

Near-final version (February 2011): under copyright and that the publisher should be contacted for permission to re-use or reprint the material in any form

Brown, Esther L. & Javier Rivas. 2011. Subject ~ Verb word-order in Spanish interrogatives: a quantitative analysis of Puerto Rican Spanish. *Spanish in Context* 8.1, 23–49.

## **Subject-verb word-order in Spanish interrogatives: A quantitative analysis of Puerto Rican Spanish<sup>1</sup>**

Esther L. Brown and Javier Rivas

We conduct a quantitative analysis of conversational speech from native speakers of Puerto Rican Spanish to test whether optional non-inversion of subjects in *wh*-questions (*¿qué tú piensas?*) is indicative of a movement in Spanish from flexible to rigid word order (Morales 1989; Toribio 2000). We find high rates of subject expression (51%) and a strong preference for SV word order (47%) over VS (4%) in all sentence types, inline with assertions of fixed SVO word order. The usage-based examination of 882 *wh*-questions shows non-inversion occurs in 14% of the cases (25% of *wh*-questions containing an overt subject). Variable rule analysis reveals subject, verb and question type significantly constrain interrogative word order, but we find no evidence that word order is predicted by perseveration. SV word order is highest in rhetorical and quotative questions, revealing a pathway of change through which word order is becoming fixed in this variety.

**Keywords:** word order, language change, Caribbean Spanish, interrogative constructions

### **1. Introduction**

In typological terms, Spanish is characterized as a flexible SVO language. As has been shown by López Meirama (1997: 72), SVO is the basic word order in Spanish, with the subject preceding the verb in pragmatically unmarked independent declarative clauses with two full NPs (Mallinson & Blake 1981: 125; Siewierska 1988: 8; Comrie 1989). Word order is flexible in that the subject, when expressed, may occupy either a pre- or post-verbal position with the different permutations of sentence types (SVO, VOS, VSO) each expressing essentially equivalent semantic information with different stylistic and pragmatic effects (i.e. Silva-Corvalán 1982; Bentivoglio & Weber 1986; Ocampo 1995; Blackwell 2003, among others).

Overall grammatical limitations on subject-verb word order in Spanish are exceedingly few (Gili Gaya 1961: 83; Cifuentes Honrubia 2000; Butt & Benjamin 2000). The post-position of the subject in relation to the verb is strongly preferred in a few specific cases, namely, when the subject is modified by a relative clause, within the relative clause itself, and “when certain adverbs and adverbial phrases precede the verb” (Butt & Benjamin 2000: 513-4). In addition, subject-verb inversion is required when pronouns such as *qué* (‘what’), *cómo* (‘how’) or *cuánto* (‘how much’) introduce an exclamation or a question (Cifuentes Honrubia 2000: 360; Butt & Benjamin 2000: 517). Perhaps the general rarity of a specifically required word order contributes to the saliency of word order ‘violations’ of the type found in some varieties of Caribbean Spanish.

[Type text]

Prescriptively, when expressed, the subject must follow the verb in *wh*-questions as in example (1).

- (1) ¿Oye, P., y... qué **piensas tú**? ¿tú crees que Puerto Rico es un país bilingüe? (Interview 1, page 24)

‘Listen, P, and... what do you think? do you think Puerto Rico is a bilingual country?’

Several varieties of Spanish (i.e. Puerto Rican, Cuban, Dominican), however, optionally violate this word order requirement for direct questions containing an interrogative pronoun (Bergen 1976; Lipski 1994; Ordóñez & Olarrea 2006). This can be seen in example (2).

- (2) ...¿qué **tú piensas** de todo esto que ha pasado en Estados Unidos, allá en Nueva York con las torres y todo ese tipo de cosas? (Interview 6, page 41)

...‘what do you think about all that has happened in the U.S., there in New York with the Towers and all those types of things?’

Some researchers cite this lack of subject inversion in direct questions as part of a larger syntactic shift within the language. Morales (1989) and Toribio (2000) suggest that word order is losing its flexibility in these varieties and, as a result, they are in the process of becoming fixed SVO systems, as has occurred with other Romance languages such as French. The co-occurrence of phenomena such as high overall usage of overt pronominal subjects, use of subject pronouns with infinitival forms (*para yo hacer esto* ‘for me to do this’) and non-inversion of subject and verb in questions all point to this trend.

In order to investigate this notion, we conduct a quantitative analysis of word order in direct *wh*-questions in Puerto Rican Spanish. Results are in line with proposals that word order in this dialect of Spanish is, “relatively fixed – subject verb-object – irrespective of sentence type or verb class” (Toribio 2000: 322). Further, results of a variable rule analysis of non-inverted questions suggest a possible diachronic pathway through which word order changes are likely occurring. We first provide a summary of previous accounts of non-inverted questions, followed by a description of the data and methods, a presentation of results and discussion.

## 2. Background

Perhaps the most salient non-phonological feature of Caribbean Spanish is subject-verb word order in *wh*-questions. Examples (3), (4) and (5), taken from our corpus, illustrate this phenomenon.

- (3) ¿Qué fue? ¿Qué yo hice? (I3, 33)

‘What was it? What did I do?’

[Type text]

- (4) No, es que- ¿ah? ¿cómo tú te llamas? (I12, 121)

‘No, the thing is- ah? what is your name?’

- (5) ¿Adónde él ha ido? (I4, 25)

‘Where has he gone?’

These examples violate the rule of Subject-Verb Inversion in questions (D’Introno 2001: 144) which is obligatory in other varieties of Spanish.

Varieties that demonstrate lack of inversion in questions are precisely those with high overt pronominal subject usage in general (e.g. Otheguy, Zentella & Livert 2007; Camacho (2008:419) for Dominican Spanish; Hochberg 1986; Cameron 1993; Morales 1997; Flores-Ferrán (2004) for Puerto Rican Spanish). However, not all subjects are equally acceptable in non-inverted questions. The second person pronoun *tú* (‘you’) is the most frequent type of subject found in these constructions (Davis 1971: 331<sup>2</sup>; Quirk 1972: 303; Lipski 1977: 61; Lantolf 1980: 116; Heap 1990: 62; Ordóñez & Olarrea 2006: 68). Other pronouns and lexical subjects have been reported, although occasionally speakers reject non-inverted questions with pronouns other than *tú* (‘you’, informal) and *usted/ustedes* (‘you’, formal singular/plural), especially when the pronoun involved is *nosotros* (‘we’) (Heap 1990: 62).

Other studies, such as Lantolf (1980: 120), Heap (1990: 62) and Toribio (2000: 322-323), provide examples of non-inverted questions in which the subject is a proper name or a simple noun phrase. We find examples of non-pronominal cases in our corpus. This can be seen in examples (6) and (7):

- (6) ¿Y cómo J. reaccionó en ese momento? (I3, 7)

‘And how did J. react at that moment?’

- (7) ¿Por qué la gente aquí se asusta cuando guía aquí? (I15, 92)

‘Why do people here get scared when they drive here?’

Heap (1990: 64) points out that lexical subjects can only occur in non-inverted constructions when they lack quantifiers and modifiers. This is also corroborated by our data. In the few examples we have of non-inverted lexical subjects, none are modified by anything other than a determiner (see example (7)). In addition, Ordóñez and Olarrea (2006: 68) indicate that most of the examples involving a lexical noun phrase as subject also contain a complex *wh*-word, such as *por qué* (‘why’), *en qué lugar* (‘in what place’) and similar structures, which is also reflected in our data.<sup>3</sup>

According to previous studies, then, non-inversion in Caribbean Spanish is possible with all pronominal subjects, *tú* (‘you’, informal) being the pronoun most frequently attested. In addition to pronouns, proper nouns and simple lexical noun phrases (noun phrases lacking quantifiers and modifiers) may also take the preverbal position in *wh*-

[Type text]

questions, especially in contexts in which there is a complex *wh*-word. These results, however, have been based either on the linguist's own intuitions or on questionnaires given to native speakers. To the best of our knowledge, previous studies of this phenomenon have never been large-scale quantitative analyses of natural speech.

