The Role of Discourse Topic in Evidentiality Marking: Variable *(De)queísmo* in Caracas[1]

Matthew Kanwit
University of Pittsburgh

1. Introduction

In Spanish, the majority of verbs require only the use of the complementizer *que* between them and their declarative complement, although there is also a subset of verbs that require the preposition *de* prior to *que* (e.g., *enterarse de que* "to find out"). Interestingly, "nonstandard" use of *de* preceding *que* in verbs that do not prescriptively require *de* is not uncommon (traditionally known as *dequeísmo,* as seen in [1]), while a parallel phenomenon exists in which the prescriptively required *de* is omitted prior to *que* (*queísmo,* as seen in [2]).

(1) …No se sabía **de que** eso iba a pasar…
"It wasn't known that that was going to happen."
(*Estudio Sociolingüístico de Caracas)*

(2) …Ni nos enteramos (**Ø**) **que** hay pelea…
"We didn't find out that there was a fight either."
(*Estudio Sociolingüístico de Caracas)*

Following Schwenter[2], these two phenomena can be considered under the all-encompassing term (*de)queísmo*, which here will be used to refer to both types of variation and which signals the optionality of the preceding *de*, as seen in (1) and (2).

While (*de)queísmo* has been investigated from many different perspectives, studies that have focused on the correlation between discourse topic and the variable use of *de* are notably lacking. Given the fact that pronominal *de* has been shown to function as an evidentiality marker that distances the speaker from the propositional content that follows it (Schwenter) and the view that certain types of discourse topics may require more distancing than others (e.g., Flores-Ferrán), the current study aims to analyze this possible correlation within the *Estudio Sociolingüístico de Caracas* corpus (Bentivoglio & Sedano). Furthermore, since 3rd person subjects are said to favor the use of *de,* as speakers mark that the following propositional content comes from an external source (Schwenter), the current study will also consider the possible role of discourse topic within the number of the main-clause subject (i.e., 1st, 2nd, or 3rd person). The paper begins with a brief overview of evidentiality, which is followed by a review of relevant research on early descriptive studies and true variationist studies of *(de)queísmo*, along with a brief review of studies that have analyzed the effect of discourse topic on linguistic production in Spanish. We then provide a detailed description of our participants, our linguistic and sociolinguistic

[1] I am thankful to Manuel Díaz-Campos, Viola Miglio, and two anonymous reviewers for their helpful suggestions. All errors, of course, are my own. An earlier version of this project appeared in the *Indiana University LC Working Papers.*

[2] Note that the format of the current journal is to only include the year of publication for authors who are cited in two or more studies. Otherwise, the year of publication can be found in the reference list.

variables, and the analysis conducted. Next, we present our results for both *queísmo* and *dequeísmo* and consider the implications of these findings.

2. Review of the Literature

First, we begin with an overview of some of the basic notions associated with evidentiality. Next, prior research of *(de)queísmo* is presented according to early descriptive studies and later variationist studies, which will receive more emphasis based on the variationist nature of the current study. Finally, a brief overview of the effect of discourse topic on the production of utterances in Spanish will be presented, as it will be central to the current analysis.

2.1. Background on Evidentiality

Evidential markers are used to indicate some characteristic of the source of the information when a speaker makes an assertion (Bybee, Perkins, & Pagliuca; Chafe & Nichols; Schwenter; Willett). Furthermore, there is a subtype of evidential grams known as indirect evidentials, which indicate that the speaker has received the information only via indirect evidence (Bybee et al.; Willet). Indirect evidentials thus typically indicate, by virtue of the speaker's having only indirect knowledge of the asserted proposition, that the speaker is therefore not completely committed to the truth value of the proposition, which adds an epistemic value to the utterance (Palmer). In other words, the speaker does not claim unconditionally that the information is accurate (Bybee et al.). The connection between evidentiality marking and (*de)queísmo* will be further considered in the current section in the review of previous work by Schwenter.

2.2. Early Descriptive Studies of *(De)queísmo*

Among the first studies to pursue the topic of (*de)queísmo* was Bentivoglio (1976), which analyzed data from the *Habla de Caracas* corpus. Results indicated that there were zero cases of *dequeísmo* with non-pronominal verbs (e.g., *hablar*). Of the eleven pronominal verbs included in the study, only five demonstrated alternation between *que* and *de que* (i.e., *acordarse, convencerse, enterarse, recordarse* and *olvidarse*), whereas the other six always co-occurred with *de que* (i.e., *admirarse, impresionarse, ocuparse, preocuparse, quejarse* and *tratarse*). The study was a descriptive first analysis of the contexts in which the alternation does and does not occur, but it was not a true variationist study, since there was no analysis of the linguistic and social factors that condition the alternation.

A follow-up study by the same author (1980-1981) re-examined the issue, proposing the explanation that *dequeísmo* had its origins in the fact that in Spanish pronominal verbs usually co-occur with prepositions, while non-pronominal verbs typically are constructed without prepositions. The investigator hypothesized that speakers may be confused by these two different processes, which may lead them to also use *de* with non-pronominal verbs, in a form of hypercorrection. Although the follow-up study provided an enhancement on earlier work in that it attempted to explain the motivations for speaker inclusion (or lack thereof) of *de*, it still did not take into account the additional linguistic and sociolinguistic factors considered in subsequent studies that implemented variationist methodologies.

2.3. True Variationist Studies of *(De)queísmo*

The first true variationist methodology applied to the study of *(de)queísmo* occurred in Bentivoglio and Galué, who investigated the phenomena in the *Estudio sociolingüístico del habla de Caracas* corpus. The corpus contained 96 thirty-minute recordings and the participants all were native to the city of Caracas and were stratified equally for gender (48 women and 48 men), social class (32 high, 32 middle, 32 lower), age (24 from ages 14-29, 24 from 30-45, 24 from 46-60, and 24 60 or older). All subordinate clauses which were headed by *que* (728 in total) were extracted and it was found that 58% were constructed without *de*, while 42% contained *de*, which corroborated earlier findings that the *queísta* variant was the more frequent (e.g., Rabanales' Chilean data). The analysis included five linguistic variables: the structural context, interference, phono-grammatical similarity, syllabic distance, and iconicity. Within the variable structural context, the categories coded were pronominal verb (e.g., *enterarse [de] que*), non-pronominal verb (e.g., *hablar [de] que*), verb with noun phrase (e.g., *tener la suerte [de] que*), *ser/estar* with an adjective (e.g., *estar consciente [de] que)*, a noun with *(de) que* (e.g., *una advertencia [de] que)*, conjunctions and prepositions with *(de) que* (e.g., *además [de] que*), and adverbs with *(de) que* (e.g., *lejos [de] que*). The variable interference considered whether or not there was intervening material between the nucleus and the subordinate clause. Phono-grammatical similarity was based on whether or not there was a phonic segment similar to *de* in the syllables prior to the subordinate clause (i.e., another voiced dental sound or another preposition *de*). This variable was based on previous work that found that phonetically similar prior segments favored the presence of *de* in Brazilian Portuguese (e.g., Mollica). The authors coded for three possibilities: absence of a similar segment, presence of a similar segment (e.g., *además*), and presence of the preposition *de*. The factor group syllabic distance was based on the number of syllables between a phonically similar segment to *de* and the subordinate clause, and there were four categories: juxtaposition, 1-2 syllables, 3 syllables, and 4 or more syllables. The final linguistic variable, iconicity, was based on the view that the presence of the preposition *de* distances the subordinate clause from the subject. The authors coded for three possibilities: the referent of the subject is the speaker (i.e., 1st person singular), the referent of the subject is the interlocutor or a third person, and there is no specific referent.

Three factor groups were selected as significant in the predictive model created by the Goldvarb binomial regression in Bentivoglio and Galué. First, grammatical context was selected, with pronominal verbs strongly favoring *queísmo* (i.e., the absence of *de*) and non-pronominal verbs and verbs with nominal phrases strongly disfavoring it. The second factor group selected was phono-grammatical similarity, with elements similar to *de* strongly favoring *queísmo* and the absence of similar elements and the prior presence of the preposition *de* both strongly disfavoring *queísmo*. The third and final independent variable selected was speaker socioeconomic level, with the lower class strongly favoring *queísmo*, the middle class neither favoring nor disfavoring it, and the upper class strongly disfavoring the phenomenon. Overall, the main contribution of the study was that it went beyond the descriptive and attempted to explain what linguistic and extra-linguistic factors conditioned the use of *de*. Despite the study's contributions, a methodological critique is worth mentioning. In the coding of iconicity, referents in which the addressee was the subject and referents in which a third person was the subject were coded together. Since it is quite possible that greater care would need to be taken when referring to an inerlocutor who was

present as opposed to someone who was not, these categories do not seem to be collapsed justifiably into one, and thus, an analysis in which they are considered separately may show that this variable is significant, unlike in the aforementioned study.

