Time Reference and Lexical Effects in Mood Choice Following Spanish Epistemic Adverb quizá(s): A Dialectal Comparison

Elizabeth Finanger
The Pennsylvania State University

1. Introduction

Throughout history, the Romance languages have purportedly experienced loss of Subjunctive temporal forms (Camús Bergareche 1990; Harris 1974; Poplack 1992; Silva-Corvalán 2001:146) while experiencing the extension of the Indicative and Conditional to contexts where previously the Subjunctive was exclusively used (Klein-Andreu 1991; Silva-Corvalán 1985). French is particularly known for this shift. Having eliminated the imperfect and pluperfect subjunctive forms from everyday speech, vernacular French uses only the present and present perfect subjunctive (que je chante, que j’ait chanté) (Woehr 1972:323). Many have wondered whether Spanish is headed in the same direction. For example, one scholar points out what he believes is the waning of past Subjunctive forms in Spanish.

The Spanish language has virtually reduced the four tenses of the subjunctive to two with which we express the five tenses of the indicative. Now it seems to be trying to get along with just one, the present subjunctive, not unlike spoken or non-literary French.” (Obaid, 1967:119)

However, in the context of epistemic adverbs in Spanish (tal vez, quizá, posiblemente, etc.) there is rising suspicion of an increase rather than a decrease in the use of Subjunctive at least as measured by its rate of use relative to the Indicative. In a diachronic study of Spanish literary works dating back to the thirteenth century, it was observed that since the fifteenth century, when the Subjunctive rate was 31% following epistemic adverbs quizá(s), there has been a rise in its use (Houle & Martínez-Gómez 2009). In the twentieth century (Woehr 1972), 63% of verbs following quizá(s) are Subjunctive, as shown in Table 1.

Table 1: Rate of Subjunctive relative to Indicative in Verbs associated with Quizá(s) (from Houle & Martínez-Gómez 2009:11)

<table>
<thead>
<tr>
<th>Century</th>
<th>Subjunctive, %</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>thirteenth</td>
<td>0%</td>
<td>4</td>
</tr>
<tr>
<td>fifteenth</td>
<td>31%</td>
<td>50</td>
</tr>
<tr>
<td>sixteenth</td>
<td>6%</td>
<td>19</td>
</tr>
<tr>
<td>seventeenth</td>
<td>15%</td>
<td>86</td>
</tr>
<tr>
<td>nineteenth</td>
<td>59%</td>
<td>24</td>
</tr>
<tr>
<td>twentieth (Woehr 1972)</td>
<td>63%</td>
<td>176</td>
</tr>
</tbody>
</table>

The variable the current paper addresses is variation in choice of mood following the epistemic adverbs quizá and quizás in modern Argentine and Peninsular Spanish. More specifically, I focus on
whether the verb following *quizá* or *quizás* surfaces in an Indicative (as in 1a. and 2a.) or Subjunctive form (as in 1b. and 2b.) and the linguistic factor(s) that influence this choice in mood.

1a. QUIZÁ eso es lo más duro del oficio.
   ‘perhaps that is (ind.) the most difficult (part) of the job’ (W.130.A)\(^3\)-Arg.\(^3\)

1b. QUIZÁ esto sea lo mejor que podemos lograr
   ‘perhaps this is (subj.) the best that we can achieve’ (W.91.A) – Arg.

2a. porque QUIZÁS puede haber un invento destinado a cambiar
   ‘because perhaps there may (ind.) be an invention destined to change’ (O.24.S) – Arg.

2b. QUIZÁS pueda pintarse, papá!
   ‘perhaps he can (subj.) paint himself, father!’ (W.7.S) – Arg.

Among prescriptive accounts, there is discord as to whether a difference in mood necessarily corresponds to a difference in meaning intended by the speaker. According to the *Gramática descriptiva de la lengua española*, the use of the Subjunctive vs. the Indicative with adverbs *quizá* and *quizás* does not imply a difference in certainty – unlike other uses of the Subjunctive (Bosque 1999: 3320). In contrast, *Modern Spanish Grammar: A Practical Guide* states that following *quizá* and *quizás*, mood depends on the degree of possibility involved – “indicative indicating a higher degree of possibility than the subjunctive” (Kattán-Ibarra 1997:303).

In place of these opposing views, the neutralization-in-discourse hypothesis (Sankoff 1988), which is the working hypothesis for the study of grammatical variation in a variationist framework, posits that although nuanced meanings may exist in some contexts, they need not be operating every time a variant form surfaces. And furthermore, this neutralization-in-discourse serves as the “fundamental mechanism” for grammatical change. In other words, even if the different readings play a role in variant selection, the distinctions they imply need not be operative every time one of the variant forms is used. Even if there are semantic differences, they can be operationalized and tested, and their contribution to mood choice relative to other (e.g., structural, lexical) considerations can be evaluated.

The contribution of different factors to mood choice is assessed through multivariate analysis.

