

UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE HUMANIDADES Y EDUCACIÓN  
COMISIÓN DE ESTUDIOS DE POSTGRADO  
ÁREA DE LINGÜÍSTICA



ESTUDIO DEL SUJETO VARIABLE DE TERCERA PERSONA EN EL  
ESPAÑOL DE CARACAS 2004-2010

Autor: José Alejandro Martínez Lara  
Tutora: Paola Bentivoglio

Caracas, octubre de 2012

UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE HUMANIDADES Y EDUCACIÓN  
COMISIÓN DE ESTUDIOS DE POSTGRADO  
ÁREA DE LINGÜÍSTICA

**Maestría en Lingüística**

ESTUDIO DEL SUJETO VARIABLE DE TERCERA PERSONA EN  
EL ESPAÑOL DE CARACAS 2004-2010

AUTOR: JOSÉ ALEJANDRO MARTÍNEZ LARA

Trabajo de grado que se presenta para  
optar al grado de *Magister Scientiarum*  
en Lingüística

TUTORA: PAOLA BENTIVOGLIO

CARACAS, OCTUBRE DE 2012

MAESTRÍA EN LINGÜÍSTICA

APROBADO EN NOMBRE DE LA UNIVERSIDAD  
CENTRAL DE VENEZUELA POR EL SIGUIENTE

JURADO EXAMINADOR:

---

Prof.<sup>a</sup> Paola Bentivoglio  
Coordinadora

---

Prof.<sup>a</sup> Irania Malaver

---

Prof. Sergio Serrón

Dedico esta investigación a Dios Todopoderoso,  
y a la Virgen María del Valle del Espíritu Santo

## AGRADECIMIENTOS

Quiero agradecer, en primer lugar, a Dios por los dones que me ha dado.

Mi agradecimiento afectuoso a la Prof. Paola Bentivoglio por ser una maestra ejemplar, su tutoría ha sido una excelente clase magistral de cómo investigar y, sobre todo, una lección de vida.

Quiero agradecer, además, a todos mis colegas, en especial a los del Instituto de Filología “Andrés Bello”, quienes me apoyaron en la realización de este estudio y me hicieron observaciones y correcciones acertadas y oportunas en el arduo camino de la investigación.

A Krístel Guirado, jefa, colega y amiga. Si por ella fuera, yo estaría haciendo el segundo tomo de esta tesis.

A María José Gallucci, por las correcciones de la forma.

A Natalia Pericchi y Mariela Cisneros, por el ánimo que me dieron para seguir adelante.

En fin, gracias a quienes, directa o indirectamente, aportaron su granito de arena en el desarrollo de este estudio.

## RESUMEN

En español –al igual que en italiano, en portugués y en gallego, así como en otras lenguas románicas– es posible que el sujeto gramatical sea explícito o implícito; este fenómeno, que ha sido objeto de muchos estudios lingüísticos, se conoce como *la expresión o manifestación variable del sujeto*. Los gramáticos tradicionales (Bello 1986 [1847]; Gili Gaya 1973; Alcina y Blecua 1975; Alarcos Llorach 1994; Seco 2001) consideran que la expresión explícita del sujeto se debe a que los hablantes desean enfatizar el referente del sujeto o bien evitar posibles ambigüedades. Sin embargo, las investigaciones hechas desde la perspectiva sociolingüística (ver Silva-Corvalán 1977; Bentivoglio 1987; Paredes Silva 1993; Travis 2007, entre otros) coinciden en que la presencia explícita es un fenómeno más complejo de lo que señalan las gramáticas. El objetivo de la presente investigación es describir, dentro del marco teórico y metodológico de la sociolingüística variacionista, la alternancia entre la presencia y ausencia del sujeto de 3ª persona, singular y plural, en el español hablado en Caracas, con el fin de determinar qué factores lingüísticos y extralingüísticos pueden propiciar, en conjunto, que los hablantes mencionen explícitamente el sujeto cuando la lengua española permite su omisión. Para esto he analizado una muestra de habla compuesta por treinta y seis (36) entrevistas, pertenecientes al Corpus PRESEEA-CARACAS 2004-2010, hechas a hablantes caraqueños estratificados equitativamente según el nivel de instrucción, la edad y el sexo; de cada una de estas entrevistas he analizado cincuenta (50) cláusulas con sujeto variable. Los casos han sido codificados según siete variables lingüísticas (turno de la conversación, ambigüedad, conectividad discursiva, tipología verbal, efecto repetición, distancia referencial y número gramatical) y dos extralingüísticas (sexo y edad). Todos los casos han sido sometidos a dos tipos de análisis: uno descriptivo (frecuencias absolutas, relativas y prueba de  $\chi^2$ ); y otro de inferencias (prueba de regla variable) por medio del programa GOLDVARB X (Sankoff, Tagliamonte y Smith 2005). Este último análisis me ha permitido determinar el efecto simultáneo que las variables independientes tienen sobre la dependiente (presencia~ausencia del sujeto de 3ª persona). Para el análisis de inferencia he propuesto dos modelos teóricos, uno en el que se excluye la variable *distancia clausular*, y otro en el que se excluye la variable *conectividad discursiva*. Según el programa de regla variable, ambos modelos dan cuenta de la presencia~ausencia del sujeto. Las variables lingüísticas *conectividad discursiva* y *distancia clausular* son las que tienen los pesos probabilísticos más altos, lo que indica que la manifestación variable del sujeto está condicionada por factores internos de la lengua sensibles al discursos. La comparación entre los resultados del presente estudio y el de estudios anteriores sobre los dialectos de Caracas (Quijada 2006) y Mérida, Venezuela, (Balasch 2008), indican que el sujeto variable de 3ª persona está en variación estable.

## ÍNDICE

0. INTRODUCCIÓN	1
1. PLANTEAMIENTO DEL PROBLEMA	5
1.1. El problema	5
1.2. Justificación	9
1.3. Objetivos	10
1.3.1. Objetivo general	10
1.3.2. Objetivos específicos	10
2. MARCO TEÓRICO	11
2.1. Gramáticas	11
2.2. La cohesión textual y el sujeto de 3ª persona	13
2.3. Perspectiva: sociolingüística variacionista	17
2.4. El sujeto variable: categorías de análisis	20
2.4.1. Número gramatical	25
2.4.2. Ambigüedad	26
2.4.3. Conectividad discursiva	27
2.4.4. Tipología verbal	31
2.4.5. Distancia clausular	33
2.4.6. Turno de la conversación	33
2.5. Antecedentes	35
2.5.1. El sujeto variable: investigaciones cuantitativas que incluyen la 3ª persona	35
2.5.1.1. Barrenechea y Alonso 1977	36
2.5.1.2. Silva-Corvalán 1977	36
2.5.1.3. Enríquez 1984	37
2.5.1.4. Morales 1986	38
2.5.1.5. Paredes Silva 1993	38
2.5.1.6. Lu 1997	39
2.5.1.7. Hurtado 2001	40
2.5.1.8. Quijada 2006	40
2.5.1.9. Dumont 2006	41
2.5.1.10. Balasch 2008	42
2.5.2. Investigaciones sobre la manifestación variable del sujeto de 1ª persona	43
2.5.2.1. Bentivoglio 1987	43
2.5.2.2. Travis 2007	44
2.5.2.3. Pérez Brabandere	45
2.5.3. Resumen de los antecedentes	47

3. METODOLOGÍA	49
3.1. El corpus	49
3.2. Las exclusiones	51
3.3. Las variables de análisis	53
3.4. Procedimientos de análisis de los datos del corpus	54
3.4.1. Selección y codificación de los casos	54
3.4.2. Análisis estadísticos: descriptivo e inferencial	58
3.4.2.1. Estadística descriptiva: frecuencias absolutas y relativas y prueba del $\chi^2$	58
3.4.2.2. Estadística inferencial	58
3.4.2.2.1. Análisis de regla variable	58
3.4.2.2.2. Diseño de los modelos de análisis	61
4. RESULTADOS Y EJEMPLOS	63
4.1. Frecuencias, pruebas de $\chi^2$ y ejemplos	65
4.1.1. Número gramatical	65
4.1.2. Ambigüedad	68
4.1.3. Conectividad discursiva	70
4.1.4. Tipología verbal	75
4.1.5. Distancia clausular	78
4.1.6. Turno de la conversación	81
4.1.7. Variables extralingüísticas: Sexo y Edad	84
4.1.8. Resumen de los resultados del análisis descriptivo	86
4.2. Análisis con GoldVarb X	87
4.2.1. Modelo con la variable Conectividad discursiva	88
4.2.2. Modelo con la variable Distancia clausular	91
4.2.3. Comparación de los resultados de los dos modelos de análisis propuestos	94
4.2.4. Comparación de los resultados de GoldVarb en los dos estudios del dialecto caraqueño (Quijada 2006 y Martínez Lara 2012)	96
5. CONCLUSIONES	99
6. REFERENCIAS BIBLIOGRÁFICAS	105
7. ANEXOS	113

## 0. INTRODUCCIÓN

El español –al igual que el italiano, el portugués, el catalán y el gallego– es una lengua en la que el sujeto gramatical puede ser explícito o implícito; cuando es explícito puede aparecer antes o después del verbo.

La posibilidad de que el sujeto<sup>7</sup> sea explícito o implícito es lo que se denomina *la expresión o manifestación variable del sujeto*, fenómeno que ha suscitado mucho interés en los estudios sociolingüísticos, lo que puede corroborarse en los señalamientos que hacen autores como Silva-Corvalán 2001, López Morales 2004 y Moreno Fernández 2005.

En los estudios sobre la manifestación variable del sujeto se ha tratado de explicar por qué en una lengua *pro-drop*,<sup>8</sup> como el español, los hablantes mencionan explícitamente este constituyente oracional. Para dar cuenta de la expresión variable del sujeto los investigadores han utilizado un conjunto de factores lingüísticos, entre los que se pueden mencionar el *cambio de referencia*, el *cambio de turno*, la *tipología verbal*, la *ambigüedad* y la *conectividad discursiva*, así como algunos factores extralingüísticos (edad, nivel socioeconómico y sexo).