Dequeísmo (but not *queísmo*) has also been studied in the Spanish of the Canary Islands (Serrano), where correlations were noted between the phenomenon and three sociolinguistic factors: socioeconomic level, gender, and age. Specifically, males were more likely to be *dequeístas* than females, and speakers of intermediate age (35-54 years) used the most *de*, followed by young speakers (ages 20-34), whereas older speakers (age 55 and older) were least likely to use *de*. Upper class speakers used *de* the least, whereas lower-middle class speakers used it the most, followed closely by lower and upper-middle class speakers.

Perhaps the most thorough variationist study of *dequeísmo* to date is that of Schwenter. The study included data from a spoken corpus of Spain (Madrid) from formal speech events, such as university lectures and two written corpora (one from Argentina and one from Chile) from technical and scholarly writings. All instances of *de que* were searched in the corpora and all verbs that co-occurred with *de que* were noted; finally, all main clause uses of those verbs were coded in order to capture uses both with and without *de*. Separate Varbrul analyses of *queísmo* and *dequeísmo* were performed in order to see whether both phenomena were sensitive to the same contextual factors.

Linguistic factors considered by Schwenter included subject of the main clause (i.e., first, second, or third person), tense of the main clause verb (i.e., past, present, or future), intervention (i.e., no intervening words or one or more intervening words), mood of the main verb (i.e., indicative or subjunctive), aspect of the main verb (i.e., perfective or imperfective), previous instances of *de* (i.e., absent or present), and mode (i.e., written or spoken). For the *dequeísmo* analysis, there were four significant factor groups. The first group selected was the subject of the main clause verb, with 3rd person strongly favoring the use of *de*, 2nd person not particularly favoring nor disfavoring it, and 1st person strongly disfavoring it. The next group selected was tense of the main clause verb, with the past strongly favoring the use of *de*, and the present and the future neither favoring nor disfavoring it. The third group selected was mode, with spoken data favoring *de* and written data slightly disfavoring it. Lastly, the factor group intervention was selected, with one or more intervening words favoring *de* and no intervening words disfavoring it. The results for *queísmo* were essentially the exact inverse, which indicated that both phenomena are regulated by the same factors. Also, the investigator explained that the co-occurrence of 3rd person subjects with *de* more so than other subjects was due to the fact that *de* serves as an evidential marker, marking propositional information when the speaker and the person to whom the information is being attributed do not coincide. Schwenter's study was important in that it empirically confirmed previous indications that *queísmo* and *dequeísmo* could justifiably be studied together. It also served as an improvement over Bentivoglio and Galué in that 2nd and 3rd person subjects were coded separately instead of being grouped together, which could explain why the previous authors did not find significance for that factor group. A shortcoming of Schwenter, however, is that no extra-linguistic factors were considered. Also, since the author compared written corpora and a spoken corpus from different countries, it is possible that differences which he attributes to mode may also be attributable to region.

The final and most recent study which will be discussed is Guirado. Also utilizing the Caracas corpus, the author found that four factor groups were selected as significant by Goldvarb. The most significant group selected was source of the evidence, a new group created by the author. Within this group, information from the speaker and information directed toward the addressee both favored *queísmo* (the absence of *de*), whereas information directed toward the speaker himself or from another person disfavored it. The second group selected was verbal context, with (one word) verbs favoring *queísmo,* verbal phrases (e.g., *darse cuenta*) neither favoring nor disfavoring it, and attributive structures (e.g., *estar consciente*) disfavoring it. Socioeconomic level was the next group selected, with the lower class strongly favoring *queísmo* and the upper class disfavoring it. Lastly, verbal tense was selected, with non-past favoring *queísmo* and past disfavoring it. Unlike Schwenter, Guirado found greater use of *de* with 1st person subjects because, according to her, speakers used *de* to project themselves as the epistemic center of the information that they were providing and thus to indicate the directness and immediateness of what was expressed. Guirado's study offers the strengths of considering both linguistic and extra-linguistic factors (unlike Schwenter's) and of offering a new explanation for the function of *de*, that of speaker projection into the proposition of the subordinate clause. Although Guirado offers an interesting analysis, Schwenter's explanation of the use of *de* to offer separation and mitigation from the material presented seems more plausible and congruent with the data presented, since use of the preposition *de* enables the speaker to distance herself from the following propositional content.

2.4. The Role of Discourse Topic

A brief explanation of the role of discourse topic in the formation of Spanish utterances is beneficial for the purposes of situating the new linguistic variable considered in the current study. Past research has demonstrated, for instance, that many techniques can be used to mitigate in Spanish, including the use of impersonal protectors (i.e., *uno,* the generic *nosotros*, and the null subject), temporal deixis (i.e., the conditional of politeness and the imperfect) and parenthetic verbs (Flores-Ferrán). The researcher also found that these strategies increased as potentially emotionally-charged topics emerged between a psychiatrist and his client, while they were less necessary during more harmless topics. In other words, there is a correlation between emotionally-charged discourse topics and the need for greater mitigation.

Previous research also includes an in-depth analysis of the linguistic measures used by politicians to take ownership, avoid fault, and assign blame via personal deixis and other tactics (Blas Arroyo). Specifically, the author notes the importance of the role of discourse topic in pronominal expression and in the force of the linguistic devices used.

The importance of discourse topic has also been considered in that subject pronoun use varies in direct correlation to speaker opinions, ideas, or feelings toward a topic (Davidson). For example, the researcher notes that subject pronouns are used to signal propositional content that is less abstract and more personally relevant.

Since discourse topic has proven to affect the occurrence of linguistic material (e.g., subject pronouns, verb tenses and moods, etc.) and since Schwenter has shown the role that *de* plays as an evidentiality marker, it is important to consider the use of this marker in terms of how it co-occurs with certain types of topics. In fact, the inclusion of discourse topic advances our knowledge of *dequeísmo* in that, since discourse topic has been

noticeably absent in the study of the phenomena up to the present, and since a partial function of including *de* is to distance the speaker from the propositional content that follows, it might be hypothesized that use of *de* should increase within the types of topics in which it is more important for the speaker to provide the buffer of additional distance from information about which it may be difficult for the interlocutors to converse.

Given Schwenter's findings regarding the use of *de* to distance the speaker from the content that follows, it is hypothesized that *de* will co-occur more with emotionally-charged topics. Further information about the coding of discourse topic in the current study is provided in the methodology section, as follows.

3. Methodology

The current study compares variation in the use of *de* by speakers in the *Estudio Sociolingüístico de Caracas* corpus (Bentivoglio & Sedano), following previous documentation of this form of variation in Caracas (Bentivoglio 1976, 1980-81). The corpus contains 160 thirty-minute recordings of Caracas natives and is stratified equally for gender (80 women and 80 men), social class (32 upper, 32 upper middle, 32 middle, 32 lower middle, and 32 lower-class speakers), and age (40 speakers from ages 14-29, 40 from 30-45, 40 from 46-60, and 40 of age 60 or older). All 160 participants were included in the current study. The dependent variable is the presence or absence of *de* prior to a subordinate clause headed by *que.* The four linguistic factors considered are adapted from Schwenter and include the subject of the main clause (i.e., first, second, or third person), the tense of the main clause verb (i.e., present indicative, present subjunctive, present perfect, imperfect, imperfect subjunctive, preterit, future, or pluperfect), intervention (i.e., no intervening words, 1-2 intervening words, or 3 or more intervening words), and discourse topic (i.e., charged, reaching out, or neutral), which will be described in detail, as follows. The three extra-linguistic factors considered include the aforementioned age, sex, and socioeconomic level.

