

Paola Bentivoglio and Mercedes Sedano
Universidad Central de Venezuela

THE *LIGHT SUBJECT CONSTRAINT* IN SPOKEN SPANISH

Abstract

The purpose of this study is to explore Chafe's hypothesis (1994) that grammatical subjects obey the light subject constraint (LSC). The data consists of six half-an-hour semiformal interviews with Caracas native speakers recorded in 1987. Chafe's hypothesis is based on the claim that in any given clause one of the referents receives the unique and special status of grammatical subject. Subjects show two types of restrictions: i) from the information load perspective, they tend to code given or accessible rather than new information; and ii) from a referential viewpoint, they can have three degrees of importance: primary, secondary or trivial (1994: 88). Information load (high or low) and referential importance combined constitute the LSC: Chafe's findings prove that in English subjects tend, on the one hand, to be given or accessible and, on the other, of primary or secondary importance. Heavy subjects, on the contrary, tend to be of trivial importance. Broderick 1999 confirms the LSC's existence in a conversational English corpus. All subjects in the corpus have been coded according to information load and referential importance. The results of the analysis show that spoken Spanish also conforms to LSC.

Keywords

Grammatical subject, given / accessible / new information, spoken Spanish.

1. Introduction

In this article, we analyze a corpus of six sample speeches, in order to determine whether the *light subject constraint* (LSC) proposed by Chafe (1994)–based on the analysis of English conversational data–also exists in spoken

Spanish.¹ Chafe argues that the grammatical subject of any given sentence is pragmatically adequate if and only if the information transmitted is either given, accessible, or new. Though the pragmatic characteristics of subjects have been studied by many linguists, empirical research is scarce, especially in spoken Spanish. Thus, validating Chafe's *light subject constraint* will allow us to prove with more certainty and thoroughness its existence in a language other than English.

In world languages, subjects are important from the following points of view: i) syntactically, because—at least in Western languages—the subject controls subject-verb agreement in terms of grammatical person and number;² and ii) semantically and pragmatically, because—as Tomlin (1985) points out—the subject generally relates to both the semantic agent and the pragmatic topic.³ Dik (1978: 143) ratifies the subject-topic association when he writes that “Topic function will frequently be assigned to Subj(ect) constituents.” Givón (1995), studying the properties attributed by Keenan (1976) to subjects in different languages,⁴ states that “the clause's grammatical subject tends to code the current discourse topic at the time when the clause is being processed” (1995: 230), i.e., the most important property is topicality. Even though the subject-topic association is not obligatory—some subjects that are neither agents nor topics may exist—there is a tendency to consider that subjects generally coincide both with agent and topic. It is this tendency that allows to predict certain characteristics of the subject to be proved or disproved by future empirical studies. The main characteristic arising from the mentioned subject-topic association concerns the kind of information transmitted by the subject: if it coincides with the topic, which normally conveys given information (Lambrecht 1987), then, it is likely that the subject also transmits given information. Chafe (1994) goes one step further when he proposes the LSC⁵ which is based on the following notions: i) the relevance of a discourse entity; and ii) the kind of information conveyed by the entity in question.

With respect to relevance, a discourse entity is of primary relevance if its referent is the protagonist of the story. It is of secondary relevance if the referent,

¹ The article is a fragmentary report on a more complete study about grammatical subjects in spoken Spanish.

² Subjectless clauses deserve a separate treatment.

³ We consider topic the referent talked about.

⁴ The properties are the following: independent existence; indispensability; absolute, presupposed or persistent reference; definiteness; topicality and agentivity.

⁵ Chafe (1994) does not associate subject and topic, because he believes that the concept of “topic” should only be used for certain phenomena in Asian languages. According to him, subjects are metaphorically *hitching posts*, i.e., starting points “to which a new contribution is attached.” Independently from his definition of subject, Chafe also says that subjects, because of being hitching posts, do not transmit new information.

in spite of not being the protagonist, is necessary to the development of events. Finally, an entity is of trivial importance if it only fulfils a minor function for which the referent's identity or specific characteristics are irrelevant. The discourse relevance of an entity is normally measured by the number of times this entity is mentioned: entities of primary relevance are proportionally mentioned many more times than others, whereas a trivial entity is scarcely mentioned in a text.