Este fenómeno –también conocido como la alternancia presencia vs. ausencia del sujeto– ha sido estudiado en diferentes variedades del español: Barrenechea y Alonso 1977 en Buenos Aires; Cifuentes 1980-81 en Santiago de Chile; Orozco y

---

<sup>7</sup> En esta investigación, cada vez que mencione el sujeto, me estaré refiriendo al sujeto gramatical, argumento externo del verbo (conjugado) con el cual el sujeto concuerda en persona y número.

<sup>8</sup> Según Serra, Serrat, Solé, Bel y Aparici (2000:345), se trata de una “etiqueta que se utiliza para denominar a las lenguas que permiten la omisión del sujeto”.

Guy 2008 en Colombia; Morales 1982 en San Juan de Puerto Rico; Montenegro 1982 en Santiago de los Caballeros; Enríquez 1984 y Blanco Canales 1999 en Madrid y Alcalá de Henares, respectivamente.

Igualmente, la expresión variable del sujeto se ha estudiado en hablantes de español en contacto con el inglés: Silva-Corvalán 1982, en inmigrantes mexicano-americanos de la parte este de Los Angeles; Bayley y Pease-Alvarez 1997, en el habla de niños chicanos en California; Hurtado 2001, en el habla de colombianos residentes en Miami; Travis 2007, en hablantes de Cali y New México; y Otheguy y Zentella 2007, en hablantes de Cuba, República Dominicana, Puerto Rico, Colombia, Ecuador y México, todos residentes en Nueva York.

En Venezuela, la manifestación variable del sujeto ha sido estudiada en los dialectos de Caracas y de Mérida: por Bentivoglio 1980 y 1987, Quijada 2006 y Pérez Brabandere 2010 en el dialecto caraqueño, y por Balasch 2008 en el merideño. Bentivoglio y Pérez Brabandere analizan la variación entre la presencia y la ausencia del sujeto pronominal de 1ª persona, mientras que Quijada y Balasch lo hacen sobre los sujetos de 3ª persona.

Pese a la amplia bibliografía sobre el tema, hay pocos estudios que den cuenta exclusivamente de la alternancia entre la presencia y la ausencia del sujeto de 3ª persona. En esta investigación estudio la expresión variable del sujeto de 3ª persona, singular y plural, en cláusulas con verbos conjugados, como se ilustra en el ejemplo (1); los casos identificados con la letra (a) ilustran la presencia y los señalados con (b), la ausencia:

- (1) a. /presencia/  
ELLA<sup>9</sup> incluso me *vio* llorando (CARA\_H13\_073)<sup>10</sup>
- b. /ausencia/  
y Ø *se acercó* a mí (CARA\_H13\_073)

Los datos analizados proceden de hablantes caraqueños, distribuidos equitativamente según el sexo, la edad y el nivel de instrucción (ver la descripción del corpus en §3.1), lo que me permite agregar nuevos datos sobre el sujeto variable de 3ª persona, singular y plural, en el habla de Caracas.

La investigación está organizada en cinco partes. En la primera, presento el problema objeto de estudio, la justificación y los objetivos; en la segunda, expongo las teorías en las que me baso para el estudio, así como los antecedentes; en la 3ª parte, describo el corpus y los procedimientos de análisis; en la cuarta, presento las variables de estudio con sus respectivas variantes, el análisis y los resultados; y, en la quinta parte, las conclusiones y las sugerencias derivadas de mi análisis.

---

<sup>9</sup> En todos los ejemplos de este estudio, señalo con versales el sujeto de la cláusula bajo análisis y en cursiva el verbo conjugado. Cuando el sujeto está implícito o tácito, lo indico con el símbolo del cero fónico (Ø).

<sup>10</sup> Cada ejemplo se presenta con un código alfanumérico que identifica al hablante del corpus utilizado. Este código está debidamente explicado en el apartado 3.1 de la metodología.



## 1. PLANTEAMIENTO DEL PROBLEMA

### 1.1. El problema

Uno de los constituyentes más importantes de la oración es el sujeto, al que la mayoría de las gramáticas tradicionales dedica secciones importantes (Bello 1986 [1847]; Alcina y Blecua 1975; Alarcos Llorach 1994; Seco 2001; Real Academia Española –en adelante RAE– 2009). Los sujetos explícitos no son obligatorios en español, pero sí en lenguas como el francés o el inglés,<sup>11</sup> tal como señala Fernández Ramírez 1951. En español, los verbos conjugados poseen morfemas que indican la persona (1<sup>era</sup>, 2<sup>da</sup> y 3<sup>era</sup>) y el número gramatical (singular y plural); estos morfemas ayudan a identificar cuál es el referente de un sujeto no explícito, como se ilustra en (2a y 2b):

- (2) a. /presencia/  
 MI PAPÁ estudi-ó muchos libros de historia (CARA\_H23\_090)  
           3<sup>a</sup> Sg.               3<sup>a</sup> Sg.
- b. /ausencia/  
 Ø escribi-ó biografías (CARA\_H23\_090)  
                           3<sup>a</sup> Sg.

En algunos casos, sin embargo, los morfemas de persona y de número gramatical no son suficientes para identificar el sujeto de una cláusula si este es implícito, puesto que la flexión de persona y de número de la 1<sup>a</sup> y de la 3<sup>a</sup> persona singular son idénticas en doce casos de la conjugación (ver cuadro 2 §2.4.2.), por lo que en la *Nueva gramática de la lengua española* –en adelante NGLE– (RAE

<sup>11</sup> En el caso de la lengua anglosajona, algunos autores indican que “en inglés los pronombres personales siempre se expresan, debido a la falta de terminaciones especiales que indiquen la persona y número del verbo” (Dixson y Andújar 1967:29).

2009:2557) se señala que: “Unas veces son razones morfológicas las que inducen a su presencia [del sujeto] [...]. En efecto, el imperfecto de indicativo presenta la misma forma en la primera y en la 3ª persona del singular”.

Además, vale la pena tomar en consideración que si bien la flexión verbal en español ofrece información sobre la persona y el número gramatical del referente sujeto, no lo hace sobre el género, que sí aparece en los adjetivos o participios que modifican el sujeto del que se predica, sea este explícito o implícito: “En las lenguas romances, a diferencia de las semíticas y las de otras familias, el verbo no posee flexión de género. El sujeto concuerda, en cambio, en género y número con los atributos adjetivales” (RAE 2009:2528). Otro dato importante del análisis de sujetos en contextos ambiguos lo constituye el hecho de que los verbos en español no poseen morfemas que sirvan para reconocer rasgos semánticos del referente como: [± humano] y/o [± animado], entre otros.

Según lo señalado en las líneas precedentes, hay contextos en los que no es posible identificar el referente del sujeto implícito. Observemos el ejemplo (3), en el que hay dos referentes, uno de 1ª y otro de 3ª persona, ambos en singular; el de 1ª persona es un hombre y el de 3ª persona es una mujer. En las cláusulas de (3a) a (3c) la referencia se expresa claramente en los dos sujetos (*yo* y *la tita*); sin embargo, en (3d) el contexto es ambiguo; puesto que el destinatario podría suponer que el referente del sujeto implícito ( $\emptyset$  *no podía*) es el mismo emisor (1ª persona) o *la tita* (3ª persona), sobre todo en una situación de habla rápida:

- (3) a. yo la agarré por un brazo  
 b. y Ø traté de nadar arrastrándola y nada/  
 c. LA TIPPA *se hundía* pues/  
 d. y Ø no podía/ (CARA\_H23\_085)

Este problema no se presenta en los textos escritos, en los que el lector puede recurrir al contexto previo.

En virtud de lo presentado en los párrafos anteriores, algunos gramáticos tradicionales consideran que la expresión explícita del sujeto en español se debe a la intención que puede tener el hablante respecto del referente del sujeto: i) enfatizarlo; ii) contrastarlo con otra entidad; y/o iii) eliminar la posible ambigüedad cuando ese referente es indefinido o no identificable. En las investigaciones hechas sobre la manifestación variable del sujeto, sin embargo, se ha determinado que la presencia de un sujeto expreso se debe a razones más complejas (ver Morales 1986, 1997; Hochberg 1986; Montes Miró 1986; Cameron 1992, 1994, 1995). Me pregunto, entonces, lo siguiente:

- ¿la coincidencia en las marcas verbales de persona y de número gramatical entre la 1ª y la 3ª persona singular condicionan la presencia explícita del sujeto de 3ª persona en el habla?
- ¿La presencia explícita del sujeto es una estrategia lingüística que usan los hablantes para evitar la ambigüedad discursiva?

Más allá de las características morfológicas de los verbos en español, también hay que señalar la importancia que juega la cohesión en la producción adecuada y eficaz de cualquier tipo de texto, sea este escrito u oral. Calsamiglia y Tusón

(1999:217) señalan que: “Cualquier unidad de discurso se compone de elementos verbales que están organizados y relacionados entre sí de manera explícita o implícita”. La manera como cada elemento lingüístico está unido y/o relacionado con otros en el texto se logra a través de marcas cohesivas, que pueden ser léxicas y/o gramaticales (Calsamiglia y Tusón 1999), lo que da lugar a un texto bien estructurado en cada una de sus partes. En tal sentido, los sujetos correferenciales en el texto deben estar relacionados por un conjunto de marcas –explícitas o implícitas–, las cuales responden a la intención del emisor de introducir un nuevo referente o mantenerlo en el discurso. A partir de este planteamiento, me hago las siguientes preguntas:

- ¿Las relaciones discursivas que existen entre todas las cláusulas y/o enunciados en el texto pueden condicionar que el sujeto sea explícito o implícito?
- ¿La distancia clausular entre dos sujetos correferenciales puede ser la razón de que este sea explícito o implícito?

