

Orden de palabras en español: un análisis sintáctico-semántico-pragmático del sujeto

Paola Bentivoglio
Universidad Central de Venezuela

1. Introducción*

La posición del sujeto en la cláusula ha sido ampliamente tratada en los estudios sobre el español a partir de Kahane y Kahane, quienes en 1950 dieron inicio a una larga serie de investigaciones sobre la posición del sujeto en nuestra lengua, entre las cuales deben mencionarse –en orden cronológico de publicación– Bolinger (1954), Hatcher (1956), Contreras (1976), Silva-Corvalán (1977), Delbecque (1979, 1987 y 1991), Morales (1982a y b) y Bentivoglio y Weber (1986). Casi todos los autores mencionados, pese a sus diferentes enfoques teóricos, han dado gran importancia a los valores funcionales de la posición de ese elemento gramatical. Mi propósito es analizar algunas de las características del sujeto gramatical en un corpus del español hablado en Caracas y comparar los resultados con los de un análisis del español escrito, realizado por Delbecque (1991).¹ Con el fin de lograr una comparación adecuada entre los resultados del corpus hablado y los del escrito, he seleccionado para

* Agradezco a Rebecca Beke las observaciones y sugerencias que han mejorado la versión original. Huelga añadir que todos los errores son de mi entera responsabilidad.

¹ Delbecque ha tratado profunda y extendidamente el tema de la posición del sujeto, más bien del *regente* según la terminología por ella adoptada, en su tesis doctoral de 1979, publicada en francés en 1987 y, luego, aparecida en versión española en 1991.

mi análisis las mismas variables lingüísticas utilizadas en su análisis por la lingüista belga.

Los objetivos del presente estudio son múltiples: i) ampliar la descripción de las características del sujeto en el habla; ii) determinar si esas características se correlacionan con una determinada posición (ante o posposición) respecto del verbo y en qué medida; iii) comparar los resultados obtenidos con los de Delbecque (1991), para establecer si, pese a la diferente naturaleza de los dos corpus analizados, existen algunas tendencias similares, y, finalmente; iv) contribuir a una visión más precisa sobre el orden de las palabras en general. Cabe observar que, pese a los múltiples estudios sobre la posición del sujeto en el español hablado en Venezuela,² falta aún analizar el problema de la posición del sujeto respecto del verbo en un corpus de lengua escrita. Por esta razón, la investigación realizada podría considerarse como un estudio piloto sobre el problema mencionado, con el fin de analizar luego de manera similar un corpus del español venezolano escrito y así poder comparar datos de habla más homogéneos en cuanto a su origen dialectal.

El hecho de que los datos de habla caraqueña se analicen exactamente como los del corpus escrito utilizado por Delbecque (1991) garantiza una comparación fidedigna entre los resultados de ambos estudios. Si la comparación permitiera constatar tendencias similares, tendríamos la oportunidad de contribuir a una mejor comprensión y descripción del estudio del orden de las palabras en nuestra lengua, lo que se ajustaría a lo afirmado por la misma Delbecque (1991: 66):

Si [los parámetros] se definen unívocamente, queda asegurado que, al manejarlos otros investigadores, siempre se obtendrán, indefectiblemente, los mismos resultados; o, por lo menos, resultados *equivalentes*, ya que es realista contar con posibles fluctuaciones, reflejo del margen de errores, previsible en todo manejo de grandes conjuntos por un investigador humano.

2. El análisis

2.1. El corpus del español escrito. La obra de Delbecque (1991) constituye el análisis más amplio y detallado sobre la posición del

² Bentivoglio (1988 y 1989), Bentivoglio y D'Introno (1989), Hernández (1993).

sujeto en el español escrito del que disponemos hasta ahora. Sus méritos son muchos y, desde mi punto de vista, los más importantes son los siguientes: i) el análisis se ha realizado sobre un amplio corpus de ensayos escritos entre 1921 y 1976 por diecinueve autores, trece españoles y seis hispanoamericanos; ii) los datos obtenidos han sido sometidos a un riguroso tratamiento estadístico; iii) la descripción tanto de los datos como de los procedimientos utilizados son exhaustivos y claros; y iv) los planteamientos y explicaciones de la autora son abarcadores y convincentes. Por todo ello, se trata de un estudio particularmente apto para ser tomado como referencia, ya que los datos, los procedimientos analíticos y la interpretación permiten hacer justamente lo que me había propuesto: analizar un corpus oral aplicando los mismos procedimientos utilizados en el análisis de un corpus escrito.³

2.2. El corpus del español hablado

Este análisis se basa en un corpus de cuatro horas de habla caraqueña grabadas en 1987,⁴ con un total de aproximadamente cuarenta mil palabras. Los hablantes seleccionados para integrar la muestra son cuatro hombres y cuatro mujeres, entre los 14 y 30 años de edad, pertenecientes a tres niveles socioeconómicos (alto, medio, bajo), aun cuando, dado el carácter comparativo del presente trabajo, las variables extralingüísticas no se han tomado en cuenta. De esa muestra se han extraído todos los casos de sujeto explícito, en cualquiera de sus formas (desde un pronombre hasta una cláusula sustantiva compleja), que se han codificado de acuerdo con un conjunto de variables lingüísticas: la morfología del sujeto, la composición (o extensión) del sintagma nominal, los rasgos semánticos del referente, la función gramatical (o caso subyacente), y la presencia/ausencia del sujeto en el contexto precedente. Todos los casos codificados se han sometido a dos tipos de análisis: uno para establecer las frecuencias absolutas y relativas necesarias para la comparación entre los dos corpus, y otro, mucho más complejo, el progra-

³ La misma Delbecque afirma que "el método diseñado [para su investigación] es apto para ser aplicado a cualquier otro texto, incluso al habla oral" (1991: 42).

⁴ Para más detalles sobre el corpus utilizado en el presente análisis, véase Bentivoglio y Sedano 1993.

ma GoldVarb, especialmente diseñado por Rand y Sankoff para analizar la variación lingüística, en la reciente versión de Lawrence, Robinson y Tagliamonte (2001).