In terms of the type of information conveyed by the subject, Chafe (1994: 72) holds that entities transmit different kinds of information,⁶ depending on what the speaker thinks there is in the addressee's mind at the moment of speaking. Hence, an entity is: i) "new" if the speaker thinks that s/he is introducing—i.e., activating—it in the addressee's consciousness during the conversation; ii) "given" if the speaker believes that it is already active in the addressee's consciousness, either because it has been talked about in the immediate previous discourse, or because its referent is actively present in the conversational situation;⁷ iii) "accessible" if it coincides with one of the three following possibilities: a) the entity was active in some previous stretch of discourse; b) it is directly related to an active or recently activated idea; c) it "is associated with the nonlinguistic environment of the conversation and has for that reason been peripherally active but not directly focused on" (1994: 86).⁸

Chafe (1994) mentions that given information is generally codified by means of elliptical elements, weakly stressed pronouns and, occasionally, weakly stressed nouns, whereas new information is generally codified by lexical noun phrases. He also adds that accessible information normally appears linguistically codified in the same way as new information does.

Chafe's LSC is based on a study carried out on a corpus of about 10,000 words of conversational American English. The data obtained in 1980-81 consists of informal teacher-student conversations, in which personal narratives are predominant. The results indicate that the information transmitted by the entity subject is distributed as follows: 81% of subjects convey given information, generally codified by pronouns; 16% convey accessible information, and only 3% transmit new information. Chafe observes that nominal entities transmitting new information are mentioned only once and so deserve to be considered "trivial."

Another linguist who has ratified the existence of LSC in spoken English is Broderick (1999), whose research is based on a sample of 4,200 words. In contrast

⁶ Literature about new or given information is abundant (see, among others, Chafe 1976 and Prince 1981).

⁷ The author considers references to the speaker and the hearer (in Spanish, *yo* "I" and *tú* "you") to be always given information.

⁸ Chafe (1976) writes that it is difficult to establish when a piece of active (given) information becomes semiactive (accessible), but he supposes that the time is relatively short if the item has not been mentioned for a while in the discourse.

to Chafe's mainly narrative data, Broderick's data are related to "perceiving, acting on, and evaluating what is in the environment of a conversation" (1999: 143). His analysis does not include imperative clauses, elliptical subjects of declarative clauses nor yes/no interrogative clauses. The results indicate that out of 488 subjects, 388 (79,5%) transmit given information,⁹ and 100 (20,5%), accessible information. In his data no subjects conveying new information were found, confirming therewith Chafe's constraint.

2. Corpus and methodology

The present research corpus includes six samples of recorded careful speech (Labov 1972), selected from the "Estudio sociolingüístico del habla de Caracas, 1987."¹⁰ In our sample, speakers were asked to talk about their personal experiences, their childhood memories, school or university years, jobs, sports, etc. The recorded materials are, therefore, more similar to Broderick's than to Chafe's. Two interviewers and the interviewee were present during the recording session. From each sample, we extracted the first 300 subjects, for a total of 1,800; the number of words is approximately 15,000. All 1,800 subjects were codified according to the following variables: i) subject type; ii) lexical subjects' information type; and iii) pragmatic relevance of lexical subjects transmitting new information.

2.1. Subject type

A subject may be: i) elliptical as in (1a); ii) a relative pronoun as in (1b); iii) a non-relative pronoun as in (1c); iv) a headless relative clause as in (1d); v) a nominal clause as in (1e); and, finally, vi) a lexical noun phrase as in (1f):

- (1)
 a. /elliptical subject/
 Ø tuvimos que cambiar de colegio (ca1fa.87)¹¹
 "we had to change school"
 b. /relative pronoun/
 de esas [gripes] *que* están dando últimamente (ca3fd.87)

⁹ Out of the 388 cases of given subjects, 344, i.e., 88,7%, are pronouns.

¹⁰ "Sociolinguistic study of Caracas spoken Spanish, 1987." Details can be found in Bentivoglio and Sedano 1993.

¹¹ The code which appears between parentheses at the end of each example corresponds to the identification of each speaker within the 1987 Caracas corpus.

- “of those [flus] that are hitting lately”
c. /non-relative pronoun/
yo antes sufría de la cabeza (ca5mc.87)
“I used to have headaches before”
d. /headless relative clause/
lo que tiene es un fin político (ca1fa.87)
“what he has is a political purpose”
e. /nominal clause/
a mí me gustaría trabajar por lo menos por dos años (ca2fb.87)
“I’d like to work for at least two years”
f. /lexical noun phrase (lexical subject)/
mi abuela iba allí al seguro (ca5fc.87)
“my grandmother used to go there to the Social Security Service”

It is essential to clarify that we only analyzed subjects codified as lexical noun phrases (lexical subjects), because they may transmit new information, contrary to elliptical and (relative and non-relative as well) pronominal subjects which usually convey only given information.¹² Furthermore, headless relative and nominal clauses do not function as entities in themselves but rather as clauses.