Formal attempts at explaining the causes of non-inversion have been varied. Some authors consider pronominal subjects in non-inverted questions to be clitics (Lipski 1977: 64; Heap 1990: 60). According to this analysis, in Caribbean dialects the pronominal subject and the verb behave as a single word in phonetic and phonological terms. Therefore, when it comes to interrogative structures, inversion does not take place. However, Ordóñez and Olarrea (2006: 69-70) contend that these pronouns do not behave as clitics because the negative element *no* ('not') can intervene between the subject and verb (cf. *¿qué tú no comes?* 'what do you not eat?' with \**¿qué no tú comes?* '\*what do not you eat?'). Instead, Ordóñez and Olarrea (2006) classify them as weak pronouns. Unlike clitics, weak pronouns may carry word stress and are obligatorily placed before negation. However, they can neither precede a left-dislocated constituent, nor can they be modified, coordinated or focused, as is the case with strong pronouns.

An alternative explanation, which has been offered by Suñer (1986), Morales (1989) and Toribio (2000), is that non-inverted questions are evidence of the fact that the basic SVO order typical of Spanish is becoming the dominant order<sup>4</sup> in Caribbean Spanish. Apart from non-inverted questions, other syntactic phenomena suggestive of a change to a rigid SVO order are the following (Suñer 1986: 196-197; Morales 1989: 147-148; Toribio 2000: 319-324): the use of preposed pronominal subjects in infinitive constructions and the overall higher frequency of pronominal subjects, each discussed in turn below.

In Caribbean Spanish the subject of infinitive constructions headed by *al* ('to-the'), *para* ('for'), *sin* ('without') and *por* ('for') may occur in preverbal position, as illustrated by example (8), taken from our corpus.

(8) Verdad, N., cállate un rato para yo hablar. (I1, 83)

'Really, N., be quiet for a while so I can talk'

In other varieties of Spanish, the subject of the infinitive occurs postverbally (*para hablar yo*) or, alternatively, the subjunctive would be used in a finite clause (*para que yo hable*). Taking into account Puerto Rican Spanish data, Morales (1989: 148) points out that in almost 70% of infinitive contexts, the subject precedes the infinitive, and the two most frequent types of preverbal pronominal subjects in these constructions are *yo* ('I') and *uno* ('one').

Caribbean varieties also exhibit a higher number of overtly expressed pronominal subjects than other varieties of Spanish (i.e. Hochberg 1986; Cameron 1993). Morales (1997: 155) compares preposed pronominal subjects in three dialects: San Juan, Madrid and Buenos Aires. Results show that the dialect of San Juan presents a higher number of overt pronominal subjects than the other two varieties, with first and second person pronouns taking the lead as the most frequent overt pronominal forms (Morales 1989: 147).<sup>5</sup>

The attributes present in the San Juan dialect (preposed subjects in infinitival clauses,

[Type text]

high overt pronominal subjects), also present in the dialect under investigation here, are argued to be symptoms of a rigid SVO order. This shift may be conditioned by syntactic and pragmatic pressures such as the need to fill the topical position of the clause.<sup>6</sup> In this respect, Toribio (2000: 321) provides evidence from Dominican Spanish in which this position is filled by a non-referential pronoun (*ello* ‘it’) which lacks semantic and thematic meaning, as can be seen in example (9).

(9) Ello llegan guaguas hasta allá (Toribio 2000: 321)

‘Buses arrive there’ (Lit.: it arrive-PL buses until there)

Morales (1989: 149) argues the change to rigid SVO order may have started with first person pronoun *yo*, non-specific uses of second person pronouns *tú* (informal) / *usted* (formal) and *uno* (‘one’), since these are the most frequent pronouns in both infinitive constructions and in the overall high occurrence of overt subject pronouns.

### 3. Data and methods

In order to determine if Puerto Rican non-inversion was, indeed, as predicted by Morales (1989) and Toribio (2000), indicative of a shift in Spanish to a rigid word order, we conducted a quantitative analysis of approximately 370,000 words of spoken Puerto Rican Spanish representing roughly 27 hours of conversation with eighteen native speakers. These conversations were collected and transcribed by a native speaker in Caguas, Cayey and San Juan, Puerto Rico in 2000 (Cortés-Torres 2005). Speakers range in age from 24-90 years old. Interviews ranged in duration from one half hour to three hours in length.

In order to obtain data on overall subject-verb word order, we coded 1000 consecutive verbs appearing in the first, second and third persons singular from three separate interviews selected at random from the collection of 18 speakers. We chose only singular verb forms in order to limit the scope of our research.<sup>7</sup> For the total 1000 conjugated verbs appearing in all sentence types, we analyzed whether the subject was expressed or null. If expressed we noted whether the subject was a lexical or pronominal subject. We also determined whether the subject, when expressed, preceded or followed the verb. We excluded commands from analysis because the type of modality they convey entails that the subject is generally left unexpressed.

From the total corpus of 370,000 words, we examined all direct *wh*-questions with a singular subject (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> persons). We coded only the verb in the main clause and excluded from analysis verbs in subordinate clauses, as well as verbs in echo, tag and yes/no questions. In the 27 hours of conversation, this totaled 882 questions. Each question was coded for the following variables:

a) word order: ‘subject-verb’, ‘verb-subject’ or ‘not applicable’ in the case of a null subject.

b) type of subject pronoun: ‘first person’ (*yo*), ‘second person’ (*tú*, *usted*), ‘third person’ (*él*, *ella*), ‘lexical NP’ or ‘null subject’. Following Ordóñez and Olarrea (2006: 67), we consider *usted* to be in the second person category. *Usted* is a speech act pronoun, and grammaticality judgments have indicated (Ordóñez & Olarrea 2006:71) that *usted* patterns more closely with second person pronouns than with third as far as non-

[Type text]

inversion is concerned. Null subjects were further divided into first, second and third person according to verb form. The questions with overt pronouns presented the following rates of subject-verb word order for each pronoun:<sup>8</sup> first person pronouns 86% (total overt *yo* tokens (N) = 15), second person pronouns 86% (total overt *tú/usted* tokens (N) = 105), third person pronouns 19% (total overt *él/ella* tokens (N) = 42) and lexical subjects 4% (total lexical subjects (N) = 330). For the Varbrul analysis, the first and second person pronouns were combined.

c) type of question word: In order to determine if the non-inversion was lexically limited, we coded for each type of question word. The following interrogative pronouns were found in our data; *qué* ('what') / *por qué* ('why') 38% (N = 215), *dónde* ('where') 13% (N = 76), *cómo* ('how') 16% (N = 119), *cuánto/a/os/as* ('how much/how many') 31% (N = 29), *quién* ('who') 14% (N = 21), *cuándo* ('when') 20% (N = 10), *cuál/cuáles* ('which') 0% (N = 21). For the Varbrul analysis, *qué/por qué* were contrasted with the other interrogative words.

d) verb-type: Previous quantitative analyses (Enríquez 1984: 240; Bentivoglio 1987: 60; Morales 1997: 156; Travis 2005: 341) have shown that verb type plays a role in the occurrence of overt pronominal subjects in Spanish. All these studies agree that overt subject pronouns are especially favored with psychological verbs such as *pensar* ('to think') and *creer* ('to believe') because they carry epistemic modality, and therefore the occurrence of the subject contributes to specify the speaker's degree of commitment to the truth of his/her message. Morales (1997: 156) and Travis (2005: 341) also found that another verb-type that tends to occur with overt pronominal subjects is speech verbs. Travis (2005: 342) relates this category to psychological verbs by pointing out that *decir* ('to say'), the most frequent verb in the speech class, is also used as a marker of epistemic stance.

