

Variable Subject Position in Main and Subordinate Clauses in Spanish: A Usage-Based Approach

Javier RIVAS
University of Colorado, Boulder

ABSTRACT: The purpose of this study is to determine whether subject position in main and subordinate clauses in Spanish is conditioned by the same factors. To achieve this goal, we conduct quantitative, multiple regression analyses using oral corpora of Peninsular Spanish. We restrict our analysis of subordinate clauses to time clauses. Our results show that the number of postverbal subjects is significantly higher in time clauses than in main clauses and that, in contrast to main clauses, pragmatic factors do not significantly constrain subject position in time clauses. We argue that these results may be attributed to the different functions that subjects play in main and time clauses. In main clauses, subjects are starting points (Chafe 1994) and have primary or secondary referential importance for the subject matter of the conversation. In contrast, the function of subjects in time clauses is to help the speaker anchor the event expressed by the main clause by indicating the performer of the event with which the main clause is temporally linked. We also suggest that, in line with typological findings, the preponderance of postverbal subjects in time clauses may be the result of the more conservative character of subordinate clauses as compared to main clauses.

KEYWORDS: word order, subject position, main clauses, time clauses.

1. INTRODUCTION AND BACKGROUND

Although there is a tendency in language for main and subordinate clauses to conform to the same word order (Givón 2001: 246), it has also been shown that in some languages main clauses present a different word order from subordinate clauses. For example, in some Germanic languages such as Dutch, German, and Old English (Mallinson & Blake 1981: 129) the verb occurs in final position in subordinate clauses, whereas in main clauses the verb appears in second position. In the same line, Givón (2001: 246) also points out that in Kru main clauses follow a S(ubject)-V(erb)-O(bject) order, whereas some types of subordinate clauses are SOV. The reason for the mismatch regarding constituent order found in these languages is that subordinate

clauses are generally more conservative than main clauses (Givón 2001; Bybee 2002; Hinterhölzl & Petrova 2010) regarding processes of word order change.¹

In typological studies (e.g., Givón 2001), Spanish is described as a VO language with flexible position for subjects. This is shown in examples (1) and (2), in which the subject respectively occurs in preverbal and postverbal position:

- (1) COREC <cinta 007, ACON007A.ASC>
 Y *tu marido* tiene un poco de mal genio SVO
 ‘And your husband is a little moody’
- (2) COREC, <cinta 006, ACON006D.ASC>
 Habían ocupado *los hippies* la cala aquella VSO
 ‘The hippies had occupied that cove’

Previous studies (Hatcher 1956; Bentivoglio & Weber 1986; Bentivoglio 1988; Ocampo 1990, 1995; López Meirama 1997, 2006; Morales 2006; Rivas 2008; Sánchez 2008) have shown that the position of the subject in Spanish is pragmatically determined. Factors that have been shown to condition the position of the subject in this language include activation state (whether the subject conveys given or new information), identifiability (whether the hearer may or may not recognize the referent of the subject), agentivity (whether the subject is the agent or the patient of the verb), verb type, pragmatic functions (e.g., focus of contrast, counter-expectation), and textual functions such as change in a locational or temporal scene and conclusion of an episode. Some of these studies (Hatcher 1956, Bentivoglio & Weber 1986, Bentivoglio 1988, Morales 2006, and Sánchez 2008) consider both main and subordinate clauses in their analyses. However, as is noted by Ocampo (1995: 426), “it has not been established for Spanish if main and subordinate clauses share the same word order distinctions.”

In fact, in the literature there is some evidence that suggests that at least some types of subordinate clauses tend to occur with a postverbal subject. Morales (1982), for example, points out that lexical subjects occur postverbally in *según* ‘according to’-clauses, as is shown in example (3). In this example, the subject *la historia* ‘history’ appears after the verb *cuenta* ‘says’ in the *según*-clause:

- (3) COREC <cinta 005, ccon005b.asc>
 [Según cuenta *la historia*] <ininteligible> poblados sí que hubo
 ‘According to what history says, there were indeed settlements’

In this same line, Fernández Ramírez (1986: 435-6) and Butt & Benjamin (2004: 538) indicate that in relative clauses the subject also tends to occur postverbally. Consider example (4):

- (4) Butt & Benjamin (2004: 539)
 Son innumerables las dificultades [que plantea *la lucha contra el terrorismo*]
 ‘The difficulties posed by the struggle against terrorism are innumerable’

¹ In fact, as Bybee (2002) shows, the conservatism of subordinate clauses over main clauses not only affects word order but also other morphosyntactic changes.

In example (4), the subject of the relative clause is *la lucha contra el terrorismo* ‘the struggle against terrorism’ and it appears after the verb *plantea* ‘pose’.

Previous quantitative analyses also corroborate this tendency. Deguchi (1983), who bases his results on a corpus of written Spanish taken from the magazine *Cambio 16*, indicates that the verb occurs more frequently in initial position in subordinate clauses than in main clauses, especially when the subordinate clause is a content or an adverbial clause (Deguchi 1983: 58-9). Similarly, although Hallebeck (2002), using data from a corpus of literary texts, points out that in subordinate clauses he finds similar tendencies to the ones found in main clauses regarding subject position, he also notes that in subordinate clauses “quedan algunos casos de postposición que no tienen una explicación clara e inequívoca [there remain some cases of postposition that cannot be explained clearly and unmistakably].”

The purpose of this study is to address the issue of whether the position of the subject in Spanish is conditioned by the same factors in subordinate clauses as in main clauses. To achieve this goal, we conduct quantitative analyses of both main and subordinate clauses using corpora of spoken Peninsular Spanish. Our study focuses on one specific type of a subordinate clause: time clauses. We discuss the data and methods of these analyses in § 2. In § 3, we present the results of our quantitative analyses. We show that the number of postverbal subjects in subordinate clauses of this type is significantly higher than the number of postverbal subjects in main clauses, and that, in contrast to main clauses, pragmatic factors do not significantly constrain subject position for subordinate clauses. We also sustain that these results may be attributed to the function of time clauses in discourse. Finally, we present our conclusions and suggestions for future studies in § 4.

