Mark Waltermire*

Mood variation after expressions of uncertainty in Spanish

DOI 10.1515/shll-2017-0011

Abstract: According to prescriptive grammars of Spanish, subordinate clauses after expressions of uncertainty require the subjunctive when introduced by the complementizer que ‘that’ (e.g., No creo que venga (S). ‘I don’t believe that he’ll come.’) but require the indicative when introduced by the complementizer si ‘if/whether’ (e.g., No sé si viene (I). ‘I don’t know if he’ll come.’). The use of these moods in actual speech, however, is highly variable. Using a database of conversational Spanish from 10 cities in Latin America and Spain, 428 tokens of subjunctive and indicative mood use were examined according to the following linguistic and social factors – subject agreement, intervening material, frequency of the subordinate verb, age, sex, and education level. Of these variables, only intervening material is statistically significant in the conditioning of mood variation. The use of the subjunctive is preferred after the complementizer que when there is no intervening material separating it from the subordinate verb. In order to further explore possible interactions between individual factors, a Joint Multiple Correspondence Analysis using R was conducted. Results from this exploratory analysis suggest that the use of the subjunctive correlates with young, male, and highly educated speakers whereas the indicative correlates with older, female, and less educated speakers. The variable use of mood after expressions of uncertainty in Spanish may be socially motivated, as these results suggest. As such, the impact of social variables on mood variation should be explored in future studies.

Keywords: mood, variation, subjunctive, morphosyntax, sociolinguistics, exploratory analysis

1 Introduction

1.1 Prescriptive rules for the use of mood in Spanish

In Spanish, as is well known to scholars of Romance Languages, subjunctive mood is overtly marked morphologically for subordinate clause verbs introduced

*Corresponding author: Mark Waltermire, New Mexico State University, Las Cruces, NM, USA, E-mail: markw@nmsu.edu
by the complementizer *que* ‘that’ provided that the matrix clause construction corresponds to one of several semantic categories that “demand” the use of the subjunctive such as desire, volition, comment, epistemicity, and uncertainty. Beginning with Bello (1902: 119–126), prescriptivists have claimed that the use of the subjunctive is obligatory in said contexts. According to this view, the subjunctive form has no real meaning and serves a primarily structural function of co-occurrence with the matrix clause. In other words, the matrix clause lexically determines whether or not the subjunctive will be used. Though this “automatic subjunctive” (a term taken from Connors 1978) may apply frequently to subordinate clause verbs of many dual clause constructions in Spanish with matrix clauses belonging to certain semantic categories (such as those listed above), it is by no means categorical. The claim that these semantic categories require the obligatory use of the subjunctive simply does not explain its actual usage in the Spanish language and fails to explain the variation that exists between the use of subjunctive and indicative mood forms in subordinate clauses. Informal language data reveal that variation exists between the two moods in actual speech. Of the semantic categories listed above, one that demonstrates considerable variation is that of uncertainty, both in bilingual dialects (De La Puente-Schubeck 1992; Lynch 1999; Silva-Corvalán 1994; Studerus 1995; Waltermire 2014) and monolingual dialects (Blake 1985; DeMello 1995; Gudmestad 2010; 2012; Studerus 1995), though the two exhibit substantial differences (Gudmestad 2014; Studerus 1995). The current study focuses solely on the category of uncertainty since many studies (e.g., Blake 1985; De La Puente-Schubeck 1992; Gallego 2016; Gudmestad 2010; 2012; Lynch 1999; Waltermire 2014) show that different categories and constructions condition variation differently. Examples (1) and (2) from the *Corpus del proyecto para el estudio sociolingüístico del español de España y de América* (PRESEEA 2014) demonstrate this type of variation.

(1) *Cuando llega a la casa tiene cinco mil tiques pero no tiene ni un vasito de leche. Entonces creo que no es, no pensar con la cabeza o darle la importancia a lo que no amerita. No creo que un triple te va (I) a resolver la vida.* ‘When he gets home he has 5,000 tickets but doesn’t even have a glass of milk. So I believe that it’s not, not thinking or giving importance to something that doesn’t merit it. I don’t believe that triple (that number) is going to solve your problems.’ (CARA_M13_079)

(2) *Tú puedes tener mucha suerte pero si vas y te metes en cualquier sitio, por más suerte que tengas, no creo que la suerte te vaya (S) a salvar.* ‘You can have a lot of luck but wherever you go, no matter how much luck you have, I don’t believe that luck is going to save you.’ (CARA_M13_079)
Clearly, the use of indicative and subjunctive moods is variable, since they are used after the same matrix clause, even by the same speaker. While prescriptive grammarians erroneously claim that the semantic category of the matrix clause determines whether or not the subjunctive will be used by the speaker, they also state that the indicative must be used when the subordinate clause is introduced by the complementizer *si* ‘if/whether’ (even if the matrix clause falls under one of the semantic categories that supposedly requires the subjunctive). For example, Butt and Benjamin (1988: 295) state that “*si* is never followed by the present subjunctive (except occasionally, in formal literary style after *saber*: *no sé* *si* *sea cierto* ‘I do not know whether it be true’ for *no sé* *si* *es cierto*).” As such, the use of the subjunctive after *no saber* *si* is considered non-standard (Butt and Benjamin 1988; DeMello 1995; Solé and Solé 1977). Since, according to this account of mood use after matrix clauses expressing uncertainty (like *no saber* ‘to not know’), the complementizer in conjunction with the matrix clause determines the use of either mood (*que* = subjunctive; *si* = indicative), it seems highly ironic that the use of mood would have been explained by Bello (1902: 119–126) as determined solely by certain lexical expressions within the matrix clause. Examples (3) and (4) (also from the PRESEEA Corpus) show that the use of indicative and subjunctive moods is also variable for subordinate clause verbs introduced by the complementizer *si* ‘if/whether.’