Finalmente, trataré de establecer si algunos factores extralingüísticos pueden o no condicionar la presencia explícita del sujeto de 3<sup>a</sup> persona en el habla actual de Caracas.

Asimismo, tomando en consideración que hay estudios previos similares sobre la alternancia entre la presencia y la ausencia del sujeto de 3<sup>a</sup> persona en el habla de dos dialectos de Venezuela (Caracas y Mérida), me pregunto ¿habrá alguna

semejanza y/o diferencia en la frecuencia de uso de los sujetos variables de 3ª persona en el habla de Caracas de 1987, Mérida 1990 y Caracas 2004-2012?

## 1.2. Justificación

El objetivo de esta investigación es describir los factores lingüísticos y extralingüísticos que, en conjunto, favorecen la mención explícita del sujeto de 3ª persona –singular y plural–, cuando el español admite la omisión de este constituyente oracional. Para esto he basado el análisis en una muestra de habla de la ciudad de Caracas perteneciente al corpus PRESEEA-Caracas 2004-2010.<sup>12</sup>

En Venezuela solo se han hecho dos estudios sobre la manifestación variable del sujeto de 3ª persona: el primero, de Quijada 2006, basado en un corpus del habla de la ciudad de Caracas grabado en 1987; y el segundo, de Balasch 2008, sobre un corpus del habla de Mérida, Venezuela, de 1990. Hacer una investigación sobre este fenómeno en un nuevo corpus sobre el habla de Caracas me permitirá: i) aportar nuevos datos sobre este fenómeno lingüístico; ii) hacer una comparación diacrónica de los resultados obtenidos en la presente investigación con los de Quijada 2006; y iii) comparar los resultados de este estudio con los obtenidos por Balasch 2008 en el dialecto de Mérida, Venezuela.

En la investigación he usado variables lingüísticas que no han sido empleadas en conjunto en ninguno de los estudios previos consultados: la *ambigüedad*, la *clase semántica de los verbos*, el *número gramatical*, el *turno de la conversación*, la

---

<sup>12</sup> Proyecto para el estudio sociolingüístico del español de España y América. Disponible en: <http://www.linguas.net/Default.aspx?alias=www.linguas.net/portalpreseea>

*conectividad discursiva* y la *distancia clausular*. Los resultados obtenidos me han permitido describir más detalladamente las posibles razones por las que el sujeto de 3ª persona aparece de forma explícita en ciertos contextos comunicativos.

### 1.3 Objetivos

#### 1.3.1. Objetivo general

Describir, desde la sociolingüística variacionista, los factores lingüísticos y extralingüísticos que influyen en la expresión variable del sujeto de 3ª persona en el corpus PRESEEA-Caracas 2004-2010.

#### 1.3.2. Objetivos específicos

- Determinar la frecuencia de uso del sujeto (explícito o implícito) de 3ª persona en el habla actual de Caracas.
- Identificar los factores lingüísticos y extralingüísticos que pueden influir conjuntamente en la expresión variable del sujeto de 3ª persona.
- Determinar, con base en la comparación de los resultados de los tres estudios (Quijada 2006, Balasch 2008 y el presente), si la variación entre la presencia y la ausencia de un sujeto de 3ª persona es estable o si hay algún cambio lingüístico en curso.

## 2. MARCO TEÓRICO

### 2.1. Gramáticas

Las gramáticas tradicionales tratan acerca del sujeto y, en su mayoría, hay apreciaciones concordantes, como puede apreciarse en los párrafos siguientes.

Bello (1986 [1847]) afirma que tanto el sujeto como el predicado son los elementos necesarios de la proposición:

podemos reconocer en cada una de estas dos frases dos partes diversas: la primera significa una cosa o porción de cosas, *el niño, los árboles*; la segunda da a conocer lo que acerca de ella o ellas pensamos, *aprenden, crecen*. Llámase la primera SUJETO o SUPUESTO, y la segunda ATRIBUTO, denominaciones que se aplican igualmente a las palabras y a los conceptos que declaramos de ellas. El sujeto y el atributo unidos forman la PROPOSICIÓN (175).

Gili y Gaya (1975:21) afirma que “un verbo en forma personal lleva consigo, en español, una relación entre dos conceptos explícitos: sujeto y predicado”, en efecto, todo verbo conjugado<sup>13</sup> indica claramente cuál es el sujeto de la cláusula por medio de los morfemas de persona (1<sup>a</sup>, 2<sup>a</sup> y 3<sup>a</sup>) y número (singular y plural). El autor señala, además, que los sujetos aparecen de forma explícita en dos circunstancias: i) “Cuando se quiere hacer resaltar la participación del sujeto en la acción, como insistiendo en que es aquél y no otro”(23) y ii) “Cuando pueda haber ambigüedad” (24).

Alcina y Blecua 1975 definen el sujeto según su relación de concordancia con el verbo y no según una perspectiva lógico-semántica. Además, solo mencionan la

---

<sup>13</sup> No ocurre lo mismo con las cláusulas cuyos verbos son impersonales y/o unipersonales.

concordancia de número, pero no de la de persona, en la que también concuerdan el sujeto y el núcleo del sintagma verbal:

Al margen de su significado y su relación semántica con el verbo, el elemento al que se conviene llamar sujeto se distingue por su concordancia con el verbo, de tal manera que el cambio de número en el verbo exige el cambio de número en el sujeto para mantener el enunciado con sentido (1975:853).

En la NGLE (RAE 2009:2527) encontramos que:

el término SUJETO puede designar un concepto semántico, cuando se opone a PREDICADO, pero también una función sintáctica que contrasta con otras posibles que pueden incidir sobre el verbo (complemento directo, indirecto, etc.) [...]. Al igual que las demás funciones sintácticas, la noción ‘sujeto’ designa una función gramatical, pero también –por extensión– el elemento que la contrae.

De acuerdo con las definiciones anteriores, se puede decir que el sujeto y el núcleo del sintagma verbal son los elementos necesarios y más importantes de la oración personal, pero no es así en las oraciones impersonales, en las que “el sujeto no se quiere o no se puede determinar” (Quilis, Esgueva, Gutiérrez y Ruiz-Va 1988:399). En este apartado no haré una descripción detallada de las oraciones impersonales, las cuales sobrepasan el ámbito de mi investigación.

La relación entre el sujeto y el verbo es estrecha, puesto que en la misma morfología verbal se encuentran las marcas de persona y número del referente que solemos llamar *sujeto*. Al respecto, Alarcos Llorach (1999:137) sostiene que esta relación es imprescindible, ya que el verbo debe contener un signo de referencia léxica (el predicado) y otro de referencia gramatical (el sujeto). En español, así como en otras lenguas, los verbos permiten que el sujeto pueda aparecer implícito en las desinencias verbales, según se lee en la NGLE: “Al igual que sucede en otras lenguas

románicas, las propiedades de la flexión verbal permiten que puedan formarse en español oraciones sin sujeto expreso” (RAE 2009:2547).

Algunos autores señalan que la expresión explícita del sujeto se debe a razones enfáticas. Pottier (1975:21), por ejemplo, indica que “El sujeto está generalmente incorporado a la forma verbal [...] de ahí el valor enfático de «yo canté»”. Además, *El esbozo para una nueva gramática de la lengua española* indica otras razones por las que los hablantes mencionan explícitamente el sujeto (RAE 1973:422):

el sujeto pronominal se emplea correctamente en español por motivos de énfasis expresivo, o para evitar alguna ambigüedad posible, según las circunstancias particulares de cada caso. Tales circunstancias hacen que el hablante, sintiendo como insuficiente la expresión del sujeto contenido en la forma verbal, necesite determinarlo más.

En la NGLE (RAE 2009:2557) se confirman los señalamientos ya citados:

Existen, no obstante, numerosos casos en los que un sujeto pronominal puede hacerse expreso, aun cuando se interprete como tema no necesariamente contrastivo y su contenido se recupere del contexto inmediato. Unas veces son razones morfológicas las que inducen a su presencia, pero otras veces los factores que intervienen son de naturaleza léxica. En efecto, el imperfecto de indicativo presenta la misma forma en la primera y en la tercera persona del singular.

## 2.2. La cohesión textual y el sujeto de 3ª persona

La cohesión es el mecanismo mediante el cual se unen superficialmente todos los elementos lingüísticos del texto, según afirman Calsamiglia y Tusón (1999:230): “se da en el orden interior del texto y funciona como un conjunto de enlaces intratextuales para establecer las relaciones semánticas que precisa el texto para constituirse como unidad de significación”.

El texto es el producto de las prácticas discursivas, tanto orales como escritas, en las que los interlocutores (re)construyen sus realidades y el mundo de las ideas. Así, en la interacción, el emisor menciona un conjunto de referentes (animados y/o inanimados; humanos y/o no humanos, etc.), que van apareciendo en todo el entramado discursivo. Estos referentes se mencionan a través de un conjunto de formas y/o estrategias (enfaticar, desambiguar, focalizar y otras) que pueden variar en cada lengua. Montolío (2000:20) denomina estas formas de presentación, mantenimiento y recuperación de los referentes en el texto como *expresiones referenciales*, las cuales atañen al *referente*, término que sobrepasa a las personas y/o objetos del mundo extralingüístico y abarca también “la representación que mentalmente se hacen los hablantes, a lo largo de la comunicación, de ese objeto o de esa persona” (2000:21). En tal sentido, las formas, tanto explícitas como implícitas, de un sujeto pueden llamarse *expresiones referenciales*.