2. DATA AND METHODS

In order to determine if subject position in Spanish is conditioned by the same factors in both main and subordinate clauses, we use two corpora of spoken Spanish. One of the corpora is the Oral Peninsular Spanish: Conversation section of *Corpus oral de referencia del español contemporáneo (COREC)*. This corpus (Marcos Marín 1994) was collected between 1991 and 1992, and consists of casual face-to-face conversations, phone conversations, and conversations taken from radio and TV programs recorded mainly in Madrid and Segovia (Spain). It has approximately 241,000 words. The other corpus is the Habla culta: Madrid section of *Corpus del español* (Davies 2002). This corpus consists of conversations between the researcher and one or two informants, free conversations between two informants, secret recordings of casual conversations, and formal language taken from, among others, lectures, classes and speeches (DeMello 1991: 446 n. 6). These conversations were recorded between the end of the 60s and the beginning of the 80s. The total number of words in this corpus amounts to approximately 140,000.

We extract all the examples of time clauses (N= 725) from both of these corpora. We restrict our analysis to time clauses in order to make our sample as homogeneous as possible, since it has not been established that all types of subordinate clauses behave in the same way regarding subject position. In order to extract time clauses, we search for time conjunctions (*cuando* ‘when’, *en cuanto* ‘as soon as’, *después de que* ‘after’, *antes de que* ‘before’, *mientras* ‘while’ and *una vez que* ‘once’). Of the initial 725 examples found, 243 (33%) occur with an overt subject. From these, we extract those examples in which *cuando* ‘when’ and *mientras* ‘while’ are not time conjunctions (N=13).² The remaining 230 examples constitute the data on which we base our analysis of time clauses. In addition, we also extract the first 800 examples of main clauses from the conversation section of *COREC* in order to compare the position of the subject in both clause types. Two types of clauses that typically occur without an overt subject, (clauses expressing a direct command through the imperative or subjunctive, and clauses with unipersonal verbs expressing weather) are excluded from the analysis at hand. Interrogative clauses are not included either because in clauses of this type subject position is grammatically determined. *Wh*-questions, for example, obligatorily trigger subject-verb inversion (D’Introno 2001: 144). Of the initial 800 examples, 300 (38%) occur with an overt subject. These are the data on which we base our analysis of main clauses.

Each clause with an overt subject is coded for the following variables:

a) Position of the subject: we distinguish between preverbal and postverbal positions.

b) Animacy of the subject: we use this variable to operationalize agentivity. Previous studies (Hatcher 1956; López Meirama 1997, 2006; Alonso-Cortés 2001) argue that agentive subjects tend to occur in preverbal position, whereas patientive subjects typically occur in postverbal position. As noted by Fox (1995: 159), there is a strong correlation between agentivity and animacy. 97% of the examples of agentive subjects found in her corpus of conversational English are also human. Therefore, according to their referent, we distinguish between human and non-human subjects.

c) Verb type: Bentivoglio & Weber (1986: 30) and Ocampo (1990: 96, 1995: 435) show that postverbal subjects are preferred with existential/presentative verbs such as *haber* ‘there (be)’, *estar* ‘be’ and *existir* ‘exist’. Bentivoglio & Weber (1986: 33) also point out that the subject occurs in postverbal position with movement verbs (e.g., *venir* ‘come’ and *llegar* ‘arrive’) and with *flip verbs*, that is to say, verbs such as *gustar* ‘like’, in which the experiencer occurs in indirect object function and the theme occurs as the subject (*me gustan las películas de terror* ‘I like horror movies’). All other types of verbs favor preverbal subjects. Considering these results, we code each verb as presentative, movement or other. We do not establish a category for *gustar*-type verbs due to the low number of tokens found in the corpus.

² An example is the following: *Pero me acordé de cuándo ésta era pequeña y dije: “Ah, pues es verda<(d)> que los niños hacen esas cosas”* ‘but I remembered when she was little and said: “little children certainly do those things”’ [COREC, <cinta 011, ccon011a.asc>].

d) Activation state: activation state is concerned with what the speaker assumes is in the hearer's focus of consciousness. Chafe (1987: 26-31, 1994: 72) distinguishes between active concepts (also called *given information*) and inactive concepts or *new information*.³ Noun phrases convey given information when they are active in the hearer's mind because they have been previously mentioned in the discourse or because their referent is active in the communicative situation (e.g., *yo* 'I' and *tú* 'you'). In excerpt (5), *Teresa* in line 5 conveys given information because it has been mentioned in the preceding discourse (line 2):

- (5) COREC <cinta 023 ACON023A.ASC>
1 <H1> <fático=duda> lo vamos a preguntar, si se puede. Si se puede, pues en junio
2 pondremos a Teresa...
3 <H3> Claro.
4 <H1> En te<palabra cortada>... y luego ya te ponemos a ti.
5 <H3> Y si no, pues cuando se vaya *Teresa* yo uso su pase.
'<H1> <phatic=hesitation> we'll ask if it is possible. If it is possible, we will sign
Teresa up for June...
<H3> Sure.
<H1> In te <broken word>... and then we will sign you up.
<H3> Or when Teresa leaves I can use her pass'

Noun phrases convey new information when they are first introduced in the discourse, and therefore were inactive in the hearer's focus of consciousness. This is exemplified in (6) with *Michele* in line 2, who has never been mentioned in the previous discourse:

- (6) COREC <cinta 006, ACON006D.ASC>
1 <H3> Es que yo creo que en todos sitios las coca-colas son grandes. Aquí las pe-
2 queñitas esas... cuando vino *Michele* se pedía de 2 en 2.
'< H3> I think coke containers are big everywhere. Here, those little ones... when
Michele came, she used to order two at a time'

Previous studies (Bentivoglio & Weber 1986: 26; Bentivoglio 1988: 13; Ocampo 1990: 96, 1995: 435; Morales 1997: 92; Casielles-Suárez 2004: 4; Rivas 2008: 897, *inter alia*) report that given subjects tend to occur in preverbal position, whereas new subjects appear postverbally. We therefore code each subject as given or new.

e) Identifiability: Du Bois (1980: 218) indicates that a noun phrase is identifiable when the addressee can recognize the concept it makes reference to. When this

³ Chafe (1987, 1994) distinguishes a third type of an activation state: semi-active concepts (also called *accessible information*). A subject is accessible if it was mentioned at an earlier point in the discourse (e.g., in the previous paragraph) but it has not been mentioned for some time. A subject may also be accessible if it belongs to a schema that has been evoked in the previous discourse (e.g., the schema of a class makes accessible noun phrases as *the teacher*, *the classroom*, *students*, among others) or if it is associated with the non-linguistic context in which the conversation is taking place. Since, as Chafe (1994: 75) points out, accessible and new noun phrases tend to be expressed in the same way, in this study we will classify subjects conveying both types of information as *new*.

recognition is not possible, the noun phrase is characterized as non-identifiable. An example of an identifiable noun phrase can be seen in excerpt (7) with *la niña* (lit. 'the girl'), which refers to the speaker's daughter:

- (7) Habla culta: Madrid
cuando me vino *la niña* se me vi... se me vino el mundo encima, pero totalmente.
'when my girl was born, the world came completely crashing down'

A non-identifiable noun phrase is *una cosa* 'one thing' in (8):

- (8) Habla culta: Madrid
1 Entonces, cuando *una cosa* te llena mucho, e... lo demás se te queda... como apar-
2 te, no?
'Then, when something is really fulfilling to you e... everything else falls into the background, right?'

As Du Bois (1980: 216) notes, the contrast between identifiable and non-identifiable referents is neutralized in generics. An example of a generic noun phrase in subject function is *una persona* 'a person' in excerpt (9):

- (9) Habla culta: Madrid
Ahora realmente choca cuando *una persona* te habla de usted.
'Now it is really surprising when a person addresses you using *usted*'

Bentivoglio & Weber (1986: 26) and Rivas (2008: 906) report that most preverbal subjects are identifiable, whereas non-identifiable subjects tend to occur in postverbal position. As a result, we code each subject as identifiable, non-identifiable or generic.⁴

f) Syntactic construction: As indicated by Du Bois (1987: 827, 2003: 34), the preferred clause structure found in discourse typically avoids new information in transitive subject function. This tendency is corroborated for Spanish by Ashby & Bentivoglio (1993: 70). Since, as noted above, postverbal subjects typically convey new information, we predict that most transitive subjects will occur in preverbal position. In fact, this tendency has already been reported in previous studies. Using a corpus of Mexican-American Spanish, Silva-Corvalán (1982) shows that whereas intransitive subjects present similar rates in both preverbal and postverbal positions, transitive subjects tend to occur preverbally. In this same line, Bentivoglio & D'Introno (1989: 53) report that in their corpus of Caracas (Venezuela) the subject occurs before the verb in almost 80% of the transitive clauses they analyze. In addition, Morales (2006: 494) shows that regardless of whether the verb is transitive or intransitive, the presence of an adverbial or prepositional phrase complement in postverbal position generally correlates with a preverbal subject. Therefore, we code each clause as tran-

⁴ The factor groups *activation state* and *identifiability* are closely related. In fact, Thompson (1997: 68) points out that, whereas new noun phrases may be identifiable or non-identifiable, given noun phrases necessarily have to be identifiable. However, there are some contexts in discourse in which a speaker may choose to present a noun phrase as non-identifiable even if it is given, that is to say, even if it has been mentioned in the preceding discourse. One of these contexts is in climatic points of the narration (Rivas 2008: 907).

sitive if the clause has a direct object, intransitive with two or more lexical constituents, and intransitive with one lexical constituent.

h) Type of subject: As Du Bois (1987: 829) notes, in discourse there is a strong correlation between lexical and new arguments. In other words, new arguments are generally encoded by means of lexical noun phrases.⁵ In contrast, pronouns typically convey given information. Since activation state has been shown to play a role in subject position in Spanish, we code each subject as pronominal or lexical. We predict that, being generally given, pronominal subjects will favor preverbal position.

3. RESULTS AND ANALYSIS

In this section we will outline the results of our quantitative analyses. These results will allow us to answer the main research question of this study: is subject position conditioned by the same factors in both main and subordinate clauses? Table 1 shows the number and percentages of preverbal and postverbal subjects in the main and time clauses we analyze for this study. In 68% of the main clauses the subject appears before the verb. Therefore, preverbal position is the preferred position for subjects in main clauses. In contrast, there is a slight preference for postverbal subjects in time clauses. As can be seen in Table 2, 52% of time clauses take a subject in postverbal position. In fact, the percentage of postverbal subjects in subordinate clauses is significantly higher ($p = 0.0000$, *Chi-square* 23.02962) than the percentage of postverbal subjects in main clauses.