2.3. Diferencias entre los dos corpus

Es necesario recordar las diferencias fundamentales entre los dos corpus utilizados para la comparación, ya que ésta podría verse afectada por esas diferencias, aun cuando, en esta etapa de la investigación, será muy difícil –si no imposible– determinar cuáles de ellas serían las más directamente implicadas. La principal diferencia, desde luego, es el tipo de corpus que se compara: i) un corpus de habla *vs.* un corpus de lengua escrita; ii) la época de las dos muestras: relativamente reciente (1987) la hablada, muy anterior (1921-1976) la mayor parte de la escrita; iii) el estilo: semiformal –a veces hasta semiespontáneo– el del corpus de habla; formal, el del corpus escrito; iv) la cantidad de datos analizados: el corpus hablado consta de 1.666 sujetos, lo que corresponde aproximadamente al 40% del corpus escrito (4.273 casos); y v) los dialectos, diferencia esta que si bien es la última en orden de mención no lo es en cuanto a su importancia: por una parte, una muestra dialectalmente homogénea de hablantes caraqueños, hijos de caraqueños y, por la otra, una muestra que consta de diecinueve diferentes idiolectos, de los cuales seis son de escritores hispanoamericanos. El análisis comparativo, sin embargo, mostrará que las diferencias mencionadas no afectan a las tendencias generales observadas en la proporción que cabría esperar.

2.4. Las variables

2.4.1. La primera variable es la posición del sujeto respecto del verbo, con dos variantes: anteposición (A) y posposición (P). Se han excluido los casos de pronombres relativos e interrogativos con función de sujeto, debido a que estos elementos ocupan una posición no variable al inicio de las respectivas cláusulas, así como los pronombres, siempre pospuestos al verbo, en las construcciones seudohendidas inversas (*el único que se salvó fui yo*).⁵ Los sujetos analizados

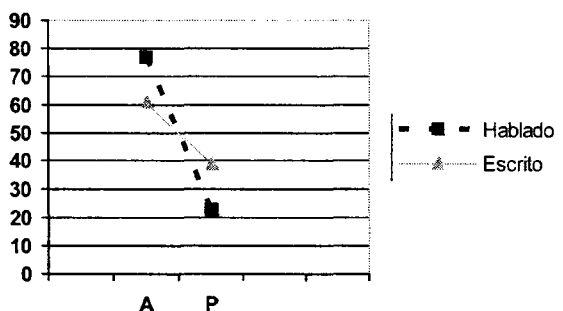
⁵ Sedano (1990) ofrece una detallada descripción de estas construcciones en el habla de Caracas.

son en total 1.666, de los cuales el 77% está antepuesto al verbo y sólo el 23%, pospuesto. La comparación de la distribución de casos en los dos corpus, hablado y escrito, se presenta en el cuadro 1 y en el gráfico 1:

**Cuadro 1: Distribución del sujeto en los dos corpus
(hablado y escrito)**

		ANTEPUESTO	POSPUESTO	TOTAL
Corpus de habla (Caracas 1987)	N	1,291	375	1,666
	%	77	23	
Corpus escrito (Delbecque 1991)	N	2,601	1,672	4,273
	%	61	39	

Gráfico 1: Posición del sujeto en los dos corpus



Como lo ilustra el gráfico 1, pese a que la tendencia es la misma, ya que los casos de anteposición son mucho más frecuentes en ambos corpus (77% en el hablado y 61% en el escrito), hay diferencias cuantitativas bastante acentuadas, como lo muestran claramente los porcentajes del cuadro 1: en el corpus escrito la posposición del sujeto se da en dos casos de cada cinco, mientras que en el corpus hablado sólo hay un caso de cada cuatro.⁶

⁶ Los resultados del análisis de Bentivoglio y D' Intronio (1989) sobre un corpus de habla de Caracas grabado en 1977 son muy parecidos: 80% de sujetos antepuestos y 20% de pospuestos.

2.4.2. Para codificar la segunda variable, se ha tomado en consideración la forma del sintagma nominal que desempeña la función de sujeto gramatical, siempre de acuerdo con las especificaciones de Delbecque (1991: 109-115). Cabe señalar que, de cada una de las variables, he tomado en cuenta los resultados de lo que Delbecque llama *análisis filtrado* y no los del análisis previo al filtrado, en que las variables son siempre más numerosas.⁷ La variable morfológica tiene cuatro variantes, que se ilustran con ejemplos extraídos del corpus hablado, en los que la parte (a) corresponde a la posición antepuesta del sujeto, que aparece en letras cursivas, y la (b), a la pospuesta. En cuanto a la forma, el sujeto puede ser:

- a. un pronombre (P) personal, demostrativo o indefinido, ver ejemplo (1);
 - b. todos los sujetos con un sustantivo como base, solo o modificado por un artículo definido o por un "singularizador" (*un, este*, entre otros; S), o por otros modificadores, como en el ejemplo (2);
 - c. una forma verbal (infinitivo con o sin clíticos, cláusulas completivas o hendidas; V), como en (4);
 - d. un nombre propio (*Lope de Vega, Simón Bolívar*, entre otros; N), ver ejemplo (5).
- (1) a. *Yo* llegué aquí en el setenta y ocho (ca5fc.87)⁸
 b. Y me acuerdo *yo* que eso fue todo un problema con la directora (ca3fb.87)
- (2) a. *los niños* son un reflejo de sus padres (ca3mb.87)
 b. mi representación la hizo *un compañero* (ca1fa.87)

⁷ En el estudio de Delbecque (1991), un ejemplo de la diferencia entre el primer análisis y el análisis filtrado es el siguiente: las variables propuestas para la morfología del regente son inicialmente nueve, según puede apreciarse en el cuadro de la página 111, mientras que las que quedan en el análisis filtrado son solamente cuatro, como lo muestra el cuadro de la página 114. Los resultados del análisis filtrado son los que se comparan con los del presente estudio.

⁸ El código alfanumérico que aparece al final de cada ejemplo remite a la identificación del respectivo hablante en el corpus de Caracas (c), así como a su grupo etario (a = 14-30 años), al nivel socio económico (1 = alto; 2 = medio alto; 3 = medio; 4 = medio bajo; 5 = bajo), al género (f = femenino; m = masculino), y, finalmente, al lugar que ocupa el hablante dentro de su propio grupo (a, b, c y d).