Two methodological differences from Schwenter bear mentioning. First, the linguistic factor group "discourse topic" has been added to the current study. Since it has been found that *de* functions as an evidential marker to distance the speaker from propositional information, it is hypothesized that it should more frequently co-occur with more opinionated / controversial topics for reasons of mitigation. To code for discourse topic, the researcher read the two sentences prior to the extracted token, in order to have an understanding of the surrounding context that was beyond the sentential level. If determinations about the discourse topic could not easily be made two sentences in advance of the token, then previous sentences were consulted until the nature of the topic could adequately be pinpointed. Within this factor group, three factors were included. The first factor was emotionally-charged contexts in which it was possible that what the speaker was saying could be offensive to the interviewer (i.e., "charged" contexts), as in (3), in which a woman is talking about her troubled childhood, being reported by a classmate's mother, and being accused of pushing another girl. Note the use of *dijo de que* in this charged context.

(3) Yo era rebeldísima, rebeldísima, rebeldísima…Tenía una compañera, era mi vecina, entonces la muchacha, nosotros nos mudamos y ... a un sitio mejor. Entonces la muchacha empezó que... ella siguió viviendo allá, y este se empezó ... ya ella decía

> que yo era superior a ella y esas tonterías, ¿no? y un buen día su mamá … fue a reclamar a la casa por qué ... el porqué yo me portaba así. Entonces, incluso, la señora le dijo **de que** yo la empujaba y cosas así.

The second factor within discourse topic included contexts of reaching out, in which a difficult topic had emerged but the particular token was produced during a section of speech that was not particularly controversial or offensive (i.e., "reaching out"), as exemplified in (4) below. In this example, the speaker is talking about what scares her and mentions her fear of guns. What she is saying does not implicate herself in any way or admit to any wrongdoings, but the topic is certainly not neutral, and thus, it is an example from the intermediate context, in which themes are more difficult but it is unlikely that the speaker's comment will be interpreted as offensive or controversial.

(4) bueno, me da miedo ... Yo creo **que** sí hay una razón, este ... bueno, que debe ser más o menos lógica, a la hora de un despelote de esos de ... de policías, estudiantes ... tiros, olvídate, cualquier persona que se altere con una ... con un arma en la mano, yo ... yo desconfío, y si ... Yo pienso **que** yo, por ejemplo, yo no soy ... este ... partidaria de las personas que usan ... que tienen porte de armas

The third and final factor was neutral contexts: generic, inoffensive topics were discussed (i.e., "neutral" topics), in which emotionally-difficult or charged themes did not emerge, as seen in (5). This was the most common context, given the nature of rather surface-level sociolinguistic interviews between interlocutors who did not know each other intimately. In the following example, a woman is discussing an innocent memory from her childhood, of eating mangoes in the country as a small child. Following this anecdote, she later goes on to discuss riding on a swing. This neutral, non-controversial context includes *recuerdo..que*.

(5) En el corral de la casa habían unas maticas de higo, entonces yo me iba con una sillita, como estaba tan pequeña…a coger los higos y a comerme los ... mangos, porque mi hermano mayor, G., como era más ... alto que yo, por supuesto, él no tenía que buscar la sillita y se comía los higos primero que yo… Recuerdo también **que** me montaba en una mata de cují… Y tenía un trapecio, que era un columpio…

The second major methodological difference from Schwenter is that the factor group "intervention" has been slightly modified to allow for a more nuanced analysis, expanding Schwenter's dichotomous distinction of intervention / no intervention to the tripartite no intervention, 1-2 intervening words, and 3 or more intervening words.

Following Schwenter, in order to capture all potential contexts of *queísmo* and *dequeísmo*, searches were performed for all instances of the verbs listed in Table 1, following which all main clause uses of such verbs co-occurring with *que* or *de que* were extracted as tokens.

Table 1. Verbs That Demonstrate *De/Ø* Variation (Adapted from Schwenter)

Verbs That Occur Prescriptively with *de (*included in the *queísmo* analysis)	*acordarse*, *alegrarse, darse cuenta, dudar, enterarse, lamentarse, olvidarse, preocuparse, quejarse, sorprenderse*
Verbs That Occur Prescriptively with Ø *(*included in the *dequeísmo* analysis)	*afirmar, agradecer, comentar, confesar, considerar, contar, decir, declarar, esperar, indicar, informar, necesitar, negar, observar, ocurrir, oír, pedir, pensar, probar, reconocer, recordarse, resultar, saber*

Following token extraction, binomial, step-up / step-down regressions were run within Goldvarb X (Sankoff, Tagliamonte, & Smith) in order to see which independent variables in which order form the predictive model that best selects the occurrence of *de*. The regression analysis measures the probabilistic weight of each independent variable in relation to the application value (i.e., *de* in the case of the *dequeísmo* analysis and Ø in the *queísmo* analysis). A weight above .50 indicates that the particular factor in question favors the selected application value, whereas a weight below .50 indicates a disfavoring effect. See Geeslin for an in-depth discussion of predictive regression models.

4. *Dequeísmo* Results

The results from the *dequeísmo* analysis will now be presented, followed by the results from the *queísmo* data. There will first be a focus on the significant factors selected by the Goldvarb binomial regression, prior to a brief mention of the factor groups that were not selected.

A general view of how the *dequeísmo* data were distributed can be seen in Table 2, in which the factor weights (i.e., probabilities) and percentages of inclusion of *de* are presented for each factor group (for more on factor weights and Varbrul analyses, see Tagliamonte 2006, 2012).

Overall, *dequeísmo* occurred in a total of 10.1% of the cases (51/504), in comparison with the prescriptive lack of *de* in the other 89.9% of the cases (453/504). Thus, for the 160 participants included in the corpus, the absence of *de* was the more common variant, although there was variation according to the participant and the linguistic context. Although *dequeísmo* was relatively rare in the current dataset[3], we can still note conditioning of the occurrence of *de* according to linguistic and social factors, the addition of which significantly improves the predictive power of the model in each case (and therefore foments their inclusion in the final model), as will be seen in the following sections. It is also worth noting that the *dequeísmo* dataset is the larger of the two in the current study, so a 10% rate here still corresponds with occurrence across 51 tokens.

[3] An anonymous reviewer accurately highlighted the relatively low rate of *dequeísmo* in the corpus.

Table 2. Linguistic and extra-linguistic factors selected as significant to the occurrence of *dequeísmo*; factor groups not selected as significant in square brackets

Factor Group	Probability	% (N)
FG5: Discourse Topic		
Charged	.90	46 (44)
Reaching Out	.53	10 (169)
Neutral	.40	5 (291)
Range	*50*	
FG8: Age		
46+ years	.75	20 (115)
30-45 years	.71	16 (148)
14-29 years	.26	2 (241)
Range	*49*	
FG7: Socioeconomic Level		
Upper	.67	16 (128)
Middle	.50	8 (212)
Lower	.37	8 (164)
Range	*30*	
FG3: Verbal Tense		
Present Perfect	.91	55 (11)
Imperfect Subjunctive	.77	40 (5)
Present Subjunctive	.66	25 (4)
Present Indicative	.53	9 (302)
Preterit and Imperfect	.41	9 (182)
Range	*50*	
FG4: Intervention		
1-3+ Words	.78	36 (22)
None	.49	9 (482)
Range	*29*	
FG2: Subject of Main Clause		
3rd Person	[.55]	12 (212)
1st Person	[.47]	9 (257)
2nd Person	[.46]	9 (35)
FG6: Sex		
Female	[.52]	11 (256)
Male	[.48]	9 (248)
Log likelihood = -113.395		p = .012

Note: Total N (/variant) = 504 (51); Corrected mean = .04
One-level analysis: Total $\chi^2 = 242.50$; χ^2 / cell = 1.08

4.1. Significant Factor Groups for *Dequeísmo*

According to the binomial regression run via Goldvarb, there were five significant factor groups that were chosen for inclusion in the predictive model: discourse topic, speaker age, socioeconomic level, verbal tense, and intervention. The fact that the step-up and step-down regressions have the same log likelihood (-113.893) and choose the same groups in the same order with the same factor weights indicates that the regressions are quite trustworthy. One critique of the variable rule program has been the difficulty in teasing apart possible interactions among factor groups (e.g., Johnson; Tagliamonte 2012). Nevertheless, in addition to the fact that the factor weights do not show great fluctuation across runs, another indication that the regression is likely free of interactions is that the

ranking of the factor weights and percentages lines up (e.g., Tagliamonte 2012), as we can note that higher percentages correspond with higher factor weights within each factor group in the current study.

