn-medial and -final positions. However, when it follows a negative anchor, it is more probable to be used in the turn-final position.

On the other hand, for the extralinguistic factors, all the predictions based on previous literature on English tag question usage were proved not to be compatible with the present analysis. It could possibly be due to the difference between the two languages or the type of the data (i.e. interview data). However, it is also plausible that there is another factor affecting the usage of tag questions in Spanish but not in English. In order to get a further understanding of this issue in question, a future study with a different type of data (e.g. spontaneous dialogues between acquaintances)

measuring the same factors would give us more insights into the possible reasons behind the discrepancies between the two languages.

References

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan(1999), *Longman grammar of spoken and written English*, Harlow: Pearson Education.

Blas Arroyo, José Luis(1995), "La interjección como marcador discursivo: el caso de *"eh"*," *Anuario de lingüística hispánica*, Vol. 11, pp. 81-118.

Carvalho, Ana M. and Kern, Joseph(2019), "The permeability of tag questions in a language contact situation: The case of Spanish-Portuguese bilinguals," *Pragmatics*, Vol. 29, No. 4, pp. 463-492.

Cuenca, María Josep(1997), "Form-use mappings for tag questions,". Wolf-Andreas Liebert, Gisela Redeker and Linda R. Waugh(eds.), *Discourse and perspective in cognitive linguistics*, Amsterdam/Philadelphia: John Benjamins, pp. 3-19.

García Vizcaíno, María José(2005), "El uso de los apéndices modalizadores *¿no?* y *¿eh?* en español peninsular," Lotfi Sayahi and Maurice Westmorland(eds.), *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics*, Somerville, MA: Cascadilla Proceedings Project, pp. 89-101.

Gómez González, María de los Ángeles(2014), "Canonical tag questions in English, Spanish and Portuguese: A discourse-functional study," Marie-Aude Lefer and Svetlana Vogeleer(eds.), *Genre- and register-related discourse features in contrast*, Amsterdam/Philadelphia: John Benjamins, pp. 93-126.

Johnson, Daniel Ezra(2009), "Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis," *Language and Linguistics Compass*, Vol. 3, No. 1, pp. 359-383.

Kimps, Ditte(2018), *Tag questions in conversation: A typology of their interactional and stance meanings*, Amsterdam/Philadelphia: John Benjamins.

Kimps, Ditte and Kristin Davidse(2008), "Illocutionary force and conduciveness in imperative constant polarity tag questions: A typology," *Text & Talk*,

Vol. 28, No. 6, pp. 699-722.

Lakoff, Robin(1973), "Language and woman's place," *Language in Society*, Vol. 2, No. 1, pp. 45-80.

Montañez Mesas, Marta Pilar(2007), "Marcadores del discurso y posición final: La forma *¿eh?* en la conversación coloquial española," *Estudios de Lingüística Universidad de Alicante (ELUA)*, Vol. 21, pp. 261-280.

Moore, Emma and Robert Podesva(2009), "Style, indexicality, and the social meaning of tag questions," *Language in Society*, Vol. 38, No. 4, pp. 447-485.

R Core Team(2020), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramírez Gelbes, Silvia(2003), "La partícula "eh" y la Teoría de la Relevancia: Un ejemplo de contenido procedimental," *Estudios Filológicos*, Vol. 38, pp. 157-177.

Rodríguez Muñoz, Francisco José(2009), "Estudio sobre las funciones pragmadiscursivas de *¿no?* y *¿eh?* en español hablado," *Revista de Lingüística Teórica y Aplicada (RLA)*, Vol. 47, No. 1, pp. 83-101.

Sankoff, David(1988), "Sociolinguistics and syntactic variation," Frederick Newmeyer(ed.), *Linguistics: The Cambridge survey*, Cambridge: Cambridge University Press, pp. 140-161.

Tagliamonte, Sali(2012), *Variationist sociolinguistics: Change, observation, interpretation*, Oxford: Wiley-Blackwell.

Tottie, Gunnel and Sebastian Hoffmann(2006), "Tag questions in British and American English," *Journal of English Linguistics*, Vol. 34, No. 4, pp. 283-311.

Jihee Hwang

Seoul National University
jiheeh@snu.ac.kr

Submission: November 28, 2021
Revision Date: December 21, 2021
Approval Date: December 21, 2021

Análisis variacionista de las preguntas de etiqueta en español – El caso de *¿no?* y *¿eh?*

Jihee Hwang
Universidad Nacional de Seúl

Hwang, Jihee(2021), "Análisis variacionista de las preguntas de etiqueta en español – El caso de *¿no?* y *¿eh?*", *Revista Asiática de Estudios Iberoamericanos*, 32(3), 1-25.

Resumen Los dos formularios de preguntas de etiqueta de uso frecuente en español, *¿no?* y ¿eh?, se consideran variables lingüísticas porque son intercambiables en un discurso sin causar diferencias de significado. Sin embargo, hasta ahora, no ha habido suficiente análisis sobre estas dos formas desde la perspectiva variacionista. El presente estudio intenta llenar este vacío mediante la realización de un análisis cuantitativo sobre el uso de preguntas de etiqueta en español con datos de los hablantes madrileños en el marco de la sociolingüística variacionista. Los resultados demonstraron que los factores lingüísticos analizados en el presente estudio muestran patrones de uso similares a los de la literatura previa, en general: (i) la mayoría de las etiquetas se utiliza con anclajes declarativos, (ii) los hablantes favorecen las etiquetas en la posición medial del turno y (iii) es más probable que se utilicen preguntas con etiquetas para mantener el contacto con el interlocutor o para obtener sus respuestas. Por otro lado, los resultados de los factores extralingüísticos no se alinearon con los patrones informados anteriormente: (i) los hombres usan más preguntas de etiqueta que las mujeres, (ii) las generaciones más jóvenes tienden a usar preguntas de etiqueta con más frecuencia y (iii) más preguntas de etiqueta fueron utilizados por hablantes con un nivel de educación medio.

Palabras clave Preguntas de etiqueta; Enfoque variacionista; Análisis cuantitativo; Análisis de corpus; Español madrileño