- c. *la diferencia de sueldo* era grandísima para aquella época (ca2mb.87)
- d. de niñita siempre me gustaban *los cuentos de la Caperucita Roja* (ca2fb.87)
- (3) a. *el que vaya por dos días* se va de Milán diciendo que es una ciudad horrenda (ca1ma.87)
- b. fue superfabuloso *compartir con grupos de todos los continentes* (ca3mb.87)
- (4) a. *la Universidad de Cornell* es totalmente aislada (ca1ma.87)
- b. me salió *Ley de Repitientes* (ca1fa.87)

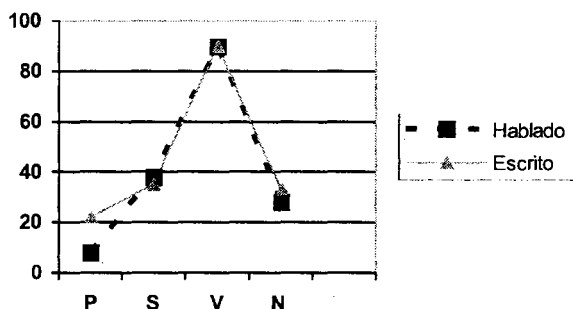
Los resultados del análisis pueden apreciarse en el cuadro 2a, mientras que en el cuadro 2b se presentan los resultados obtenidos por Delbecque (1991: 302); el gráfico 2 ilustra la comparación entre los dos corpus, con atención a la posposición que representa en ambos la posición marcada. En todos los cuadros, en la última fila, se señalan los resultados obtenidos por medio de la prueba del χ^2 , sin hacer comentarios al respecto ya que todos ellos son estadísticamente significativos y muestran que la hipótesis nula ha sido rechazada en todos los casos.

Cuadro 2a: Posición del sujeto según su forma en el corpus de habla

		ANTEPUESTO	POSPUESTO	TOTAL	%
Pronombre (P)	N	921	85	1.006	60
	%	92	8		
Sustantivo (S)	N	329	202	531	32
	%	62	38		
Forma verbal (V)	N	8	75	83	5
	%	10	90		
Nombre propio (N)	N	33	13	46	3
	%	72	28		
Total	N	1.291	375	1.666	100
	%	77	23		
$\chi^2 = 407,41$		3 g. d. l.	$p < 0,0000$		

Cuadro 2b: Posición del sujeto según su forma en el corpus escrito

		ANTEPUESTO	POSPUESTO	TOTAL	%
Pronombre (P)	N	400	116	516	12
	%	78	22		
Frase nominal (S)	N	1,900	1,009	2,909	68
	%	65	35		
Forma verbal (V)	N	44	418	462	11
	%	10	90		
Nombre propio (N)	N	257	129	386	9
	%	67	33		
Total	N	2,601	1,672	4,273	100
	%	61	39		
$\chi^2 = 600,85$		3 g. d. l.	$p < 0,0000$		

Gráfico 2: Posposición del sujeto según su forma en los dos corpus

La comparación de los porcentajes obtenidos muestra que, cuando se analizan separadamente los diferentes tipos de formas, la mayor diferencia se debe a la categoría de los pronombres, ya que con éstos la posposición sólo alcanza el 8% en el corpus hablado frente al 22% en el corpus escrito. La diferencia de 14 puntos porcentuales parece ser así la mayor responsable de la diferencia global de 16 puntos entre las dos muestras, según lo hemos podido apreciar en el cuadro 1 y gráfico 1. La posposición de los pronombres parece ser

una característica de la muestra escrita, mas no de la hablada. No hay grandes diferencias en las otras tres categorías, como lo ilustra muy bien el gráfico 2, donde las líneas correspondientes a los dos corpus casi llegan a sobreponerse.

2.4.3. La tercera variable atañe a las unidades lingüísticas (una o más) que integran el sujeto; puede hablarse de una variable relacionada con la extensión del sujeto. Esta variable, así como en el análisis filtrado de Delbecque (1991: 115-123), tiene las siguientes cinco variantes:

- a. unidad mínima: un sustantivo, un pronombre, un infinitivo (con clíticos) o un nombre propio (UM), como en el ejemplo (5);
 - b. sustantivo precedido por un morfema introductor (artículo o singularizador) (MS), véase el ejemplo (6);
 - c. elementos nominales (pronombres, sustantivos, nombres propios) yuxtapuestos y/o coordinados, o pronombres modificados por adverbios; también infinitivos, nombres propios o sustantivos, sin morfema introductor, pero acompañados de uno o varios complementos adjetivales o nominales y/o preposicionales (SNC), como en (7);
 - d. sujeto introducido por un morfema y ampliado por un elemento adjetivo, adverbial u otro elemento sustantivo coordinado; también dos elementos nominales, ambos precedidos por algún morfema, o bien un sustantivo + morfema introductor + complemento preposicional o atributivo (MSNC), como en (8);
 - e. sujeto + subordinada (relativa u otra) o bien acompañado de tres o más complementos preposicionales; también regente en forma de completiva (SNCC), véase el ejemplo (9).
- (5) a. entonces *Florenxia* es una ciudad bellísima (ca1ma.87)
 b. tiene un pasillo inmenso donde no transitan *carros* (ca3fb.87)
- (6) a. entonces *las sobremesas* son muy variadas y entretenidas (calma.87)
 b. a mí me fascina *un bochinche* (ca3mb.87)
- (7) a. *muy pocos grupos* utilizan muppets hoy en día (ca3mb.87)
 b. Y también me gusta mucho *trabajar con jóvenes* (ca5fc.87)

- (8) a. *Abuela, las fiestas a las cinco de la tarde son de niños* (ca5fc.87)
 b. *que me gustaba una cancioncita de Alí Primera* (ca3fb.87)
- (9) a. *tener un negocio [en] que haiga [sic] real de por medio y mercancía es una responsabilidad* (ca5mb.87)
 b. *no estaba el problema de que [...] tenías que esconderte del director* (ca3fb.87)