4.1.1. Discourse topic

According to the Goldvarb regression, discourse topic was the first factor group chosen. "Charged" discourse topics, as hypothesized, strongly favored the inclusion of *de* (factor weight .90) with *dequeísmo* occurring in 46% of such contexts (20 out of 44 cases). As illustrated previously in example (3), the use of *de* within emotionally-charged discourse topics for which the speaker shares information that may not reflect particularly positively on him might foster a certain distancing from the propositional content included. Thus, it is not surprising that occurrences of *de* are strongly favored for such topics. *Dequeísmo* was also slightly favored in contexts of "reaching out" (factor weight .53) (16/169 cases). As previously indicated in example (4), "reaching out" topics include emotional involvement and are more "conceptually heavy" than neutral contexts, yet the information provided is not particularly offensive or face-threatening. Thus, a slight, yet not extremely strong, favoring of *de* is not surprising. The use of *de* was disfavored in neutral contexts (factor weight .40), occurring just 5% of the time in such contexts (15/291), which were the most common overall. As expected for sociolinguistic interviews, neutral topics were the most common, and since the information provided within such discourse topics tends to be more surface level and not emotionally-charged or personally difficult, the fact that *de* was used within such topics the least was not unexpected. Overall, the strong significance of this factor group is indicated via its large range of 50. This factor group will be further considered in conjunction with the subject of the main clause in Section 4.3

4.1.2. Speaker age

Speaker age was the second factor group chosen by the regression as significant. Within this group, the oldest (i.e., 61 and older) and second oldest (i.e., 46-60) speaker groups were combined due to the similarity of their results. The data indicate that the oldest speaker group (i.e., 45 and older) was the most likely to produce *de* (factor weight .75), doing so 20% of the time (23 out of 115 cases). The middle speaker group (i.e., 30-45) also strongly favored *dequeísmo* (factor weight .71) (23/148 cases). The youngest group strongly disfavored it with a factor weight of .26 and produced tokens with *de* just 2% of the time (5/241). This finding differs somewhat from those of previous studies. That the middle speaker group favored *de* is in line with the Canary Island participants of Serrano, although in that study younger speakers used more *de* than older ones, which could be a possible difference between this phenomenon in Caracas and the Canaries.

4.1.3 Speaker socioeconomic level

Socioeconomic level was the third factor group selected by Goldvarb. Within this group, upper middle class and lower middle class speakers were combined with middle class speakers, based on similar behavior. The upper class group was the most likely to produce tokens with *de* and strongly favored this variable factor weight of .67), producing such tokens 16% of the time (21 of 128 cases). A possible motivation for the greater production of *de* by the upper class group is due to hypercorrection in order to maintain more formal, higher-class sounding speech, in order to show evidence of formal education,

which was a conclusion of Bentivoglio (1980-81). The middle class, in conjunction with the upper middle and lower middle class groups, (factor weight .50) (16/212 cases), neither favored nor disfavored the use of *de*, while the lower class group (factor weight .37) (14/164 cases) disfavored its use. The range for this factor group was 30. These results confirm Guirado, who found that the upper class strongly favored the use of *de*, whereas the lower class strongly disfavored it.

4.1.4. Verbal tense

Verbal tense was the fourth factor group selected as significant. Three tenses strongly favored the inclusion of *de*: the present perfect (factor weight .91), the imperfect subjunctive (factor weight .77), and the present subjunctive (factor weight .66). The present indicative (factor weight .53) (26 of 302 cases) did not particularly favor nor disfavor the occurrence of *de*. The past tenses (preterit and imperfect) patterned together and were then combined, disfavoring the use of *de* (factor weight .41), as *dequeísmo* occurred in only 9% of past indicative contexts (16 of 182 cases). That the compound perfect and imperfect subjunctive favored *dequeísmo* and co-occurred with *de* much more than the present tense did confirms Schwenter's finding that *de* was favored in the past tenses, but not in the present. However, the preterit and imperfect's disfavoring of *de* differs from Schwenter's findings about the past. This could be due, in part, to Schwenter's use of two (of three) written corpora and regional differences (using corpora from Spain, Chile, and Argentina). The range for verbal tense was 50, which is quite high, partially due to the small number of tokens that occurred in the present perfect (11), imperfect subjunctive (5), and present subjunctive (4). However, even disregarding those tenses, there was still a range of 12 separating the present indicative and the past tenses (i.e., preterit and imperfect)[4].

4.1.5 Intervention

Intervention was the fifth and final factor group selected. Within this group, the factors 1-2 words and 3 or more words were combined due to the "knockout" produced by the "3+ words" variable. Here, when one or more words separated the main clause verb from (*de*) *que*, the inclusion of *de* was strongly favored (factor weight .78), whereas when no words intervened, *de* was very weakly disfavored (factor weight .49). These results support Schwenter and his view that *de* provides a mechanical function that aids in processing, helping to connect the information that succeeds *de que* with the separated information that preceded it. In fact, the finding that intervening words even more strongly favored the use of *de* in the current data than in Schwenter's (with a factor weight of .78 in the current study as opposed to .60 for Schwenter) could be due to the mode of communication, as the oral data here benefit more from the mechanical function of *de* than the two (of three) written corpora of Schwenter, which enable the reader to search visually for desired information and thus, put less communicative burden on the writer to include additional elements to explicitly connect sentential components.

[4] An anonymous reviewer noted that in addition to verb tense, it is possible that the verb type (i.e., the verb lexeme itself) plays a role here. In following the previous work of Schwenter, across the two datasets combined I searched for the 33 verbs listed in Table 1. I only retained types that included tokens with and without *de* (i.e., verbs that permitted variation). Nevertheless, subsequent analyses will do well to analyze the possible effect of verb type. As noted, a random-effect for verb type in a subsequent investigation would enable further analysis of this topic, should results be significant.

4.2. Non-significant Factors

In this section, the factor groups mentioned were not selected as significant by GoldVarb and thus will receive somewhat less emphasis.

4.2.1. Subject of the main clause

As one might expect, first person subjects were most frequent in the current study, given the nature of interview data. *Dequeísmo* occurred in 9% (23/257) of clauses with a first person subject (factor weight .47), 9% (3/35) of those headed by second person subjects (factor weight .46) (which were relatively infrequent due to the nature of the interview), and 12% (25/212) of those with third person subjects (.55 factor weight). Even though this group was not significant in the present data, Schwenter's finding that *dequeísmo* occurred more with third person subjects is borne out in the present data. Further, the fact that main-clause subject was not chosen to be significant may also illustrate the complexity of this factor group, for which Schwenter and Guirado found opposing results, with the former finding more use of *de* with the third person and the latter with the first person. This variable will be further considered in conjunction with discourse topic in Section 4.3.

4.2.2 Speaker sex

Speaker sex was not chosen as significant, as females only very slightly favored the production of *de* (factor weight .52), doing so 11% of the time (28 of 256 cases), whereas males only very slightly disfavored it (.48), producing *de* 9% of the time (23 of 248 cases). These results compare to Guirado in the sense that speaker sex was not found to be significant either.

4.3 Cross-tabulating Discourse Topic and Subject of the Main Clause

Given the fact that discourse topic was selected as the most significant independent variable in the current data and that Schwenter found the subject of the main clause to play a crucial role in evidentiality marking, the two variables have been cross-tabulated in order to see in detail the intersection of the two variables (Table 3). With respect to differences across the two variables, we begin by making comparisons within 1st and 3rd person subjects and then make comparisons across person.

The cross-tabulation indicates a general pattern in which the highest rates of use of *de* are in charged contexts, which is the case for both 1st person and 3rd person subjects of main clauses[5] (also demonstrated graphically in Figure 1 below). In fact, in the case of 1st person subjects, the rate of use of *de* was approximately nine times greater in charged contexts (44%) than in neutral contexts (5%), which was a statistically significant difference, χ^2 (1, N = 183) = 31.9, $p < .001$. Interestingly, the distribution was nearly identical for 3rd person subjects, as *de* was used significantly more frequently in charged contexts (44%) than in neutral contexts (6%), χ^2 (1, N = 130) = 26.0, $p < .001$. Additionally, within 1st person subjects, there is significantly more use of *de* in charged contexts than contexts of reaching

[5] This is also the case for 2nd person subjects, which are not included in any of the chi-square comparisons due to the fact that any comparison with 2nd person subjects would include at least one cell which has a value of less than 5, which renders such comparisons inappropriate. Low use of 2nd person subjects is typical of sociolinguistic interviews in which the interviewee makes infrequent reference to the interviewer.

out (9%), χ^2 (1, N = 92) = 13.0, p < .001. Once again, 3rd person subjects demonstrate a similar distribution as 1st person subjects, as the rate of *de* use is significantly higher in charged contexts than reaching-out contexts (10%), χ^2 (1, N = 107) = 15.4, p < .001. Nevertheless, not all comparisons were significantly different. For 1st person subjects, although the use of *de* was higher in contexts of reaching out (9%) than neutral contexts (5%), differences were not statistically significant, χ^2 (1, N = 249) = 1.85, p = .174. Once again, the pattern was very similar for 3rd person subjects, as slightly higher use of *de* in reaching-out contexts (10%) than neutral contexts (6%) did not yield significant differences, χ^2 (1, N = 187) = 1.09, p = .297.