Table 1: Subject position in main and time clauses

	MAIN CLAUSES		TIME CLAUSES	
	N	%	N	%
preverbal	203	68	108	47
postverbal	97	32	122	53
total	300	100	230	100

There is some evidence in the literature that suggests that postverbal subjects were more frequent in earlier stages of the language than at present. Myhill (1985: 194), who bases his analysis on 2000 consecutive clauses taken from *Don Quixote* (17th century Spanish), reports a higher percentage of postverbal subjects than in present-day Spanish: 44%.⁶ In the same line, Gordillo & Hamilton (2012) indicate that in his corpus of spoken New Mexican Spanish, generally regarded as a conservative variety (Bills 1997; Bills & Vigil 2008), 52% of lexical subjects occur in postverbal position. This evidence tentatively suggests that time clauses are more conservative than subordinate clauses as far as subject position is concerned. This fact ties in with the

⁵ The converse, however, is not true. That is to say, lexical arguments can also convey given information.

⁶ Similarly, using data taken from *Don Quixote*, Givón (2001: 275) also reports that the percentage of postverbal subjects is higher (53%) than in present-day Spanish. However, he bases his results on only 73 examples.

typological tendency, described above, for subordinate clauses to be more conservative than main clauses (Givón 2001; Bybee 2003).

In order to determine if subject position is conditioned by the same factors in both main and subordinate (time) clauses, we submit our data to quantitative analyses using Varbrul (Rand & Sankoff 1990). This statistical program is used to identify which of the factor groups described in the previous section (animacy of the subject, verb type, activation state, identifiability, syntactic construction, and type of subject) makes up a statistically significant effect, determined by a p value of 0.05 or less, for the occurrence of a postverbal subject, regardless of the corpus being used for the analysis. Varbrul also shows the relative strength of each factor group by means of the range. The higher the range, the greater the magnitude of effect of the factor group on the occurrence of postverbal subjects. Finally, within each factor group, the individual factors are ranked according to their factor weights using values within the interval 0-1. If a factor weighs more than 0.5, it favors subject postposition, whereas a factor weighing less than 0.5 disfavors it. Table 2 presents the linguistic factors that Varbrul selects as significant for main clauses.

Table 2: Linguistic factors favoring postverbal position in main clauses

Input:	.300		
Total N:	300		
	% postverbal	Factor Weight	% Data
Verb type			
Presentative	96	.90	16.7
Movement	50	.47	7.3
Other	16.7	.39	76
	Range	51	
Syntactic construction			
Intransitive with 1 constituent	69.5	.77	27.3
Intransitive with 2 or more constituents	28.2	.44	34.3
Transitive	9.6	.34	38.3
	Range	43	
Identifiability			
Non-identifiable	93.2	.85	14.7
Generic	45.7	.47	15.3
Identifiable	16.7	.42	70
	Range	43	
Activation state			
New	75	.76	32
Given	12.3	.37	68
	Range	39	

Log likelihood = -92.419, Chi-square per cell = 12221, $p=0.017$

As can be seen in this table, of all the factor groups analyzed, Varbrul selects only 4 as being significant for main clauses: verb type, syntactic construction, identi-

fiability, and activation state. The factor group of verb type has the greatest magnitude of effect on subject position. Presentative verbs heavily favor postverbal subjects with a factor weight of .90, whereas the “other” category disfavors them (factor weight .39). Since the category of movement verbs has a factor weight of .47 and this weight is close to .5, we can conclude that they neither favor nor disfavor subjects in postverbal position. The following two factor groups that Varbrul selects as significant, syntactic construction and identifiability, have the same magnitude of effect. Within the factor group of syntactic construction, intransitive clauses with only one constituent highly favor postverbal subjects with a factor weight of .77, whereas intransitive clauses with 2 or more constituents and transitive clauses disfavor subjects in this position (factor weights .44 and .34 respectively). Within the factor group of identifiability, postverbal position is highly favored by non-identifiable subjects (factor weight .85), and disfavored by identifiable subjects with a factor weight of .42. As was the case with movement verbs above, generic subjects also have a factor weight so close to .5 (.47) that we can say they neither favor nor disfavor postverbal position. Finally, activation state is the last factor group that significantly constrains subject position in main clauses. New subjects have a factor weight of .76, and therefore, favor postverbal position, whereas given subjects disfavor it with a factor weight of .37.

Table 3: Linguistic factors favoring postverbal position in time clauses

Input:	.565		
Total N:	230		
	% postverbal	Factor Weight	% Data
Syntactic construction			
Intransitive with 1 constituent	80.2	.73	48.3
Intransitive with 2 or more constituents	38.2	.40	29.6
Transitive	13.7	.17	22.2
	Range	56	
Verb type			
Presentative	96.2	.83	11.3
Movement	65.5	.58	23.9
Other	40.9	.40	64.8
	Range	43	
Type of subject			
Lexical	75.7	.67	59.1
Pronominal	20.2	.26	40.9
	Range	41	
Animacy			
Non-human	85.7	.70	30.4
Human	38.8	.40	69.6
	Range	30	

Log likelihood = -90.584, Chi-square per cell = 1.3580, p=0.038

Table 3 provides the results of the Varbrul run for time clauses. This statistical program selects the following factor groups as significant: syntactic construction, verb type, type of subject, and animacy. The factor group with the greatest magnitude of effect on subject position is syntactic construction. Intransitive clauses with 1 constituent favor the occurrence of subjects after the verb (factor weight .73), whereas both intransitive clauses with 2 or more constituents and transitive clauses disfavor it with factor weights of .40 and .17 respectively. The next factor group that significantly constrains subject position in time clauses is verb type. Presentative and movement verbs favor subjects in postverbal position (factor weights .83 and .58 respectively), whereas other types disfavor it (factor weight .40). Type of subject is another factor group which Varbrul selects as significant. Postverbal position is favored by lexical subjects (factor weight .67) and disfavored by pronominal subjects, with a factor weight of .26. Finally, animacy is the last factor group that significantly conditions subject position in time clauses. Non-human subjects have a factor weight of .70, and hence favor postposition, whereas human subjects disfavor it (factor weight .40).