(3) *Me gustaría que para los jóvenes, e, tuvieran un lugar sano donde divertirse. No sé si tú me entiendes (I).*

‘I would like for young people, um, to have a safe place to have fun. I don’t know if you understand me.’ (LHAB_M21_019)

(4) *Hay que ir hasta la Cuidad Deportiva aquí arriba en Villa Marista para buscar a esos profesores que se reúnen en algunas escuelas. Ahí en el estadio, e, se están dando clases y esas cosas, pero que tengan mejor, e; no sé si me entiendes (S); mejor preparación porque los profesores están preparados.*

‘You’ve got to go to Ciudad Deportiva here above Villa Marista to look for those teachers who get together in some schools. There in the stadium, um, they’re giving classes and that kind of thing, but they need to have better, um; I don’t know if you understand me; better preparation because the teachers are prepared.’ (LHAB_M21_019)

Despite claims to the contrary, examples (3) and (4) suggest that the use of mood is most likely productive and is certainly not determined by semantic
category alone and not even by the complementizer. Given this variation in the linguistic manifestation of mood in Spanish, newer approaches to the understanding of mood have been established by linguists in order to explain the use of the subjunctive (1) as determined by the truth value of the entire proposition as considered by the speaker and (2) as inherently variable and, as such, being governed by certain structural factors.

1.2 Semantically-based explanations of mood variation

The first semantically-based theory was proposed by Lenz (1944) and further discussed in Gili Y Gaya (1969), Hadlich (1971), Rivero (1971), Terrell and Hooper (1974), Garcia and Terrell (1977), Bergen (1978), Lantolf (1978), King (1992), DeMello (1995), Studerus (1995), and Bosque and Demonte (1999). The common denominator in each of these analyses is that variation in the use of indicative and subjunctive mood is contingent upon the speaker’s level of commitment to the truth of the entire proposition. In other words, the matrix clause construction alone does not lexically determine the use or non-use of the subjunctive, but rather conditions it in conjunction with the speaker’s previous knowledge of what is being described. For example, in (3) above, a possible explanation for the speaker’s use of the indicative could be that, even though there is doubt about whether the other interlocutor understands him, he has a greater sense that the interlocutor understands him than not concerning the need for a safe place for young people to play. In this sense, the use of the indicative pragmatically conveys more of an affirmation than uncertainty, while the use of the subjunctive in (4) would be seen as conveying a greater level of doubt as to the interlocutor’s understanding of students’ need for better academic preparation. Within this framework, the subjunctive form is semantically significant and not determined lexically by the matrix clause. Though this clause surely expresses uncertainty on the part of the speaker, the choice of mood signals a semantic nuance that is pragmatically determined by the speaker with regards to his perspective on the interlocutor’s understanding.

The problem with this type of semantic-pragmatic analysis is that assumptions are made about the speaker’s thought processes and intentions, which cannot be measured empirically and are highly speculative. Somewhat ironically, the majority of studies of mood variation in Spanish do not focus exclusively on the analysis of conversational discourse, with several notable exceptions (e.g., DeMello 1995; Finanger 2011; García 2011; Lynch 1999; Silva-Corvalán 1994). Instead, they rely (wholly or partly) on the use of
grammaticality judgments, questionnaires, surveys, and elicitation tasks that do not reflect the actual usage of the subjunctive as it naturally occurs in speech. These approaches, which have been utilized in the vast majority of studies of mood variation in Spanish (e.g., Gallego and Alonso-Marks 2014; García and Terrell 1977; Gudmestad 2010; 2012; Guitart 1982; Lantolf 1978; Montrul 2007; Studerus 1995), have the benefit of yielding a large number of tokens of indicative and subjunctive forms but have many more disadvantages. The vast majority of elicitation tasks clearly reveal the object of study (i.e., it would be obvious even to the non-linguist that mood forms were being elicited from participants). Furthermore, the use of isolated examples of speech is incapable of reflecting the actual variation that exists between the indicative and subjunctive in subordinate clauses since the use of these moods is a discourse-level phenomenon according to the semantic-pragmatic framework utilized in these studies. A fill-in-the-blank prompt simply does not capture the range of complexity presented by the diverse contexts in which subjunctive and indicative moods manifest themselves in Spanish. Further complicating matters is the fact that participants are surely aware that their language choices are being scrutinized in some way, which typically yields formalized results as participants aim to choose the “correct” form.

If we are to approach an understanding of the semantic significance of the subjunctive and the pragmatic factors that account for its use, we must use naturalistic data that accurately portray the forms that speakers choose in natural settings rather than delimiting possible uses of mood according to narrowly defined contexts that are not produced by the speakers themselves. The irony of this type of data elicitation is that if the speaker’s willingness to commit to the truth of the entire proposition affects the choice of mood and is naturally conditioned by his prior knowledge of the situation being described, how can he choose mood when the examples in a questionnaire do not reflect his previous knowledge or experiences in any way?

1.3 Sociolinguistic approaches to the explanation of mood variation

A growing number of recent studies on mood variation in Spanish (Finanger 2011; Gallego and Alonso-Marks 2014; García 2011; Gudmestad 2010; 2012; King et al. 2008; Waltermire 2014) have utilized a variationist sociolinguistic framework (described in Guy 1993; Labov 1994; 2001; Tagliamonte 2011) to explain the use of mood in this language. According to Tagliamonte (2013: 129), “one of the foundations of variationist analysis is its attempt to discover not
individual occurrences, not even overall rates of occurrence, but patterns of variability in the body (or bodies) of material under investigation.” This approach to mood variation was pioneered by Shana Poplack (1992), who utilized a variationist framework in order to establish a hierarchy of constraints that govern the use of mood in French. In Poplack’s study of the conditioning of mood variation in French, she claims that for lexical heads in the matrix clause other than falloir ‘have to’ (which is lexically conditioned for the subjunctive), the use of the subjunctive is determined by several linguistic factors found within the utterance, which are: 1) appearance of the imperfect, passé composé, present, or future verb tenses in the matrix clause, 2) no distance between the two clauses, 3) morphologically suppletive/frequent forms of the subordinate verb, and 4) presence of the complementizer que ‘that.’ According to this analysis, the use of the subjunctive in subordinate clauses in which the matrix clause head is not falloir ‘have to’ is determined primarily by linguistic factors. This approach is desirable due to the fact that it explains variation in the use of the indicative and subjunctive in contexts that prescriptively require the subjunctive and does not make any assumptions about the speaker’s intent or prior knowledge.