On this basis, we coded each token for the type of verb. We classified all verbs with ten or more occurrences separately, and grouped all others into a category 'other'. The rate of subject-verb word order with each of the verb types is: *creer* ('to think') 100% (N = 10), *decir* ('to say') 54% (N = 24), *estar* ('to be') 9% (N = 53), *hacer* ('to do') 42% (N = 38), *ir* ('to go') 43% (N = 23), *llamar(se)* ('to call, to be called') 5% (N = 34), *querer* ('to want') 80% (N = 15), *ser* ('to be') 1% (N = 168), *tener* ('to have') 21% (N = 33), *ver* ('to see') 20% (N = 10), and other 55% (N = 84). In the Varbrul analysis, copula verbs *ser*, *estar*, and *llamarse* are contrasted with all other verbs combined.

e) type of question (rhetorical, quotative, interrogative): Rhetorical questions are generally defined as interrogative constructions that are not expecting an answer. Rather, they have the illocutionary force of an assertion. Previous studies (Igalada Belchí 1994: 335; Pascual Olivé 1998: 808; Cid Uribe & Ortiz-Lira 2000: 47) point out that there is no phonological, grammatical or lexical difference between rhetorical and non-rhetorical questions in Spanish. Rather, the difference between both types lies in the communicative function they convey. As is noted by Igalada Belchí (1994: 341), rhetorical questions are used as a communicative strategy to reinforce an assertion that the speaker assumes to be shared by the hearer. This can be seen illustrated in example (10). Here the speaker is explaining that, as a pilot, she is entitled to a bonus at the end of the year if a certain amount of her flights leave on time. The rhetorical question *¿qué tú crees que yo voy a hacer?* ('what do you think I am going to do?') is a strategy the speaker uses to reinforce the statement that follows it. In fact, the speaker makes no pause after she utters this

[Type text]

question, and does not seem to be seeking/requesting information:

- (10) si el vuelo sale temprano y llega dentro de catorce minutos, eso va a cuenta a un banco, si ese banco excede creo que son ciento veinte días al año, te dan a ti como bono, como piloto te dan un bono de cuatro por ciento, ciento sesenta te dan seis, ciento ochenta te dan ocho por ciento. Pues, **¿qué tú crees que yo voy a hacer?** Yo salí media hora tarde pero yo cojo el reloj, ‘salió a tiempo el vuelo’ y me cojo un ocho por ciento de, de bono cuando llega el final de año (I12, 28)

‘If the flight leaves early and arrives less than fourteen minutes later, that goes to a data bank, if that data bank has an excess of I think it is one hundred and twenty days a year, you are given a bonus, as a pilot you are given a bonus of four per cent, if one hundred and sixty, you are given six, if one hundred and eighty, you are given eight per cent. So, **what do you think I’m gonna do?** I left half an hour late but I take the clock, ‘the flight left on time’ and I get a bonus of eight per cent at the end of the year.’

Rhetorical questions were coded based on the researchers’ best interpretation of the semantic intent of the speaker.

Questions coded as ‘quotative’ are taken from portions in which the speaker is quoting another. This is illustrated in (11).

- (11) Y no solamente eso, tú vas allí, ¿y qué tú ves?, **¿a qué tú vienes hoy?** ¿cuál es el problema tuyo?, y yo, ‘un momentito, yo no tengo ningún problema, sabes, tú no estás atendiendo a uno de la calle, tú estás atendiendo a un profesional.’ (I12, 109)

‘And not only that, you go there, **¿and what do you see?** ‘what do you want today? what is your problem?’ And I’m like ‘wait a minute, I don’t have any problem, you know what?, you are not helping someone from the street, you are helping a professional.’

In this example, the speaker is criticizing the way people are treated in the administration in Puerto Rico. She uses several quotations in her retelling or re-enactment of the story.

Questions in which the speaker did not seem to be using a rhetorical device, unclear cases, or instances in which s/he was not using another speakers’ words, were coded as ‘interrogative’. Consider example (12):

- (12) [speaking with the child] ¿Qué, mi amor? **¿Qué tú quieres?**, dale, habla. (16, 78)

‘What, honey? **What do you want?** come on, talk’

[Type text]

Here, the semantic content of the question coupled with discursive cues were used to identify the question type. An answer to the question by an interlocutor was not required for it to be considered 'interrogative'.

The rate of subject-verb word order for rhetorical questions with overt subjects was 54% (N = 44). Subject-verb order occurred at a rate of 56% (N = 83) in quotations, and at a rate of 15% (N = 364) in interrogative questions.

The remaining variables (f-i) were included in order to test the notion of whether perseveration played a role in the use of non-inverted questions. Perseveration (also known as *priming*, *persistence*, *perseverance* and *bird-of-a-feather effect*) is a psycholinguistic phenomenon that entails the repetition of a linguistic form due to its occurrence in the immediately preceding contexts. Cameron and Flores-Ferrán (2004) show that the occurrence of overt pronominal subjects triggers the use of more pronominal subjects, whereas null subjects lead to the use of more null subjects. Given the clear perseveration of form and meaning, we hypothesized that we could find evidence of perseveration of word order. We hypothesized that SV word order in a previous clause would favor SV word order in the following question, and VS word order in the preceding utterance would favor VS word order in questions.

In order to test this, in our analysis of *wh*-questions, we considered linguistic variables derived from Cameron and Flores-Ferrán (2004): subject expression, word order, form of the subject, verb type in the immediately preceding context (at a distance of one or two clauses) and switch in reference. Consequently, all examples were also coded according to the following variables:

f) subject expression in the previous one or two clauses: We identified the conjugated verb that most immediately or closely preceded the question and coded the subject as 'overt' or 'null'.

g) word order of the previous one or two clauses: The clause with a conjugated verb that most closely preceded the question was analyzed to determine word order. These clauses were coded as 'subject-verb', 'verb-subject' or 'not applicable' in the case of null subjects.

h) form of *overt* subject in the previous one or two clauses: In order to determine if, independent of meaning, the use of an overt pronominal *yo*, for example, would favor use of an overt *yo*, even in cases in which the referents are different, we coded the immediately preceding clauses for whether the subject expressed was the same *form* of the subject in the *wh*-clause, different from the one in the *wh*-clause, and, again, not applicable in the case of null subjects.

i) verb in the previous one or two clauses: Cameron and Flores-Ferrán (2004) found that with increased similarity between clauses, greater was the influence of the perseveration effect. In order to examine this in our data, we coded the conjugated verb in preceding clauses as 'same as the verb expressed in the *wh*-clause' or 'different from the one in the *wh*-clause'. We did not distinguish different forms of the paradigm, but rather relied upon lemmas.

j) switch reference: As is noted by Cameron and Flores-Ferrán (2004: 49), the perseveration effect is more prevalent in those contexts in which there is not a switch in the reference of the subject than in those contexts in which switch reference takes place. We therefore also coded for switch reference in our data. In order to do this, we followed methods presented in previous research (Cameron 1993; Flores Ferrán 2004; Cameron &



[Type text]

Flores Ferrán 2004).

We identified the subject of the target verb. In our analysis, these are the verbs occurring in all direct *wh*-interrogative sentences. We then determined the subject (the trigger) of the preceding conjugated verb. The subject of the trigger clause could be expressed or null. If the subject of the target and trigger clauses had the same referent, we coded the target question as ‘no switch’. In (13) we have an example of no switch because the referent of *ella* ‘she’, the subject of the target question, coincides with the (covert) subject of the preceding conjugated verb (*está trabajando* ‘she is working’) which in turn has the same referent as *una de las hijas de O.* ‘one of O.’s daughters’, the subject of the previous clause.

- (13) I: Una de las hijas de O. es doctora, está trabajando en Florida.  
E: ¿Qué hace ella?  
I: Pediatra creo que es o algo así.  
(I1, 18-19)

‘I: One of O.’s daughters is a doctor; she’s working in Florida.  
E: What does she do?  
I: She’s a pediatrician or something like that.’

If the referents were different between the trigger and the target, these questions were coded as instances of ‘switch reference’. Consider example (14).

- (14) J: ¿quién fue la que tuvo la experiencia esa de-?  
V: ¿Quién tú crees?  
(I4, 84)

‘J: Who was the one that had that experience of-?  
V: Who do you think?’

Example (14) is an instance of switch reference because the referent of the target question (*tú* ‘you’) does not coincide with the referent of *la que tuvo la experiencia esa de-* (‘the one that had that experience of-’), which is the subject of the previous conjugated verb (the trigger).

The data collected and coded for the 882 direct *wh*-questions were then submitted to Varbrul (Rand & Sankoff 1990) for analysis. We analyzed two distinct dependent variables in two separate sets of variable rule analyses. To test for perseveration of form and meaning, we analyzed all 882 questions to determine which linguistic factor groups favored overt subjects vs. null subjects. We also set out to determine which factors constrained word order. Only questions with an overt subject were submitted to the variable rule analysis in order to determine factors that constrained word order. Each factor group (a-i above) was used in the variable rule analyses. The results are presented in the following sections.