En español, el sujeto de 3ª persona puede aparecer, recuperarse y/o mantenerse en el texto a través de un conjunto de marcas formales cohesivas: las frases nominales léxicas (en adelante FNL), que pueden ser definidas e indefinidas; los pronombres personales (tónicos y átonos); las marcas flexivas de los verbos; y la elipsis,<sup>14</sup> puesto que la 3ª es la forma no marcada del paradigma pronominal y abarca un grupo ilimitado de referentes (materiales o abstractos; animados o inanimados;

---

<sup>14</sup> *Elipsis* es un término utilizado en los estudios de Análisis del Discurso para referirse a la presentación, a través del cero fónico, de un referente mencionado previamente en el texto; puede decirse que equivale a los términos *implícito* o *tácito* que utilizo como sinónimo en esta investigación.

imaginarios o reales, etc.), los cuales, generalmente, no están presentes en la interacción.

Montolío (2000:28) señala que las FNL indefinidas se utilizan cuando el referente se menciona por primera vez en el discurso; lo contrario ocurre con las FNL definidas empleadas para remitir a un referente ya mencionado en el discurso previo, que, por lo tanto, es conocido y está activo en la memoria del destinatario. Así, las FNL definidas indican que el emisor cree o tiene la certeza de que el destinatario está familiarizado o tiene un conocimiento sobre el referente y sirven para recuperar el referente previo. Los pronombres, de naturaleza gramatical, por su parte, deben estar próximos al referente mencionado a través de una FNL (definida o indefinida); de lo contrario, puede haber problemas de ambigüedad y/o de correcta interpretación del mensaje. Montolío (2000:65) señala que: “Debe existir poca distancia entre el pronombre y la expresión que ha mencionado por primera vez el referente; en caso contrario, se corre el riesgo de que el lector no pueda interpretar el pronombre”.

Calsamiglia y Tusón 1999 señalan que la elipsis es, quizá, el recurso lingüístico cohesivo más común de la lengua y reduce el peso de la repetición léxica y pronominal. Afirman, también, que los sujetos implícitos, en lenguas como el español, son la forma más habitual de elipsis:

La *elipsis* es otro fenómeno de cohesión textual [...]. Una forma de elipsis muy habitual atañe al referente sujeto, que una vez [que] ha aparecido por primera vez, en las lenguas como el español en las que su expresión es opcional, no aparece hasta que se introduce un nuevo sujeto (238).

La elipsis se utiliza cuando el referente sujeto ha sido mencionado ya en el texto y es conocido por el destinatario. En la NGL (RAE 2009:2556) se indica que: “Los sujetos tácitos de tercera persona son característicos de las oraciones en las que se ha introducido un tema discursivo y existen ciertas garantías de que la predicación que se va a establecer se aplica a esa misma entidad”; por esto, los sujetos tácitos de 3ª persona pueden identificarse en el texto a través de las anáforas y las catáforas. Por ende, la aparición de un sujeto implícito es indicio de que este es correferencial con el sujeto de la(s) cláusula(s) previa(s) (Montolío 2000:67); la elipsis puede utilizarse de forma secuencial hasta que haya un cambio de referente sujeto en el texto. En este caso, no es válido que el emisor la mantenga, pues se podría dar un caso de ambigüedad y el emisor debería, entonces, optar por un sujeto pleno, sea este una FNL definida o indefinida.<sup>15</sup>

En otro orden de ideas, Briz Gómez 2001 señala que el hecho de que el emisor y el destinatario compartan conocimientos y tengan una cercanía vivencial puede permitir que el emisor no mencione explícitamente un sujeto, por considerar que es innecesario e inadecuado para el mensaje que quiere transmitir.

Antes de finalizar este apartado, es conveniente subrayar que el texto no es un producto comunicativo estático e inerte, sino que, por el contrario, va desarrollándose a través de un eje espacio-temporal según la linealidad y la intencionalidad del

---

<sup>15</sup> “El sujeto elíptico permite mantener el mismo referente sujeto a lo largo de varias frases. En caso de que el sujeto de la nueva frase sea distinto al de la precedente (es decir, se produzca un cambio de tema de una frase a otra), el escritor debe abandonar el sujeto elíptico y emplear una expresión que le permita hacer referencia a una entidad distinta a la del sujeto de la frase anterior” (Montolío 2000:68).

mensaje que se transmite. Según Calsamiglia y Tusón (1999:220), un texto, para ser adecuado y coherente, debe cumplir con una serie de requisitos:

a) lo que aparece primero orienta lo siguiente, b) a lo largo del texto es necesario marcar las relaciones existentes en su interior, de modo que el mundo de referencias se vaya manteniendo, recuperando y proyectando hacia delante, y c) [...] tiene un importante papel que la secuencia de enunciados progrese hacia un fin o una meta determinada.

Puede decirse, entonces, que en el texto deben aparecer las marcas formales necesarias y pertinentes para ir orientando al destinatario sobre cuál es el tema de la comunicación, acerca de quién se predica y cuál es el objetivo que se persigue.

En resumen, el sujeto en español está siempre presente –explícita o implícitamente– en las oraciones o cláusulas personales. El sujeto de 3ª persona puede presentarse en varias formas: FNL indefinidas y definidas, pronombres y cero fónico. Cada una de estas responde a una necesidad o estrategia discursiva elegidas por el emisor (desambiguar, topicalizar o recuperar un referente en el discurso, entre otras).

### 2.3. La perspectiva: sociolingüística variacionista

Debido a la naturaleza del fenómeno que he analizado, esta investigación se inscribe dentro de la sociolingüística variacionista, entendida como la disciplina que estudia la lengua tanto en su entorno social como en el nivel lingüístico (Silva-Corvalán 2001). Uno de los principios básicos de esta disciplina es que la lengua es variable, pues: “los hablantes recurren a elementos lingüísticos distintos para expresar cosas distintas, naturalmente, pero, a la vez tienen la posibilidad de usar elementos

lingüísticos diferentes para decir una misma cosa” (Moreno Fernández 2005:21). El sociolingüista, por lo tanto, debe analizar la lengua hablada, en el hecho interactivo, para determinar qué factores pueden causar la selección de los elementos lingüísticos empleados por los hablantes.

El fundador de la sociolingüística variacionista es William Labov, quien observó, en el nivel fonético-fonológico, que la variación lingüística no es arbitraria, sino que está condicionada por factores inherentes a la lengua y, que esa variación puede estudiarse cuantitativamente. Uno de los principios básicos de los postulados de Labov 1972 es que las variantes de una variable lingüística pueden tener el mismo significado referencial, pero adquieren otro significado tanto social (según la edad, el nivel social, la instrucción y el sexo del hablante) como estilístico.

Silva-Corvalán 2001 afirma que el éxito logrado por Labov 1966 y Sankoff 1978 en los estudios fonético-fonológicos y en el desarrollo de técnicas de análisis cuantitativas para los estudios sociolingüísticos hicieron posible que algunos lingüistas extendieran el método variacionista a otros niveles de la lengua (morfológico, sintáctico o léxico). La autora aclara, sin embargo, que los estudios de variación sintáctica no pueden desarrollarse de la misma manera que los del nivel fonológico por las siguientes razones:

- a. Hay menos variación sintáctica que fonológica en una variedad determinada de lengua [...].
- b. La variación sintáctica es más difícil de estudiar y, especialmente, de cuantificar, debido a la poca frecuencia con que se dan los contextos de ocurrencia de una variante determinada [...].
- c. Los contextos de ocurrencia de una variante sintáctica son en general más difíciles de identificar y definir que aquellos de una variable fonológica.

- d. Aún más importante, la variación sintáctica plantea el problema de las posibles diferencias de significado que pueden estar asociadas con cada variante [...] (Silva-Corvalán 2001:129-130).

Varios lingüistas (Lavandera 1978, Romaine 1981 y García 1985, entre otros) han cuestionado la posibilidad de extender el método variacionista desde el nivel fonético-fonológico a los otros niveles de la lengua, debido a la posibilidad de que en los demás niveles las distintas variantes de una variable tengan significados aun levemente diferentes. Al respecto, Labov 1978 y otros autores (Sankoff 1978, Silva-Corvalán 2001 y López Morales 2004) sustentan que sí es posible extender el método a otros niveles de análisis lingüístico, siempre que se considere la equivalencia funcional de dos o más formas más allá de su valor de igualdad o verdad.

En el presente estudio, sin embargo, el valor de igualdad de las variantes analizadas no es objeto de discusión, puesto que un sujeto explícito (pronombre o FNL) tiene el mismo valor funcional que un sujeto implícito ( $\emptyset$ ) en los distintos contextos en los que esas forman aparezcan. En tal sentido, desde la perspectiva de la sociolingüística variacionista adoptada en este estudio, solo me queda señalar que, según Moreno Fernández 2005, hay que tener en cuenta que: i) las variantes de una variable son dos o más formas de decir lo mismo o, en el caso de la variación morfosintáctica, dos o más formas funcionalmente equivalentes en distintos contextos discursivos; ii) la variable dependiente debe tener alguna correlación con algunos de los factores extralingüísticos –por ejemplo, la edad, el grado de instrucción y el sexo de los hablantes– del corpus utilizado; y iii) los datos deben analizarse con pruebas

estadísticas y multivariantes que permitan determinar cuáles factores contribuyen a la aparición del fenómeno lingüístico bajo análisis, ya que uno de los objetivos del método variacionista es calcular la probabilidad de aparición de un fenómeno lingüístico particular en contextos lingüísticos, extralingüísticos y/o estilísticos específicos.