2 Data and methodology

The purpose of the current research is to determine whether the use of the indicative or subjunctive in subordinate clauses when the matrix verb expresses uncertainty is determined by linguistic factors such as those described in Poplack (1992), social factors, or both. The current study is the first to date to focus solely on the sociolinguistic conditioning of mood variation after expressions of uncertainty in monolingual Spanish. Following Gudmestad (2010: 41), “existing research does not provide conclusive evidence on the role of social variables.” This is partly due to the fact that the role of social variables (aside from nationality) in the use of mood has not

---

1 DeMello (1995) does focus solely on mood variation after the expression no sé si ‘I don’t know if’ in multiple dialects of Spanish, but (1) does not analyze mood variation after other expressions of uncertainty nor (2) does he attempt to explain the conditioning of mood. Instead, frequencies of use of both moods in multiple monolingual dialects after this expression are compared according to several linguistic factors. Though de la Puente Schubeck (1992) and Studerus (1995) examine social factors related to mood variation after expressions of uncertainty in this language, their analyses focus on bilingual speech and mood variation after a multitude of different expressions.
been explored for monolingual varieties of Spanish. There are sound reasons for examining social variables, especially given that the role of generation has been found to be a significant determinant of mood use in bilingual varieties of Spanish, with speakers of the youngest generation tending to use the subjunctive less often than speakers of older generations (De La Puente-Schubeck 1992; Lynch 1999; Silva-Corvalán 1994). Though generation is socially constructed while age is biological, there is overlap between the two. This distinction is important with regards to language contact since generation is often based on immigration (i.e., the first generation is the first to have lived in a new country). To avoid complications regarding these terms, the current study will focus on age, not generation. The variables that seem to have the greatest conditioning effect on the choice of mood forms are a speaker’s education level, age, and sex. Though these factors may not condition the variable use of mood in monolingual Spanish, the fact that they have not been pursued in previous research on mood variation in these varieties is altogether surprising and represents a gap in our understanding of this phenomenon.

Multivariate analysis is well suited for the current study in that it quantitatively shows which factors favor the use of the subjunctive within the given functional domain (i.e., after the complementizer of a matrix clause expressing uncertainty). I will examine the same linguistic variables that Poplack analyzed in her 1992 study, with minor modifications. Like Poplack, I will analyze the following three variables that seem likely to influence the choice of mood – 1) form of the complementizer (either *que* ‘that’ or *si* ‘if/whether’), 2) distance (i.e., intervening material) between the matrix and subordinate verbs, and 3) lexical frequency of the subordinate verb. I will also add a fourth variable that seems of particular importance – the presence or absence of the same subject of the matrix and subordinate clauses, as this was found to be a significant variable influencing the use of mood in DeMello (1995). Another variable shown to be of importance in the conditioning of mood in certain contexts in the work of several authors (e.g., Finanger 2011; García 2011; Gudmestad 2012) is time reference. Though this variable was included upon initial analysis, speakers’ use of the indicative was categorical in past reference contexts and, for this reason, will not be pursued here further. Tokens were coded for these independent factors in the following way:

1. form of the complementizer – (a) presence of *que* ‘that’ vs. (b) presence of *si* ‘if/whether’;
2. distance between the matrix and subordinate verbs – (a) separated by a complementizer vs. (b) separated by more than just the complementizer (Note: clitics of any type were not counted as intervening material);
3. lexical frequency of the subordinate verb – (a) frequent vs. (b) infrequent\(^2\).
4. subjects of the matrix and subordinate clauses – (a) same subject, (b) different subjects, and (c) ambiguous (due to the fact that Spanish is a pro-drop language and the verb forms for yo 'I' and él/ella/Usted 'he/she/you (formal)' are the same for the imperfect, conditional, present subjunctive, and past subjunctive).

The data for the current study come from the online version of the PRESEEA Corpus (http://preseea.linguas.net), which consists of semi-guided interviews of native Spanish speakers from ten major cities throughout Latin America (Caracas, Havana, Lima, Medellín, Monterrey, Montevideo, Santiago) and Spain (Alcalá de Henares, Madrid, Valencia). The number of participants varies from city to city, but there are more participants (N = 72 to 108) from cities with populations that exceed one million than there are from cities of fewer than one million residents (N = 54) (Moreno-Fernández 2005: 128). Dialect was not included as a variable for the current study since these cities are not representative of the national dialects spoken in their respective countries. As part of the PRESEEA Project, these interviews were transcribed and entered into a searchable digital database. Participants for the project were chosen according to specific residence requirements and are equally represented by sex, age, and education level (Moreno-Fernández 2005). The age groups utilized in the PRESEEA Corpus, which appear as “generation,” are the following: 1\(^{\text{st}}\) generation = 20 to 34 years of age; 2\(^{\text{nd}}\) generation = 35 to 54 years of age; 3\(^{\text{rd}}\) generation = 55 years of age and older (Moreno-Fernández 2005). As just discussed, age may or may not be significant in the conditioning of mood choice in monolingual Spanish. This factor has not been explored for mood variation in monolingual varieties of this language. However, it does seem to be important in bilingual dialects, as discussed previously. Likewise, the role of sex and education level have only been investigated in bilingual varieties. The inclusion of these factors is important given previous findings that suggest that women create more opportunities for the use of the subjunctive, which is more likely among highly educated speakers (Waltermire 2014) of bilingual Spanish. In

\(^2\) The frequency with which the subordinate verb occurs overall within the Spanish language was determined using Mark Davies’ *A frequency dictionary for Spanish* (2006). This dictionary, which contains the 5,000 most frequently occurring words in the Spanish language, was based on a 20-million word corpus evenly divided among spoken language samples and written (both fiction and non-fiction) texts from Spain and Latin America. For the purposes of the current study, it was determined that there should be a distinction between frequent and infrequent words. Frequent words are those actually listed in the Davies dictionary while infrequent words are not included.
other words, women use mood-based expressions (i.e., those expressing volition, personal comment, uncertainty, future intention, etc.) to a greater degree than men do. Participants fall into three categories with respect to education level – low, medium, and high. Any instance of language use, including multi-word collocations, can be searched according to these factors within the corpus.