#### 4. Results

##### 4.1 Rates of subject expression and subject position

Overall in the speech of these native Puerto Ricans, subjects are null in approximately half of the cases (49%), as can be seen in Table 1. The expressed subjects are found in both pre- and postverbal positions. When a subject is expressed, however, it is overwhelmingly placed in preverbal position, with just 46 expressed subjects being found in postverbal position. As results in Table 2 illustrate, when the subject is found in postverbal position, it is overwhelmingly a lexical NP and not a subject pronoun. Of the postverbal subjects 74% are either a proper name (N = 8) or any other type of lexical NP (N = 26). Data from Tables 1 and 2 indicate, therefore, that in conversational speech, a mere 4% of subjects of singular verbs are expressed overtly after the verb, and 74% of the overtly expressed post-verbal subjects are non-pronominal. These findings corroborate the notion of Puerto Rican Spanish word order as predominantly SVO in conversational speech.

**Table 1.** Subject expression and position with regard to verb in all sentences

	<b>Percent</b>
<b>Null</b> (N = 526)	49
<b>Overt</b>	
Preverbal (N = 503)	47
Postverbal (N = 46)	4
<b>Total</b> (N = 1075)	100

**Table 2.** Subject expression and position with regard to verb in all sentences

<b>Subject</b>	<b>% preverbal</b>	<b>% postverbal</b>
lexical	13 (N = 66)	57 (N = 26)
proper	3 (N = 14)	17 (N = 8)
non-pronominal	16 (N = 80)	74 (N = 34)
<i>yo</i>	23 (N = 117)	7 (N = 3)
<i>tú</i>	17 (N = 85)	4 (N = 2)
<i>usted</i>	0 (N = 1)	0 (N = 0)

[Type text]

<i>ella</i>	12 (N = 59)	4 (N = 2)
<i>él</i>	14 (N = 69)	2 (N = 1)
<i>uno</i>	3 (N = 14)	0 (N = 0)
<i>esto / este</i>	2 (N = 9)	0 (N = 0)
<i>ese</i>	1 (N = 4)	0 (N = 0)
<i>esa</i>	2 (N = 9)	2 (N = 1)
<i>eso</i>	11 (N = 56)	7 (N = 3)
Total	100% N = 503	100% N = 46

Results in Table 1 differ from the ones provided by previous research on other Spanish dialects, in which the percentage of postverbal subjects is considerably higher (from 22% to 33%, depending on the study). In her analysis of the Spanish of Caracas, Bentivoglio (1988: 14) reports that VS word order amounts to 30% of all cases of overt subjects. Bentivoglio and Weber (1986) provide similar results (32%) from their corpus of the Spanish of Caracas, Santiago de Chile and Mexico. Likewise, using a corpus of Peninsular conversational Spanish, Rivas (2008: 896) finds that VS order accounts for almost 33% of all overt intransitive subjects. In contrast, Morales (2006: 489), whose results are based on a corpus of the Spanish of San Juan, Madrid and Buenos Aires, provides a slightly lower percentage of postverbal subjects, namely, 22.48%. Although Morales (2006) does not provide separate data for each dialect, we suggest that the lower overall percentage of postverbal subjects reported by that researcher may be due to the inclusion of the Puerto Rican variety in the analysis.

Subject expression is higher in *wh*-questions as can be seen in Table 3. In a direct *wh*-question, subjects are null just 42% of the time,<sup>9</sup> and most *wh*-questions (44% of them) have an overt, post-verbal subject as is prescriptively required in Spanish. As discussed above, non-inversion in questions is a salient feature of Caribbean Spanish. However, as the quantitative analysis reveals, subjects fail to invert in just 14% of the 882 questions analyzed.

**Table 3.** Null and overt subjects in *wh*-questions

	N	%
<b>Overt subject</b>		
preverbal	124	14
postverbal	392	44
<b>Null subject</b>	366	42
<b>Total</b>	882	100

Non-inverted questions, therefore, of the type *¿qué tú piensas?* ('what do you think?') are comparatively infrequent. If we exclude the 366 cases of null subjects, however, and just examine cases in which a subject is overtly expressed, the proportion of SV questions is, of course, greater. Roughly 25% of *wh*-questions with expressed subjects are SV. These results suggest that non-inversion is perhaps not as prevalent in Puerto Rican Spanish as is generally assumed, and the association of non-inverted questions with Puerto Rican and Caribbean dialects may constitute a sociolinguistic stereotype.

A comparison of rates of subject expression in postverbal position suggests that VS word order predominates as a way to mark syntactically an utterance as a question. In these interrogative sentences, 76% of expressed subjects appear postverbally. This contrasts markedly with overall rates of subject expression in which, when expressed, only 4% of subjects occur in postverbal position (Table 1).

Results of our quantitative analysis clearly demonstrate that the intuitions expressed in previous research regarding the prevalence of the second person singular pronoun are substantiated. As Table 4 illustrates, when the subject precedes the verb in questions (SV) the pronoun is overwhelmingly *tú* or *usted* at 73%. Expressed at significantly lower rates (10%) are the first person *yo* pronoun and other pronominal subjects such as *eso* ('that'), *uno* ('one') or a proper noun. Results in Table 4 are parallel to the ones provided by Ordóñez and Olarrea (2006: 71) based on grammaticality judgments of native speakers.

**Table 4.** Subject pronouns in SV and VS questions

	SV		VS		Total
	N	%	N	%	
1 <sup>st</sup> person <i>yo</i>	13	10	2	1	15
2 <sup>nd</sup> person <i>tú, Ud.</i>	91	<b>73</b>	14	4	105
3 <sup>rd</sup> person <i>él, ella</i>	8	5	34	9	42
lexical	12	10	336	<b>87</b>	342
<b>Total</b>	124	100	386	100	504

This result can be contrasted with the questions in which the expressed subject is in the expected position: post-verbal or VS. As Table 4 illustrates, in VS questions the overt

[Type text]

subject is overwhelmingly lexical (87%). Only 4% of VS questions have *tú* or *usted* as the subject. For the 105 instances of *tú* and *usted* used in a question, 87% (N = 91) are expressed pre-verbally, a point to which we will return. Although the token number is quite low, of the 15 instances of *yo* used in a question, like the instance with *tú/usted*, 87% (N = 13) have SV word order. We now turn to what linguistic factors seem to favor or disfavor the appearance of a an overt subject in conversational Puerto Rican Spanish.

**4.2 Linguistic factors that favor / disfavor overt subject usage in Puerto Rican Spanish**  
 Previous research (summarized above) has addressed factors that constrain overt subject usage vs. null subjects in Spanish. The aforementioned studies (Hochberg 1986; Cameron 1993; Morales 1997, 2006; Flores-Ferrán 2004) do not seem to specifically address direct *wh*-questions as we do in this current analysis.<sup>10</sup> In order to determine what may or may not favor the appearance of overt subjects in direct *wh*-questions, we submitted our data to a variable rule analysis using Varbrul (Rand & Sankoff 1990). This enables us to determine the independent contribution of each factor group while controlling for all the other independent variables (Guy 1993).

Through this analysis, we are able to determine the independent statistical significance of each factor group – determined by both a ‘p’ value and by the log likelihood (Sankoff 1988). Further, Varbrul enables us to determine the relative strength of each factor group. The greater the range of the factor group, the greater the magnitude of effect. The factor group with the greatest range, therefore, is the group that contributes most significantly to constraining the occurrence of an overt or null subject in these questions. Lastly, we can determine a constraint hierarchy through the Varbrul analyses. Within each factor group, the individual factors are ranked according to their factor weight. These weights reflect the degree to which they favor (> .50) or disfavor (<.50) the application of the dependent variable.

The results of these analyses are summarized in Table 5. In direct *wh*-questions, the factor group to most significantly constrain overt subject usage was the verb type. The copula verbs (*ser, estar, llamarse*) all heavily favor the presence of an overt subject with a factor weight of .67. Questions containing all other verb types strongly disfavor overt subjects with a factor weight of .38. This result is inline with previous research in which verb type significantly constrains pronominal subject use.