In both main and subordinate clauses, then, subject position is constrained by verb type and syntactic construction and, as can be seen in Tables 2 and 3, both factor groups present similar tendencies in both types of clauses. Presentative verbs (*haber* 'there (be)', *estar* 'to be') highly favor postverbal subjects. These results tie in with findings of previous research. Both Bentivoglio & Weber (1986: 30) and Ocampo (1995: 428) show that presentative verbs occur with postverbal subjects regardless of whether they convey given or new information. Other types of verbs disfavor postposition with the exception of movement verbs. Whereas in main clauses movement verbs neither favor nor disfavor postposition, they slightly favor it in time clauses. As is shown in Sánchez (2006), movement verbs are also part of the presentative lexicon in Spanish. Therefore, with verbs of this type, there is also a tendency to place their subject in postverbal position, although this tendency is not as strong as with presentative verbs.

The other factor group that significantly constrains subject position in both main and subordinate clauses is syntactic construction. In both types of clauses, the sole syntactic construction that favors postverbal subjects is intransitive clauses with one participant. Therefore, these results corroborate Morales's (2006) analysis that transitive clauses and intransitive clauses with two or more constituents behave similarly regarding subject position.⁷ Both of these types disfavor postposition in main

⁷ In fact, Morales (2006: 494) notes that in intransitive clauses with an adverbial or a prepositional phrase complement, the position of this complement plays a crucial role in the determining subject position. Adverbial or prepositional complements in preverbal position correlate with postverbal subjects, whereas adverbial or prepositional complements in postverbal position typically occur with preverbal subjects. An example of the former is (Morales 2006: 493) [...] *en el mismo avión llegaba Luis* 'Luis was arriving in that same plane' and an example of the latter is *los jóvenes de hoy se encuentran con una civilización mil veces más abierta* 'young people nowadays find a much more open-minded civilization'. In our corpus, we only find 10 examples of an intransitive verb with an adverbial or prepositional phrase complement in preverbal position (8 in main clauses and 2 in time clauses). In all of these cases, the subject does indeed occur after the verb. In all the other examples

and time clauses. However, as can be seen in Tables 2 and 3, the factor weight of transitive clauses is lower than the factor weight of intransitive clauses with two or more constituents. To certain extent, then, these results also corroborate previous findings (e.g., Silva-Corvalán 1982, Bentivoglio & D’Introno 1989) that suggest transitive clauses tend to occur with preverbal subjects.

The other factor groups Varbrul selects as significant on subject position differ in main and subordinate clauses. In main clauses, subject position is constrained by two additional factors: identifiability and activation state. Both of these factors are pragmatic. Again, these results corroborate the results found in previous studies as mentioned above. Preverbal position is favored by identifiable and given subjects, whereas non-identifiable and new subjects tend to occur postverbally. These tendencies can be attributed to cognitive reasons (Ocampo 1995: 444). Non-identifiable and new information is harder to process and hence it will tend to occur later in the clause.

In contrast, in time clauses the other two factor groups that significantly constrain subject position are type of subject and animacy. Pronominal subjects favor preverbal position, whereas lexical subjects tend to occur in postverbal position. These results tie in with the end-weight principle (Quirk, Greenbaum, Leech and Svartvik 1985: §18.9), which establishes that the heavier or more grammatically complex a constituent is, the later it will occur in the clause. Lexical NPs are heavier elements than pronouns, and therefore the former are more likely to occur after the verb than the latter. Consider in this respect the following examples:

- (10) COREC <cinta 021, ccon021b.asc>
1 Después ya cuando *yo* me casé ya, ya había, ya comprábamos de eso pero antes
2 no
‘Afterwards, when I got married, there was already... we already bought that type but not before’
- (11) COREC <cinta 019, ccon019a.asc>
Cuando se casó *tío Antonio*, pues ya... Ya nada. Se acabó...
‘When Uncle Antonio got married, then, nothing. It was over....’

The subject of the time clause in these examples has similar characteristics: it is human, given, identifiable and it occurs in an intransitive clause with only one constituent. We also have the same verb in both clauses: *casarse* ‘get married’. However, in (10) the subject occurs in preverbal position because it is a pronoun (*yo* ‘I’) and pronouns are light elements, whereas in (11) it occurs in postverbal position because it is a lexical noun phrase (*tío Antonio* ‘Uncle Antonio’) and therefore, it is heavier than the pronominal subject in (10).

The other factor group that Varbrul selects as significant for time clauses is animacy. Non-human subjects favor postverbal position, whereas human subjects tend

we have of intransitive clauses with one or more non-subject constituents (103 in main clauses and 68 in subordinate clauses), they occur in postverbal position. Therefore, in the corpus, intransitive clauses with 2 or more constituents disfavor postverbal subjects because these additional constituents occur in postverbal position in 94% of cases.

to occur preverbally. Previous studies (Hatcher 1956; López Meirama 1997, 2006; Alonso-Cortés 2001) have also found this same tendency in their corpora, although some of them (e.g., López Meirama 1997) include only main clauses in their analysis. However, our statistical analyses do not select animacy as a significant factor in main clauses. As we will show below, the reason why animacy plays a significant role in subject position in time clauses is related to the function that time clauses play in discourse.