The instances of subjunctive and indicative used in subordinate clauses were analyzed for the following matrix clauses which express uncertainty: *dudar (de) que* ‘to doubt that,’ *es dudoso que* ‘it is doubtful that,’ *hay duda (de) que* ‘there is doubt that,’ *no saber que/si* ‘to not know that/whether,’ *no se sabe que/si* ‘it is not known that/whether,’ *no creer que* ‘to not believe that,’ *no pensar que* ‘to not think that,’ *no estar seguro/a (de) que/si* ‘to not be sure that/whether,’ *no es seguro que/si* ‘it is not certain that/whether,’ *no es cierto que/si* ‘it is not true that/whether,’ *es incierto que/si* ‘it is uncertain that/whether,’ *no es verdad que* ‘it is not true that,’ and *¿quién sabe si? ‘Who knows if?’3 Many of these clauses were not found. For clauses including *si* ‘if/whether,’ only the clause *no saber si* ‘to not know if’ was encountered. For those including the complementizer *que* ‘that,’ I found only four clauses - *no creer que* ‘to not believe that,’ *no pensar que* ‘to not think that,’ *no saber que* ‘to not know that,’ and *no estar seguro/a que* ‘to not be sure that.’ These matrix clauses yielded a total of 428 tokens, all of which were found by searching the PRESEEA Corpus, which shows the context in which tokens are used and allows direct access to the location of tokens in the contexts of the interviews themselves. All tokens were divided according to indicative or subjunctive forms and coded for the linguistic and social factors just described.

3 Variable use of mood after expressions of uncertainty

As stated thus far, the use of mood in clauses expressing uncertainty is variable for native speakers of Spanish. Of the 428 tokens in the current data set, the subjunctive was used 127 times (29.7%) while the indicative was used 301 times (70.3%). Given the prescriptive claims that the complementizer *si* ‘if’ does not collocate with the subjunctive (see Section 1.1), it is expected that there will be a general tendency for speakers to use the indicative more frequently after the

---

3 Negated clauses of certainty (e.g. *no creer que* ‘to not believe that,’ *no pensar que* ‘to not think that,’ etc.) have been included in the current study since they are expressions of uncertainty that demonstrate variation in the use of mood in Spanish just like non-negated expressions of uncertainty.
matrix clause *no saber si* ‘to not know if.’ To see if the form of the complementizer (either *si* ‘if/whether’ or *que* ‘that’) has any effect on the choice of mood, each of the matrix clauses corresponding to these complementizers were grouped together and analyzed separately for the choice of mood. The distribution of mood by each complementizer form is shown in Table 1.

Table 1: Distribution of mood in subordinate clauses preceded by a matrix clause expressing uncertainty according to the form of the complementizer.

<table>
<thead>
<tr>
<th></th>
<th>Indicative</th>
<th>Subjunctive</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Si</em></td>
<td>252/287 (87.8%)</td>
<td>35/287 (12.2%)</td>
</tr>
<tr>
<td><em>Que</em></td>
<td>49/141 (34.8%)</td>
<td>92/141 (65.2%)</td>
</tr>
</tbody>
</table>

\( \chi^2 = 127.53; df = 1; p < 0.0001. \)

As expected, use of the subjunctive is greater for subordinate clauses introduced by *que* while use of the indicative is more prevalent after the complementizer *si*. Variation in the choice of mood after *que* clearly shows that the semantic class of uncertainty followed by this complementizer does not determine the use of what has been called the “automatic subjunctive” by Conners (1978). The reverse could also be said of subjunctive use following *si*. Though use of the subjunctive after this complementizer is limited to 12.2%, speakers still use this mood in this context, despite prescriptive claims to the contrary. The highly frequent use of the indicative for subordinate clauses introduced by *si* is not surprising since it does not occur in constructions “demanding” the use of the subjunctive (with the exception of “if clauses” which suppose a hypothetical commentary, e.g., *Si fueras presidente ... ‘If you were president ... ’*). There may be an association between the complementizer *que* and the subjunctive, however, given the great number of subordinate clauses introduced by this complementizer that take the subjunctive (e.g., in clauses of volition such as *Quiero que sepas* ‘I want you to know,’ of epistemicity such as *Es posible que llueva* ‘It’s possible that it will rain,’ and of subjective commentary such as *Es una lástima que no puedas venir* ‘It’s a shame that you can’t come’).\(^4\) In other

---

\(^4\) Gudmestad (2010; 2012) shows that native Spanish speakers of varied nationality exhibit variation in the choice of mood for these types of matrix clauses as well (which also prescriptively call for the use of the subjunctive). Furthermore, Garcia and Terrell (1977) show that native speakers accept the use of the indicative after clauses of these semantic categories. The participants of their study accepted use of the indicative after clauses of volition at a rate of 8%, clauses of uncertainty and epistemicity at a rate of 25%, and clauses of subjective comment at a rate of 48%. 
words, an association has been made between the use of the subjunctive and the complementizer *que* due to the relatively high frequency with which these constructions occur.

Given the statistically significant distribution of mood use shown in Table 1, it will be necessary to analyze the use of mood separately for each complementizer. As far as I am aware, this is the first study to examine mood variation in Spanish according to the form of the complementizer. Examining the use of mood regardless of the complementizer would yield inaccurate results since the use of the indicative would appear to be more common than the subjunctive after all expressions of uncertainty while this is really only true for the matrix clause *no saber si* ‘to not know if.’ Though there is less variation after this clause than for clauses introduced by *que*, use of mood after *no saber si* is still variable and, for this reason, will be analyzed further. This variability may be the result of linguistic factors, social factors, or both. First we will turn to the role of linguistic factors in the conditioning of mood variation.