In hopes of contributing to ongoing research with regard to mood choice following the epistemic adverbs *quizá* and *quizás*, this study investigated potentially influential linguistic factors in determining the mood in this epistemic adverbial context in Argentine and Peninsular Spanish. More specifically, there are two previous findings I investigated. First, time reference has been found to be the most influential factor in determining mood in epistemic adverbial contexts (Schwenter et al. 2008). Second, Schwenter (in press) observed dialectal variation in mood in these contexts. More specifically, Schwenter claims Argentine Spanish is “significantly different” than Peninsular Spanish in that Argentine Spanish shows a higher Subjunctive rate following epistemic adverbs. Thus, I investigated both overall rates of Subjunctive and influential factors in determining mood in Peninsular Spanish as compared to Argentine Spanish. The remainder of the paper is organized as follows. In section 2 the corpus is presented, followed by discussion of the current hypotheses and details regarding the coding of the tokens in section 3. Then, section 4 presents a discussion of the results. Lastly in section 5, I conclude the paper outlining practical implications for the classroom and thoughts for the linguistic community to address in future studies.

\(^*\) A special thanks to Rena Torres-Cacoullos for her mentorship in carrying out this study.

\(^1\) In the current study no differences were found in rate of Subjunctive following *quizá* and *quizás*. Potential differences are most likely considerations of the phonological context. Note that in a large number of tokens, the difference between *quizá* and *quizás* is neutralized (e.g. *quizá(s)* sea).

\(^2\) (W.130.A) indicates the source identification of the token. W/O – written/oral, # - line number, A/S – *quizá*/*quizás*.

\(^3\) Arg. – example sentences taken from Argentine data; Pen. – from Peninsular data.
2. Corpus and Data

The data for the present study were taken from CREA (Corpus de Referencia del Español Actual). CREA is an online database of more than 200 million Spanish words extracted from books, newspapers, emails, speeches, advertisements, interviews, radio, etc., from many Spanish-speaking countries. The database is organized by source, country, topic, formality, and audience. In order to be sensitive to dialectal variation in the relative frequency of Subjunctive use with epistemic adverbs, (cf. Schwenter et al. 2008), I focused on Argentine and Peninsular Spanish. A total of 581 oral and written tokens were extracted from CREA from Argentine Spanish and 727 oral tokens from Peninsular Spanish. For Argentina, all available oral tokens were from Senate proceedings (1998, 1999) of the Republic of Argentina which totaled 148 tokens, and 433 tokens were extracted from written Argentine fiction, published between 1975 and 2002 – including novels, stories, and theater. The Peninsular data analyzed is exclusively oral data from debates, interviews, and radio discourse. Although it would be preferable to analyze a single genre/mode rather than having to combine spoken and written Argentine data sources, we expect parallel distribution patterns between modes and dialects despite differences in overall rates of subjunctive (e.g., for subject expression, Cameron 1993, Travis 2007).

2.1. Exclusions

The following contexts of quizá(s) were excluded from the quantitative analysis. Illustrative examples are taken from both Argentine and Peninsular data.

3. parenthetical quizá(s); because scope of adverb is unclear
   …no entiende su lenguaje, QUIZÁ, porque no le interese su mensaje
   ‘…she doesn’t understand his language, perhaps because she is not interested (subj.) in his message.’ (10) – Pen.

4. competing subjunctive trigger: these tokens were excluded due to a prior word or phrase that may also trigger subjunctive mood in Spanish
   no le interesa ir a consultar a pesar de que QUIZÁS tenga medios para consultar en un lugar privado
   ‘he is not interested in going to the docor unless perhaps he has (subj.) the means to go to a private place’ (W.127.S) – Arg.

5. “por eso” : it was unclear whether the scope of quizá(s) extended to the verb or solely to the expression por eso and these occurred exclusively in Indicative mood.
   QUIZÁS por eso tienen buen carácter.
   ‘perhaps for that reason they have (ind.) good character.’ (W.98.S) – Arg.

6. pre-posed verb – since all cases occurred in indicative mood.
   fue QUIZÁ un poco más problemático por el aspecto social
   ‘it was (ind.) perhaps a little more problematic due to the social aspect.’ (44) – Pen.

These parameters excluded 217 cases and yielded a dataset of 483 Argentine tokens (115 oral tokens and 369 written tokens) and 593 Peninsular tokens.

3. Hypotheses and Coding of Tokens

In addition to Woehr (1972) and Houle & Martinez-Gómez (2009), Schwenter et. al (2008) also analyzed epistemic adverbs and mood focusing on modern Spain, Mexico, and Argentina. The following five hypotheses will be based primarily on these three works as well on my own

\[\text{The results for Peninsular Spanish are based on joint work with Arthur Wendorf.}\]
observations. To test the hypotheses that follow, I coded for source (oral vs. written), polarity, person/number, adjacency, verb tense/aspect, time reference, and lexeme.