**Table 5.** Linguistic factors favoring overt subjects in *wh*-questions in Puerto Rico

	% overt	Factor weight	% data
<b>Verb type</b>			
copula ( <i>ser, estar, llamarse</i> )	73	<b>.67</b>	40
other	47	<b>.38</b>	60
	<i>Range</i>	29	
<b>Question type</b>			
quotative	76	<b>.72</b>	13
rhetorical	66	<b>.63</b>	7
‘true’ interrogative	55	<b>.45</b>	79
	<i>Range</i>	27	

[Type text]

<b>Switch reference</b>			
switch	63	<b>.54</b>	72
no-switch	43	<b>.39</b>	27
	<i>Range</i>	<i>15</i>	
<b>Previous clause subject expression</b>			
overt	64	<b>.56</b>	53
null	50	<b>.43</b>	46
	<i>Range</i>	<i>13</i>	

---

Input: .160, Total N = 842

Log likelihood = -512.101, Chi-square per cell = 1.0423

The next factor group selected as significant was the question type. In both quotative and rhetorical questions, overt subjects are preferred (factor weights .72 and .63 respectively). Conversely in interrogative questions, null subjects are favored (factor weight .45). These results seem to reflect pragmatic functions of the subjects. Davidson (1996), for example, in an analysis of first and second person pronouns highlights several discourse - pragmatic functions that these overt subject (pronouns) have in Spanish. Davidson's (1996) study does not include third person subjects, but some explanation of our results can be found in his work.

Overt subjects add 'pragmatic weight' to an utterance. As noted by Davidson, utterances with overt pronominal subjects "are 'weightier', in the sense of more personally relevant and more invested with emotion, than other types of utterances in a conversation" (1996: 556). In this way, by using overt subjects in quotative questions (and hence adding pragmatic weight), the speaker ensures a more dramatic retelling of his or her story. Davidson also finds that subject pronouns can be used to mark utterances as "more personally relevant" (1996: 553) to the speaker. In this way the rhetorical questions, which are questions that express ponderings on the part of the speaker more than they are expressions of lack of knowledge about someone or something, seem to reflect the same tendency.

The summary of our findings in Table 5 also reveals the importance of switch reference in our data – even when considering questions in isolation. Consistent with findings reported in previous studies, when there is a switch in reference between the previous clause and the target question subject, overt subjects are preferred (with a factor weight of .54). When there is no switch in reference, and the subject of the trigger clause and the target clause are the same, overt subjects are disfavored (factor weight .39). This finding suggests perseveration of meaning.

There is also perseveration of form, however. The notion that overt subjects yield more overts and nulls breed more nulls has been highlighted in other studies (Cameron and Flores-Ferrán 2004). In our analysis, we find that when there is an overt subject in the preceding clause, overt subjects are favored (factor weight .56), and conversely when the preceding subject is null, the target subject tends to be null. This effect suggests perseveration of form (independent of meaning), where irrespective of a change in referent or not, overt yields overt and null yields null. Clearly, there is perseveration of both form and meaning with regard to overt subject expression in our data. Does

[Type text]

perseveration account for word order? The following section outlines the results of our variable rule analyses on word order in questions.

### 4.3 Linguistic factors that favor / disfavor SV word order in *wh*-questions

Of all the variables that we used in our quantitative analysis described in methods, Varbrul selected as significant just three; the type of subject expressed, the type of question, and the verb category. The results are summarized in Table 6.

**Table 6.** Linguistic factors favoring SV word order in *wh*-questions in Puerto Rico

	% SV	Factor weight	% data
<b>Subject type</b>			
1 <sup>st</sup> , 2 <sup>nd</sup> ( <i>yo, tú, Ud.</i> )	86	<b>.96</b>	24
3 <sup>rd</sup> ( <i>él, ella</i> )	19	<b>.49</b>	8
lexical	4	<b>.24</b>	67
<i>Range</i>		<i>72</i>	
<b>Question type</b>			
rhetorical, quotative	55	<b>.81</b>	25
interrogative	15	<b>.38</b>	74
<i>Range</i>		<i>43</i>	
<b>Verb type</b>			
non-copula	49	<b>.72</b>	48
copula ( <i>ser, estar, llamarse</i> )	3	<b>.30</b>	51
<i>Range</i>		<i>42</i>	

Input: .10, Total N – 492

Log likelihood = -101.601, Chi-square per cell = 1.4828

The linguistic factor group with the greatest magnitude of effect in constraining realization of SV or VS word-order in Puerto Rican questions is the subject type (range = 72). When the subject is an overt *yo, tú* or *usted*, non-inversion is highly favored with a factor weight of .96. When the subject is third-person pronoun (*él, ella*), lack of inversion is very slightly disfavored with a factor weight of .49. This effect is close to negligible as the weight so closely approximates .50. If the subject is a lexical NP, however, SV word order is strongly disfavored with a factor weight of .24. Lexical subjects account for just 4% of the non-inverted questions. These tendencies are inline with previous research.

Verb type is also selected as significant. If the verb is not a copula, that is to say either *ser* ('be'), *estar* ('be'), or *llamarse* ('be called'), SV word order is favored. Verbs contrasted with the copulas include the frequent verbs *creer* ('think'), *decir* ('say') *pensar* ('think'), *querer* ('want') often cited as contributing to this process. However, it also includes 61 other verb types – not all of which have a high token frequency. This result suggests that the occurrence of non-inverted questions is not lexically limited to one verb class such as psychological verbs *creer* and *pensar*. If, however, the verb is a copula, SV word order is highly disfavored with a factor weight of just .30.

The verb *ser* ('be') makes up approximately a third of the data reported upon here

[Type text]

(34%). Of the 168 tokens of *ser*, only one had SV word-order [*¿y cuándo E. no ha sido así?* ‘and when has E not been like that?’(I15, 13)]. The remaining tokens are predominantly the question *¿Qué es eso?* (‘what’s that?’). If *ser* is excluded from analysis (owing to the near categorical VS word order), SV word order in questions is greater (38%) yet the same factor groups are selected as significantly constraining word order variation in Puerto Rican questions.

Importantly, our analysis identifies type of question as a factor group that significantly constrains word order in these questions. Both quotative and rhetorical questions strongly favor the use of non-inversion with a factor weight of .81. In these question types SV word order is used 55% of the time (N = 126). On the other hand, interrogatives, or questions in the classical sense, strongly disfavor SV word order with a weight of .38. In these cases, SV word order is used in just 15% of the cases (N = 364).

As we noted previously (section 4.1), SV word order predominates in Puerto Rican Spanish, and VS word order continues to be a syntactic marker for questions (Table 3). Both rhetorical questions and questions found in quotations share an important commonality, however. In neither case is the speaker trying to elicit an answer from his or her interlocutor. The inexistence of a marker to syntactically denote a question (VS word order), therefore, does not conflict with communicative goals.

Rhetorical and quotative questions make up approximately a quarter of the data reported upon in Varbrul (25%). Despite this, more than half (57%) of the non-inverted questions are of this type, as can be seen summarized in Table 7. We feel that this finding suggests a pathway by which rigid word order is permeating the grammar of Spanish, which is discussed in the next section.

**Table 7.** Percent of data with SV word order for each question type

	<b>% data SV</b>
<b>Interrogative questions</b> (N = 54)	44
<b>Quotative and rhetorical questions</b> (N = 70)	56

Variable rule analyses ruled out any perseverative effect as a source of or a motivating factor for SV word order in these direct *wh*-questions. None of the factor groups that we coded in order to test this (summarized in methods) was selected as significant. Cameron and Flores-Ferrán (2004) described interactions in their data. That is, when there was a switch in reference, the perseverative effect was diminished, and when there was not a switch, the perseveration effect was greater. We tested for this in our data by running independent analyses on questions with no-switch and on questions with switch reference. In neither case did we find evidence of perseveration of form (SV word order) and in those analyses the same factor groups highlighted in Table 6 were selected as significant, with the same magnitude of effect and ordering of factors within each group. It is apparent, therefore, that perseveration is not a process active in the selection of SV word order in questions for the Puerto Rican data. just to the probability



[Type text]

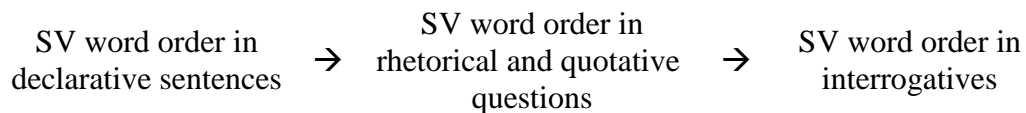
of overt subjects vs. null.