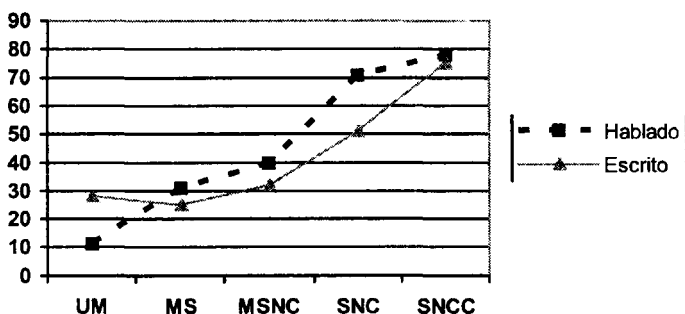
Los resultados del análisis pueden apreciarse en los cuadros 3a y 3b; el gráfico 3 ilustra la comparación entre los dos corpus:

Cuadro 3a: Posición del sujeto según su extensión en el corpus de habla

		ANTEPUESTO	POSPUESTO	TOTAL	%
Unidad mínima (UM)	N	955	118	1,073	64
	%	89	11		
Morfema + sustantivo (MS)	N	225	101	326	20
	%	69	31		
SN + complementos (s) (SNC)	N	19	47	66	4
	%	29	71		
Morfema + SN + compl. (MSNC)	N	75	49	124	7
	%	60	40		
SN + cláusula completiva (SNCC)	N	17	60	77	5
	%	22	78		
Total	N	1,291	375	1,666	100
	%	77	23		
$\chi^2 = 340,81$		4 g. d. l.	p < 0,0000		

Cuadro 3b: Posición del sujeto según su extensión en el corpus escrito

		ANTEPUESTO	POSPUESTO	TOTAL	%
Unidad mínima (UM)	N	518	199	717	17
	%	72	28		
Morfema + sustantivo (MS)	N	709	240	949	22
	%	75	25		
SN + complementos (s) (SNC)	N	173	177	350	8
	%	49	51		
Morfema + SN + compl. (MSNC)	N	840	396	1,236	29
	%	68	32		
SN + cláusula completiva (SNCC)	N	361	660	1,021	24
	%	35	75		
Total	N	2,601	1,672	4,273	100
	%	61	39		
$\chi^2 = 439,62$		4 g. d. l.		p < 0,0000	

Gráfico 3: Posposición del sujeto según su extensión

Si se examinan los porcentajes de cada una de las cinco variantes, nos damos cuenta de que las diferencias más relevantes se encuentran en: i) la de las unidades mínimas, en la que en el corpus escrito la posposición alcanza el 28% de los casos frente al 11% del corpus hablado; es obvio que los pronombres son unidades mínimas en cuanto a extensión, de lo que se puede inferir que la mayor parte del porcentaje de esta variante se debe a las formas pronominales;

ii) la de las unidades SNC (constituidas por elementos nominales yuxtapuestos y/o coordinados, o pronombres modificados por adverbios; infinitivos, nombres propios o sustantivos, sin morfema introductor, pero acompañados de uno o varios complementos adjetivales y/o preposicionales), en la que los sujetos de la muestra de habla se posponen con mayor frecuencia (71%) que los sujetos de la muestra escrita (51%). Las otras tres variantes muestran diferencias mucho menores, que varían entre los 8 puntos porcentuales de la variante MSNC y los 3 de la variante SNCC.

2.4.4. La cuarta variable refleja los rasgos semánticos del referente del sujeto gramatical y tiene cinco variantes (véase también Delbecque 1991: 123-133):

- a. el referente del sujeto es un ser animado, portador de vida (A), como lo ilustra el ejemplo (10);
- b. el referente del sujeto es una cosa (objetos concretos, instituciones, lenguas, elementos del entorno natural, (O), ver el ejemplo (11);
- c. el sujeto expresa el resultado de un proceso, una situación, un estado de cosas, un hecho; simboliza un conglomerado de elementos posiblemente diferentes, a los que, mentalmente, se confiere coherencia interna (R), como en el ejemplo (12);
- d. el sujeto menciona un proceso o una acción y se caracteriza por la presencia de la perspectiva dinámica (P), como en (13);
- e. el sujeto expresa una idea, una noción abstracta (I), ver el ejemplo (14).

A continuación, se ofrecen los ejemplos ilustrativos de las categorías semánticas mencionadas en el párrafo anterior:

- (10) a. a veces *la mamá* tiene que pegarle a uno (ca3mb.87)
b. E., se llamaba la... *la primera maestra de primer grado* (ca5mb.87)
- (11) a. Porque *Venezuela* vivía también una situación totalmente irreal ¿no? (calma.87)
b. empezaron a agrietarse *las paredes* (calfa.87)
- (12) a. *este amor por la música* es fantástico (calma.87)

- b. y, por supuesto, eran muchas *las expectativas en cuanto a eso* (ca3fb.87)
- (13) a. *acercarse a algunas realidades* fue lo mejor que me pudo haber pasado (ca3fb.87)
- b. era obvio *que ella te hablara de esa información*, sí. (ca1ma.87)
- (14) a. no fue exactamente que *la política* determinara eso (ca3fb.87)
- b. No hay ejemplos en el corpus.