3.1. Source (Written vs. Oral)

First, Schwenter et al. (2008) found that Subjunctive is favored in written mode. Given that written form and Subjunctive are often associated with prestige (Poplack 1997) I would predict that mood will be sensitive to oral vs. written mode; however the given data may complicate this matter. Whereas oral language generally tends to display more informal language than written data, the only available Argentine oral data in CREA is taken from Senate proceedings – a very formal setting in which a high register dominates. Thus we may expect to find more Subjunctive mood in these oral data. Likewise, while written language tends to be associated with higher registers and prestige, I chose low-register written data from theater, stories, and novels which is likely to display more casual speech features than other written genres. Biber (2009) has shown this for English. With these genre differences in mind, I predict Subjunctive mood will surface more frequently in the high-register oral data (Senate proceedings) as opposed to written fiction which is likely to display less prestigious language than governmental proceedings. However, given it was necessary to combine oral and written genres to obtain sufficient Argentine tokens, the difference in the source of the token was noted for Argentine data only while all Peninsular tokens were from oral sources.

7a. Oral source:
QUIZÁS sea la última vez que hable en este recinto.
‘perhaps this is (subj.) the last time that he’ll talk in this enclosure’ (O.6.S) – Arg.

7b. Written source:
QUIZÁ sea esa cosa indefinible que trasciende.
‘perhaps it’s (subj.) that indefinable thing that transcends’ (W.237.A) – Arg.

3.2. Polarity

Second, it has been shown that polarity also plays a role in choice of mood following epistemic adverbs, with negative polarity favoring Subjunctive (Houle & Martínez-Gómez 2009). This might be a way of operationalizing a meaning difference. Thus, I coded for polarity as affirmative, as in 8a, or negative as in 8b.

8a. Affirmative
QUIZÁS le vengan a visitar
‘perhaps they will (subj.) visit him’ (31) – Pen.

8b. Negative
QUIZÁS no hiciera falta agregarlos a su informe…
‘perhaps it was (subj.) not necessary to add them to your report…’ (W.38.S) – Arg.

Note that negative polarity includes negative particles such as nunca ‘never’ and nadie ‘no one/nobody’ etc, as illustrated in 9.

9. Negative polarity includes nunca ‘never’, nadie ‘no one/nobody’, etc
QUIZÁ nunca pensó ser prudente con Marcela…
‘perhaps he never thought (ind.) to be prudent with Marcela…” (W.54.A) – Arg.

In line with previous findings, I also predict that negative polarity will favor Subjunctive.

5 Note that in speech, quizá(s) sea and any other /s/-initial verbs would be cases of phonetic neutralization.
3.3. Adjacency

Third, I predict that distance between *quizá(s)* and the verb will play a role in determining whether Subjunctive or Indicative mood surfaces. I coded for adjacency as adjacent or distant (≥ 1 intervening word between adverb and verb), and then further broke it down according to the following categories:

10a. Adjacent - no intervening material between adverb and verb
   Corrientes QUIZÁS es una de las jurisdicciones más castigadas…
   ‘Corrientes perhaps is (ind.) one of the most punished jurisdictions…’ (O.4.S) – Arg.

10b. Intervening subject
   Bueno, QUIZÁS yo creo que con la edad se me ha pasado.
   ‘Well, perhaps I think (ind.) that with age it has passed me by.’ (163) – Pen.

10c. Intervening clitic(s)
   QUIZÁ nos haga millonarios a los tres.
   ‘Perhaps he will make (subj.) all three of us millionaires.’ (W.65.A) – Arg.

10d. Intervening negative particle(s)
   Ustedes, QUIZÁ no han visto esa cantidad de números de lotería
   ‘You (pl), perhaps have not seen (ind.) that quantity of lottery numbers’ (97) – Pen.

10e. Intervening clitic and negative particle
   …pues QUIZÁ no le agrada a la Reina…
   ‘…well perhaps he didn’t please (subj.) the Queen…’ (W.268.A) – Arg.

10f. Other intervening material
   QUIZÁ mi admisión de lo inevitable me hizo bajar en Perú
   ‘perhaps my admission of the unavoidable made (ind.) me get off (the train/bus) in Peru’ (W.210.A) – Arg.

I predict that with intervening material, association with *quizá(s)* weakens and the speaker/writer reverts to the default Indicative, which is much more frequent in Spanish overall.

3.4. Verb Tense/Aspect and Time Reference

Although there is dissention as to whether mood implies a difference in meaning when following the adverbs at hand, there is widespread agreement among prescriptive grammarians on two accounts. The use of either the subjunctive or indicative form is permitted when the event refers to the present or past, whereas when the event refers to the future, the present subjunctive, or less commonly, the future indicative may be used, but not the present indicative (Butt & Benjamin 2004:251; Kattán-Ibarra 1997:301, 313).