## 5. Discussion

The solidifying of subject-verb word order in this variety of Spanish is not such an improbable notion. Word order in modern Spanish is more rigid than that of earlier periods. In the history of the Spanish language (Penny 1991; Lapesa 2000) examples abound in which, over the centuries, certain grammatical constructions cease to have flexible word order in favor of a rigidly prescribed ordering of constituents (e.g. in the creation of the synthetic future and conditional verb forms, formation of perfect tenses, position of verbal clitics). In all these examples, word order was once variable but is now syntactically rigid. It seems the non-inversion we highlight in this study follows a general trend in Spanish of becoming less flexible and more rigid in the ordering of constituents.

Results of the present analysis show that, in this variety of Spanish, SVO is the dominant word order. In our analysis of 1000 consecutive first, second and third person singular verbs, just 4% had a subject expressed in postverbal position. The tendency for the subject to precede the verb is so pervasive that it seems SV word order is encroaching upon one of the few instances in which VS word order is grammatically required; in direct interrogative questions. Approximately a quarter of these questions with an expressed subject have SV word order.

This variety of Spanish, therefore, appears to be solidifying a subject-verb word order even in contexts specifically requiring the inverse. The quantitative analysis coupled with results of the variable rule analysis suggest the following pathway of change through which this is possible:



The first step in this change is an increase in overt (pronominal) subject usage overall in the language. As previous studies show (Hochberg 1986; Cameron 1993; Morales 1997), subjects are expressed in Puerto Rican Spanish at a significantly higher rate than in other varieties of Spanish. Further, these overt subjects are overwhelmingly placed in preverbal position, strengthening the SV construction through frequent repetition. The SV order is in this way extended to constructions in which other varieties of Spanish require subject-verb inversion such as in infinitival clauses (*para yo hacerlo* ‘for me to do it’) and *wh*-questions such as *¿qué tú piensas?* (‘what do you think?’).

Our data show that direct *wh*-questions continue to be the most frequently occurring contradiction to the SV word order. In fact, *wh*-questions display subject-verb inversion in 75% of cases. However, our data also reveal that the strong preference for SV word order begins to make inroads in questions in which the speaker is not seeking immediate input in order to fill an information gap (rhetorical uses and quotative cases). In these uses, the lack of a syntactic marker (VS word order) to indicate a question is not functionally necessary owing to the nature of the utterance. That is to say, the speaker is free to not mark the question syntactically as an interrogative since the communicative goal is not necessarily to obtain information in response. Allowing SV word order in these non-information-seeking questions weakens the overall VS tendency in

[Type text]

interrogatives and brings this variety of Spanish one step closer to being a rigid SV language.

The frequent occurrence of SV word order in the rhetorical and quotative questions allows the SV order to encroach upon the information-seeking interrogatives. Leading the way in this area is the most frequent pronoun *tú* ('you', informal), accounting for 83% (n = 45) of the 54 interrogatives of this type (Table 7). It seems, then, that when an information-seeking interrogative fails to invert, the subject is overwhelmingly the overt pronoun *tú*. Morales (1989: 149) already regarded *tú* as one of the pronominal forms that contributed to the solidifying of SV order. She based this idea on the frequent appearance of this pronoun in infinitive constructions with a preverbal subject and in the overall occurrence of (non-specific) *tú* generally. However, unlike what happens in the contexts described by Morales (1989), in non-inverted questions second-person *tú* is overwhelming specific (91% of cases). This result suggests that second person pronouns take the lead in the change towards a rigid SV language, regardless of whether their referent is specific or non-specific.

Why would it be the case that *tú* is leading the way? We suggest one possible explanation is the likelihood of 'chunk' status of the subject pronoun and the verb in lexical representation. Within the exemplar model of lexical representation we adopt in our usage-based approach (Bybee 2001), it is argued that with increased frequency of a word pair (such as *tú vives* 'you (informal) live') there is increased likelihood of lexical storage as a unit or 'chunk'. So with increased string frequency there is increased possibility of chunk status.<sup>11</sup> Further, the probability of co-occurrence of two lexical items given the independent lexical frequency of one of the words, or, the transitional probability (Jurafsky, Bell, Gregory & Raymond 2001; Bybee 2001; Bush 2001), also contributes to the probability of lexical storage as a chunk.

In our data, we calculate for the 370,000 words the string frequency (Bush 2001) of each verb type (individual verb forms such as *creo, crees, cree; digo, dices, dice*, etc. plus the corresponding subject pronoun) found in SV interrogative questions (N = 57) [see Appendix]. For example, the combination *yo creo* appears 188 times in the 370,000 words we analyze. Further, we calculate the transitional probabilities for each of these forms. To continue with the same example, the verb form *creo* appears 358 times. Of those uses, 188 (the string frequency of *yo creo*) are directly preceded by *yo*. This yields a transitional probability of .53 (over half the occurrences of *creo* being preceded by *yo*). Despite the subject pronoun *tú* not being as frequent as the first person singular pronoun *yo* in the whole corpus (*yo* = 5571, *tú* = 2493), the string frequencies of *tú* + the verb form is the highest (*tú* = 1006, *yo* = 969, *ella* = 190, *él* = 151). Further, *tú* + the corresponding verb form has the highest transitional probability of all pronominal forms (*tú* + verb = 0.45, *yo* + verb = 0.27, *ella* + verb = 0.03, *él* + verb = 0.03).

If, indeed, increased string frequency as well as increased transitional probability can be interpreted as suggestive of 'chunk' status in lexical representation, then the cases of *tú* in non-inverted information-seeking questions finds at least partial support in this explanation. When speakers access verbs in memory, it is possible for the lexical selection to be the verb accompanied by the subject pronoun (i.e. the chunk). The implication is that the overall use of *tú* is bolstered by the chunk status of the combinations of this pronoun plus the verb. The subject pronoun is more apt to be used given its dual representation in memory – as a separate lexical entry *tú* and as part of

[Type text]

chunks such as *tú vives*. The same is not true for other subject pronouns + verbs with lower string frequencies and transitional probabilities which make them less likely to be stored as a chunk.

The high rate of overt (pronominal) subjects reinforces the SV construction in the grammar of this Puerto Rican variety. The SV order spreads to rhetorical and quotative questions, increasing the frequency, and perhaps acceptability, of the construction [interrogative pronoun + subject + verb]. Lastly, we see the advent of acceptability of this construction in information-seeking interrogative utterances, with the subject pronoun *tú* leading the way. A postulated last step could be the generalization of all subject types in these SV interrogative sentences. Future diachronic analyses may be able to confirm or disconfirm this notion.

## Notes

1. We are grateful to Renee Payne for help on some initial coding of data. We are indebted to Mayra Cortés-Torres for the use of the interviews that she conducted and transcribed and for her constructive suggestions. We also received helpful comments on an earlier version of this paper presented at LASSO XXXVII. We would like to thank William Raymond for providing us with raw frequency values for words in our corpus. We would also like to acknowledge useful suggestions for improvements provided by two anonymous reviewers. Any errors are our own.
2. In fact, Davis (1971: 331) indicates that non-inversion is restricted to the pronoun *tú*.
3. In this respect, Torrego (1984: 105-106) maintains that only when the *wh*-word is an argument of the verb is inversion compulsory in Spanish. Therefore, non-inversion is possible outside of the Caribbean with both non-argument *wh*-words such as *cómo* ‘how’, *cuándo* ‘when’, *dónde* ‘where’, and with complex *wh*-words such as *por qué* ‘why’ and *en qué medida* ‘to what an extent’. In Italian, non-inversion is also possible with *perché* ‘why’ (Rizzi 1996: 87, n. 16). In our data, all the examples with a preverbal lexical subject occur with non-argument *wh*-words. We have found no examples non-inverted lexical subjects with *qué* or *quién*.
4. We understand the term ‘dominant order’ as the statistically most frequent, in the sense proposed by Siewierska (1988: 8): “By definition the term ‘dominant’ implies statistical prevalence”.
5. This high frequency of overt pronominal forms in Caribbean dialects has been accounted for by the need to functionally compensate for the loss of /s/ in syllable-final position (Hochberg 1986). Since /s/ is a marker of second person, its elision leads to double (*tú/él tiene* ‘you/he have/has’) or triple (*yo/tú/él tenía* ‘I/you/he had’) ambiguity in some tenses. However, later studies (e.g. Cameron 1993) find that the occurrence of subject pronouns is not causally related to the loss of implosive /s/. In fact, as Toribio (2000: 320) points out, in language lack of verbal agreement does not always entail the obligatory occurrence of subject pronouns.
6. In this same line, D’Introno (2000: 122-123) argues that Caribbean Spanish may be in the process of changing from a pro drop language to a non-pro drop language. For this scholar, the parameter of pro-drop includes a number of subparameters: number-person, case, animacy, specificity, thematic role (D’Introno 2000: 138). In the transition from pro-drop to non-pro-drop some of these subparameters undergo a weakening that renders the system as unstable. The use of overt pronominal subjects in Caribbean Spanish