Table 3. Cross-tabulation of discourse topic and person in *dequeísmo* analysis

Topic	Presence of *de*	1st person		2nd person		3rd person		Total	
		N	%	N	%	N	%	N	%
Charged	*De*	8	44	1	100	11	44	20	45
	Ø	10	56	0	0	14	56	24	55
	Total	18	100	1	100	25	100	44	100
Reaching out	*De*	7	9	1	8	8	10	16	9
	Ø	67	91	12	92	74	90	153	91
	Total	74	100	13	100	82	100	169	100
Neutral	*De*	8	5	1	5	6	6	15	5
	Ø	157	95	20	95	99	94	276	95
	Total	165	100	21	100	105	100	291	100
Total	*De*	23	9	3	9	25	12	51	10
	Ø	234	91	32	91	187	88	453	90
	Total	257	100	35	100	212	100	504	100

Moving to comparisons across the person of the subject of the main clause, none of the comparisons between 1st person and 3rd person subjects indicate significant differences. Within charged contexts, *de* co-occurred with 1st person (44%) and 3rd person subjects (44%) at the same rate, which unsurprisingly resulted in differences that were not statistically significant, χ^2 (1, N = 43) = .000, p = .977. Similarly, within contexts of reaching out, *de* was used with 1st person subjects at a rate that was highly similar to that for 3rd person subjects (10% and 9%, respectively), χ^2 (1, N = 156) = .004, p = .950. Finally, as with the first two types of discourse topics, rates of use of *de* were very similar for both 1st person (5%) and 3rd person subjects (6%) in neutral contexts, χ^2 (1, N = 270) = .098, p = .754.

Thusly, for the *dequeísmo* data, the more detailed analysis also reveals that discourse topic generally plays a more important role than the person of the subject, since we see similar rates across person and large differentiation across discourse topic. The regression model (Table 2) indicated this to generally be the case (by virtue of its selection of discourse topic, but not subject of the main clause), and now we see that sub-dividing our data leads us to a similar conclusion for the more nuanced distributions of the results as well.

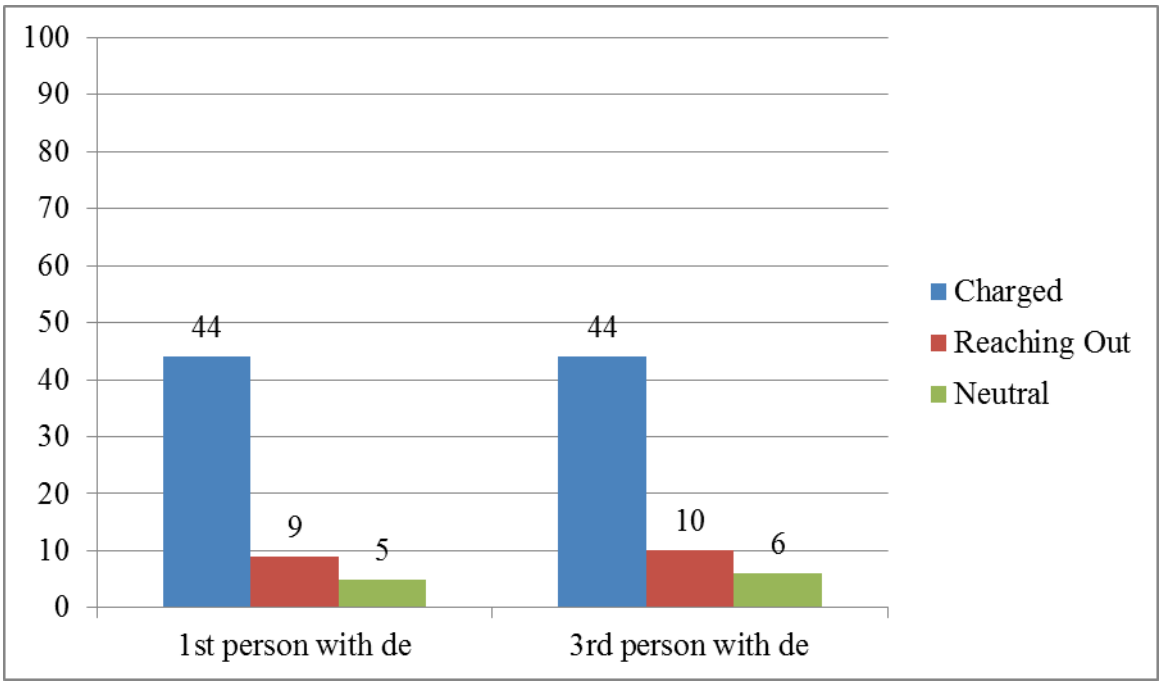

Figure 1. Percentage of *de* use according to subject type and topic in *dequeísmo* analysis

Note: Rates represent the percentage of *de* use according to the subject type (e.g., 44% use of *de* with 1st person subjects in charged topics, as opposed to 56% omission of *de* in the same context)

5. *Queísmo* Results

The results from the *queísmo* analysis will now be presented. Once again, there will first be a focus on the significant factors selected by the Goldvarb binomial regression, followed by a brief mention of the factor groups that were not selected.

A general view of how the *queísmo* data was distributed can be seen in Table 4, in which the factor weights and percentages of omission of *de* are presented for each factor group. Please note that for the *queísmo* analysis, percentages of the omission of *de* are discussed, the opposite of what was discussed for *dequeísmo*, in which the inclusion of *de* was analyzed.

Table 4. Linguistic and extra-linguistic factors selected as significant to the occurrence of *queísmo*; factor groups not selected as significant in square brackets

Factor Group	Probability	% (N)
FG5: Discourse Topic		
Neutral	.66	79 (82)
Reaching Out	.34	50 (46)
Charged	.27	32 (22)
Range	*39*	
FG3: Verbal Tense		
Present Perfect	.75	83 (6)
Present Indicative	.58	74 (89)
Imperfect Subjunctive	.47	50 (4)
Preterit	.39	49 (41)
Imperfect	.15	20 (10)
Range	*60*	
FG2: Subject of Main Clause		
2nd Person	[.77]	86 (7)
1st Person	[.55]	68 (88)
3rd Person	[.39]	53 (55)
FG4: Intervention		
None	[.51]	64 (141)
1-3+ Words	[.42]	56 (9)
FG6: Sex		
Female	[.54]	67 (86)
Male	[.44]	58 (64)
FG7: Socioeconomic Level		
Middle	[.59]	71 (35)
Upper and Upper Middle	[.51]	64 (70)
Lower and Lower Middle	[.42]	56 (45)
FG8: Age		
30-45 years	[.56]	69 (51)
14-29 years	[.53]	66 (59)
46+ years	[.39]	53 (40)
Log likelihood = -82.731		p = .049

Note: Total N (/variant) = 150 (95); Corrected mean = .63
One-level analysis: Total χ^2 = 128.76; χ^2 / cell = 1.23

Overall, *queísmo* occurred in a total of 63.3% of the cases (95/150), in comparison with the prescriptive inclusion of *de* in the other 36.7% of the cases (55/150), which indicates that for the 160 participants a lack of *de* was once again the more common variant (as seen in the analysis of *dequeísmo)*. It is worth noting that the inclusion of *de* is considered prescriptive and yet occurred in the minority of cases here. Based on the current findings, *queísmo* does not appear to be particularly stigmatized, since it was so widespread. It is also

interesting to note that it was much more common than the previously-analyzed *dequeísmo*, occurring 63.3% of the time, as opposed to the 10.1% rate of *dequeísmo*.