### 3.1 Linguistic factors

The following three linguistic factors (mentioned in Section 2) will now be analyzed - distance between the matrix clause and subordinate verb, lexical frequency of the subordinate verb, and presence or absence of the same subjects of the matrix and subordinate clauses. For clauses introduced by either complementizer, there are several factors that seem likely to favor use of the subjunctive. No distance between the matrix clause and subordinate verb (i.e., no intervening material), in other words when they are separated only by the complementizer (e.g., *No sé si viene* ‘I don’t know if he’s coming’ as opposed to *No sé si su otra hermana viene* ‘I don’t know if his other sister is coming’), would seem to favor the subjunctive for reasons of linguistic processing. Given that subjunctive forms are much less frequent in Spanish than indicative forms and are therefore less efficiently accessed from stored memory, the ability of speakers to recall these forms and use them accurately is more facile when no intervening material exists. Hawkins (1999; 2001) has shown that adjacent forms (for which there is no intervening material) are retrieved more efficiently than non-adjacent forms given that there is less information to process. Though the subjunctive is not commonly used after the complementizer *si* (at a rate of merely 12.2%; see Table 1) it is still expected that the use of the subjunctive will be greater in this context when there is no intervening material between the matrix clause and subordinate verb. For similar reasons of processing, it seems that the use of the subjunctive would also be preferred for highly frequent verbs.
That is, recall of frequently used forms is more efficient than that of infrequently used forms since they are accessed more readily from stored memory (Bybee 2001; 2010). Following Bybee (2001: 28), “since tokens of use map onto existing representations, high-frequency items grow strong and therefore are easier to access.” Lastly, the presence of different subjects would seem to favor the subjunctive. DeMello (1995: 561) found that the subjunctive was used after no sé si ‘I do not know if’ with non-coreferential subjects at a far greater rate (88\%) than the indicative in the Spanish of Bogotá, Caracas, Mexico City, and Santiago. This is most likely due to the fact that the subjunctive is generally described as a reservation on the part of the speaker regarding the truth of the utterance and/or a way to express lack of previous knowledge with regards to the truth of the proposition (which ultimately means he must be describing something that was done by someone or something other than himself).^5^5

Given the small data set for the current study, which is due to the narrow variable context (matrix clauses expressing uncertainty) and the nature of the data itself (conversational speech), statistical significance for each factor group will be assessed using Fisher’s exact tests of probability, which are more reliable than other statistical tests when using small data sets. Following Pederson (1996: 4), a Fisher’s exact test “computes the significance of an observed table by exhaustively computing the probability of every table that would lead to the marginal totals that were observed in the sampled table. The significance values obtained using Fisher’s exact test are reliable regardless of the distributional characteristics of the data sample.” The two-tailed p-values are provided for each variable directly below all reported frequencies, as in Table 2.

Table 2 shows that none of the posited predictions hold for the use of the subjunctive in subordinate clauses introduced by si. The use of the subjunctive according to these factors never exceeds a rate of 12.2\%, with the exception of infrequent subordinate verbs at 14\%, which is actually contrary to expectations, though hardly a great enough difference to make any claims that this is an important factor in the use of the subjunctive mood. None of these factors is significant in the determination of mood use for subordinate clauses introduced by si. But perhaps they do play a role in the determination of mood use for

^5^ There were 11 instances for which the subjects of the two clauses were ambiguous (6 after si; 5 after que) and, as such, unidentifiable. This is due to the fact that Spanish is a pro-drop language and that verb forms coincide for first and third person singular forms of the imperfect, conditional, present subjunctive, and past subjunctive as stated in Section 2. No other discourse cues served to indicate the nature of the subjects in these instances. As a result, and since they were only identifiable to the participants themselves, the category of “ambiguous subject agreement” cannot and should not be considered a separate factor and will, therefore, not be included in calculations.
subordinate clauses introduced by *que*, which show much more variance in the use of mood forms and a preference for the subjunctive (at 65.2%; see Table 1). The distribution of mood for these clauses is shown in Table 3.

As Table 3 shows, our predictions hold for the first two factors, but not the third. The subjunctive is used more commonly when the matrix and subordinate clauses have different subjects (at a rate of 67%) and when there is no intervening material between the matrix clause and subordinate verb (at a rate of 70.1%). Of course, these rates are not much greater than the average rate of subjunctive use for subordinate clause verbs following *que*. Like the use of the subjunctive after *si*, the use of this mood is also more common for infrequent subordinate verbs following *que*. Unlike the findings of Poplack (1992), in which frequent verbs in French favor the subjunctive while infrequent verbs do not. Of these variables, only intervening material is statistically significant in the distribution of mood after expressions

<p>| Table 2: Distribution of mood according to linguistic factors for all subordinate clauses introduced by the complementizer <em>si</em>. |
|-----------------|-----------------|-----------------|-----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Factor group</th>
<th>Factor</th>
<th>Indicative</th>
<th>Subjunctive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clause subjects</td>
<td>Same</td>
<td>38/43 (88.4 %)</td>
<td>5/43 (11.6 %)</td>
<td>43/281 (15.3 %)</td>
</tr>
<tr>
<td></td>
<td>Different</td>
<td>211/238 (88.6 %)</td>
<td>27/238 (11.4 %)</td>
<td>238/281 (84.7 %)</td>
</tr>
<tr>
<td>Intervening material</td>
<td>Absent</td>
<td>194/221 (87.8 %)</td>
<td>27/221 (12.2 %)</td>
<td>221/287 (77 %)</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>58/66 (87.9 %)</td>
<td>8/66 (12.1 %)</td>
<td>66/287 (23 %)</td>
</tr>
<tr>
<td>Subordinate verb frequency</td>
<td>Frequent</td>
<td>166/187 (88.8 %)</td>
<td>21/187 (11.2 %)</td>
<td>187/287 (65.2 %)</td>
</tr>
<tr>
<td></td>
<td>Infrequent</td>
<td>86/100 (86 %)</td>
<td>14/100 (14 %)</td>
<td>100/287 (34.8 %)</td>
</tr>
</tbody>
</table>

Clause subjects: *p = 1*; Intervening material: *p = 1*; Subordinate verb frequency: *p = 0.57.*