With regard to verb tense/aspect morphology and time reference, Schwenter et al. (2008) found that time reference showed the greatest magnitude of effect. Furthermore the following observations have been made: present time reference favors Subjunctive and past time reference disfavors Subjunctive (Woehr 1972; Schwenter 2008). Where these two studies differ is with regard to future time reference. Woehr found that future generally favors Subjunctive with epistemic adverbs; whereas Schwenter found future to disfavor Subjunctive, except in the case of *quizás* and *tal vez*. To explore the relationship between time reference and mood, I initially coded verbs based on morphology, according to Table 2.
Table 2: Verb tense/aspect (exemplified with tener ‘to have’) and time reference

<table>
<thead>
<tr>
<th>Verb morphology</th>
<th>Example</th>
<th>Time reference (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>present indicative</td>
<td>quizá tiene</td>
<td>present (171)</td>
</tr>
<tr>
<td>present subjunctive</td>
<td>quizá tenga</td>
<td>present &amp; future (278)</td>
</tr>
<tr>
<td>conditional</td>
<td>quizá tendría</td>
<td>present &amp; future (83)</td>
</tr>
<tr>
<td>preterit</td>
<td>quizá tuvo</td>
<td>past (67)</td>
</tr>
<tr>
<td>imperfect indicative</td>
<td>quizá tenia</td>
<td>past (101)</td>
</tr>
<tr>
<td>imperfect subjunctive</td>
<td>quizá tuviera/tuviese</td>
<td>past (66)</td>
</tr>
<tr>
<td>synthetic future</td>
<td>quizá tendrá</td>
<td>future (2)</td>
</tr>
<tr>
<td>periphrastic future</td>
<td>quizá va a tener</td>
<td>future (13)</td>
</tr>
<tr>
<td>present perfect indicative</td>
<td>quizá ha tenido</td>
<td>past (47)</td>
</tr>
<tr>
<td>present perfect subjunctive</td>
<td>quizá haya tenido</td>
<td>past (11)</td>
</tr>
<tr>
<td>pluperfect indicative</td>
<td>quizá habría tenido</td>
<td>past (18)</td>
</tr>
<tr>
<td>pluperfect subjunctive</td>
<td>quizá hubiera tenido</td>
<td>past (25)</td>
</tr>
<tr>
<td>conditional perfect</td>
<td>quizá habría tenido</td>
<td>past (6)</td>
</tr>
<tr>
<td>future perfect</td>
<td>quizá habrá tenido</td>
<td>past (1)</td>
</tr>
<tr>
<td>periphrastic conditional</td>
<td>quizá iba a tener</td>
<td>past (2)</td>
</tr>
</tbody>
</table>

Then, based on contextual cues, I further coded the verbs in terms of time reference as past, present, future, and non-past\(^6\). Examples 11-13 provide tokens of clear past, present, and future time reference\(^7\).

11. Past time reference: imperfect morphology
   QUIZÁ sólo \textit{era} una forma de escapar
   ‘Perhaps \textit{it was} (ind.) only a way to escape’ (W.108.A) – Arg.

12. Present time reference: present indicative morphology
   QUIZÁ \textit{son} hoy más necesarias que nunca…
   ‘\textit{perhaps they are} (ind.) more necessary today that ever…’ (236) – Pen.

13. Future: conditional morphology expressing clear future time reference
   QUIZÁ el sábado \textit{aceptaríamos} su invitación a comer.
   ‘\textit{Perhaps on Saturday we would accept} your invitation to eat.’ (W.361.A) – Arg.

While these tokens are easily distinguished with regard to time reference with the help of context (ex. adverbs: \textit{hoy} ‘today’), examples 14-16 exemplify the versatile nature of present subjunctive verb morphology in referring to both present and future temporal reference. Example 16 specifically illustrates the common scenario of present subjunctive verb morphology where making a replicable distinction between present and future time reference is unfeasible. This distinction is especially problematic within a limited context – as is the case with the data from CREA (which both Schwenter et al. 2008 and I used). As a consequence of the substantial number of ambiguous tokens in time reference with present subjunctive and conditional verb morphology, temporal reference was analyzed as a two-factor group opposing past time reference and non-past time reference. The non-past category includes present, future, and hypothetical time reference.

14. Present time reference
   el laboratorio QUIZÁ \textit{sea} mejor ahora...
   ‘the laboratory perhaps \textit{is} (subj.) better now…’ (567) – Pen.

\(^6\) Inter-rater reliability for time reference was 49/50 = 98%.

\(^7\) Preterit and Imperfect verb forms were coded automatically as “past”.
15. Future time reference
   y QUIZÁS no vuelva nunca más…
   ‘and perhaps he will never return (subj.) again…’ (W.15.S) – Arg.

16. Non-past time reference: present or future time reference
   ¿Hay otros problemas que QUIZÁ nos puedan afectar?
   Are there other problems that can/will be able to (subj.) affect us? (157) – Pen.