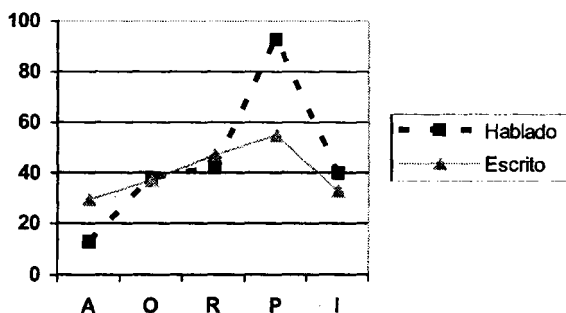
Los resultados se ofrecen en los cuadros 4a y 4b, y la comparación entre las dos muestras se ilustra en el gráfico 4:

Cuadro 4a: Posición del sujeto según rasgos semánticos en el corpus de habla

		ANTEPUESTO	POSPUESTO	TOTAL	%
Animado (A)	N	1,036	153	1,189	71
	%	87	13		
Objeto (O)	N	131	80	211	13
	%	62	38		
Resultado/situación (R)	N	117	84	201	12
	%	58	42		
Proceso/acción (P)	N	4	56	60	4
	%	7	93		
Idea (I)	N	3	2	5	0
	%	60	40		
Total	N	1,291	375	1,666	100
	%	77	23		
$\chi^2 = 308,34$		4 g. d. l.	p < 0,0000		

Cuadro 4b: Posición del sujeto según rasgos semánticos en el corpus escrito

		ANTEPUESTO	POSPUESTO	TOTAL	%
Animado (A)	N	1,036	153	1,189	71
	%	87	13		
Objeto (O)	N	131	80	211	13
	%	62	38		
Resultado/situación (R)	N	117	84	201	12
	%	58	42		
Proceso/acción (P)	N	4	56	60	4
	%	7	93		
Idea (I)	N	3	2	5	0
	%	60	40		
Total	N	1,291	375	1,666	100
	%	77	23		
$\chi^2 = 308,34$		4 g. d. l.	p < 0,0000		

Gráfico 4: Posposición del sujeto según rasgos semánticos

Los dos cuadros y el gráfico reflejan que solamente dos categorías se diferencian sustancialmente: por una parte, la proporción de sujetos que codifican el resultado de un proceso o de una acción (codificado como P) en el corpus hablado es muy superior a la de los sujetos en el corpus escrito: 93% *vs.* 55%, respectivamente; y, por la otra, en el corpus escrito, los sujetos con referentes semánticamente animados se posponen al verbo con una frecuencia que

redobla la de los sujetos del corpus hablado: 29% *vs.* 13%, respectivamente. Las otras tres categorías se comportan de manera muy parecida en las dos muestras.

2.4.5. La ausencia (Au) o presencia (Pr) de los referentes del sujeto en el contexto precedente se analiza por medio de la quinta variable, que tiene dos variantes (véase también Delbecque 1991: 143-150): la primera corresponde a la presencia de ese referente en el contexto precedente y la segunda, a su ausencia. Estas categorías, establecidas por Delbecque (1991), corresponden al tipo de información codificada por un elemento lingüístico, en este caso, por el que desempeña la función gramatical de sujeto: información nueva *vs.* información dada (o no-nueva). A este respecto parece conveniente leer la explicación ofrecida por la misma autora:

[...] definiciones generales que hacen corresponder la *información dada* con el conjunto de elementos que el autor considera o supone presentes en la conciencia del lector en el momento del enunciado y a propósito de los cuales tiene algo que decir, no son de ninguna utilidad en un estudio aplicado. Para que las distinciones sean válidas, tienen que ser operantes. Esta es la razón por la que el análisis de la *presencia en el contexto* del elemento expresado por el regente se restringe a apuntar hacia la presencia de uno o varios elementos correferentes con el regente, bajo forma idéntica, sinonímica o perifrástica, en el segmento de texto que va del principio del párrafo precedente al final del párrafo siguiente. La propia organización del texto sirve, pues, para delimitar el contexto inmediato de la proposición (1991: 145).

Parece inútil añadir que en el análisis de un corpus hablado no se pueden aplicar de manera tan estricta los parámetros adoptados por la lingüista belga en cuanto a la delimitación del *contexto inmediato de la proposición* analizada. Creo, sin embargo, que el conocimiento detallado del texto hablado (oído y analizado ya en varias ocasiones) me ha permitido establecer los límites contextuales con bastante precisión.

Las dos variantes mencionadas se ilustran en el ejemplo (15):

- (15) Ah, bueno, bueno, yo nací en Caracas el diez de octubre de mil novecientos sesenta... eh... Mis padres vivían en un apartamento en... en Las Mercedes, mientras estaba... en proyecto de construcción *una casa* en La Castellana. Después que

estuvo lista *la casa* –soy sietemesino– eh... después que estuvo lista *la casa* nos mudamos para La Castellana y... donde viví hasta los diez años de edad. (ca1ma.87)

En (15) se reproducen los primeros enunciados de un hablante caraqueño, quien empieza hablando de su nacimiento y de dónde vivió en su infancia. El sintagma nominal *una casa* representa información nueva, no mencionada ni presente en la situación,⁹ mientras que las dos menciones sucesivas del mismo referente, ahora con un artículo definido (*la casa*), codifican información no-nueva. El cambio de un tipo de información a otra se manifiesta de manera inequívoca en los modificadores del sustantivo *casa*: de *una*, cuando el referente se menciona por primera vez, no está presente en el contexto precedente, a *la casa*, cuando el referente ya ha sido mencionado, es decir, está presente en el contexto. La frecuencia de los sujetos que transmiten los dos tipos de información se muestran en los cuadros 5a y 5b y se ilustran en el gráfico 5:

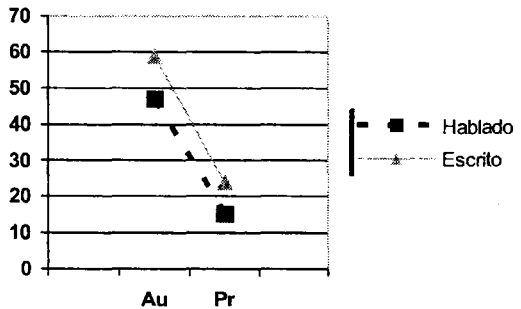
Cuadro 5a: Posición del sujeto según presencia/ausencia en el contexto, en el corpus de habla

		ANTEPUESTO	POSPUESTO	TOTAL	%
Ausencia del sujeto en el contexto (Au)	N	210	187	397	24
	%	53	47		
Presencia del sujeto en el contexto (Pr)	N	1,081	188	1,269	76
	%	85	15		
Total	N	1,291	375	1,666	100
	%	76	24		
χ^2 corregido por Yates = 178,90			p < 0,0000		

⁹ En un corpus de habla, el referente del sujeto podría estar presente no solamente en el discurso precedente sino también en la situación, posibilidad que, obviamente, no se da en un corpus escrito.