5.1. Significant Factor Groups for *Queísmo*

According to the *queísmo* binomial regression run via Goldvarb, two significant factor groups were chosen for inclusion in the predictive model: discourse topic and verbal tense. Once again, the fact that the step-up and step-down regressions have the same log likelihood (-82.731) and choose the same groups in the same order with the same factor weights indicates that the regressions are trustworthy. That only two groups were chosen this time is likely partially due to the fact that there were fewer overall tokens of *queísmo* contexts than *dequeísmo* ones.

5.1.1. Discourse topic

According to the Goldvarb regression, discourse topic was the first factor group chosen (as was also the case with the *dequeísmo* data). As expected, we note the inverse of the *dequeísmo* factor weights, with neutral topics strongly favoring (factor weight .66) the occurrence of *queísmo* and therefore co-occurring with a lack of *de* 79% of the time (65/82 cases). This was hypothesized, since neutral topics do not require the speaker to distance herself from the propositional content that follows and, thus, should not need to co-occur with *de*. As expected, contexts of "reaching out" disfavored the use of *queísmo* (factor weight of .34), as did "charged" discourse topics (factor weight.27). Thus, the three factors occurred in the inverse order of the *dequeísmo* data, indicating that the omission of *de* in contexts where it would be prescriptively expected (i.e., *queísmo*) is constrained by discourse topic in the same way that its inclusion in contexts where it was not prescriptively expected (i.e., *dequeísmo*) was. The strong significance of this factor group is indicated via its large range of 39. This variable will be further considered in conjunction with the subject of the main clause in Section 5.3.

5.1.2. Verbal tense

Verbal tense was the second and final factor group selected as significant by the *queísmo* binomial regression. The present perfect (factor weight .75) once again strongly favored non-prescriptive behavior, frequently co-occurring with *queísmo.* The present indicative (factor weight .58) slightly favored *queísmo*, occurring with the omission of *de* in 74% of the cases (66/89). That the present favored the omission of *de* confirms Guirado's finding that non-past contexts favored *queísmo.* The imperfect subjunctive (factor weight .47) very slightly disfavored *queísmo*, which is the inverse of the *dequeísmo* data, in which it strongly favored *dequeísmo*. This means that these data are quite compatible, with the imperfect subjunctive favoring *de* for both phenomena, which would be logical given the subjunctive's use in casting doubt and the ability of *de* to further distance the speaker from the propositional content included. However, since there were only a handful of tokens of this form, it would be interesting to track in future studies with more tokens. The preterit (factor weight .39), with 49% omission of *de* (20 of 41 cases), and especially the imperfect (20% omission, factor weight .15) both strongly disfavored *queísmo*, which supports Guirado's finding that past tenses disfavored *queísmo*. In the present study, in other words, these two past tenses favored the use of *de*, which would make sense given the greater distance from the propositional content that they convey, as compared to the present (which

here favored the omission of *de*). This also is consistent with Schwenter's finding that *de* is used more in the past than in the present. The range for this factor group was 60 (interestingly, the same range that occurred for *dequeísmo*), and, once again, even if we exclude the less frequently-occurring tenses, there was still a range of 19 separating the present indicative and the preterit.

5.2. Non-significant Factors

In this section, the factor groups mentioned were not selected as significant by Goldvarb and thus will once again receive less emphasis.

5.2.1. Subject of the main clause

Although the subject of the main clause was not chosen as a significant factor group, *queísmo* was more likely to co-occur with first person subjects (68%) (60 of 88 cases), which slightly favored it (factor weight .55), than with third person subjects (53%) (29 of 55 cases), which disfavored it (.39 factor weight). Second person subjects favored *queísmo*, but there were only seven such tokens in these interview data. Although this group was not significant in the present data, Schwenter's finding that *de* occurred more with third person subjects is borne out and the current percentages mirror the percentages of the *dequeísmo* data in the present study, in which *de* co-occurred with third person subjects at a higher percentage than for other persons.

This factor group will be further considered in conjunction with discourse topic in Section 5.3.

5.2.2. Intervention

Intervention was not selected as significant for *queísmo*, but, as expected, the results were the inverse of the *dequeísmo* data. Within this group, the factors 1-2 words and 3 or more words were combined due to a "singleton" produced by the infrequently-occurring "3+ words" variable. Here, once again, when one or more words separated the main clause verb from (*de*) *que*, the omission of *de* was disfavored (factor weight .42), whereas when no words intervened between the main clause verb and (*de*) *que*, the omission of *de* was neither favored nor disfavored (factor weight .51), occurring 64% of the time (90 of 141 cases). These results support Schwenter and mirror those of the *dequeísmo* data in the current analysis.

5.2.3. Speaker sex

Females slightly favored the omission of *de* (factor weight .54), doing so 67% of the time (58 of 86 cases), whereas males slightly disfavored it (.44), omitting it 58% of the time (37/64), though this factor group was not significant. In this case, we do not see the inverse of the *dequeísmo* data, in which women favored the production of *de,* indicating that females favored the innovative (i.e., non-prescriptive) forms in both cases. The current *queísmo* data, in which females omitted *de* more than males, supports Guirado, even though, as with her study, speaker sex was not found to be statistically significant. These findings also somewhat support Serrano's, who reported less use of *de* by females than males, albeit within the phenomenon of *dequeísmo*.

5.2.4. Speaker socioeconomic level

Upper middle class speakers were combined with upper class speakers, based on similar production, and lower middle class speakers were combined with the lower class for the same reason. This variable was not significant, but interestingly, the middle class group most strongly favored *queísmo* here (factor weight .65), producing it 71% of the time. The upper / upper middle class group neither favored nor disfavored *queísmo,* with a factor weight of .51 (64% *queísmo*, 45/70 cases). The lower / lower middle class group (factor weight .42) disfavored *queísmo*, which is interesting, since they also disfavored *dequeísmo* in the previous data, indicating more prescriptive behavior than the other, higher class speakers in both contexts. The use of more formal speech by working classes, which may be an attempt at upward social mobility, has been reported in previous work (for an overview on the effect of socioeconomic level on linguistic behavior, see Díaz-Campos).

5.2.5 Speaker age

Within this group, the oldest (i.e., 61 and older) and second oldest (i.e., 46-60) speaker groups were once again combined due to the similarity of their results. It is interesting to note that, similar to the *dequeísmo* data, where the oldest speaker group (i.e., 45 and older) was the most likely group to produce *de*, here once again they were the least likely to produce tokens with only *que* (factor weight .39). This indicates that, of all the groups, the oldest speaker group most preferred *de* in both contexts. The middle speaker group (i.e., 30-45) slightly favored *queísmo* (factor weight .56), which serves as a possible additional indicator that *queísmo* is the norm in the speech of Caracas, along with the aforementioned finding that the phenomenon occurred in a total of 63% of the data overall. The youngest group neither favored nor disfavored *que,* with a factor weight of .53. Similar to Guirado, this factor group was not statistically significant.

5.3 Cross-tabulating Discourse Topic and Subject of the Main Clause

As was the case with the *dequeísmo* data, since discourse topic was selected as the most significant independent variable in the *queísmo* data and since the subject of the main clause is known to play an important role in evidentiality marking, the two variables have been cross-tabulated for further analysis (Table 5). We begin with a comparison within 1st and 3rd person subjects and then move to comparisons across person.

Cross-tabulating person and discourse topic in the *queísmo* analysis indicates a general pattern in which there was greater absence of *de* in neutral contexts than in charged or reaching-out contexts. It will be recalled that overall there were fewer tokens in the *queísmo* data, and chi-square comparisons of charged contexts with other contexts were not possible due to small cells. Statistical comparisons of neutral and reaching-out contexts were possible and indicated that for both 1st person and 3rd person subjects there was more omission of *de* in neutral contexts. For 1st person subjects, the rate of absence of *de* was more than 25% greater in neutral contexts (83%) than in contexts of reaching out (57%), which was a statistically significant difference, χ^2 (1, N = 75) = 5.69, p = .017. For 3rd person subjects, the differentiation between contexts was even greater, as the rate of absence of *de* was 40% greater in neutral contexts (73%) than in reaching-out contexts (33%), which was also a statistically significant difference, χ^2 (1, N = 47) = 7.42, p = .006.