<p>| Table 3: Distribution of mood according to linguistic factors for all subordinate clauses introduced by the complementizer <em>que</em>. |
|-----------------|-----------------|-----------------|-----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Factor group</th>
<th>Factor</th>
<th>Indicative</th>
<th>Subjunctive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clause subjects</td>
<td>Same</td>
<td>10/21 (47.6 %)</td>
<td>11/21 (52.4 %)</td>
<td>21/136 (15.4 %)</td>
</tr>
<tr>
<td></td>
<td>Different</td>
<td>38/115 (33 %)</td>
<td>77/115 (67 %)</td>
<td>115/136 (84.6 %)</td>
</tr>
<tr>
<td>Intervening material</td>
<td>Absent</td>
<td>32/107 (29.9 %)</td>
<td>75/107 (70.1 %)</td>
<td>107/141 (75.9 %)</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>17/34 (50 %)</td>
<td>17/34 (50 %)</td>
<td>34/141 (24.1 %)</td>
</tr>
<tr>
<td>Subordinate verb frequency</td>
<td>Frequent</td>
<td>37/101 (36.6 %)</td>
<td>64/101 (63.4 %)</td>
<td>101/141 (71.6 %)</td>
</tr>
<tr>
<td></td>
<td>Infrequent</td>
<td>12/40 (30 %)</td>
<td>28/40 (70 %)</td>
<td>40/141 (28.4 %)</td>
</tr>
</tbody>
</table>

Clause subjects: *p = 0.22*; Intervening material: *p = 0.04*; Subordinate verb frequency: *p = 0.56.*
of uncertainty followed by the complementizer *que*. This may be a processing phenomenon whereby retrieval of the subjunctive form in adjacent contexts is more efficient since it requires less cognitive processing than when it occurs in non-adjacent contexts. It should also be noted here that adjacent contexts are more frequent than non-adjacent contexts (see Tables 2 and 3) which also likely results in more efficient retrieval of the subjunctive. That said, this should also be the case after *no saber si*, but it is not (see Table 2). It could be, however, that the two expressions behave differently with regards to distance or that, in the case of non-adjacent contexts, the semantic complexity of intervening material drives the use of either mood. These possibilities should be considered in future research on mood variation in Spanish.

### 3.2 Social factors

There have been very few studies on the relation between social factors and their possible conditioning role in the use of mood in Spanish, practically all of which (with the exception of DeMello (1995), who looked at education level) have focused on bilingual varieties of this language (see Section 2). This may largely be due to the fact that the use of the subjunctive and indicative moods has not been seen as a social issue. This is altogether surprising since the use of mood seems to be conditioned at the very least by education level. Though DeMello (1995) did not find any difference in the use of the indicative after *no sé si* among educated and uneducated Mexican speakers, prescriptive forms usually see greater use among highly educated speakers, especially when the use of these forms is highly complex. The social prestige attached to the “correct” use of prescriptive forms may be a prime motivator for their use; and the use of mood forms may be no different. Waltermire (2014), who examined the variable use of subjunctive and indicative forms in the bilingual Spanish of Albuquerque, New Mexico, found higher rates of subjunctive use among speakers with greater levels of formal education in this language. Though this factor was not found to be significant in de la Puente-Schubeck (1992: 133), she found that several other social variables are significant in the conditioning of the subjunctive, namely that of age. In her study, also of a bilingual speech community (of northern New Mexico), de la Puente-Schubeck found that older speakers (over 46 years of age) used the subjunctive 32% of the time, while middle age and young speakers used the subjunctive only 19% and 13% of the time, respectively. This is an important finding, though it may be related to the bilingual situation in northern New Mexico, where older
speakers tend to speak Spanish more often than younger speakers (Bills 1997; Bills and Vigil 1999). Like De La Puente-Schubeck (1992), Waltermire (2014: 124) did not find any significant differences in the use of mood between men and women, though it was shown in this study that women create almost three times as many mood-based constructions in speech. Though rates of use of the subjunctive may be lower in bilingual communities, whether due to advanced internal simplification (Silva-Corvalán 1994) or to a preference for English (Waltermire 2014), there is no reason to assume that social factors such as education level, age, and sex do not play a role in the conditioning of mood use in monolingual varieties of Spanish as well. Their potential impact on the use of mood forms will now be analyzed.

As with the analysis of linguistic factors on the use of mood after expressions of uncertainty in Spanish, the analysis of mood use according to social factors will be analyzed separately by complementizer. The distribution of mood for subordinate clauses introduced by si is shown first, in Table 4.

Table 4: Distribution of mood according to social factors for all subordinate clauses introduced by the complementizer si.

<table>
<thead>
<tr>
<th>Factor group</th>
<th>Factor</th>
<th>Indicative</th>
<th>Subjunctive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education level</td>
<td>Low</td>
<td>60/71 (84.5 %)</td>
<td>11/71 (15.5 %)</td>
<td>71/287 (24.7 %)</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>83/90 (92.2 %)</td>
<td>7/90 (7.8 %)</td>
<td>90/287 (31.4 %)</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>109/126 (86.5 %)</td>
<td>17/126 (13.5 %)</td>
<td>126/287 (43.9 %)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>132/151 (87.4 %)</td>
<td>19/151 (12.6 %)</td>
<td>151/287 (52.6 %)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>120/136 (88.2 %)</td>
<td>16/136 (11.8 %)</td>
<td>136/287 (47.4 %)</td>
</tr>
<tr>
<td>Age</td>
<td>20–34</td>
<td>77/86 (89.5 %)</td>
<td>9/86 (10.5 %)</td>
<td>86/287 (30 %)</td>
</tr>
<tr>
<td></td>
<td>35–54</td>
<td>95/105 (90.5 %)</td>
<td>10/105 (9.5 %)</td>
<td>105/287 (36.6 %)</td>
</tr>
<tr>
<td></td>
<td>≥ 55</td>
<td>80/96 (83.3 %)</td>
<td>16/96 (16.7 %)</td>
<td>96/287 (33.4 %)</td>
</tr>
</tbody>
</table>

Education level: $p = 0.27$; Sex: $p = 0.86$; Age: $p = 0.28$.

Though none of the social variables just analyzed is statistically significant in the distribution of mood forms following the complementizer si, there are two interesting observations to make concerning Table 4. First, older speakers ($≥ 55$) use the subjunctive after no saber si more often than do younger speakers (20–54). This is expected given that older speakers tend to be less conservative in the use of standard forms in speech since they are no longer engaged in the “linguistic marketplace” (Sankoff and Laberge 1978) and do not feel pressures to conform to standard usage for economic and social reasons. Second, there is a lower rate of use of the subjunctive among speakers of a medium education level
than among speakers with a high level of education. Since no saber si normatively takes the indicative, these speakers may be conforming to a normative use of mood in this context as a means toward greater social ascendancy. It is interesting to note, also, that more opportunities for the use of mood were created with each level of education (from 71 to 90 to 126). This is also true for subordinate clauses introduced by the complementizer que, as shown in Table 5.