Cuadro 5b: Posición del sujeto según presencia/ausencia en el contexto, en el corpus escrito

		ANTEPUESTO	POSPUESTO	TOTAL	%
Ausencia del sujeto en el contexto (Au)	N	750	1,103	1,853	43
	%	41	59		
Presencia del sujeto en el contexto (Pr)	N	1,851	569	2,420	57
	%	76	24		
Total	N	2,601	1,672	4,273	100
	%	61	39		
χ^2 corregido por Yates = 67,07			p < 0,0000		

Gráfico 5: Posposición del sujeto según ausencia/presencia del sujeto en el contexto

Tanto en los cuadros como en el gráfico, más que los porcentajes diferentes de anteposición y posposición en las dos muestras, llama la atención la diferencia de 19 puntos porcentuales entre la ausencia del sujeto en el contexto: sólo 397 casos sobre 1.666, el 24%, en el habla, frente a 1.853 casos sobre 4.273, el 43%, en la muestra escrita (ver última columna a la derecha en los dos cuadros). Este resultado confirma lo que todos –aunque de manera intuitiva– sabemos: la proporción de ausencia del sujeto en el contexto, lo que corresponde a la información nueva, es mucho mayor en el discurso escrito que en el hablado, pues en este último caso, la capacidad de procesamiento del oyente está severamente limitada por las varias presiones de la comunicación oral, presiones a las que no está sometido el lector del discurso escrito y, mucho menos, el lector del género ensayo académico. En cuanto a la posposición, la diferencia entre los

dos corpus se manifiesta bastante acentuadamente tanto en los sujetos ausentes en el contexto como en los presentes o mencionados previamente. En los sujetos de los dos tipos los resultados, si bien diferentes, van en la misma dirección: en ambos corpus la ausencia del sujeto en el contexto precedente favorece la posposición.

2.4.6. El análisis multivariado del corpus de habla. En este último apartado, sólo presentaré los resultados obtenidos al someter los datos del corpus de habla al programa GoldVarb_2001 (Lawrence, Robinson y Tagliamonte 2001). Los resultados obtenidos por medio de este programa no pueden cotejarse con los de Delbecque (1991), ya que la autora no presenta en su obra los resultados obtenidos, pese a que en un capítulo introductorio la autora (1991: 49-51) menciona haber utilizado una versión del programa estadístico “del matemático D. Sankoff, diseñado en primer lugar para dar cuenta de fenómenos de variación sociolingüística”.

El análisis multivariado arroja los resultados que se presentan en el cuadro 6 y en el gráfico 6:

Para interpretar los resultados del cuadro 6, es necesario tener presente que en este análisis los pesos probabilísticos superiores a 0,500 contribuyen significativamente a la variante *posposición del sujeto*, mientras que los pesos inferiores a 0,500 no lo hacen. La contribución de una variante a la posposición es tanto mayor cuanto más alto es ese número.

El cuadro 6 muestra que todas las variables propuestas para el análisis han resultado ser estadísticamente significativas: la contribución más importante se debe a la primera variable (*forma del sujeto*), cuya variante *frase verbal* (V, ver ej. 3) es la que ha obtenido el mayor peso probabilístico (0,781), seguida a cierta distancia por la variante *sustantivo* (S, ej. 2) con el peso de 0,658. Las variantes *nombre propio* (N, ej. 3) y *pronombre* (P, ej. 1) apenas alcanzan el peso de 0,381 y 0,191, respectivamente. Las variantes de segunda variable (*extensión del sujeto*) con peso superior a 0,500 son tres: con un peso de 0,632 la que corresponde a un sintagma nominal modificado por una cláusula completiva (SNCC, ej. 9), con uno de 0,589 la variante correspondiente a un sintagma nominal modificado por un complemento pospuesto (SNC, ej. 7) y, finalmente, con un peso de 0,549 la variante *unidad mínima* (UM, ej. 5). Los pesos probabilísticos de las otras dos variantes (MSNC, ej. 8; MS, ej. 6) son bastante inferiores a

Cuadro 6: Contribución de las variables independientes a la posición del sujeto en el corpus de habla

VARIABLES	Nº DE CASOS	% POSICIÓN	PESO PROBABILÍSTICO
<i>Forma del sujeto</i>			
Frase verbal (V)	75/83	90	0,781
Sustantivo (S)	202/531	38	0,658
Nombre propio (N)	13/46	28	0,381
Pronombre (P)	85/1.006	8	0,191
<i>Extensión del sujeto</i>			
SN + cláusula completiva (SNCC)	60/77	78	0,632
SN + complemento(s) (SNC)	47/66	71	0,589
Unidad mínima (UM)	118/1.073	11	0,549
Morfema + SN + compl. (MSNC)	49/124	40	0,398
Morfema + sustantivo (MS)	101/326	31	0,335
<i>Rasgos semánticos</i>			
Proceso/acción (P)	56/60	93	0,807
Resultado/situación (R)	84/201	42	0,491
Objeto (O)	80/211	38	0,448
Idea (I)	2/5	40	0,399
Animado (A)	153/1.189	13	0,314
<i>Presencia /ausencia en el contexto</i>			
Ausencia (A)	187/397	47	0,553
Presencia (P)	188/1.269	15	0,447
N = 1.666	Input 0,434	Significancia = 0,013	
		Verosimilitud máxima = -611.371	
		p = 0,0000	

0,500 (0,398 y 0,335). La tercera variable, que atañe a los rasgos semánticos caracterizadores del referente del sujeto, sólo tiene una variante significativa, la de los sujetos que codifican un proceso o acción (P, ej. 13), y que tiene un peso de 0,807. Las otras cuatro variantes sólo han obtenido pesos comprendidos entre 0,314 (el más bajo) y 0,491 (el más alto). La cuarta y última variable, relacionada con la presencia o ausencia del referente del sujeto bajo análisis en el contexto precedente, consta de dos variantes, de las cuales la que corresponde a la ausencia (*una casa*, en el ej. 14) muestra un peso probabilístico de 0,553, mientras que la presencia (*la casa*, ej. 14)

tiene un peso de 0,447. Todos los datos del análisis multivariado (input, probabilidades por variantes, verosimilitud máxima y significación) son satisfactorios y comprueban que el conjunto de variables seleccionadas es adecuado para estudiar el fenómeno de la variación entre anteposición y posposición del sujeto en el español hablado en Caracas.