Table 5. Cross-tabulation of discourse topic and person in *queísmo* analysis

Topic	Presence of *de*	1st person		2nd person		3rd person		Total	
		N	%	N	%	N	%	N	%
Charged	Ø	3	23	1	100	3	38	7	32
	De	10	77	0	0	5	62	15	68
	Total	13	100	1	100	8	100	22	100
Reaching out	Ø	12	57	4	100	7	33	23	50
	De	9	43	0	0	14	67	23	50
	Total	21	100	4	100	21	100	46	100
Neutral	Ø	45	83	1	50	19	73	65	79
	De	9	17	1	50	7	27	17	21
	Total	54	100	2	100	26	100	82	100
Total	Ø	60	68	6	86	29	53	95	63
	De	28	32	1	14	26	47	55	37
	Total	88	100	7	100	55	100	150	100

With respect to comparisons across the person of the subject of the main clause, none of the comparisons between 1st person and 3rd person subjects indicate significant differences, which is also what we saw in the *dequeísmo* analysis. However, even though statistical significance was not reached, some differences in patterning are worth noting. Within contexts of reaching out, the absence of *de* co-occurred with 1st person subjects (57%) at a greater rate than it did for 3rd person subjects (33%), although differences did not reach statistical significance, χ^2 (1, N = 42) = 2.40 p = .121. The fact that differences were not significant is also likely related to the rather low token count for 1st and 3rd person subjects in contexts of reaching out in the *queísmo* data. Similarly, within neutral contexts, the absence of *de* co-occurred with 1st person subjects at a rate that was higher than that for 3rd person subjects (83% and 73%, respectively), although differences again did not reach significance, χ^2 (1, N = 80) = 1.15, p = .283. Thusly, although differences were not significant, we see that the absence of *de* occurs more for 1st person subjects than 3rd person subjects in both contexts of reaching out and in neutral contexts. This is a pattern that might be expected given the connection of evidentiality marked by *de* and 3rd person subjects, which indicate information from a source other than the speaker. Nonetheless, we note again that differences were not significant across person and that discourse topic appears to play a larger role in the current data.

Overall, for the *queísmo* data, the more detailed analysis reveals that discourse topic generally plays a larger role than the person of the subject, since we see more similar rates across person and more differentiation across discourse topic. It will be recalled that a similar pattern existed for the *dequeísmo* data. Nevertheless, there is an important difference between the *queísmo* data and the *dequeísmo* data in that person appears to play more of a role in the *queísmo* data (although comparisons still did not reach significance), such that across reaching-out and neutral contexts the rates of omission of *de* were higher for 1st person subjects than they were for 3rd person subjects. On the other hand, it will be recalled that such rates did not show differentiation in the *dequeísmo* data, for which the inclusion of *de* was nearly identical for 1st and 3rd person subjects across contexts. Possible reasons for differences across the two datasets will be proposed in the following section.

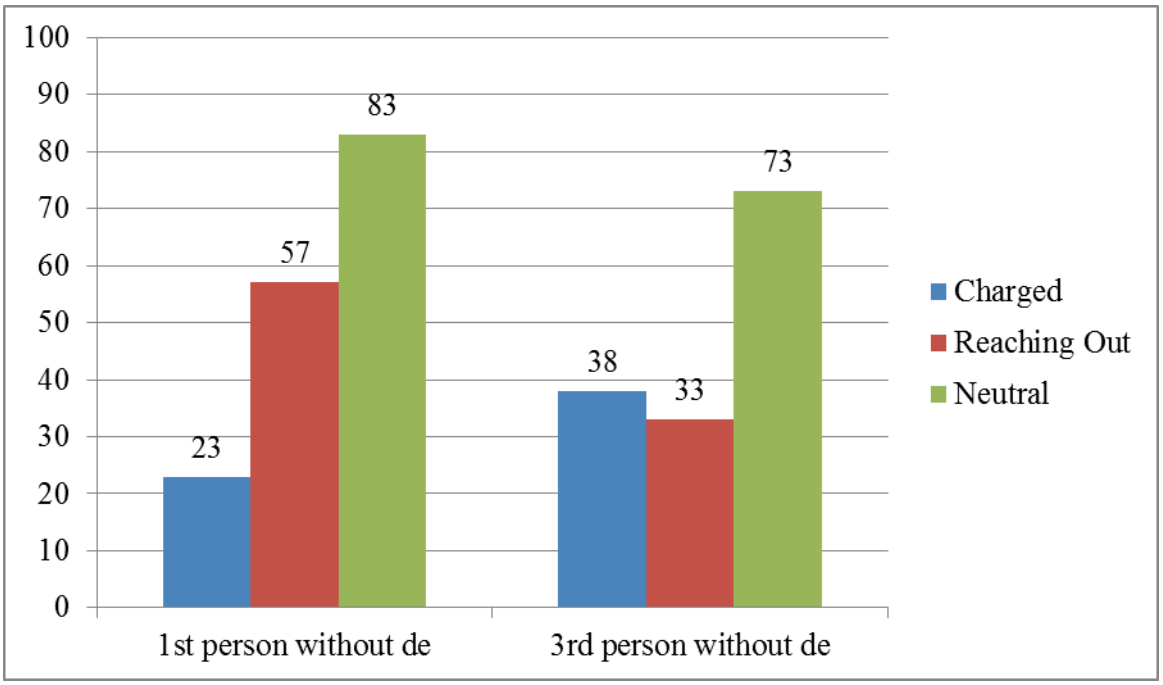

Figure 2. Percentage of *de* omission by subject type and topic in *queísmo* analysis

Note: Rates represent the percentage of *de* omission according to the subject type (e.g., 23% omission of *de* with 1st person subjects in charged topics, as opposed to 77% inclusion of *de* in the same context)

6. Conclusions and Future Studies

With respect to overall frequencies, *queísmo* (63.3%) was shown to be much more prevalent than *dequeísmo* (10.1%), indicating a general preference to omit *de* in the speech of Caracas. This result is in line with the general tendency for simplification in language systems, such that reduction and omission tend to be more common and less marked than fortition and insertion (e.g., Penny).

In the *dequeísmo* analysis, it was shown that discourse topic, speaker age, socioeconomic level, verbal tense, and intervention were all selected as statistically significant. In the *queísmo* analysis, discourse topic and verbal tense were chosen by the regression model due to their significance.

One of the main objectives of the current study was to investigate the possible correlation between the new variable included, discourse topic, and the use of pronominal *de*. Findings indicated that this variable indeed plays a pivotal role in the conditioning of the *(de)queísmo* phenomena. In fact, discourse topic was chosen as the first significant factor group in both regressions. A clear difference was seen between "charged" contexts and neutral contexts. Interestingly, intermediate situations in which contexts were not neutral but attempts to reach out to the interlocutor or to diminish a potentially controversial statement behaved somewhat differently in the two analyses. First, *de* co-occurred with such contexts of "reaching out" in higher percentages than it did with neutral contexts in both analyses. However, "reaching out" contexts (10% use of *de*) patterned more closely with neutral contexts (5% *de*) in the *dequeísmo* analysis, and more closely

with "charged" contexts in the *queísmo* analysis (with factor weights of .34 and .27 for charged and reaching out, respectively).

The strong significance of discourse topic lends further support to Schwenter's view that *de* serves an attenuating function of evidentiality marking. Schwenter highlighted the ability of the evidential marker *de* to weaken the amount of certainty and the connection that the speaker had with the following propositional content, which usually occurred in conjunction with third person subjects, when the speaker was not the original source of the information. On the other hand, the current study indicates that *de* is also conditioned by the discourse topic to the extent that more emotionally-charged topics require greater mitigation / attenuation than less potentially offensive, neutral topics. The multi-faceted nature of discourse topic was further revealed in that even within an emotionally-charged topic, a speaker may reach out to the interlocutor, which obviates the need for further mitigation via *de*, as was seen with the intermediate discourse topic of "reaching out," which co-occurred with less use of *de* than charged contexts and more use of *de* than neutral contexts in both analyses.

In further consideration of the ability of *de* to serve as a marker of evidentiality, another main objective of the study was to provide a detailed analysis of the possible relationship between discourse topic and the person of the main-clause subject. Accordingly, cross-tabulations were run for the two aforementioned independent variables within both the *dequeísmo* and *queísmo* analyses. Both analyses generally pointed to greater differentiation according to discourse topic than person. Additionally, in both analyses there was a general pattern of lower rates of use of *de* in neutral contexts than in the other discourse topics, and this generally held across the person of the main-clause subject. For comparisons across the person of the subject of the main clause (e.g., the use of *de* in 1st person subjects as compared to 3rd person subjects in neutral contexts), differences were not significant between subjects for any of the discourse topics in either the *dequeísmo* or *queísmo* analyses. However, although statistical significance was not reached in the *queísmo* analysis, it offered some trends that are worth exploring further in the future. In the latter analysis, the rate of use of *de* was greater with 3rd person subjects than with 1st person subjects in both contexts of reaching out and in neutral contexts.