The most important difference between subject position in main and time clauses as demonstrated by these results is that pragmatic factors do not seem to significantly constrain subject position in time clauses, whereas they play an important role in determining subject position in main clauses. The reason for this difference is that time clauses, and subordinate clauses in general, play a different role in discourse from main clauses. In discourse there are two main pragmatic categories: foreground and background. Foregrounded clauses constitute the skeleton of the text and they are ordered temporally. In contrast, backgrounded clauses supply the material that extends or comments upon the core of the text (Hopper & Thompson 1980: 280-281). As has been shown in previous studies (Thompson 1987; Bybee 2002), there is a strong correlation between foregrounding and main clauses on the one hand, and backgrounding and subordinate clauses on the other. In this respect, the role of time clauses is to make information about time relevant to the hearer. That is to say, the subordinate clause anchors the event that is expressed in the main clause by establishing a temporal link such as precedence, subsequence, or simultaneity with the main clause. One way of providing this anchoring is by means of the subordinating conjunction (Givón 2001: 330):

(12) Precedence

COREC <cinta 006, PCON006A.ASC>
 si vuelves *antes de que* yo vaya para allá espérame
 ‘if you come back before I leave for there, wait for me’

(13) Subsequence

COREC <cinta 013, ccon013g.asc>
 se tenían que quedar *después de que* terminara veinte minutos o media hora
 ‘they had to stay for 20 minutes or half an hour after it finished’

(14) Simultaneity

Habla culta: Madrid

- 1 Nosotros íbamos a la sierra. ¡Bueno!, de pequeños, a la sierra, a... De pequeños,
- 2 a Valencia, *mientras* vivió mi abuela.
 ‘We used to go to the mountains, well, when we were little, to the mountains, to...
 when we were little (we used to go) to Valencia, while my grandmother was alive’

Another way of providing anchoring between the time and the main clause is by means of the tense/aspect/mood of the verb, especially with *cuando* ‘when’, which is the more general / frequent subordinating conjunction:⁸

⁸ Diessel (2008), using a corpus of British English, shows that the position of time clauses is iconic, that is to say, it mirrors the sequential ordering of the events. If the event expressed by the

- (15) Precedence
Habla culta: Madrid
Cuando nosotros ya nos dimos cuenta, ya mi padre *se había retirado* de la Marina
'When we came to realize it, my father had already retired from the navy'
- (16) Subsequence
COREC <cinta 018, ccon018d.asc>
Cuando yo me compré ese tarro de naranja me *costó* 325
'When I bought that orange jar, it cost me 325'
- (17) Simultaneity
COREC <cinta 006, ACON006A.ASC>
Cuando yo leía esto, *costaba* quinientas pelas
'When I read this, it cost 500 pesetas'

The anchoring function of time clauses also determines the nature of the subject that appears in these constructions. Chafe (1994) maintains that subjects in general can be regarded as the *starting point* of the clause. Each clause has "a referent from which it moves on to provide its own new contribution" (Chafe 1994: 83). Subjects are the result of the grammaticalization of this starting point. Starting points are usually given (Chafe 1994: 85) and they typically have primary or secondary referential importance to the subject matter of the conversation, depending on their frequency of mention. The subject of time clauses, however, is not a starting point. Regardless of its activation state status and degree of referential importance, the subject of time clauses occurs when it allows the speaker to help temporally anchor the event expressed by the main clause, by indicating the performer of the event expressed by the time clause. This is the reason why pragmatic factors do not significantly constrain subject position in time clauses. As Bybee (2002: 4-5) points out, functions such as topicalization or presenting new noun phrases are typical of main clauses, subordinate clauses being very rarely used for these functions.

Chafe (1994) points out that starting points are generally given information. As our results for main clauses show, given subjects tend to occur in preverbal position in Spanish. This fact suggests that preverbal position is the preferred position for starting points in this language. Since subjects of time clauses are not starting points, they will tend to occur in postverbal position. Grammatical factors, however, may revert this tendency. As can be seen in Table 3 and in the discussion of this table, transitive and intransitive clauses with 2 or more constituents, clauses with verbs not belonging to the presentative and movement types, and pronouns will favor preverbal position for

time clause occurs prior to the event expressed by the main clause, the time clause will precede the main clause. In contrast, if the event expressed by the time clause occurs after the event of the main clause, the time clause will occur after the main clause. This tendency is also observed in our corpus: when the event expressed by the time clause takes place before the event expressed by the main clause, the time clause occurs before the verb in 85% of cases. In order to find out if this phenomenon had any impact on subject position, using Varbrul, we conducted a quantitative analysis of all time clauses starting with *cuando* 'when' according to whether they expressed an event that occurred prior to, simultaneously, or after the event expressed by the main clause. We also coded for position of time clause (before or after the main clause). However, Varbrul did not select either of the two factor groups as significant for determining subject position.

subjects in time clauses. The fact that subjects of time clauses are not starting points also accounts for the results regarding animacy. As has been shown in typological studies (Comrie 1981: 1999; Mallinson & Blake 1981: 86; Givón 2001: 200), human noun phrases are more likely to be starting points than non-human noun phrases.⁹ Therefore, as Table 3 shows, in time clauses human subjects will tend to occur in preverbal position, because they share a characteristic with starting points, which typically occur preverbally in Spanish. In contrast, non-human subjects will prefer the postverbal position.

4. CONCLUSIONS

Although it has often been implicitly assumed in previous studies that subject position in Spanish is conditioned by the same factors in both main and subordinate clauses, the quantitative and multiple regression analyses we present in this study reveal that there are important differences between them. In order to keep our data as homogeneous as possible, we focus our analysis on both main clauses and only one specific type of a subordinate clause, namely, time clauses. Our results show that the number of postverbal subjects in time clauses is significantly higher than the number of postverbal subjects in main clauses. The multiple regression analyses also show that pragmatic factors do not significantly constrain subject position in time clauses, even though they play an important role in determining the position of the subject in main clauses.