El análisis multivariado del corpus de habla revela tendencias que valdrá la pena explorar más exhaustivamente tanto desde el punto de vista cuantitativo (mayor cantidad de datos) cuanto estadístico.

3. Discusión de los resultados y conclusión

La comparación entre los resultados del análisis llevado a cabo por Delbecque (1991) sobre un corpus escrito, constituido por ensayos de autores españoles e hispanoamericanos, y los resultados del presente análisis de un corpus del habla de Caracas, grabado en 1987, muestra sin lugar a dudas que, pese a las diferencias notables entre las dos muestras (ver sección 2.3), las tendencias hacia la posposición del sujeto son mayoritariamente las mismas, y de manera especial, las que pueden apreciarse en el siguiente cuadro-resumen:

Cuadro 7: Comparación de las tendencias hacia la posposición en los dos corpus (sólo %)

VARIABLES	CORPUS	HABLADO %	ESCRITO %
<i>Forma del sujeto</i>			
Frase verbal (V)		90	90
<i>Extensión del sujeto</i>			
SNCC		78	75
<i>Presencia/ausencia en el contexto</i>			
Ausencia		47	59

Los resultados resumidos en el cuadro 7 muestran que, en términos de porcentajes, las diferencias son pequeñas cuando se trata de sujetos morfológicamente “pesados”, que son también extensos. En efecto, en el corpus de habla,¹⁰ los sujetos pospuestos, cuya forma es

¹⁰ No dispongo de la información análoga en cuanto al corpus escrito.

V (77 en total, ver cuadro 2a) también pertenecen en su gran mayoría ($44/77 = 57\%$) al grupo de sujetos con extensión SNCC (ver cuadro 3a), mientras que casi todos los demás ($27/77 = 35\%$) son del tipo SNC. Estas dos categorías reúnen las cláusulas completivas que desempeñan la función de sujeto y que están constituidas por infinitivos con complementos o bien por cláusulas encabezadas por *que*. Tanto en el corpus hablado como en el escrito, las tendencias en cuanto al tipo de información transmitida son muy equilibradas: la posposición se da en menor grado en el corpus hablado (47%) que en el escrito (59%), diferencia que puede explicarse debido a la tipología de los textos ensayísticos, más elaborados y planificados, frente a la mayor “espontaneidad” y falta de planificación de los textos hablados.

Los resultados más interesantes del presente análisis comparativo, sin embargo, no son tanto las semejanzas mencionadas cuanto a las diferencias, que pueden observarse en el cuadro 8:

Cuadro 8: Comparación de las mayores diferencias en cuanto a la posposición en los dos corpus (sólo %)

VARIABLES	CORPUS	HABLADO %	ESCRITO %
<i>Forma del sujeto</i>			
Pronombre (P)		8	22
<i>Extensión del sujeto</i>			
Unidad mínima (UM)		11	28
SN + complemento(s) (SNC)		71	51
<i>Rasgos semánticos</i>			
Proceso/acción		93	55

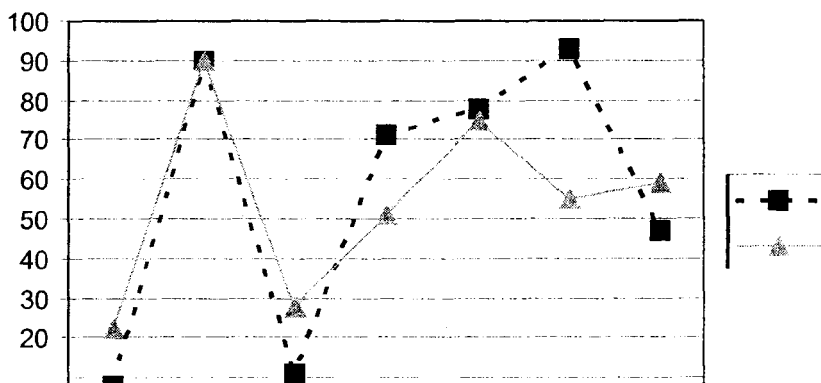
El cuadro 8 permite reflexionar sobre los aspectos siguientes: la muestra escrita se caracteriza, por una parte, por la mayor abundancia de pronombres (22% *vs.* 8%) y unidades mínimas (28% *vs.* 11%) pospuestos.¹¹ Por otra parte, en la muestra de habla se privilegia la

¹¹ Los casos de pronombres pospuestos que, a su vez, son unidades mínimas alcanza el 68% (80/118) en el corpus de habla, del que dispongo de los datos necesarios para hacer estos cálculos. De éstos casos, trece casos son marcadores del discurso (*¡fíjate tú, ¡figúrate tú, ¡imagínate tú,* etc.), en los que la variación, si bien posible, es muy infrecuente. Cabe también observar que estas expresiones son propias del habla y no ciertamente del género ensayo.

posposición del sujeto SNC (71% *vs.* 51%), lo que puede explicarse por el gran número de infinitivos con complemento(s) presentes en esta categoría: 26 sobre 47 casos, el 55% de la totalidad.¹² Los enunciados que expresan procesos/acciones favorecen la posposición en ambos corpus, pero hay una divergencia notable entre uno y otro: en el habla la posposición alcanza una frecuencia altísima (93%) –lo que casi podría considerarse como obligatoriedad– frente a la frecuencia mucho menor (55%) de la muestra escrita. Esto revela que –aun cuando, en la muestra hablada, este tipo de sujetos es escaso (60 casos sobre 1.666 = 4%),¹³ mientras que su uso es mucho más frecuente (784 casos sobre 4.273 = 18%) en el corpus escrito– el habla casi no permite variación en cuanto a la posición del sujeto cuando éste expresa proceso u acción, variación que, por el contrario, parece ser una característica de la muestra escrita y que puede deberse a la mayor variedad de recursos estilísticos empleados por los autores de los ensayos.