Table 5: Distribution of mood according to social factors for all subordinate clauses introduced by the complementizer que.

<table>
<thead>
<tr>
<th>Factor group</th>
<th>Factor</th>
<th>Indicative</th>
<th>Subjunctive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education level</td>
<td>Low</td>
<td>6/19 (31.6%)</td>
<td>13/19 (68.4%)</td>
<td>19/141 (13.5%)</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>14/53 (26.4%)</td>
<td>39/53 (73.6%)</td>
<td>53/141 (37.6%)</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>29/69 (42%)</td>
<td>40/69 (58%)</td>
<td>69/141 (48.9%)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>25/64 (39.1%)</td>
<td>39/64 (60.9%)</td>
<td>64/141 (45.4%)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>24/77 (31.2%)</td>
<td>53/77 (68.8%)</td>
<td>77/141 (54.6%)</td>
</tr>
<tr>
<td>Age</td>
<td>20–34</td>
<td>16/57 (28.1%)</td>
<td>41/57 (71.9%)</td>
<td>57/141 (40.4%)</td>
</tr>
<tr>
<td></td>
<td>35–54</td>
<td>17/51 (33.3%)</td>
<td>34/51 (66.7%)</td>
<td>51/141 (36.2%)</td>
</tr>
<tr>
<td></td>
<td>≥ 55</td>
<td>16/33 (48.5%)</td>
<td>17/33 (51.5%)</td>
<td>33/141 (23.4%)</td>
</tr>
</tbody>
</table>

Education level: $p = 0.20$; Sex: $p = 0.38$; Age: $p = 0.14$.

The greater number of opportunities for the choice of mood found in educated speech regardless of the complementizer may reflect the fact that these speakers have been taught that a significant semantic difference exists between the two forms (Blake 1991) and that they have acquired less frequent forms in this way since acquisition of mood use after expressions of uncertainty occurs after that of more frequent expressions (Pereira Rodríguez 1997). This result supports the observation that the use of the subjunctive is learned, but that educated speakers may simply choose not to use prescriptively correct forms, preferring rather to tease out the semantic nuances between the different moods according to sociolinguistic context. Again, as we saw with no saber si, speakers of a medium education level follow normative uses of mood more closely than speakers of low and high education levels. They use the subjunctive after matrix clauses with que (the normative form in this context) the most frequently of all three groups. This suggests even further that these speakers may be attempting to gain prestige by adhering more closely to standard uses of mood. Also like mood use after no saber si, the use of mood among women and men is not highly divergent after expressions of uncertainty followed by the complementizer que. These findings further support those of De La Puente-Schubeck (1992), Studerus (1995), and Waltermire (2014). Further supporting the idea that older speakers...
are less conservative with the prescriptive use of mood (as shown in Table 4) is the fact that they use the subjunctive least often of any group (at a rate of only 51.5 %), presumably because they do not feel the same pressures to conform to standard speech as do younger speakers.

3.3 Exploratory analysis

The analysis of rates of use of subjunctive and indicative mood forms after expressions of uncertainty did not reveal any definitive patterns with respect to either linguistic or social factors, though a lack of intervening material between the complementizer *que* and the subordinate verb does seem to favor use of the subjunctive. Furthermore, the conditioning of mood use after expressions of uncertainty cannot be confirmed due to a relatively low number of tokens which has resulted from a narrow variable context (expressions of uncertainty) and the nature of the data (conversational speech) for the current study. Given this limitation, an exploratory analysis is needed in order to gauge the potential conditioning effects of linguistic and social factors on the use of mood forms. Since social variables have never been fully examined with respect to mood variation in monolingual Spanish, an exploratory approach is necessary. It will indicate which variables should be explored in future research and paves the way for confirmatory analyses based on larger bodies of data.

One of the most effective ways to explore statistical correlations between the use of mood and the linguistic and social factors just examined is by conducting a Joint Multiple Correspondence Analysis using R (R Core Team 2016). Joint Multiple Correspondence Analysis reveals frequency-based correlations in corpus data, which are generated graphically as biplots on a two-dimensional plane. The most explanatory dimension is Dimension 1 (horizontal axis) since it captures the greatest amount of explained inertia (i.e., degree of variation in the data) while Dimension 2 (vertical axis) captures the second greatest amount of explained inertia. The relative proximity of points indicates their relative degree of correlation when all factors are considered simultaneously. There are three types of Multiple Correspondence Analysis – indicator, Burt, and joint. Joint Multiple Correspondence Analysis was conducted for the current study as it is more effective in explaining inertia and depicting correlations graphically (Greenacre 2007). The accepted threshold for inertia is generally 80 %. For the current analysis, it is 87.6 %. The results from this analysis are displayed in Figure 1. In order to interpret Figure 1, one needs to focus on the intersection of the vertical and horizontal lines (i.e., the origin or the center) of the plot at coordinates (0.0). The importance of this intersection lies in that it is the meeting
point of the average row and column profiles. As such, the further data points are from this intersection, the more different they are from the average profile. The first observation from this analysis is that the subjunctive (MOOD:S) correlates with the complementizer *que* (COMP:Q) while the indicative (MOOD:I) correlates with the complementizer *si* (COMP:S), which was shown previously in Table 1. Correlations with the subjunctive are not as strong among factors as they are for the indicative. Unfortunately, the points converge toward the middle of both axes to such an extent that they are difficult to read. In order to home in on these factors, I zoomed in on this portion of the graph so that they would be easy to see and interpret. These points can be seen easily in Figure 2.