The reason for these discrepancies in the results lies in the different functions that subjects play in both main and subordinate clauses. In main clauses, subjects are starting points; they typically convey given information and have primary or secondary importance to the subject matter of the conversation. Therefore, pragmatic factors such as activation cost and identifiability significantly determine the position of subjects in main clauses. Identifiable and given subjects occur preverbally, whereas non-identifiable and new subjects typically occur in postverbal position. In contrast, in time clauses the function of subjects is to help the speaker anchor the event expressed by the main clause by indicating the performer of the event with which the main clause is temporally linked. For this reason, pragmatic factors do not significantly constrain subject position in time clauses. Since starting points typically occur in preverbal position in Spanish, the subject of time clauses, not being a starting point, will tend to occur postverbally. There are, however, some grammatical factors that may lead to the occurrence of the subject in preverbal position in time clauses. Some of these grammatical factors, namely, syntactic construction and verb type, also significantly constrain subject position in main clauses in similar ways. Transitive clauses and intransitive clauses with two or more postverbal participants as well as presenta-

⁹ Instead of *starting point*, in these studies they use the term *topic*. Chafe (1994: 84), however, considers that the term *topic* should only be used to describe certain grammatical constructions found in some Asian languages.

tive verbs typically occur with a preverbal subject in both clause types. Similarly, postverbal subjects are favored in intransitive constructions with one participant and with non-presentative verbs. Movement verbs also favor postverbal subjects in time clauses, whereas they neither favor nor disfavor postposition in main clauses. In addition, in time clauses pronominal subjects typically occur preverbally, whereas lexical subjects tend to occur postverbally. These tendencies tie in with the end-weight principle, which establishes that heavy elements (e.g., lexical subjects) will occur later in the clause than light elements (e.g., pronominal subjects). Our results also show that human subjects favor preverbal position in time clauses, whereas non-human subjects disfavor it. The reason is that human subjects are better candidates for starting points, and therefore tend to occupy the position starting points usually take. Non-human subjects, on the other hand, will typically occur in postverbal position. Future research should determine if the tendencies we have found for subject position in time clauses are also applicable to other types of subordinate clauses such as content clauses, relative clauses or other so-called adverbial clauses such as conditional, reason, purpose, and concessive.

The results of our statistical analyses show that, with regard to subject position, main and subordinate clauses in Spanish do not behave in the same way. Main clauses favor preverbal subjects, whereas time clauses favor postverbal subjects. In typological studies it has been shown that some languages present different word orders in main and subordinate clauses. The reason for this fact is that subordinate clauses tend to be more conservative than main clauses. In this respect, there is some evidence in previous studies (Myhill 1986, Givón 2001, Gordillo & Hamilton 2012) that suggests that postverbal subjects are also more common in earlier stages of the Spanish language as well as in conservative varieties such as New Mexican Spanish. We can, therefore, tentatively conclude that time clauses are more conservative than main clauses as far as subject position is concerned. It is the task of future research to corroborate or contradict this conclusion.

REFERENCES

- ALONSO-CORTÉS, A. (2001): "Rasgos activos y ergativos del español." *Verba* 28, 187-206.
- ASHBY, W. & P. BENTIVOGLIO (1993): "The preferred argument structure of spoken French and Spanish." *Language Variation and Change* 5, 61-76.
- BENTIVOGLIO, P. & E. WEBER (1986): "A functional approach to subject word order in spoken Spanish." In O. Jaeggli & C. Silva-Corvalán (eds.): *Studies in Romance Linguistics*. Dordrecht: Foris, 23-40.
- BENTIVOGLIO, P. (1988): "La posición del sujeto en el español de Caracas: un análisis de los factores lingüísticos y extralingüísticos." In R. M. Hammond & M. C. Resnick (eds.): *Studies in Caribbean Spanish Dialectology*. Washington DC: Georgetown Univ. Press, 13-23.

- BENTIVOGLIO, P. & F. D'INTRONO (1989): "Orden de palabras y posición del sujeto en el español de Caracas." In I. Páez Urdaneta, F. Fernández & L. Barrera Linares (eds.): *Estudios lingüísticos y filológicos en homenaje a María Teresa Rojas*. Caracas: Universidad Simón Bolívar, 51-61.
- BILLS, G. (1997): "New Mexican Spanish: Demise of the earliest European variety in the United States." *American Speech* 72/2, 154-71.
- BILLS, G. D. & N. A. VIGIL (2008): *The Spanish language of New Mexico and Southern Colorado. A linguistic atlas*. Albuquerque: the University of New Mexico Press.
- BUTT, J. & C. BENJAMIN (1988): *A new reference grammar of modern Spanish*. London: Hodder Education, 2004⁴.
- BYBEE, J. (2002): "Main clauses are innovative, subordinate clauses are conservative." In J. Bybee & M. Noonan (eds.): *Complex clauses in grammar and discourse. Essays in honor of Sandra A. Thompson*. Amsterdam/Philadelphia: John Benjamins, 1-17.
- CASIELLES-SUAREZ, E. (2004): *The syntax-information structure interface. Evidence from Spanish and English*. New York / London: Routledge.
- CHAFE, W. (1987): "Cognitive constraints on information flow." In R. S. Tomlin (ed.): *Coherence and grounding in discourse*. Amsterdam: John Benjamins, 21-51.
- CHAFE, W. (1994): *Discourse, consciousness and time*. Chicago: The Univ. of Chicago Press.
- COMRIE, B. (1981): *Language universals and linguistic typology. Syntax and morphology*. Oxford: Blackwell, 1989².
- D'INTRONO, F. (2001): *Sintaxis generativa del español: evolución y análisis*. Madrid: Cátedra.
- DAVIES, M. (2002-): *Corpus del español: 100 million words, 1200s-1900s*. <<http://www.corpusdelespanol.org>>.
- DEGUCHI, A. (1983): "Aspectos cuantitativos del orden de palabras en el español contemporáneo." *Lingüística Hispánica* 6, 55-66.
- DEMELLO, G. (1991): "Pluralización del verbo 'haber' impersonal en el español hablado culto de once ciudades." *Thesaurus* XLVI/3, 445-71.
- DIESSEL, H. (2008): "Iconicity of sequence: a corpus-based analysis of the positioning of temporal adverbial clauses in English." *Cognitive Linguistics* 19/3, 465-90.
- DU BOIS, J. W. (1980): "Beyond definiteness: The trace of identity in discourse." In W. Chafe (ed.): *The pear stories. Cognitive, cultural and linguistic aspects of narrative production*. New Jersey: Ablex Publishing Corporation, 203-74.
- DU BOIS, J. W. (1987): "The discourse basis of ergativity." *Language* 63/4, 805-55.
- DU BOIS, J. W. (2003): "Argument structure. Grammar in use." In J. W. Du Bois, L. E. Kumf & W. J. Ashby (eds.): *Preferred argument structure. Grammar as architecture for function*. Amsterdam / Philadelphia: John Benjamins, 11-60.
- FERNÁNDEZ RAMÍREZ, S. (1986): *Gramática española*. 4. *El verbo y la oración*. Madrid: Arco/Libros.
- FOX, B. (1995): "The category 'S' in English conversation." In W. Abraham, T. Givón & S. A. Thompson (eds.): *Discourse, grammar and typology. Papers in honor of John W. M. Verhaar*. Amsterdam / Philadelphia: John Benjamins, 153-78.
- GIVÓN, T. (2001): *Syntax. An introduction*. Amsterdam / Philadelphia: John Benjamins.
- GORDILLO, Y. & J. A. HAMILTON (2012): "Word order variation with NP subjects: The case of New Mexican Spanish." Paper presented at 6th International Workshop on Spanish Sociolinguistics, April 12-14, 2012, The University of Arizona, Tucson.
- HALLEBECK, J. (2002): "La posición del sujeto en la proposición subordinada." *Neophilologus* 86, 225-48.
- HATCHER, A. G. (1956): *Theme and underlying question. Two studies of Spanish word order*. *Word* 12, suplement 3. New York: The Linguistic Circle of New York.