2.2. Lexical subjects’ information type

Following Chafe (1994: 71-76), we codified three information types: given, accessible and new. Examples in (2) illustrate the different kinds of information:

- (2)
a. /given information/
Después me acuerdo de... de mi padrino.
Mi padrino murió cuando yo tenía once años (ca2fb.87)
“Then I remember... my godfather. My godfather died when I was eleven years old”
b. /accessible information/
Y la primera vez que choqué, panita, una vez que choqué... que... que... bueno, yo creía que el tipo me iba a decir de todo y todo eso (ca2fb.87)
“And the first time I had a car accident, pal, once when I hit... well, I thought the man would tell me all kind of things and so on”
c. /new information/
“Estoy perdido de mi casa, ésta no es la zona mía,” todo... entonces... hasta que me dijo un señor: “Vente conmigo, que yo te llevo hasta tu casa” (ca5mc.87)

¹² The only exception could be some indefinite pronouns, as *alguien* “someone,” *cualquiera* “anyone,” which might convey new information but, in view of their scarce presence in the corpus, we have included them among entities transmitting given information.

“I’m lost here, this is not my area, all... then... until a man told me: ‘Come with me, I’ll take you to your place’”

The subject of (2a)–*Mi padrino* “my godfather”—is deemed to transmit new information because its referent has already been introduced into the discourse in the previous clause; the subject of (2b)–*el tipo* “the guy”—conveys accessible information because, according to our world knowledge, when somebody hits another vehicle, that car’s driver is likely to show some kind of reaction; the subject of (2c)–*el señor* “the man”—transmits new information because its referent is being mentioned for the first time.

2.3. Pragmatic relevance of lexical subjects that transmit new information

We measured relevance by counting the number of times a given entity is mentioned in the text from which subjects were extracted.¹³ The values assigned to the number of mentions are arbitrary, nevertheless, we think that they may reflect the different degrees of relevance with some precision. Examples in (3) illustrate the types of information analysed: relevant (3a), semi-relevant (3b), and irrelevant (3c). This last option corresponds to subjects whose referents are mentioned from once to a maximum of three times in the text.¹⁴

(3)

a. /relevant: 9 mentions and over/

NO EXAMPLES FOUND

b. /semirelevant: 4 to 8 mentions/

Ahorita, este... me da miedo manejar de noche... o sea, esos temores así de que te vayan a asaltar y no sé que, o... o de que te hayas quedado... vayas a quedar accidentado. Yo me he quedado dos veces accidentada y las dos veces han sido de noche. Me dio un miedo espantoso.

ENTREVISTADOR 1: ¿Pero nunca has conseguido un pavo generoso que te dé la mano?

HABLANTE: No, sí, me ayudó *UNA PAREJA*₁, sí, eh... una de las veces me ayudó *una pareja*₂. Fue en la bajada de Cumbres, eh... de Cumbres de Curumo, que... eh ... se me espichó un caucho, y yo, ¡Dios mío! yo jamás en mi vida... ¡un caucho! Eran las nueve de la noche y... este... ¿cómo es? El... entonces... el ... el... esa bajada de Cumbres bajan todo el mundo, así de... a millón ¿no? y yo: “Bueno, déjame poner el

¹³ Taking into account all mentions present in a text and not only those appearing in a given episode is justified in Bentivoglio and Martínez 1998.

¹⁴ According to Chafe (1994) trivial subjects are only mentioned once in a given discourse. We have decided to extend to three the number of times a trivial subject may be mentioned.

triángulo de seguridad.” Bueno, me... pasaban y tocaban la corneta, me decían de todo. Más, *la pareja*₃ que me había... me había dicho... me... que me iba a decir de todo; pero entonces *Ø*₄ [la pareja] se pararon y entonces...este... el... *el... muchacho*₅ [de la pareja] me cambió la... l ... el caucho y eso. La otra vez tuve... le... le di a un niño cinco bolívars para que llamara a mi mamá [risas] y le... le avisara que... se me había espichado. Dos, las dos veces han sido por espichada de caucho. (ca2fb.87)

“Now, I’m afraid of driving at night... well, I’m afraid that someone is going to attack me, I don’t know... or that I’m going to stay... going to remain there. I have been in an accident twice and both times it happened at night... I was terribly afraid. INTERVIEWER 1: But did you ever find a generous guy wishing to give you a hand?