[Type text]

reveals a tendency to change the values of the factors that are included in the pro-drop parameter. The result of this instability is not clear: either some of the subparameters change or the whole parameter changes and Caribbean Spanish becomes a non-pro-drop language. Alternatively, in a study on Dominican Spanish (another variety with high overt pronoun usage and non-inversion in questions), Camacho (2008) argues that variation may be triggered by changes in the prosodic and morphological properties of words with ‘pronominal features’ [lexical designation as strong pronouns, weak pronouns, clitics or inflection] as a language changes from being a null subject language to a non-null subject language.

7. Moreover, previous studies (e.g. Flores-Ferrán 2004: 59) confirm that singular pronominal subjects occur more frequently than plural pronominal subjects.
8. In this and the following sections, the percentage corresponds with the occurrence of the variant in non-inverted constructions, whereas ‘N’ provides the total number of examples of the relevant factor in both inverted and non-inverted questions.
9. Of these cases of null subjects, 6% are first person verb forms, 35% are second person forms, and 59% of null subjects are third person verb forms. Although third person forms account for most (59%) of the null subjects in our corpus, third person subjects are more often expressed overtly, rather than null. In our data, third person subjects are overt 64% of the time, whereas second and first person subjects are overt 46% and 41% of the time, respectively. This pattern of overt-null subject expression is in contrast to that found in Brazilian Portuguese. Barbosa, Duarte and Kato (2005) point out that in Brazilian Portuguese overt third person subject expression lags behind first and second persons.
10. It is not clear from previous studies if interrogatives were excluded or if they were counted with all declarative sentences.
11. Representation of the individual lexical items within a chunk does not preclude their independent lexical representation elsewhere.

## References

- Barbosa, Pilar, María Eugenia L. Duarte and Mary A. Kato 2005. “Null subjects in European and Brazilian Portuguese”. *Journal of Portuguese linguistics* 4: 11-52.
- Bentivoglio, Paola. 1988. “La posición del sujeto en el español de Caracas: un análisis de los factores lingüísticos y extralingüísticos”. *Studies in Caribbean Spanish dialectology* ed. by Robert M. Hammond & Melvyn C. Resnick, 13-23. Washington DC: Georgetown University Press.
- Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Bentivoglio, Paola y Elizabeth Weber. 1986. “A functional approach to subject word order in spoken Spanish”. *Studies in Romance linguistics* ed. by Osvaldo Jaeggli and Carmen Silva-Corvalán, 23-40. Riverton, USA: Foris.
- Bergen, John J. 1976. “The explored and unexplored facets of questions such as ‘qué tú tienes?’”. *Hispania* 59. 93-99.
- Blackwell, Sarah. 2003. *The case of Spanish NP anaphora*. Amsterdam/Philadelphia: John Benjamins.
- Bush, Nathan. 2001. “Frequency effects and word-boundary palatalization in English”. *Frequency and the emergence of linguistic structure* ed. by Joan Bybee and Paul Hopper, 255-280. Amsterdam/Philadelphia: John Benjamins.

[Type text]

- Butt, John and Carmen Benjamin. 2000. *A new reference grammar of modern Spanish*. London: Arnold.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Cameron, Richard. 1993. "Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain". *Language Variation and Change* 5. 305-334.
- Cameron, Richard and Nydia Flores-Ferrán. 2004. "Perseveration of subject expression across regional dialects of Spanish". *Spanish in Context* 1(1). 41-65.
- Cid Uribe, M.E. and Héctor Ortiz-Lira. 2000. "La prosodia de las preguntas indagativas y no-indagativas del español culto de Santiago de Chile". *Lingüística Española Actual* 22(1). 23-49.
- Cifuentes Honrubia, José L. 2000. "El orden de palabras en la oración". *Introducción a la lingüística española* ed. by Manuel Alvar, 359-370. Barcelona: Ariel.
- Comrie, Bernard. 1989. *Language universals and linguistic typology. Syntax and morphology*. Oxford: Blackwell.
- Cortés-Torres, Mayra E. 2005. *La perífrasis estar + -ndo en el español puertorriqueño: variación dialectal o contacto lingüístico?* Ph.D. dissertation, University of New Mexico.
- Davidson, Brad. 1996. "'Pragmatic weight' and Spanish subject pronouns: The pragmatic and discourse uses of 'tú' and 'yo' in spoken Madrid Spanish". *Journal of Pragmatics* 26. 543-565.
- Davis, J. Cary. 1971. "Tú, ¿qué tú quieres?". *Hispania* 54(2). 331-333.
- D'Introno, Francesco. 2000. "Teoría generativa y estudios sintácticos sobre el español de Puerto Rico y el Caribe". *Revista de estudios hispánicos* 26(1). 107-147.
- D'Introno, Francesco. 2001. *Sintaxis generativa del español: Evolución y análisis*. Madrid: Cátedra.
- Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.
- Flores-Ferrán, Nydia. 2004. "Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact?". *Language Variation and Change* 16. 49-73.
- Gili Gaya, S. 1961. *Curso superior de sintaxis española*. Barcelona: Publicaciones y ediciones Spes, S.A.
- Guy, Gregory. 1993. "The quantitative analysis of linguistic variation". *American dialect research* ed. by Dennis R. Preston, 223-49. Amsterdam: John Benjamins.
- Heap, David. 1990. "Questions with preverbal subject in Dominican Spanish: Morphological variation and syntactic restructuring". *Papers from the annual meeting of the Atlantic Provinces Linguistic Association* 14. 59-67.
- Hochberg, Judith. 1986. "Functional compensation for /s/ deletion in Puerto Rican Spanish". *Language* 62(3). 609-621.
- Igalada Belchí, Dolores Asunción. 1994. "Estrategias comunicativas: la pregunta retórica en español". *Revista Española de Lingüística* 24(2). 329-344.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory and William D. Raymond. 2001. "Probabilistic relations between words: Evidence from reduction in lexical

[Type text]