#### 2.4. El sujeto variable: categorías de análisis

La manifestación variable del sujeto se ha estudiado en lenguas románicas –el español (Enríquez 1986; Morales 1982, 1986; y Orozco y Guy 2008); el portugués (Paredes Silva 1993) y el catalán (Mayol 2009)– y también en lenguas no románicas como el chino mandarín (Li y Thompson 1979) y el yorùbá (Lamidi 2009). Algunas investigaciones se han hecho sobre muestras orales (Cifuentes 1980-81; Montenegro 1982; Cameron 1992, 1993) y otras en textos escritos (Paredes Silva 2003; Gutiérrez Maté 2008); en varias, la atención se centra en hablantes de segunda lengua (L2) (Silva-Corvalán 1977, 1982; Bayley y Pease-Alvarez 1997; Hurtado 2001; Flores-Ferrán 2002, 2004; Morales 1982, 1997); algunos investigadores han analizado solo la 1ª persona gramatical (Bentivoglio 1987 y 1999; Cantero 1976; Blanco Canales 1999; Samper Padilla, Hernández Cabrera y Troya Déniz 2006; Pérez Brabandere 2010) y otros han estudiado las tres personas gramaticales (Barrenechea y Alonso 1977; Miró Vera y de Pineda 1990; Cameron 1992). Sin embargo, pocas son las investigaciones en las que se han tomado en cuenta solamente los sujetos de 3ª persona (Quijada 2006 y de Balasch 2008).

Desde los primeros estudios sobre la manifestación variable del sujeto hasta la actualidad, han sido varias las categorías de análisis utilizadas para describirla. Entre estas se cuentan: la *tipología verbal*, la *ambigüedad*, el *número gramatical*, el *énfasis*, el *cambio de turno*, el *cambio de referencia*, la *conectividad discursiva*, y la *distancia clausular*. Sin embargo, en ninguna de las investigaciones consultadas se analiza la influencia de todas estas variables en conjunto, como puede verse en el cuadro 1, en el cual se presentan todas las variables utilizadas en las investigaciones que he consultado. He clasificado a este conjunto de factores en tres grupos; de acuerdo a si son sintácticos, semánticos, discursivos y textuales y de lenguas en contactos. En las líneas siguientes a este cuadro describiré seis de las categorías de análisis que se han utilizado en distintos estudios sobre la alternancia entre la presencia y la ausencia del sujeto gramatical y que he usado en esta investigación.







#### 2.4.1. Número gramatical

El número gramatical “se refiere a la cantidad de individuos, de ejemplares, de piezas repetidas de un conjunto que se toma en consideración. La lengua española distingue entre la multiplicidad y la unicidad” (Lamíquiz 1972:93). En tal sentido, los morfemas flexivos de los verbos, además de indicar el modo, el tiempo, el aspecto y la persona, también marcan el número del núcleo del sujeto.

Alonso y Henríquez Ureña (1969:105) señalan que “[...] las personas gramaticales tienen singular y plural, el verbo tiene formas especiales para las tres personas en singular y plural”.

La RAE (2010:35) indica que el número “es una propiedad gramatical característica de los sustantivos, los pronombres, los adjetivos, los determinantes (en el sentido amplío, que abarca los cuantificadores) y los verbos. Se presenta en dos formas: SINGULAR [...] y PLURAL”.

En tal sentido, Lamíquiz (1972:95) señala que el paradigma verbal presenta, además de las series temporales (pasado, presente y futuro) y personales (1ª, 2ª, 3ª persona), la serie de número correspondiente a cada una de las tres personas (singular y plural).

En virtud de esto, en varios estudios sobre la presencia~ausencia del sujeto (Barrenechea y Alonso 1977; Bentivoglio 1980 y 1987; Enríquez 1984; Paredes Silva 2003; Dumont 2006; Balasch 2008; y Pérez Brabandere 2010), se ha tomado en consideración el número gramatical como una categoría de análisis, la cual consta de las dicotomía singular/plural.

### 2.4.2. Ambigüedad

Como se ha indicado en §2.1., el sujeto concuerda con el verbo en persona y número, y la desinencia verbal, en la mayoría de los casos, es clara en español. Sin embargo, la RAE (2010:54) señala que: “La coincidencia en la opción Ø de la 1ª y 3ª personas del singular da lugar a SINCRETISMO de algunas formas”. Estas formas sincréticas o similares de la 1ª y la 3ª persona singular se presentan a continuación en el cuadro 2:

Cuadro 2: Distribución de las formas verbales morfológicamente iguales en español (Seco 2001:276-278)

EJEMPLO: YO/ÉL/ELLA AMABA	CONJUGACIONES VERBALES (NOMENCLATURA DE BELLO 1847)	
	Modo	Tiempo
1. <i>amaba</i>	Indicativo	Copretérito
2. <i>había amado</i>	Indicativo	Antecopretérito
3. <i>amaría</i>	Indicativo	Pospretérito
4. <i>habría amado</i>	Indicativo	Antepospretérito
5. <i>ame</i>	Subjuntivo	Presente
6. <i>amara</i> (1ª forma)	Subjuntivo	Pretérito
7. <i>amase</i> (2ª forma)		
8. <i>haya amado</i>	Subjuntivo	Antepresente
9. <i>hubiera amado</i> (1ª forma)	Subjuntivo	Antepretérito
10. <i>hubiese amado</i> (2ª forma)		
11. <i>amare</i>	Subjuntivo	Futuro
12. <i>hubiera amado</i>	Subjuntivo	Antefuturo

En virtud de esto, Barrenechea y Alonso 1977; Enríquez 1984; Bentivoglio 1987; Paredes Silva 1993; Lu 1997; Hurtado 2001; Dumont 2006; Quijada 2006; Pérez Brabandere 2010, entre otros, afirman que el sujeto gramatical puede aparecer en tres contextos distintos según la morfología de los verbos. El primero es el contexto con formas verbales *no ambiguas*, es decir, aquellas en las que los morfemas flexivos de persona y número son distintos para cada una de las personas del paradigma (*YO*

*TENGO un carro de lujo, TÚ TIENES un perro grande y MARÍA TIENE una casa bella*). El segundo es el contexto con formas verbales *morfológicamente ambiguas*, aquellas en la que las flexiones de persona y número son similares tanto para la 1ª como para la 3ª persona singular (*YO CANTABA en el patio y ÉL CANTABA en la ducha*). El tercero son los casos *morfológicos y contextualmente ambiguos*, en los que la morfología de los verbos y el contexto de enunciación no permiten identificar el referente del sujeto bajo análisis.

#### 2.4.3. Conectividad discursiva

En §2.2., he señalado las características discursivas del texto. Atendiendo a esto, muchos investigadores han intentado determinar la manera como cada una de las unidades lingüísticas se relacionan entre sí en un texto, puesto que esas unidades no están aisladas ni son independientes las unas de las otras, sino que están interrelacionadas, desde las más pequeñas (morfemas y palabras) hasta las más grandes (sintagmas y oraciones).

En virtud de esto, Silva-Corvalán 1977 propuso la categoría *cambio de referencia o correferencialidad*, para determinar si el sujeto gramatical de una cláusula es correferencial o no con el sujeto de la cláusula inmediatamente anterior. Esta categoría ha sido “el factor estadísticamente más significativo” (Silva-Corvalán 2001:154) en los estudios acerca del sujeto variable, pero no ha sido suficiente para explicar la relación existente entre el referente sujeto de la cláusula bajo análisis y los elementos lingüísticos de la(s) cláusula(s) anterior(es). Por esta razón, Paredes Silva

1993 propuso la categoría *conectividad discursiva*, la cual toma en consideración las distintas relaciones (morfosintácticas, semánticas y discursivas) entre los referentes con función de sujeto en cláusulas adyacentes. Paredes Silva 1993 propone para esta categoría seis grados de conexión entre dos cláusulas, las cuales van desde la conexión óptima hasta la interrumpida, como puede observarse en el cuadro 3:

Cuadro 3: Descripción de la categoría conectividad discursiva de Paredes Silva 1993

GRADO DE CONEXIÓN	DESCRIPCIÓN	EJEMPLO
1° o conectividad óptima	El sujeto de la cláusula bajo análisis y el de la cláusula inmediatamente anterior son correferenciales y los verbos correspondientes son iguales en tiempo, modo y aspecto.	Ø telefonei procé, Ø <i>continuei</i> a copiar o trabalho. Ø <i>fiquei</i> de saco cheio, Ø <i>parei</i> de copiar e Ø <i>fui</i> juntar-me aos meus familiares napraia.
2° o conexión casi óptima	Los sujetos de dos cláusulas contiguas son correferenciales, pero sus verbos no tienen el mismo modo o tiempo o aspecto.	Últimamente <i>eu nao tenho tido</i> muitas novidades pra te contar, porque <i>EU estava</i> vivendo em completa alienagáo.
3° o conexión ligeramente debilitada	Entre dos cláusulas cuyos sujetos son correferenciales hay una cláusula cuyo referente sujeto es distinto al que está bajo estudio, pero no compite por ser el sujeto de la cláusula bajo análisis. El sujeto de la cláusula intermedia –que aparece subrayado– puede ser impersonal, indeterminado o inanimado.	Ø <i>fiquei</i> estudando ontem até às 3:30h. AM. <u>Ainda sao</u> <u>llhs.</u> Ø vou ficando por aqui.

(Continúa en la página siguiente)

GRADO DE CONEXIÓN	DESCRIPCIÓN	EJEMPLO
4° o conexión debilitada	El sujeto de la cláusula bajo análisis no es correferencial con el de la cláusula precedente, pero sí lo es con un elemento lingüístico –el que aparece subrayado en el ejemplo– que desempeña otra función sintáctica (objeto directo o indirecto, complemento circunstancial).	Desculpe- <u>me</u> pela minha reclamação infundada, EU <i>estava</i> brincando
5° o conexión fuertemente debilitada	Un nuevo referente –el que aparece subrayado– en la cláusula precedente puede competir con el sujeto de la cláusula bajo análisis y crear ambigüedad contextual.	Vou ter que estudar segunda-feira mesmo. A <u>Vera</u> vai ter que ser rápida e direta, pois EU <i>tenho</i> mais o que fazer.
6° o conectividad interrumpida	Hay un cambio de tópico que puede o no tener el mismo referente sujeto.	Diga-me o Prego total e envié o número de tua conta ai nos EUA para que eu possa te enviar a soma correspondente. Quanto ao livro, EU <i>seguí</i> tuas instruções.