   With these coding parameters in place, my hypothesis follows previous accounts that non-past
temporal reference favors Subjunctive, while past disfavors Subjunctive.

3.5. Person and Number

   Fifth, Schwenter et al. (2008) found that 1st person disfavors Subjunctive. Perhaps this is due to a
correlation of grammatical person with distance, speaker commitment, and/or source of information.
Such a correlation was found for use of the preposition de preceding the complementizer que, meaning
“that” (Ella se enteró de que venían para la fiesta. ‘She found out that they were coming for the party’)
(Schwenter 1999). As noted in Schwenter (1999) 1st person disfavors use of de suggesting “that its
presence or absence depends heavily on the source of information in the speaker’s assertion”. This de
particle has been identified as a distance mechanism and is more commonly used in correlation with
second and third person. Thus, I coded for person/number (note that vos and usted(es) were included
in 2nd person). My intention was to analyze the co-occurrence patterns of subject and mood and make a
case for 1st person disfavoring Subjunctive based on this notion of distance and possibly relate
Subjunctive mood to its purported relationship with uncertainty in the context of quizá(s).

3.6. Lexeme

   Lastly, I coded for the particular lexeme of the verb token – which was an additional factor group that
had not been considered in previous studies of the context at hand. Coding for lexeme was motivated
by Poplack’s 2001 study in which particular French verbs (falioir ‘must’ vouloir ‘want’ and aimer
‘like’) were highly associated with the Subjunctive as compared to other verbs. My intent was to begin
to investigate the possibility of a so-called lexeme effect in epistemic adverbial contexts in Spanish.
Note that the finite verb is coded as the lexeme as this encodes mood.

17. Coding for lexeme
   mientras aprenden sus primeras letras QUIZÁS puedan olvidar los horrores de la
   guerra – lexeme: puedan
   ‘while they learn their first lyrics perhaps they can (subj.) forget the horrors of the
   war.’ (243) – Pen.

3.7. Summary of Predictions

   Knowing that Subjunctive mood is considered more conservative than Indicative mood, and that the
nuances of uncertainty may spill over into Subjunctive contexts with quizá(s), previous studies and
current observations lead to the following predictions. The predictions for the factor groups outlined
above are that the oral/formal source, negative polarity, and non-past time reference will favor
Subjunctive while past time reference, increased distance between quizá(s) and verb, and 1st person
will disfavor Subjunctive mood.

4. Results and Discussion

   Table 3 presents the results of the variable-rule analyses of linguistic factors for mood choice
following quizá(s) ‘perhaps’ in the Argentine data and Table 4 corresponds to the Peninsular data. As
anticipated, non-past time reference favors Subjunctive in both dialects (factor weights: Argentina .66; Spain .53), while past time reference strongly disfavors Subjunctive (factor weights: Argentina .34; Spain .40). These results support previous findings that time reference is the most influential factor in determining mood with epistemic adverbs (Woehr 1974; Schwenker et. al 2008; Houle & Martinez-Gómez 2009). Overall rates of Subjunctive mood are also comparable: Argentina 46% and Spain 41%. Given the Peninsular data was oral discourse, it is possible that had we also included an analysis of formal (written) data – which may be more associated with Subjunctive mood – the overall rate of Subjunctive might have approximated that of Argentina (which included formal and informal data) more closely.

Table 3: Linguistic factors contributing to choice of Subjunctive over Indicative following quizá(s) ‘perhaps’ in Argentine Spanish (non-significant factor groups in brackets)

<table>
<thead>
<tr>
<th>Probability</th>
<th>% Subj.</th>
<th>Total N</th>
<th>% Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Past</td>
<td>.66</td>
<td>62</td>
<td>240</td>
</tr>
<tr>
<td>Past</td>
<td>.34</td>
<td>31</td>
<td>242</td>
</tr>
<tr>
<td>[Person]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[First]</td>
<td>[.55]</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>[Second/Third]</td>
<td>[.49]</td>
<td>44</td>
<td>409</td>
</tr>
<tr>
<td>[Polarity]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Affirmative]</td>
<td>[.51]</td>
<td>47</td>
<td>406</td>
</tr>
<tr>
<td>[Negative]</td>
<td>[.44]</td>
<td>41</td>
<td>76</td>
</tr>
<tr>
<td>[Adjacency]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Adjacent]</td>
<td>[.50]</td>
<td>46</td>
<td>214</td>
</tr>
<tr>
<td>[Distant]</td>
<td>[.50]</td>
<td>46</td>
<td>268</td>
</tr>
<tr>
<td>[Source]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Oral]</td>
<td>[.50]</td>
<td>56</td>
<td>115</td>
</tr>
<tr>
<td>[Written]</td>
<td>[.50]</td>
<td>43</td>
<td>367</td>
</tr>
<tr>
<td>Range</td>
<td>32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total N = 483, Input = .46
Log likelihood –308.758,