Since evidentiality (i.e., that information has come from another source) can be marked by *de* and since 3rd person subjects would provide an appropriate environment to do so, as Schwenter has noted, we might expect the former to generally favor the latter. Nonetheless, there appears to be a potentially important distinction in the function of *de* when it appears in cases of *dequeísmo* as opposed to in *queísmo*. To the extent that for *dequeísmo* the inclusion of *de* is prescriptively unexpected (e.g., *Cree de que*), whereas in the *queísmo* the inclusion of *de* is prescriptively expected (e.g., *Se alegra de que*), it might follow that *de* could have a stronger evidential function in contexts where its use is more innovative (and perhaps more pragmatically motivated), as opposed to contexts where it is more grammatically (i.e., morphosyntactically) motivated. In other words, when *de* is less expected (i.e., in instances of *dequeísmo*), we might argue that speakers are more clearly able to use the form for pragmatic, as opposed to morphosyntactic, functions. Thusly, in those cases speakers would be more obviously using the form to attenuate the propositional

content that follows and not because *de* is mandated by the syntax[6]. This pattern is borne out in the data to the extent that, for multiple discourse-topic types, we see more of a differentiation of *de* use across 1st and 3rd person subjects in the *queísmo* data than in the *dequeísmo* data, although the reader is reminded that the pattern mentioned for *queísmo* still did not reach significance.

An additional conclusion that should be gleaned from the current study is that care should be taken in the study of *queísmo* and *dequeísmo* together, as not all factors demonstrated inverse patterns in the two analyses. Although the selection of more factor groups as statistically significant in the *dequeísmo* analysis may be due to the larger number of tokens in that data set, certain, differential trends emerged in comparing the two phenomena. For example, verbal tense was a significant factor group in both regressions, and yet the preterit strongly disfavored *dequeísmo* in the first analysis, while rather strongly disfavoring *queísmo* (and thus favoring *de*) in the second analysis. Differential conditioning across the two phenomena would require caution in condensing the two together in the future, as some studies have done, such as Guirado, who placed all tokens of both phenomena into one Goldvarb regression, coding each token as including *de* or not. This also likely contributes to differential findings regarding specific factor groups in comparing Serrano, Schwenter, and Guirado.

Looking forward, given the role played by discourse topic in the current study, future studies will do well to further investigate this variable via greater control. Since in a sociolinguistic interview neutral topics are typically discussed and there is not much commentary from the interviewer, it would be interesting to see how the use of *de* might increase in a more typical conversation between two or more interlocutors. Specifically, if the speakers previously knew each other and discussed more heated topics, it is hypothesized that *de* use would be even higher than that found in the current study. Additionally, this would be a beneficial way to obtain more tokens, which may, in turn, yield more statistically significant findings. Such a format would also foster greater use of the second person, which was rarely used in the current data. Thus, it would also enable a more accurate depiction of the role of the person of the subject for these phenomena, since this factor group was not statistically significant in either regression in the current study, likely due in part to the rarity of second person subjects. This, in turn, would make possible additional statistical comparisons within the cross-tabulation of person and discourse topic.

In attempts for greater control, since percentages of *dequeísmo* were found to be quite low in the current study, future studies could also include sentence preference tasks, in which participants were required to choose between two nearly identical sentences that only differed in their inclusion (or exclusion) of *de*. For example, participants could select between "*Me acuerdo de que mi amigo tenía el día más difícil de su vida"* versus "*Me acuerdo que mi amigo tenía el día más difícil de su vida."* In such a task, researchers could also manipulate different discourse topics, different main-clause subjects, etc. The present study has contributed to our knowledge of evidentiality marking through (*de)queísmo,* the role of discourse topic, and the intersection of person and discourse topic. Future research

[6] This is not meant to imply that *de* cannot be used as an innovative morphosyntactic extension, but simply that the addition of *de* in *dequeísmo* (as opposed to the more expected, standard use of *de* in *queísmo*) could be more obviously serving an additional function beyond the morphosyntax for interlocutors and thus be a more effective attenuator of what follows.

will do well to continue to probe how evidentiality is differentially marked across discourse topics, sentential subjects, situational contexts, registers, and dialects.

References

Bentivoglio, P. "Queísmo y dequeísmo en el habla culta de Caracas." In F. M. Aid, M. C. Resnick & B. Saciuk eds. *1975 Colloquium on Hispanic Linguistics*. Georgetown: Georgetown University Press, 1976. 1-18.

---. "El dequeísmo en Venezuela: ¿un caso de ultracorrección?." *Boletín de Filología* 31 (1980-1981): 705-719.

Bentivoglio, P. & D. Galué. "Ausencia y presencia de la preposición *de* ante cláusulas encabezadas por *que* en el español de Caracas: Un análisis variacionista." *Boletín de Filología* 37 (1998-99): 139-158.

Bentivoglio, P. & M. Sedano. "Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana." *Boletín de Lingüística* 8 (1993): 3-35.

Bentivoglio, P. & M. Sedano. *Estudio sociolingüístico del habla de Caracas*. Caracas: Universidad Central de Venezuela, 1987.

Blas Arroyo, J. L. "Mire usted Sr. González… Personal deixis in Spanish political-electoral debate." *Journal of Pragmatics* 32 (2000): 1-27.

Bybee, J. L., R. Perkins & W. Pagliuca. *The evolution of grammar: tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press, 1994.

Chafe, W. & J. Nichols. *Evidentiality: The coding of epistemology in language*. Norwood, NJ: Ablex, 1986.

Davidson, B. "Pragmatic weight and Spanish subject pronouns: The pragmatic and discourse uses of 'tú' and 'yo' in spoken Madrid Spanish." *Journal of Pragmatics* 26 (1996): 543-565.

Díaz-Campos, M. *The handbook of Hispanic sociolinguistics*. Malden, MA: Wiley-Blackwell, 2011.

Flores-Ferrán, N. "An examination of mitigation strategies used in Spanish psychotherapeutic discourse." *Journal of Pragmatics* 42 (2010): 1964-1981.

Geeslin, K. "A comparison of copula choice in advanced and native Spanish." *Language Learning* 53 (2003): 703-764.

Guirado, K. "Deíxis proposicional en el habla de Caracas: Un análisis cuantitativo del *(de)queísmo*." *Boletín de Lingüística* 26 (2006): 130-156.

Johnson, D. E.. "Getting off the GoldVarb standard: Introducing Rbrul for mixed effects variable rule analysis." *Language and Linguistics Compass* 3 (2009): 359-383.

Mollica, M. C. *Queísmo e dequeísmo no português do Brasil*. Doctoral thesis. Universidade Federal do Río de Janeiro, Río de Janeiro, 1989.

Palmer, F. *Mood and modality. 2nd edition.* Cambridge: CUP, 2001.

Penny, R. *A history of the Spanish language*. *2nd edition*. Cambridge: Cambridge University Press, 2002.

Rabanales, A. "*Queísmo* y *dequeísmo* en el español de Chile." In M. J. Tejera ed. *Estudios filológicos y lingüísticos. Homenaje a Ángel Rosenblat en sus 70 años*. Caracas: Instituto Pedagógico de Caracas, 1974. 413-445.

Sankoff, D., S. Tagliamonte & E. Smith. *Goldvarb X: A variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto, 2005.

Schwenter, S. "Evidentiality in Spanish morphosyntax: a reanalysis of (*de)queísmo*." In M. J. Serrano ed. *Estudios de variación sintáctica*. Frankfurt/Main: Vervuert-Iberoamericana, 1999. 65-87.

Serrano, M. J. "Estudio sociolingüístico de una variante sintáctica: el fenómeno dequeísmo en el español canario." *Hispania* 81 (1998): 392-405.

Tagliamonte, S. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press, 2006.

---. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA: Wiley-Blackwell, 2012.

Willett, T. "A cross-linguistic survey of the grammaticalization of evidentiality." *Studies in Language* 12 (1988): 51-97.