Esta propuesta de la lingüista brasileña ha servido de base para otras categorías similares (Hurtado 2001; Paredes Silva 2003; Orozco y Guy 2008, entre otros). Quijada 2006 ha reagrupado los seis grados originales propuestos por Paredes Silva 1993 en tres, debido a que: i) en su corpus algunas de las variantes tenían muy pocos casos, lo cual afectaba estadísticamente a su análisis; y, ii) podía haber una interacción entre esta categoría y la *ambigüedad*. Las tres variantes propuestas por Quijada 2006 se presentan en el cuadro 4:

Cuadro 4: Descripción de la categoría conectividad discursiva de Quijada 2006

GRADO DE CONEXIÓN	DESCRIPCIÓN	EJEMPLO
Conectividad 1 u óptima	Este grado reúne los grados 1 y 2 de la propuesta de Paredes Silva 1993. El sujeto de la cláusula bajo análisis es correferencial con el sujeto de la cláusula precedente, independientemente de su forma así como del modo, tiempo y aspecto del verbo.	porque <i>ella sufría</i> de hemiplejía y Ø <i>me cargaba</i> so.. sobre la pierna derecha (cd3fd.87) <sup>30</sup>
Conectividad 2	Corresponde al grado 4 de la propuesta original. El sujeto de la cláusula no es correferencial con el sujeto de la cláusula precedente, pero sí lo es con otro elemento lingüístico que cumple una función sintáctica distinta a la de sujeto (objeto directo, indirecto, complemento circunstancial, atributo).	<i>Le amputaron una pierna y Ø quedó</i> con una nada más (cc4fc.87)
Conectividad 3	Este grado está compuesto por los 3, 5 y 6 de Paredes Silva 1993. El sujeto de la cláusula no es correferencial con el sujeto de la cláusula precedente ni está presente en alguna estructura lingüística que cumpla otra función sintáctica. Este grado de conectividad puede darse cuando en la interacción confluyen factores como la introducción de una aclaración, el cambio del tema de conversación o del turno de habla.	Entonces, <i>ella estaba</i> adentro escuchando música, <u>de repente cuando se fue la luz,</u> <i>ella se quedó</i> tranquila, así sentada (ca5fa)

<sup>30</sup> Estos ejemplos son los de Quijada 2006 y provienen del Corpus Sociolingüístico de Caracas de 1987. Para más detalles sobre el significado de los códigos, ver el anexo 6.

#### 2.4.4. Tipología verbal

Las gramáticas describen los verbos de distintas maneras: i) según la vocal temática (1ª, 2ª y 3ª conjugación); ii) según el complemento que los acompaña (transitivos e intransitivos); iii) según tengan o no un sujeto (personales, impersonales y unipersonales); o iv) según la clase semántica del verbo.

Bentivoglio 1980 ha propuesto una clasificación que consta de cuatro tipos de verbos según sus características semánticas: i) de volición; ii) de cognición; iii) verbales; y, iv) otros. Esta nomenclatura ha sido utilizada como categoría para el análisis del sujeto variable y ha sido el modelo para otras clasificaciones usadas en otras investigaciones (Enríquez 1984; Quijada 2006; Travis 2007; Balasch 2008, entre otros).

El grupo de investigación ADESSE ofrece una clasificación más amplia sobre los verbos, basada en los postulados de la gramática sistémico funcional de Halliday (1967 y 1994). Según este lingüista, hay seis procesos verbales: mental, relacional, material, verbal, existencial y conductual. El grupo ADESSE parte de esta clasificación y organiza los verbos en seis “macroclases”, las cuales, a su vez, tienen subclases (García-Miguel, Costas y Martínez 2005). En el cuadro 5 pueden apreciarse estos seis tipos de procesos:<sup>31</sup>

---

<sup>31</sup> Para más detalles, puede consultarse la página electrónica del grupo de investigación ADESSE: <http://adesse.uvigo.es/>.

Cuadro 5: Descripción de los tipos de procesos verbales del grupo de investigación ADESSE

TIPO DE PROCESO VERBAL	DESCRIPCIÓN	EJEMPLO <sup>32</sup>
Mental	Comprende los verbos en los que “Una entidad dotada de vida psíquica (A1) mantiene o experimenta algún tipo de estado, cambio de estado o actividad interior perceptiva, sensitiva y/o cognitiva (A2)”.	<i>Gustar, querer, ver, recordar, crear, elegir, etc.</i>
Relacional	Comprende aquellos verbos que significan posesión y atribución.	<i>Ser, estar, parecer, nombrar, etc.</i>
Material	Contiene todos los verbos de acciones físicas que no refieren a procesos mentales.	<i>Ir, señalar, mecer, fabricar, romper, destruir, etc.</i>
Verbal	Comprende todos los verbos de comunicación y de valoración u opinión.	<i>Decir, hablar, criticar, pedir, etc.</i>
Existencial	Está compuesto por los verbos que refieren a etapas de la vida y de la existencia.	<i>Ocurrir, empezar, durar, nacer, morir, etc.</i>
De modulación	Está constituido por los verbos que funcionan como auxiliares o semiauxiliares, o bien por verbos de apoyo.	<i>Ayudar, imponer, tratar, atreverse, etc.</i>

Pérez Brabandere 2010 ha sido la 1ª en utilizar la nomenclatura verbal del grupo de de investigación ADESSE en los estudios de variación morfosintáctica del habla de Caracas.

<sup>32</sup> Estos ejemplos provienen del esquema que proporciona el grupo ADESSE en <http://adesse.uvigo.es/data/clases.php>. Vea el esquema completo de las macroclases y subclases verbales en el anexo 1.

#### 2.4.5. Distancia clausular

Cameron 1995 plantea la hipótesis de que un sujeto será implícito si su correferente está entre las cinco cláusulas mencionadas previamente; por el contrario, el sujeto será explícito, si su correferente está a más de diez cláusulas de distancia.

Travis 2007, por su parte, plantea la posibilidad de analizar no solo los sujetos plurales sino también los singulares y establece una categoría de análisis a la que denomina *Distance from previous mention* ‘Distancia desde la mención previa’, la cual mide la distancia clausular en la que puede mantenerse activo un referente sujeto en el texto. Esta categoría consta de cuatro unidades de medición: i) cláusulas adyacentes; ii) cláusulas con sujetos correferenciales interrumpida por una cláusula; iii) cláusulas con sujetos correferenciales interrumpida por dos, tres o cuatro cláusulas; y, iv) cláusulas con sujetos correferenciales interrumpida por más de cinco cláusulas. Dumont 2006 y Pérez Brabandere 2010 han utilizado en sus estudios esta misma categoría.

#### 2.4.6. Turno de la conversación

La conversación es entendida como “un conjunto sucesivo y alternante de turnos de habla ocupados por diferentes hablantes” (Briz 2004:69). Albelda (2004:111) indica que la conversación “se caracteriza por una interlocución en presencia (cara a cara), inmediata, con toma de turno no predeterminada, dinámica y cooperativa”. En tal sentido, la conversación es el ejercicio o el acto por el cual los seres humanos se comunican y se relacionan y está caracterizada por la presencia de

dos o más interlocutores en un momento y un espacio específico, quienes comparten un mismo código lingüístico. Cada interlocutor interviene en la conversación a través de turnos de habla, que es “un lugar de habla relleno por emisiones informativas que son reconocidas y aceptadas por los interlocutores mediante su atención manifiesta y simultánea” (Briz 2004:69).

La conversación coloquial, sin embargo, no es la única forma de interacción entre los individuos, también existen otros tipos; uno de esos es la entrevista, la cual, según Albelda (2004:112), se organiza distintamente de la conversación coloquial y/o espontánea en, al menos, dos rasgos: i) los turnos de habla son predeterminados, puesto que el entrevistador es quien cede u otorga la palabra; y ii) los roles de los participantes son estáticos, el entrevistador no cambia su papel con el entrevistado. Además, Albelda señala que “la entrevista, en principio, tiende a cumplir muchos de los rasgos situacionales que predisponen al empleo de un registro más bien formal” (2004:114).

En los estudios sociolingüísticos, cuyos objetivos son observar formas alternantes, los corpus de estudios están compuestos, en la mayoría de los casos, por entrevistas semidirigidas.<sup>33</sup> En estas, las distintas intervenciones hechas tanto por el entrevistador como por el entrevistado son consideradas como turnos de habla.

Bentivoglio 1987, en el estudio sobre el sujeto variable de 1ª persona en el español de Caracas, reconoce que hay cambios de turnos en las entrevistas que

---

<sup>33</sup> Este es el caso de la mayoría de los corpus de habla hechos en Venezuela, tales como: el Corpus de la Norma Culta de Caracas de 1977; los Corpus Sociolingüísticos de Caracas de 1977 y 1987; el de Mérida de 1995; el corpus Preseea Caracas 2004-2010; entre otros.

analiza del Corpus Sociolingüístico de Caracas de 1977, por lo que estudia las cláusulas con sujetos explícitos o implícitos según estén: i) en un mismo turno de habla, entendido como un enunciado no interrumpido del entrevistado; o ii) en un cambio de turno, entendido como un nuevo enunciado del entrevistado después de haber sido interrumpido por una pregunta o un comentario tanto del entrevistador como el de otra persona ajena a la entrevista, o bien por una llamada telefónica o por el timbre de la casa. Bentivoglio (1987:40), sin embargo, no considera como interrupciones del turno de habla del entrevistado las pequeñas interjecciones (*ujum, ajá, ok*, etc.) hechas por uno de los investigadores o por cualquier otra persona presente en la entrevista, puesto que tales expresiones son un tipo de estímulo, pero no interrupciones o turnos de habla.