Table 4: Linguistic factors contributing to choice of Subjunctive over Indicative following quizá(s) ‘perhaps’ in Peninsular Spanish (non-significant factor groups in brackets)

<table>
<thead>
<tr>
<th>Probability</th>
<th>% Subj.</th>
<th>Total N</th>
<th>% Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjacency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjacent</td>
<td>.52</td>
<td>60</td>
<td>206</td>
</tr>
<tr>
<td>Distant</td>
<td>.35</td>
<td>44</td>
<td>387</td>
</tr>
<tr>
<td>Range</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Past</td>
<td>.53</td>
<td>44</td>
<td>457</td>
</tr>
<tr>
<td>Past</td>
<td>.40</td>
<td>31</td>
<td>136</td>
</tr>
<tr>
<td>Range</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Polarity]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Negative]</td>
<td>[.60]</td>
<td>43</td>
<td>75</td>
</tr>
<tr>
<td>[Affirmative]</td>
<td>[.49]</td>
<td>41</td>
<td>518</td>
</tr>
<tr>
<td>[Person]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Second/Third]</td>
<td>[.52]</td>
<td>43</td>
<td>491</td>
</tr>
<tr>
<td>[First]</td>
<td>[.41]</td>
<td>32</td>
<td>102</td>
</tr>
</tbody>
</table>
| Total N = 593, Input = .41
Log likelihood –390.752
Most striking from these results is that time reference is the only significant factor group in the Argentine data. In addition to time reference, in the Peninsular data, adjacency was also found to be significant in determining mood, while the Argentine data show no difference based on adjacency. Furthermore, it is interesting to note that while falling short of statistical significance, person and polarity behave differently in the two dialects at hand. In the Peninsular data, polarity and person lean toward previously mentioned hypotheses, with negative polarity favoring Subjunctive while first person disfavors Subjunctive. However in the Argentine data, negative polarity shows a tendency to disfavor Subjunctive and first person tends to favor Subjunctive.

It is by comparing these results with previous studies that we begin to suspect there may be something else at work. In cross-tabulating time reference with verb morphology (Table 5), one sees that it is the present subjunctive and conditional tenses that do not exclusively refer to one time reference – and this is where we now turn our attention. The following table displays the temporal reference to which the various verb tenses refer. Note that the category ‘non-past’ is added in addition to past, present, and future since it is often impossible to determine whether the speaker/writer is referring to the present or future with present subjunctive or conditional verb morphology in a limited context (see example 16 above).

Table 5 Verb Morphology in relation to Time Reference in Argentine data

<table>
<thead>
<tr>
<th>Total N</th>
<th>Present</th>
<th>Future</th>
<th>Non-Past</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present subjunctive (i.e. tenga)</td>
<td>148</td>
<td>58</td>
<td>28</td>
</tr>
<tr>
<td>Conditional (i.e. tendría)</td>
<td>41</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Present Perfect/pluperfect (i.e. ha tenido/había tenido)</td>
<td>30</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>Present indicative (i.e. tiene)</td>
<td>50</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>Present perfect / pluperfect subjunctive (i.e. haya tenido/hubiera tenido)</td>
<td>21</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>Imperfect indicative (i.e. tenía)</td>
<td>75</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>Preterit (i.e. tuvo)</td>
<td>53</td>
<td>53</td>
<td>0</td>
</tr>
<tr>
<td>Imperfect subjunctive (i.e. tuviera)</td>
<td>53</td>
<td>53</td>
<td>0</td>
</tr>
</tbody>
</table>

Given the variability in time reference in the present subjunctive tense, I decided to take a closer look at what may be influencing choice in mood. It is proposed by usage-based linguists that frequency plays a role in creating and further shaping grammar. Rather than base grammatical structure on innate categories – divorced from how language is used (as in Chomsky’s generative grammar) – usage-based grammar views grammar as being “built up from specific instances of use that marry lexical items with constructions…” Furthermore, “… it [grammar] is routinized and entrenched by repetition and schematized by the categorization of exemplars” (Bybee 2006:730). Thus repetition, or frequency, is part of the linguistic experience that shapes the cognitive organization of language.

With this theory of usage-based grammar in mind, I was led to examine the role of particular lexemes in determining whether the verb occurred in the Subjunctive or Indicative mood. Similar to Poplack’s 2001 study in which particular French verbs (falloir ‘must’ vouloir ‘want’ and aimer ‘like’) were highly associated with the Subjunctive as compared to other verbs, this so-called lexeme effect with the verb ser ‘to be’ is precisely what can be observed in the data at hand. Namely, ser ‘to be’ correlates strongly with Subjunctive mood compared to the overall data in the present tense. This observation begged the question of how ser compares with overall data according to time reference. Table 6 presents the results of a clear lexeme effect in mood choice in Argentina Spanish with ser following quizá(s), in the present and past temporal reference contexts. Note that even in the past, which generally favors Indicative mood, ser ‘to be’ in Subjunctive forms reaches nearly 50%. Also, ser makes up more than 1/4 of the Argentine data (120/450).
Table 6: Lexeme Effect: *ser* ‘to be’ displays higher rates of Subjunctive in past, present, and non-past time reference when compared to non-*ser* verb rates following *quizá(s)* ‘perhaps’ in **Argentine** Spanish (future temporal reference excluded due to no clear tokens with *ser*).