- HINTERHÖLZL, R. & S. PETROVA (2010): "Introduction." In R. Hinterhölzl & S. Petrova (eds.): *Information structure and language change. New approaches to word order variation in Germanic*. Berlin: Mouton de Gruyter, 1-13.
- HOPPER, P. & S. A. THOMPSON (1980): "Transitivity in grammar and discourse." *Language* 56/2, 251-99.
- LÓPEZ MEIRAMA, B. (1997): *La posición del sujeto en la cláusula monoactancial en español*. Santiago de Compostela: Univ. de Santiago de Compostela.
- LÓPEZ MEIRAMA, B. (2006): "Semantic and discourse-pragmatic factors in Spanish word order." In J. C. Clements & J. Yoon (eds.): *Functional approaches to Spanish syntax. Lexical semantics, discourse and transitivity*. New York: Palgrave MacMillan, 7-52.
- MALLINSON, G. & B. J. BLAKE (1981): *Language typology. Cross-linguistic studies in syntax*. Amsterdam: North Holland.
- MARCOS MARÍN, F. (1994): *Corpus oral de referencia del español contemporáneo*. www.llf.uam.es/~fmarcos/informes/corpus/corpusix.html
- MORALES, A. (1982): "La perspectiva dinámica oracional en el español de Puerto Rico." In O. Alba (ed.): *El español del Caribe*. Santiago, República Dominicana: Universidad Católica Madre y Maestra, 203-9.
- MORALES, A. (1997): "La manifestación del sujeto en español: tendencias universales y variación dialectal." In R. M. Hammond & M. G. MacDonald (eds.): *Linguistic studies in honor of Bohdan Saciuk*. West Lafayette, IN: Learning Systems Inc. 87-103.
- MORALES, A. (2006): "Los sujetos "ligeros" del español y su posición en la oración." In M. Sedano, A. Bolívar & M. Shiro (eds.): *Haciendo lingüística. Homenaje a Paola Bentivoglio*. Caracas: Univ. Central de Venezuela, 487-501.
- MYHILL, J. (1985): "Pragmatic and categorial correlates of VS word order." *Lingua* 66, 177-200.
- OCAMPO, F. (1990): "The pragmatics of word order in constructions with a verb and a subject." *Hispanic Linguistics* 4/1, 87-128.
- OCAMPO, F. (1995): "The word order of two-constituent constructions in spoken Spanish." In P. Downing & M. Noonan (eds.): *Word order in discourse*. Amsterdam / Philadelphia: John Benjamins, 425-47.
- QUIRK, R., S. GREENBAUM, G. LEECH & J. SVARTVIK (1985): *A comprehensive grammar of the English language*. London: Longman.
- RAND, D. & D. SANKOFF. (2001): *GoldVarb: a variable rule application for Macintosh*.
- RIVAS, J. (2008): "La posición del sujeto en las construcciones monoactanciales del español: Una aproximación funcional." *Hispania* 91/4, 897-912.
- SÁNCHEZ, I. (2006): "The presentative lexicon." ms.
- SÁNCHEZ, M. E. (2008): "Tipos de cláusula, clases verbales y posición del sujeto en español." *Lexis* XXXII/1, 83-105.
- SILVA-CORVALÁN, C. (1982): "Subject expression and placement in spoken Mexican-American Spanish." In J. Amastae & L. Elías-Olivares (eds.): *Spanish in the United States: Sociolinguistic aspects*. Cambridge: Cambridge Univ. Press, 93-120.
- THOMPSON, S. A. (1987): "'Subordination' and narrative event structure." In R. Tomlin (eds.): *Coherence and grounding in discourse*. Amsterdam / Philadelphia: John Benjamins, 435-54.
- THOMPSON, S.A. (1997): "Discourse motivations for the core-oblique distinction as a language universal." In A. Kamio (ed.): *Directions in functional linguistics*. Amsterdam / Philadelphia: John Benjamins, 59-82.