## 2.5. Antecedentes

### 2.5.1. El sujeto variable: investigaciones cuantitativas que incluyen la 3ª persona

En esta sección presento, de manera resumida y en orden cronológico, algunos estudios que se han hecho sobre la manifestación variable del sujeto. Debo acotar que no me limitaré solo a aquellos que analizan la 3ª persona exclusivamente (Quijada 2006; Dumont 2006; Balasch 2008), puesto que en otros estudios acerca de la 1ª y la 2ª persona se utilizan variables lingüísticas y procedimientos metodológicos que dan cuenta de la presencia~ausencia del sujeto y que han sido fundamentales para el desarrollo de mi investigación (Silva-Corvalán 1977; Bentivoglio 1987; Travis 2007; y Pérez Brabandere 2010).

Vale señalar que en varios de los estudios reseñados se usan los programas estadísticos VARBRUL<sup>34</sup> y/o GOLDVARB,<sup>35</sup> diseñados especialmente para las investigaciones sociolingüísticas, por lo que muchos de los resultados se expresan en pesos probabilísticos y no solamente en valores relativos.

2.5.1.1. Barrenechea y Alonso 1977 utilizan la metodología cuantitativa para analizar la expresión variable del sujeto de todo el paradigma pronominal con el propósito de determinar, a partir de resultados cuantificables, las variables lingüísticas que influyen en el uso variable del sujeto en el habla culta de Buenos Aires. Para esto las investigadoras analizaron, en un corpus de habla, las variables *persona gramatical* y *ambigüedad*. Las conclusiones son las siguientes: i) los sujetos de segundas personas (*vos, usted y ustedes*) en el tratamiento familiar de cortesía se emplean de forma explícita con más frecuencia que el resto de las formas pronominales; y ii) la ambigüedad no resulta ser una variable significativa para la explicación del fenómeno, debido a la baja frecuencia de casos completamente ambiguos (0,84%).

2.5.1.2. Silva-Corvalán 1977 es el primer estudio que se hace desde la perspectiva de la sociolingüística variacionista, por lo que representa una referencia fundamental. Las variables de su estudio son: i) la presencia vs. la ausencia del sujeto en cláusulas con verbos finitos; y, ii) la alternancia entre la posición pre y postverbal

---

<sup>34</sup> Este es el primer programa informático creado para calcular la regla variable en los estudios sociolingüísticos (ver Sankoff y Rousseau 1978).

<sup>35</sup> Para más detalles sobre este programa ver §3.5.4.

del sujeto (1977:20-21). La lingüista analiza una muestra de habla de doce mexicanos que residen en la zona Oeste de Los Angeles (West Los Angeles), distribuidos en igual proporción según el sexo (masculino y femenino) y la edad (de 14 a 20 años y más de 50 años). Los casos analizados son 1.284, de los cuales 783 (61%) son implícitos y 501 (39%), explícitos. De los 501 sujetos explícitos 236 (65%) están en posición preverbal y 175 (35%) son postverbales. La autora señala que 430 de los 501 (86%) sujetos explícitos se deben a que sus referentes difieren del referente de la cláusula inmediatamente anterior (1977:23). Silva-Corvalán concluye que los sujetos explícitos sirven para: i) establecer el referente tópico; ii) desambiguar contextos en los que hay dos referentes que compiten por ser sujeto de un verbo con morfología ambigua; y, iii) contrastar el referente tópico con otro referente (78). La importancia del estudio de Silva-Corvalán 1977 estriba tanto en analizar cuantitativamente la alternancia entre la presencia y la ausencia del sujeto como en determinar que, en español, la falta de correferencialidad entre sujetos de cláusulas adyacentes es una de las causas más relevantes para la presencia explícita de un sujeto.

2.5.1.3. Enríquez 1984 analiza la alternancia entre la presencia y la ausencia del sujeto de 1ª, 2ª y 3ª persona en el habla de Madrid, según las siguientes variables: *número, tiempo, aspecto y modo verbal, semántica del verbo, ambigüedad, polaridad, tipo de oraciones y estilo del enunciado*. Su muestra comprende 23.717 cláusulas con referentes sujetos, de las cuales 4.857 (32%) tienen un sujeto explícito. Una de las conclusiones a las que llega Enríquez es que, en el habla de Madrid, las

formas explícitas de 3ª persona, referidas a sujetos humanos, son las más frecuentes, lo que podría deberse a la necesidad que sienten los hablantes de determinar a estos sujeto, contrariamente a lo que sucede con la 1ª y 2ª persona (Enríquez 1984:197-198).

2.5.1.4. Morales 1986 analiza el sujeto variable de 1ª, 2ª y 3ª persona, singular y plural, en un corpus bilingüe estratificado según las siguientes categorías: i) los puertorriqueños con pocos conocimientos del inglés; ii) los que habían estudiado y/o trabajado en los Estado Unidos por un período menor de diez años; y, iii) los que habían nacido en USA o habían vivido y/o trabajado en ese país por un tiempo no menor de diez años. En su análisis, la autora utiliza la variable lingüística *Cambio de referente*. Los resultados muestran que la frecuencia de uso de sujetos explícitos es bastante similar para todas las personas del paradigma pronominal en los dos grupos extremos (i y iii). Morales también observa una diferencia con relación a la edad de los hablantes, pues menor es su edad, mayor es la frecuencia de uso de sujetos explícitos. La autora (1986:83) concluye que el contacto con el inglés no condiciona el uso pronominal del sujeto, aun cuando prevé que el español del Caribe podría estar cambiando lingüísticamente hacía un sistema de sujetos pronominales redundante.

2.5.1.5. Como se indicó en 2.2.1.3., Paredes Silva 1993 propone la variable *conectividad discursiva*. La lingüista brasileña analiza la presencia~ausencia del sujeto en un corpus de textos escritos informales de portugués de Río de Janeiro de acuerdo con las variables: *ambigüedad* (en sujetos de 1ª, 2ª y 3ª personas); *énfasis* (en

la 1ª y la 2ª personas); y, *conectividad discursiva* (en la 1ª persona), todas del singular. Paredes Silva 1993 utilizó el programa GOLDVARB para procesar sus datos. Los resultados obtenidos indican que los grados 5 y 6 de la *conectividad discursiva* favorecen estadísticamente la presencia de un sujeto (§2.2.1.3.), lo que confirma la hipótesis de que el debilitamiento de la conectividad es relevante para la variabilidad del sujeto.

2.5.1.6. Lu 1997 estudia el sujeto variable de 1ª, 2ª y 3ª persona en textos escritos. El autor enriquece el campo de estudios de este fenómeno lingüístico al hacer una comparación entre el español y el chino mandarín, además de sustentar sus resultados con cálculos estadísticos por medio del programa VARBRUL. El objetivo es observar si, en ambas lenguas, el sujeto se omite en los mismos contextos. El autor concluye que en el chino mandarín se prefiere la representación explícita del sujeto gramatical, pese a que en esta lengua es posible omitir este argumento verbal. Observa, también, que el pronombre de 1ª persona singular (*wǒ*) es el que más se representa de forma explícita en mandarín, mientras que el de la de 2ª de cortesía, singular y plural, es el más frecuente en español. El autor afirma que la 3ª persona presenta algunas dificultades debido a que puede remitir a cualquier persona o cosa (concreta o abstracta) de la que estén predicando los interlocutores. Esta sería la razón por la que los hablantes mencionan explícitamente el sujeto de 3ª persona con mayor frecuencia.

2.5.1.7. Hurtado 2001 analiza el habla de colombianos residentes en Miami-Dade en un corpus compuesto por 46 entrevistas, distribuidas en distinta proporción según el sexo y la edad. Las variables lingüísticas usadas son: *conectividad discursiva*, *ambigüedad*, *estilo y tipo de discurso*, *tipo de verbo*, *modo verbal*, *realización de /-s/*. Hurtado analiza 3.600 casos, de los cuales 3.168 (50,3%) tienen un sujeto implícito y 3.132 (49,7%) uno explícito. Concluye que la manifestación variable del sujeto está más condicionada por los factores lingüísticos que por los sociales. De los lingüísticos, los más significativos son: la *correferencialidad*, la *ambigüedad* y la *especificidad del referente*; respecto a la *conectividad discursiva*, Hurtado indica que “a mayor grado de conectividad menos expresión de sujetos” (232). En cuanto a la forma, señala que *usted* y *uno* son los sujetos explícitos de 3ª persona más frecuentes.

2.5.1.8 Quijada 2006 constituye el antecedente más importante y directo del presente estudio, ya que su objetivo es caracterizar la expresión variable del sujeto en el español de Caracas, y utiliza para ello una muestra de 48 entrevistas de caraqueños del Corpus Sociolingüístico de Caracas de 1987 (CSC'87), el cual está estratificado equitativamente según el sexo, la edad y el nivel socioeconómico. Las variables lingüísticas independientes son: *ambigüedad*, *turno en la conversación*, *tipología verbal* y *conectividad discursiva*. Los resultados indican que de 1.876 casos analizados, 1.253 (67%) son de sujetos implícitos y 623 (33%), explícitos. Todos los casos fueron analizados con el programa GOLDVARB 2001, el cual seleccionó como significativas las variables lingüísticas *tipología verbal*; *conectividad discursiva*,

*turno de la conversación*, así como la variable extralingüística *nivel socioeconómico*. Quijada 2006 concluye que el grado más bajo de *conectividad discursiva*, el 3, es el que más contribuye a la presencia de un sujeto explícito, ya que la interrupción de la correferencialidad entre dos cláusulas, por un cambio de turno o la aparición de un nuevo referente, hace que haya una mayor probabilidad de sujetos explícitos. Igualmente, la autora señala que los hablantes del nivel medio y las mujeres son quienes tienden a usar más frecuentemente un sujeto explícito (Quijada 2006:70). La investigadora sugiere que las investigaciones futuras se hagan siguiendo la misma metodología empleada por ella, de manera que puedan hacerse comparaciones, recomendación que he tomado en cuenta.