- production". *Frequency and the emergence of linguistic structure* ed. by Joan Bybee and Paul Hopper, 229-254. Amsterdam/Philadelphia: John Benjamins.
- Lantolf, James P. 1980. "Constraints on interrogative word order in Puerto Rican Spanish". *The Bilingual Review* 7(2). 113-122.
- Lapesa, Rafael. 2000. *Estudios de morfosintaxis histórica del español* (2 vols). Madrid: Gredos.
- Lipski, John M. 1977. "Preposed subjects in questions: Some considerations". *Hispania* 60(1.) 61-67.
- Lipski, John M. 1994. *Latin American Spanish*. London: Longman.
- López Meirama, Belén. 1997. "Aportaciones de la tipología lingüística a una gramática particular: El concepto de orden básico y su aplicación al castellano". *Verba* 24: 45-82.
- Mallinson, Graham and Barry J. Blake. 1981. *Language typology: Cross-linguistic studies in syntax*. Amsterdam: North Holland.
- Morales, Amparo. 1989. "Hacia un universal sintáctico del español del Caribe: el orden SVO". *Anuario de lingüística hispánica* 5. 139-152.
- Morales, Amparo. 1997. "La hipótesis funcional y la aparición de sujeto no nominal: el español de Puerto Rico". *Hispania* 80(1). 153-165.
- Morales, Amparo. 2006. "Los sujetos 'ligeros' del español y su posición en la oración". *Haciendo lingüística. Homenaje a Paola Bentivoglio* ed. by Mercedes Sedano, Adriana Bolívar & Martha Shiro, 487-501. Caracas: Universidad Central de Venezuela.
- Ocampo, Francisco. 1995. "The word order of two-constituent constructions in spoken Spanish". *Word order in discourse* ed. by Pamela Downing & Michael Noonan, 425-447. Amsterdam/Philadelphia: John Benjamins.
- Ordóñez, Francisco and Antxon Olarrea. 2006. "Microvariation in Caribbean/non Caribbean Spanish interrogatives". *Probus* 18. 59-96.
- Otheguy, Ricardo, Ana Zentella and David Livert. 2007. "Language and dialect contact in Spanish in New York: Towards the formation of a speech community". *Language* 83: 770-802.
- Pascual Olivé, Esther. 1998. "The counsel's report: A cognitive sociopragmatic analysis of rhetorical questions". *Estudios de lingüística cognitiva* ed. by José L. Cifuentes, 805-816. Alicante: Quinta Impresión S.L.
- Penny, Ralph. 1991. *A history of the Spanish language*. Cambridge: Cambridge University Press. Span. trans. *Gramática histórica del español* by José I. Pérez Pascual & María E. Pérez Pascual. Barcelona: Ariel, 1993.
- Quirk, Ronald J. 1972. "On the extent and origin of questions in the form *qué tú tienes?*" *Hispania* 55(2). 303-304.
- Rand, David & David Sankoff. 1990. *GoldVarb: A variable rule application for the Macintosh*. Montreal, Canada: Centre de recherches mathématiques, Université de Montréal.
- Rivas, Javier. 2008. "La posición del sujeto en las construcciones monoactanciales del español: Una aproximación funcional". *Hispania* 91(4). 897-912.
- Rizzi, Luigi. 1996. "Residual verb second and the *wh*-criterion". *Parameters and functional heads: Essays in comparative syntax* ed. by Adriana Belletti and Luigi Rizzi, 63-90. Oxford / New York: Oxford University Press.

[Type text]

- Sankoff, David. 1988. "Variable rules". *Sociolinguistics. An international handbook of the science of language and society* ed. by Ulrich Ammon, Norbert Dittmar & Klaus J. Mattheier, 984-997. New York: Walter de Gruyter.
- Siewierska, Anna. 1988. *Word order rules*. London: Croom Helm.
- Silva-Corvalán, Carmen. 1982. "Subject expression and placement in spoken Mexican-American Spanish". *Spanish in the United States: Sociolinguistic aspects* ed. by Jon Amastae & Lucía Elías-Olivares, 93-120. Cambridge: Cambridge University Press.
- Suñer, Margarita. 1986. Lexical subjects of infinitives in Caribbean Spanish. *Studies in Romance linguistics* ed. by Osvaldo Jaeggli and Carmen Silva-Corvalán, 189-203. Dordrecht, Holland / Riverton, NJ: Foris Publications.
- Toribio, Almeida J. 2000. "Setting parametric limits on dialectal variation in Spanish". *Lingua* 10. 315-341.
- Torrego, Esther. 1984. "On inversion in Spanish and some of its effects". *Linguistic Inquiry* 15(1). 103-129.
- Travis, Catherine E. 2005. "The yo-yo effect: Priming in subject expression in Colombian Spanish". *Selected papers from the 34<sup>th</sup> Linguistic Symposium on Romance Languages (LSRL), Salt Lake City, 2004* ed. by Randall Gess & Edward J. Rubin, 329-349. Amsterdam/Philadelphia: John Benjamins.

#### *Authors' addresses*

Esther L. Brown  
Department of Spanish and Portuguese  
University of Colorado at Boulder  
McKenna Languages Building  
278 UCB  
Boulder, CO 80309-0278 USA

[esther.brown@colorado.edu](mailto:esther.brown@colorado.edu)

Javier Rivas  
Department of Spanish and Portuguese  
University of Colorado at Boulder  
McKenna Languages Building  
278 UCB  
Boulder, CO 80309-0278 USA

[rivasrod@colorado.edu](mailto:rivasrod@colorado.edu)

#### **Appendix: Bigram frequencies and transitional probabilities of verbs with pronominal subject *tú* in SV interrogative questions**

<b>verb</b>	<b>yo</b>	<b>tú</b>	<b>él</b>	<b>ella</b>
<i>coger</i>	8/24= 0.33	<b>8/22= 0.36</b>	2/44= 0.05	2/44= 0.05
<i>comer</i>	<b>3/11= 0.27</b>	2/11= 0.18	1/43= 0.02	0/43= 0
<i>creer</i>	188/358= 0.53	<b>16/24= 0.70</b>	0/11= 0	0/11= 0

[Type text]

<i>dar</i>	<b>2/31= 0.06</b>	1/31= 0.03	0/277= 0	1/277= 0
<i>decir</i>	99/398= 0.25	<b>44/93= 0.47</b>	22/570= 0.04	29/570= 0.05
<i>decir</i> (past)	<b>128/356= 0.36</b>	4/29= 0.14	10/400= 0.03	14/400= 0.04
<i>estar</i>	<b>101/318= 0.32</b>	60/191= 0.31	27/1659= 0.02	41/1659= 0.02
<i>hacer</i>	15/85= 0.18	<b>16/79= 0.20</b>	10/487= 0.02	8/487= 0.02
<i>hacer</i> (past)	14/71= 0.20	<b>8/32= 0.25</b>	5/163= 0.03	7/163= 0.04
<i>hacer</i> (cond.)	<b>1/3= 0.33</b>	<b>1/3= 0.33</b>	0/3= 0	0/3= 0
<i>jugar</i>	5/25= 0.20	<b>4/6= 0.67</b>	0/9= 0	1/9= 0.11
<i>llamar</i>	6/26= 0.23	<b>4/12= 0.33</b>	1/220= 0.00	0/220= 0
<i>llevar</i>	7/28= 0.25	<b>7/21= 0.33</b>	1/68= 0.01	0/68= 0
<i>pensar</i>	13/20= 0.65	<b>6/9= 0.67</b>	2/24= 0.08	0/24= 0
<i>poner</i>	3/43= 0.07	<b>3/26= 0.12</b>	0/61= 0	1/61= 0.02
<i>querer</i>	<b>61/173= 0.35</b>	31/138= 0.22	12/152= 0.08	9/152= 0.06
<i>saber</i>	97/776 = 0.13	<b>612/880= 0.70</b>	9/147= 0.06	7/147= 0.048
<i>tener</i>	<b>162/498= 0.33</b>	108/399= 0.27	42/1115= 0.04	46/1115= 0.04
<i>vender</i>	0/2= 0	0/4= 0	0/23= 0	0/23= 0
<i>vender</i> (imp.)	0/2= 0	<b>1/1= 1</b>	0/2= 0	0/2= 0
<i>ver</i>	10/89= 0.11	<b>46/204= 0.23</b>	0/155= 0	0/155= 0
<i>ver</i> (imp.)	4/50= 0.08	<b>8/15= 0.53</b>	0/50= 0	0/50= 0
<i>vivir</i>	11/30= 0.37	<b>12/15= 0.8</b>	16/172= 0.09	19/172= 0.11
<i>vivir</i> (imp.)	10/50= 0.2	<b>4/7= 0.57</b>	2/50= 0.04	4/50= 0.08
<i>volver</i>	<b>2/13= 0.15</b>	0/6= 0	0/20= 0	1/20= 0.05
<b>Overall Bigram Frequency</b>	950	<b>1006</b>	151	190
<b>Average Transitional Probability</b>	0.27	<b>0.5</b>	0.03	0.03

Unless otherwise indicated, all verb forms are in the present tense. Due to syncretism and homophony (e.g.; *yo/él vivía* ‘I, s/he, you formal lived’, *juego* ‘game’, *juego* ‘I play’), some of these values are estimates.