Though the points for the subjunctive (MOOD:S) and the complementizer *que* (COMP:Q) are out of range in Figure 2, which now only shows the middle portion of the graph, correlations with the subjunctive can now be seen more easily. Young speakers (AGE:A), males (SEX:M), and more highly educated speakers (ECUD:M and EDUC:H) more closely correlate with use of the subjunctive than the indicative as these points are much closer in proximity to the subjunctive. Each of these predictions was made in the previous two sections, but their role is more clearly presented here as they have not been separated by complementizer. The exact opposite factors are most highly correlated with use

---

**Figure 1:** Joint Multiple Correspondence Analysis of mood and all other factors.
of the indicative (MOOD:I). That is, the points of closest proximity to this mood are less educated speakers (EDUC:L), females (SEX:F), and older speakers (AGE:C). These are important findings. Even though they are exploratory in nature and cannot be used to confirm the conditioning role of these factors in the use of mood, they do suggest that social factors do have some role in the conditioning of mood use in monolingual varieties of Spanish, which has not received any significant attention in the field. Essentially, they point the way for future research on mood variation in Spanish and suggest that social factors should be included in future analyses. These studies should be confirmatory in nature if possible, but this is made difficult at present due to a lack of large oral corpus data for Spanish that is tagged for social factors.

Another factor that may be of importance to future research on the variable use of mood after expressions of uncertainty in monolingual varieties of Spanish is Mexican nationality. During the process of coding, I observed that Mexican speakers tend to use the subjunctive often after no saber si. Since this variable only seems important after the complementizer si, it was not included as one of the variables in Table 4 (for reasons of consistency; i.e., including the same variables for both complementizers) nor was it included in the Joint Multiple Correspondence Analysis (for which factors cannot be separated by

---

**Figure 2:** Joint Multiple Correspondence Analysis of mood and all other factors (zoom).
Furthermore, dialect was not included as a variable in the current study in general given that participants for the PRESEEA Corpus come from a handful of cities (see Section 2) that are not representative of national dialect (i.e., speakers from Medellin are not representative of Colombian Spanish). High rates of subjunctive use after this clause among Mexican speakers had previously been reported in DeMello (1995), which led me to include Mexican nationality (as opposed to the seven other nationalities represented for the current study) as a factor to consider further. The rate of use of the subjunctive after the matrix clause *no saber si* among Mexican speakers is remarkably high as shown in Table 6.

The extremely high rate of use of the subjunctive after this matrix clause by Mexican nationals resembles that of DeMello (1995), who found that Mexican speakers use the subjunctive after *no saber si* 65% of the time. It is difficult to make a very clear comparison between the two studies since there were far fewer examples of this construction in DeMello’s study (N = 17) and his participants were from Mexico City rather than Monterrey.\(^6\) It does seem clear, however, that Mexican speakers display a tendency to use the subjunctive often after *no saber si*, unlike speakers from other countries. The Mexican speakers for the current study used this mood 15 out of 33 times (at a rate of 45.5%), which is almost as much as the total number of times the subjunctive was used after *no saber si* for all non-Mexican participants (from nine cities throughout Latin America and Spain). These speakers combined for a total of 20 uses of the subjunctive after *no saber si* out of a possible 254 instances (at a rate of only 7.9%). This is an interesting finding that shows that the morphosyntactic characteristics of certain dialects are subtle and subject to variation rather than merely present or absent, as they have often been described (see Lipski 1994).

\(\chi^2 = 38.5; \ df = 1; \ p < 0.0001.\)

\(^6\) Given that neither Mexico City or Monterrey can be considered representative of Mexican Spanish in general, these results should be taken with a grain of salt. Furthermore, the low numbers of use for this study (N = 33) are far from definitive even for speakers of Monterrey. That said, this is an expression that should be studied more in the future, ideally using a large body of data of Mexican Spanish that includes speakers from all over the country.
4 Conclusion

Variation in the use of indicative and subjunctive mood forms in Spanish after expressions of uncertainty is indeed a complex phenomenon that has never been fully explored. Somewhat surprisingly, the role of social factors in the conditioning of mood variation after these expressions in monolingual Spanish has never been studied until now. In fact, even for other expressions that exhibit variation in the use of mood in monolingual Spanish, these factors (with the exception of nationality) have never received much attention. The results from the current study demonstrate that there indeed is a need to study the role of social factors in the conditioning of mood variation in monolingual, not just bilingual, varieties of this language. Though semantic nuances between the indicative and subjunctive may exist, assessing a speaker’s prior knowledge and level of commitment to the truth of a proposition is seemingly impossible to prove empirically and has proven speculative in past studies. The current study adds to the growing body of empirical research on mood variation in monolingual Spanish, but presents two key differences – a focus on social factors and the exclusive use of conversational speech data as the object of analysis.

Regardless of the complexity of conditioning of mood variation, one thing is certain. The semantic category of uncertainty (which prescriptively requires the use of the subjunctive) does not categorically determine the use of this mood. Use of the subjunctive in subordinate clauses is partly determined by the complementizer, with que favoring its use; but the use of this mood is also possible after si, as has been demonstrated throughout the current study. An important finding of the current research is that this mood is preferred by Mexican speakers after no saber si, which suggests a need to characterize dialects on the basis of morphosyntactic variation rather than the presence or absence of a specific morphosyntactic characteristic. Of the linguistic and social factors analyzed in the current study, the only one that favors the use of the subjunctive after the complementizer que is the lack of intervening material. This is important since this complementizer has often been seen as a trigger for the use of the subjunctive when, in fact, it is the lack of intervening material after this complementizer that seems to trigger the use of the subjunctive. Aside from intervening material, however, not a single other variable was deemed statistically significant in the conditioning of mood variation after expressions of uncertainty.

Though no definitive patterns of conditioning were revealed when variables were analyzed separately by complementizer, correlations between both moods and social factors were revealed when all factors were analyzed together by way of a Joint Multiple Correspondence Analysis. This type of exploratory analysis
indicates that the role of social factors should be further researched in future studies. Contrary to previous claims, age, sex, and education level all correlate with the use of mood with the subjunctive correlating most highly with young, male, and more educated speakers and the indicative correlating most highly with older, female, and less educated speakers. Though these results are not confirmatory, they do point to a need to examine the role of social factors in the conditioning of mood variation in monolingual Spanish. Why would the social conditioning of mood variation be any different from that of other types of variation that have received attention from sociolinguists in the past, such as phonological variation or lexical variation?

References