SPEAKER: No, yes, a couple helped me, yes... once a couple₂ helped me. It happened while coming down from Curumo, eh... from Cumbres de Curumo, where I had a flat tire, and I, my God, never in my life... a tire! It was nine at night and well... how does it go? That... then... the... the... that descent from Curumo everybody comes down like this... very fast, don’t they? And I “Well, let me put out the security triangle.” Well, they... passing me by and honking and saying all kinds of things. It’s more, the couple₃ that told me ... that was going to say to me all kind of things, but then they₄ stopped and then... the... the boy₅ changed the tire, etc. The second time I had... I gave five bolívars to a child so that he would call my mom [laughs] and tell her that I had a flat tire. Both, both times it happened because of a flat tire.”

c. /irrelevant: 1 to 3 mentions/

Entonces estuvimos viviendo [el hablante y su ex-exposa]; nació *UNA NIÑA*₁ *que*₂ ahorita tiene cinco años ya; seguimos viviendo y, bueno, hasta que... hasta que ella [la ex-exposa] conoció a un señor más mayor que yo, un señor maduro ya, que tenía ahí de todo y, bueno, sucesivamente se llegó al caso de... de separarnos, pues, de divorcio (ca5mc.87).

“Then, we were living [the speaker and his ex-wife]; a baby girl₁ was born who₂ now is already five years old, we went on living and, well, until... until she [the ex-wife] met a man older than me, a mature man, who had everything and, well, after that we got to the point of... of separation, well, of divorce.”

3. Analysis

In Table 1 we present the results related to the first variable, *subject’s type*, hierarchically arranged in descending order:

Table 1. Distribution of subjects according to their type

Subject’s type	N of tokens	%
Ø	976	54.22
Non relative pronoun	408	22.66

Lexical noun phrase	263	14.60
Relative pronoun	109	6.05
Nominal clause	29	1.61
Headless relative clause	15	0.83
TOTAL	1,800	100.00

In Table 1, we may observe that elliptical subjects are the most frequent (54.22%), followed by subjects that are non relative pronouns (22.66%), lexical noun phrases (14.60%) and relative pronouns (6.05%). Nominal and headless relative clauses are rare (1.61% and 0.83%, respectively).

In Table 2 below, we analyze the 262 lexical noun phrase subjects in order to determine the type of information they convey, in accordance with the second variable.

Table 2. Type of information transmitted by lexical subjects

Information type	N of tokens	%
Accessible	132	50
Given	107	41
New	24	9
TOTAL	263	100

The results in Table 2 indicate that subjects which are lexical noun phrases generally transmit accessible (50%) or given information (41%). The few cases (24, i.e., 9%) of new information confirm, at least partially, Chafe's constraint. With the purpose of ratifying this finding, it is necessary to assess the lack of pragmatic relevance of the mentioned subjects which is represented in Table 3:

Table 3. Pragmatic relevance of lexical subjects carrying new information

Pragmatic relevance	N of tokens	%
Irrelevant (trivial) (1-3 mentions)	23	96
Semi-relevant (4-8 mentions)	1	4
Relevant (9 or more mentions)	–	–
TOTAL	24	100

The results show that the greatest percentage (96%) is of subjects which convey irrelevant or trivial information. This finding largely confirms that Chafe's proposed constraint is valid for spoken English as well as for spoken Spanish.

With respect to the main characteristics of subjects carrying new but trivial information, Chafe establishes three groups of such subjects in his corpus: i) the subject is an information source which contains a verb like *say* or *tell*; ii) it offers a

parallel contrast with respect to another subject;¹⁵ iii) it is used to express surprise. In the present research we have found subjects of the verb *decir* “to say” that are a source of information (see example 4a), but no cases of subjects used to express parallel contrast or surprise. Furthermore, in our data we have observed, on the one hand, fixed verb-subject combinations, as exemplified in (4b) and, on the other hand, three types of predicates considered presentative by Hatcher (1956) and Contreras (1976: 53-54): i) existence-presence, ii) occurrence, and iii) appearing predicates, as shown in (4c) to (4e):

- (4)
- a. /*decir* predicates/
hasta que me *dijo* un señor: “Vente conmigo...” (ca5mc.87)
“until a man told me: ‘Come with me...’”
 - b. /verb-subject semifixed combinations/
y me *daba un miedo* (ca3fd.87)
“and I was so afraid”
 - c. /existence-presence predicates/
también *está* una unidad estudiantil El Paraíso (ca1fa.87)
“also there is a public school called El Paraíso”
 - d. /occurrence predicates/
me *ayudó* una pareja (ca2fd.87)
“a couple helped me”
 - e. /appearing predicates/
a veces hasta *aparecían* exámenes (ca1fa)
“sometimes also (already made) tests appeared”

We considered necessary to highlight two important characteristics of the twenty three trivial subjects shown in Table 3: i) 71% of them follow the verb; and ii) 57% of their referents are inanimate. These findings seem worth noting because Spanish word order is predominantly subject-verb [-object] (see, among others, Bentivoglio and D’Introno 1989)¹⁶ and, moreover, subjects tend to represent human rather than inanimate entities (Chafe 1994: 168; Foley and Van Valin 1985: 288).