<table>
<thead>
<tr>
<th></th>
<th>% Subj.</th>
<th>Total N</th>
<th>% Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SER</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>69</td>
<td>32</td>
<td>27</td>
</tr>
<tr>
<td>Non-past</td>
<td>67</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>Past</td>
<td>46</td>
<td>67</td>
<td>56</td>
</tr>
<tr>
<td><strong>NON-SER</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>46</td>
<td>78</td>
<td>22</td>
</tr>
<tr>
<td>Non-past</td>
<td>63</td>
<td>76</td>
<td>21</td>
</tr>
<tr>
<td>Past</td>
<td>25</td>
<td>176</td>
<td>49</td>
</tr>
</tbody>
</table>

Upon finding a lexeme effect with the verb *ser* in Argentine Spanish, I was led to examine potential lexeme effects in Peninsular Spanish. Not surprisingly, I found *ser* to surface more often in Subjunctive mood (47%) when compared to the overall rate of Subjunctive (41%). Furthermore, frequent verbs *tener* ‘to have’ and existential and auxiliary verb *haber* ‘to have’ display even higher rates of Subjunctive, 50% and 54% respectively, (first column in Table 7). This finding led me to further investigate the potential relationship between mood and frequency. Looking beyond token frequency following *quizá(s)* and overall token frequency (second column Table 7), I also analyzed the potential role of token frequency of Indicative/Subjunctive forms of these verbs in all contexts – not just following the epistemic adverbs at hand (see column three Table 7). Finally, based on these counts, I considered relative frequency of the Subjunctive (with respect to the Indicative) (last column Table 7).

At first glance, I notice an overarching pattern in that *haber*, *tener*, and *ser* all display relatively higher Subjunctive frequencies when compared to other frequent verbs. However, *poder* ‘to be able’ also displays a high token frequency for the Subjunctive (n=873) and a high relative frequency (11%) and yet does not display an above average rate of Subjunctive following *quizá(s)* (overall rate = 41%) as do *haber*, *tener*, and *ser*. Furthermore, *estar* surfaces with a relative frequency of 4% and yet does not behave as *haber* (7%) and *tener* (6%) with regard to rate of Subjunctive mood. In sum, these data suggest that both relative frequency and token frequency correlate positively with Subjunctive mood, however given the fact the *poder* and *estar* behave differently despite high Subjunctive token frequency and relative frequency, it is likely that the specific constructions, such as *quizá(s) + verb* (*haya, tuviera, sea*), play more of a role in determining mood than frequency alone.

---

8 I originally attempted to determine if compound VPs with highly frequent elements would behave differently from VPs with less frequent elements and/or from simple VPs, but the result was negative.

9 Many thanks to Arthur Wendorf for his help in analyzing the overall vs. relative frequency data.
Table 7: Lexeme Effect: high frequency verbs *ser*, *tener*, and *haber* display higher rates of Subjunctive when compared to overall verb rate of 41% following *quizá(s)* ‘perhaps’ in Peninsular Spanish. Overall high token freq. and relative freq. correlate with higher rates of Subjunctive.

<table>
<thead>
<tr>
<th>FOLLOWING</th>
<th>OVERALL</th>
<th>TOKEN</th>
<th>RELATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>quizá</em> &amp; <em>quizás</em></td>
<td>% Subj.</td>
<td>Total N</td>
<td>Frequency</td>
</tr>
<tr>
<td>Haber</td>
<td>54</td>
<td>24</td>
<td>26505</td>
</tr>
<tr>
<td>Tener</td>
<td>50</td>
<td>30</td>
<td>14555</td>
</tr>
<tr>
<td>Ser</td>
<td>47</td>
<td>185</td>
<td>66877</td>
</tr>
<tr>
<td>Poder</td>
<td>40</td>
<td>54</td>
<td>7595</td>
</tr>
<tr>
<td>Deber</td>
<td>33</td>
<td>12</td>
<td>1249</td>
</tr>
<tr>
<td>Estar</td>
<td>31</td>
<td>42</td>
<td>16838</td>
</tr>
<tr>
<td>Hacer</td>
<td>23</td>
<td>13</td>
<td>6274</td>
</tr>
<tr>
<td>Decir</td>
<td>20</td>
<td>10</td>
<td>9572</td>
</tr>
</tbody>
</table>

Given that access to large samples of natural language are now available with electronic corpora, researchers have confirmed that both written and spoken discourse are “characterized by the high use of conventionalized word sequences, which include sequences that might be call formulaic language and idioms, but also conventionalized collocations (sometimes called ‘prefabs’; Erman & Warren 2000)” (Bybee 2006). Whereas idioms possess an extended semantic meaning (such as *level the playing field* or *pull strings*), prefabs occur repeatedly in discourse and are known to represent the conventional way of expressing certain notions (such as *prominent role* and *mixed messages*). While Bybee (2006) notes the line between idiom and prefab is not always clear, for the purposes of this study, it is the prefabricated word sequences that are of interest.