Las semejanzas y divergencias de resultados entre los dos corpus se reflejan claramente en el gráfico 6:

Gráfico 6: Comparación de la posposición en los dos corpus



¹² Cabría también recordar que en las construcciones donde prima la posposición hay una estrecha correlación entre el sujeto y la semántica del verbo, según lo comprobado por Morales (1982 a y b), así como por Bentivoglio 1988 y 1989.

¹³ La mayoría de sujetos de este último tipo, en el corpus de habla, es también extenso: por ejemplo, el 80% de los sujetos pospuestos pertenece a las categorías SNNC y SNC.

En el gráfico 6 se observa la coincidencia entre los puntos que corresponden tanto a los sujetos con forma V así como a los sujetos cuya extensión es SNCC. Los puntos de divergencias relativas corresponden a la forma P, a la extensión UM y a la ausencia del sujeto en el contexto (Aus), mientras que las de mayor envergadura pueden observarse en los sujetos cuya extensión es SNC y, sobre todo, en los sujetos que expresan proceso o acción (P/A). Desde un punto de vista de tipología textual, los únicos rasgos que diferencian de manera sustancial el uso de la posposición en la función gramatical sujeto son la mayor frecuencia de sujetos pospuestos del tipo SNC así como la de los sujetos que expresan proceso/acción en la muestra hablada. Los análisis posteriores que se realicen deberían, por lo tanto, investigar más a fondo la confiabilidad de los presentes resultados.

El presente artículo no agota, desde luego, el tema de la posición del sujeto en el español hablado y en el escrito. Espero, más bien, que éste pueda ser el punto de partida para los muchos otros análisis de textos tipológica y dialectalmente diferentes, ya que sólo así podremos comprender de manera más clara y exhaustiva uno de los problemas más intrigantes y, al mismo tiempo, menos estudiado, del español.

BIBLIOGRAFÍA

Bentivoglio, Paola

- 1988 "La posición del sujeto en el español de Caracas. Un análisis de los factores lingüísticos y extralingüísticos". En *Studies in Caribbean Spanish dialectology*. Eds. Robert M. Hammond y Melvyn C. Resnick. Washington, DC: Georgetown University Press. 13-23.
- 1989 "Función y significado de la posposición del sujeto nominal en el español hablado". En *Estudios sobre el español de América y lingüística afroamericana*. Bogotá: Instituto Caro y Cuervo. 40-58.

Bentivoglio, Paola y Francesco D'Introno

- 1989 "Orden de palabras y posición del sujeto en el español de Caracas". En *Estudios lingüísticos y filológicos en homenaje a María Teresa Rojas*. Comps. Iraset Páez Urdaneta *et al.* Caracas: Universidad Simón Bolívar. 51-61.

Bentivoglio, Paola y Mercedes Sedano

- 1993 "Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana". *Boletín de Lingüística* 8: 3-35.
- 2000-2001 "*El sujeto liviano: una restricción de tipo funcional*". *Boletín de Filología* XXXVIII: 9-22.

Bentivoglio, Paola y Elizabeth G. Weber

- 1986 "A functional approach to subject word order in spoken Spanish". En *Studies in Romance linguistics*. Eds. Osvaldo Jaeggli y Carmen Silva-Corvalán. Dordrecht: Foris Publications. 23-40.

Bolinger, Dwight L.

- 1954 "Meaningful order in Spanish". *Boletín de Filología* 8: 45-56.

Chafe, Wallace

- 1994 *Discourse, consciousness and time*. Chicago: Chicago University Press.

Contreras, Heles

- 1976 *A theory of word order with special reference to Spanish*. Amsterdam: North-Holland.

Díaz Campos, Manuel

- 1996 *La posición de la frase nominal-sujeto respecto del verbo: un estudio de habla infantil caraqueña*. Trabajo de grado, Maestría en Lingüística. Universidad Central de Venezuela.

Delbecque, Nicole

- 1979 *Étude quantitative d'un phénomène de variation syntaxique: la position du sujet en espagnol*. Tesis de doctorado, Katholieke Universit  t Leuven.
- 1987 *Probl  mes et m  thodes de l'   tude de la variation syntaxique. Le cas de la position du sujet en espagnol*. Leuven: Leuven University Press.
- 1991 *El orden de los sintagmas. La posici  n del regente*. Salamanca: Universidad de Salamanca.

Hatcher, Anna G.

- 1956 *Theme and underlying questions. Two studies in Spanish word order*. Supplement to *Word* 12.

Hern  ndez S., Pedro G.

- 1993 *La posici  n del sujeto: una contribuci  n al estudio del orden de palabras en el espa  ol hablado en Caracas*. Tesis de licenciatura en Letras, Universidad Central de Venezuela. Caracas.

Kahane, Henry y Ren  e Kahane

- 1950 "The position of the actor expression in colloquial Mexican Spanish". *Language* 26: 236-263.

Lawrence, Helen, John Robinson y Sali Tagliamonte

- 2001 Goldvarb_2001 [en l  nea]. Disponible en <http://www.york.ac.uk/depts/lang/webstuff/goldvarb>

Morales, Amparo

- 1982a "La perspectiva din  mica oracional en el espa  ol de Puerto Rico". En *El espa  ol del Caribe*. Ed. Orlando Alba. Santiago de los Caballeros, RD: Pontificia Universidad Cat  lica Madre y Maestra. 203-219.
- 1982b "La posici  n del sujeto en el espa  ol de Puerto Rico a la luz de la clase sem  ntica verbal, la oposici  n tema-remata y el t  pico oracional". *Ling  stica Espa  ola Actual* 4: 23-38.

Moreno Fern  ndez, Francisco

- 1994 "Status quaestionis: socioling  stica, estad  stica e inform  tica". *Ling  stica* 6: 95-154.

Sedano, Mercedes

1990 *Hendidas y otras construcciones con SER en el habla de Caracas*. Caracas: Universidad Central de Venezuela.

Silva-Corvalán, Carmen

1977 *A discourse study of some aspects of word order in Spanish spoken by Mexican-Americans in West Los Angeles*. Tesis de maestría. University of California Los Angeles.