4. Conclusions

¹⁵ Within parallel contrast there are two pragmatically related clauses in which subjects are different, as in *Él empieza a decaer y ella empieza a surgir* “He starts declining and she starts rising.”

¹⁶ In spite of the quantitatively predominant SV order, Bentivoglio (1989) observes that with some verbs—for example, flip verbs such as *gustar* “to like”—VS is the unmarked order.

The present research findings allowed us to confirm Chafe's LSC for spoken Spanish: out of 1,800 subjects analyzed, only 24 (that is, 1.33%) function as entities transmitting new information which, nevertheless, is mainly trivial (94%). The results also show that lexical subjects conveying new but trivial information have syntactic as well as semantic peculiarities. They tend to: i) appear in presentative constructions; ii) follow the verb (VS); iii) refer to inanimate entities.

References

- Bentivoglio, Paola. "Función y significado de la posposición del sujeto nominal en el español hablado." *Estudios sobre el español de América y lingüística afroamericana* 1989: 40-57. Bogotá: Instituto Caro y Cuervo.
- , and Francesco D'Introno. "Orden de palabras y posición del sujeto en el español de Caracas." In *Estudios lingüísticos y filológicos en homenaje a María Teresa Rojas*, edited by Iraset Páez Urdaneta & Luis Barrera Linares, 51-61. Caracas: Universidad Simón Bolívar, 1989.
- , and María Eugenia Martínez-Padrón. "Newness, animacy and continuity of reference in spoken Spanish." *Lingua Americana* 2 (1998): 22-37.
- , and Mercedes Sedano. "Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana." *Boletín de Lingüística* 8 (1993): 3-35.
- Broderick, John P. "Wallace Chafe's light subject constraint in conversational discourse in the immediate mode of consciousness." *Word* 50 (1999): 143-154.
- Chafe, Wallace. "Givenness, contrastiveness, definiteness, subjects, topics, and point of view." In *Subject and Topic*, edited by Charles N. Li, 27-55. New York: Academic Press, 1976.
- . *Discourse, Consciousness and Time*. Chicago: University of Chicago Press, 1994.
- Contreras, Heles. *A Theory of Word Order with Special Reference to Spanish*. Amsterdam: North-Holland, 1976.
- Dik, Simon C. *Functional Grammar*. Amsterdam: North Holland, 1978.
- Foley, William A. and Robert Van Valin, Jr. "Information packaging in the clause." In *Language Typology and Syntactic Description I*, edited by Timothy Shopen, 282-364. New York: Cambridge University Press, 1985.
- Givón, Talmy. *Functionalism and Grammar*. Amsterdam: John Benjamins, 1995.
- Hatcher, Anne G. "Theme and underlying question. Two studies of Spanish word order." *Word* 12, (1956): 14-31.
- Keenan, Edward L. "Toward a universal definition of 'subject'." In *Subject and Topic*, edited by Charles Li, 303-333. New York: Academic Press, 1976.
- Labov, William. *Sociolinguistic Patterns*. Philadelphia: University of Philadelphia Press, 1972.

- Lambrecht, Knud. "On the status of SVO sentences in French discourse." In *Coherence and Grounding in Discourse*, edited by Russell S. Tomlin, 217-260. Amsterdam: John Benjamins, 1987.
- Prince, Ellen. "Toward a taxonomy of given-new information." In *Radical Pragmatics*, edited by Peter Cole, 223-255. New York: Academic Press, 1981.
- Tomlin, Russell S. "Interaction of subject, theme, and agent." In *Beyond the Sentence: Discourse and Sentential Form*, edited by Jessica Wirth, 61-80. Ann Harbor: Karoma, 1985.

About the authors

Paola Bentivoglio is Professor of Linguistics at Universidad Central de Venezuela where she runs the Institute of Philology "Andres Bello." She specializes in modern and historical syntax of Spanish and English. She has contributed (also with Mercedes Sedano) to several internationally published volumes on esp. Romance languages. Mercedes Sedano is Professor of Linguistics at Universidad Central de Venezuela. Her scholarly interests are in Venezuelan Spanish, sociolinguistics and discourse analysis.