If *ser*, *tener*, and *haber*, in the context of the epistemic adverbs *quizá(s)*, surface more often in Subjunctive mood as opposed to Indicative mood when compared with overall rates of Subjunctive, perhaps these word sequences are being stored as units rather than individual parts, and the units are experiencing entrenchment (cf. Langacker 2000) – which is often the case for irregular verb forms. With each token of *quizá(s)* sea, *quizá(s)* haya sido, *quizá(s)* tenga, etc., these word sequences are increasingly being used and stored as prefabs, and they experience an increase in lexical strength with each token. Also noteworthy in our lexeme analysis is how *ser* behaves in terms of adjacency when compared to non-*ser* verbs. In cross-tabulating *ser* with adjacency in Peninsular Spanish, we find that *ser* surfaces immediately adjacent to *quizá(s)* 50%, whereas non-*ser* verbs only surface 35% immediately adjacent to *quizá(s)*. The fact that the verb *ser* is less likely to appear with intervening material between it and *quizá(s)* as compared to other verbs provides further support for the understanding that *quizá(s)* sea, *quizá fuera*, etc., is functioning as a prefabricated word sequence. Storing repeated behaviors as chunked units is not unique to language and allows for more efficient access and production.

Chunking (includes prefabs) occurs automatically as behaviors are repeated in the same order, whether they are motor activities such as driving a car or cognitive tasks such as memorizing a list. Repetition is the factor that leads to chunking, and chunking is the response that allows repeated behaviors to be accessed more quickly and produced more efficiently.” (Haiman 1994, as cited by Beckner & Bybee 2009)

The view that frequency strengthens linguistic mental representations is part of exemplar theory (K. Johnson 1997; Pierrchumbert 2001).

In this model, every token of experience is classified and placed in a vast organizational network as a part of the decoding process. The major idea behind exemplar theory is that the matching process has an effect on the representations.
themselves; new tokens of experience are not decoded and then discarded, but rather they impact memory representations. In particular, a token of linguistic experience that is identical to an existing exemplar is mapped onto that exemplar, strengthening it. (Bybee 2006)

5. Further Implications and Future studies

Considering these results in the L2 classroom setting, this study provides insight into the nature of quizá(s) with regard to mood and the verb ser – adding one more piece to the puzzle of mood following epistemic adverbs. Klein (1980c) notes that while some epistemic adverbs and phrases strongly favor Indicative (*es evidente que, seguramente*) and others favor Subjunctive (*dudo que, es posible que*), quizá is at the crossroads between Indicative and Subjunctive – both are relatively equally employed. Whether these semantic differences in certainty are active or not, teaching mood in the context of epistemic adverbs (*tal vez, posiblemente*, etc) can be difficult due to the variability seen in this study. However, these data offer a few guidelines.

- Both Indicative and Subjunctive mood are equally accepted following quizá(s).
- Past time reference favors Indicative, while non-past time reference favors Subjunctive.
- When quizá(s) is followed by the verbs ser, tener, and haber, the verb is more likely to surface in a Subjunctive verb form than an Indicative form.

With regard to future investigations, this study exemplifies the importance of considering lexeme effects in determining influential factors in mood choice in Spanish. Thus for future studies with perhaps more tokens, a lexeme factor group may reveal other lexemes in addition to ser, tener, and haber as having an effect on choice in mood.

In sum, this paper has provided further support for the previous observations that time reference is the primary factor in determining mood following epistemic adverbs quizá(s) in Spanish. Adjacency was also found to play a role in Peninsular Spanish. However, all other factors considered in previous studies were found insignificant. Furthermore the data make a supplementary contribution by shedding light on an additional factor that has not been considered in previous studies - a lexeme effect found with the verbs ser ‘to be,’ tener ‘to have,’ and haber ‘to have’. These high frequency verbs were found to strongly correlate with Subjunctive mood – suggesting entrenchment as prefabricated word sequences. This finding provides further support for usage-based theories of grammar and exemplar theory in that frequency plays a role in creating and shaping grammar and underscores the importance of incorporating a lexeme analysis in future studies of choice in mood in Spanish.

References


Houle, Leah and Rebeca Martínez-Gómez. 2008. La alternancia del subjuntivo e indicativo con los adverbios epistémicos. Manuscript from The University of New Mexico.


Klein, Flora. 1980c. Experimental verification of semantic hypotheses applied to mood in Spanish. Georgetown University Papers on Languages and Linguistics 17.15-34.