2.5.1.9. Dumont 2006 centra su estudio en la presencia explícita del sujeto de 3ª persona, en especial de los que tienen la forma de FNL. La investigadora usó una muestra de cinco transcripciones de hablantes del español del corpus New Mexico-Colorado Spanish Survey (NMCOS) de la University of New Mexico. Las variables utilizadas son: i) *distance from previous mention*; ii) *realization of previous mention*; iii) *information status*; iv) *specificity*; v) *grammatical number*; vi) *semantic class of verb*; vii) *animacy*; viii) *referentiality*; ix) *syntactic role*; y, x) *potential for ambiguity of morphological form*.<sup>36</sup> Dumont 2006 unificó la especificidad y el estatus de la información en una sola variable. Analiza 1.028 cláusulas, de las cuales 224 (22%)

---

<sup>36</sup> MT: i) distancia de la mención previa; ii) realización de la mención previa; iii) estatus de la información; iv) especificidad; v) número gramatical; vi) clase semántica del verbo; vii) animalidad; viii) referencialidad; ix) rol sintáctico; y, x) potencialidad de la forma ambigua.

son FNL con función de sujeto. Su análisis muestra que hay una mayor tendencia a que las FNL sujeto indefinidas remitan a referentes no específicos y que los verbos de movimiento favorezcan los sujetos explícitos. Finalmente, la autora considera importante hacer estudios en corpus de conversaciones entre dos o más interlocutores.

2.5.1.10. Balasch 2008 analiza la manifestación variable del sujeto de 3ª persona, plural y singular, desde la sociolingüística variacionista, en el habla de veinte merideños (hombres y mujeres), mayores de sesenta años, distribuidos en cinco niveles socioeconómicos (alto, medio alto, medio, medio bajo y bajo), extraída de las entrevistas del Corpus Sociolingüístico de Mérida (Domínguez y Mora 1995). La autora utiliza las variables explicativas: *número gramatical*, *clase semántica de los verbos* y la nueva versión de la *conectividad discursiva* (Paredes Silva 2003), que consta de siete grados de conexión, uno más que el de la propuesta original de 1993. Balasch 2008 concluye que la frecuencia de los sujetos explícitos aumenta a medida que se debilita la conexión, pasando del grado 1 al 7. Sus resultados concuerdan con los de Paredes Silva 2003 y, a grandes rasgos, con los de Quijada 2006, por lo que puede decirse que esta variable da cuenta satisfactoriamente de la variabilidad del sujeto y, por lo tanto, es recomendable incluirla en las investigaciones sobre este fenómeno.

## 2.5.2. Investigaciones sobre la manifestación variable del sujeto de 1ª persona

En esta sección presento algunas investigaciones que tratan acerca del sujeto variable de 1ª persona, las cuales son una referencia clave en mi estudio, puesto que, si bien no toman en cuenta la 3ª persona, utilizan variables lingüísticas que han servido para explicar la alternancia presencia~ausencia del sujeto y, que además, yo mismo he empleado en la presente investigación: *número gramatical*, *tipología verbal*, *distancia clausular* y *turno de la conversación*.

2.5.2.1. Bentivoglio 1987<sup>37</sup> es el primer estudio acerca de la manifestación variable del sujeto de 1ª persona, singular y plural, en el habla de Caracas; entre sus aciertos están: i) una metodología clara; ii) un análisis cuantitativo exhaustivo de los datos; y, iii) la propuesta de cuatro variables lingüísticas, no contempladas en Silva-Corvalán 1977: *turno de la conversación*; *tipología verbal*, *número gramatical* y *énfasis*. El objetivo de Bentivoglio es “explorar por qué y en qué medida, en un contexto determinado, una de las dos opciones [presencia~ausencia del sujeto] es la preferida” (13); para esto analiza doce muestras de habla de caraqueños, pertenecientes al Corpus Sociolingüístico de Caracas de 1977, el cual está estratificados equitativamente según la edad, el sexo y el nivel socioeconómico. Las variables lingüísticas utilizadas son: i) *cambio de referente*; ii) *cambio de turno*; iii) *ambigüedad*; iv) *énfasis*; v) *tipo de verbo*; y, vi) *número gramatical*. Sus resultados indican que de 892 cláusulas con verbos conjugados en 1ª persona, 535 tienen sujeto

---

<sup>37</sup> Esta es la traducción en español de la tesis de maestría escrita en inglés y presentada en 1980 (ver referencia bibliográfica), publicada en español siete años más tarde.

implícito, lo que representa el 60%, y 357 tienen sujetos explícitos (40%). Los datos son analizados con el programa VARBRUL 2S, el cual indicó que las variables que contribuyen significativamente a la presencia explícita del pronombre de 1ª persona son las siguientes: *cambio de referente; número gramatical; ambigüedad; tipo de verbo; sexo; y, cambio de turno*. La autora concluye que la presencia explícita de un sujeto pronominal de 1ª persona no es “fortuita” sino que responde a la imbricada relación de factores morfológicos, pragmáticos, semánticos y discursivos (Bentivoglio 1987:66).

2.5.2.2. Travis 2007 también estudia el sujeto variable de primera persona. Su objetivo es analizar el papel del *priming* en cláusulas con verbos conjugados en 1ª persona. El *priming* o *efecto repetición*, como ha sido traducido al español, es un mecanismo que se revela cuando una forma o estructura, que aparece en un enunciado, vuelve a aparecer en los enunciados siguientes. Cameron 1995 llama al referente sujeto del primer enunciado *trigger* ‘gatillo’ y el del segundo, el que está bajo análisis, *target* ‘blanco’.<sup>38</sup> Travis basa su estudio en dos dialectos y dos géneros discursivos diferentes: i) entrevistas sociolingüísticas de hablantes de español de New Mexico, del corpus NMCOS; y, ii) conversaciones espontáneas de colombianos grabadas en la ciudad de Cali, Colombia (Travis 2005). Las variables usadas son: i) *clase semántica de los verbo*; ii) *TAM*; iii) *distancia clausular*; iv) *priming* o *efecto repetición*; y, v) *tipo de cláusula*. Los resultados obtenidos con el programa

---

<sup>38</sup> Para más detalle, vea Cameron 1992; 1995 y Cameron y Flores-Ferrán 2004.

GOLDVARB 2001 indican que en el corpus de NMCOS, de los 853 casos analizados, 569 (66%) tienen sujetos implícitos y 284 (33%), sujetos explícitos, mientras que de los 878 sujetos del corpus de Colombia, 457 (52%) son implícitos y 421 (48%), explícitos. Según el análisis realizado por medio de GOLDVARB, los factores significativos, en ambas muestras, son: la *clase semántica de los verbos*, la *distancia clausular*, y el *priming (efecto repetición)*. La investigadora concluye que, en los dos corpus, la presencia explícita de un sujeto pronominal de 1ª persona está condicionada tanto por el *priming*, puesto que “a pronominal mention favors a subsequent pronominal mention” (Travis 2007:131)<sup>39</sup> como por la *distancia clausular*, ya que la frecuencia de uso de sujetos pronominales es mayor cuando la distancia entre cláusulas con referentes sujetos correferenciales también lo es.

2.5.2.3. Pérez Brabandere 2010 analiza la presencia~ausencia del sujeto pronominal de 1ª persona, con el fin de determinar los factores que influyen en su expresión variable y también de comparar sus resultados con los de Bentivoglio (1987 [1980]). Para esto, analiza una muestra de habla de 36 hablantes del corpus PRESEEA-Caracas 2004-2010, el cual está estratificado equitativamente según el sexo, la edad y el nivel educativo. Usa las variables: *número gramatical*, *ambigüedad*, *TAM del verbo*, *polaridad de la cláusula*, *cambio de turno*, *clase semántica de los verbos*, *conectividad discursiva*, *efecto repetición*<sup>40</sup> y *distancia clausular*. Sus resultados indican que de 3.600 casos de sujetos analizados, 2.528 (70%) son implícitos y, 1.072

---

<sup>39</sup> MT: una mención pronominal favorece una mención pronominal posterior.

<sup>40</sup> Este es el nombre en español que la autora le da a la variable que Cameron y Flores-Ferrán 2004 y Travis 2007 llaman *Perseveration* y *Priming*, respectivamente.

(30%), explícitos. La autora analiza los datos con el programa GOLDVARB X, según dos conjuntos de variables distintos, con el fin de prevenir una posible interacción entre las variables *conectividad discursiva*, *efecto repetición* y *distancia clausular*. Concluye que el modelo de análisis sin la variable *conectividad discursiva* es el que mejor da cuenta del fenómeno, y que las cinco variables que condicionan la presencia de un sujeto pronominal de 1ª persona son: *efecto repetición*, *distancia clausular*, *ambigüedad*, *clase semántica de los verbos* y *número gramatical*.

En el cuadro 1 puede observarse que las variables lingüísticas más usadas en los estudios sobre la presencia~ausencia del sujeto son las siguientes: *ambigüedad*, en doce (12) investigaciones; *tipología verbal*, en ocho (8); *número gramatical*, en siete (7); *cambio de referencia* y *conectividad discursiva*, en seis (6); *efecto repetición* y *distancia clausular*, cada una en cuatro (4); y, *cambio de turno*, en tres (3). También puede verse en el cuadro 1 que siete (7) investigaciones se basan en corpus de hablantes bilingües (español-inglés). Con base en los estudios reseñados, puedo decir que las ocho variables mencionadas parecen dar cuenta, en alguna medida, de la variabilidad del sujeto; por esta razón, he considerado conveniente utilizar seis de ellas en mi investigación. Según ha podido observarse tanto en el cuadro 1 como en las reseñas, los estudios del sujeto variable –desde 1977 hasta la actualidad– se han hecho más complejos tanto por el aumento de variables lingüísticas explicativas como por el empleo de programas estadísticos muy precisos.

### 2.5.3. Resumen de los antecedentes

En primer lugar, he descrito, según las gramáticas tradicionales y los manuales de estilo, las características del sujeto en español. En segundo lugar, he presentado la perspectiva de la sociolingüística variacionista, en la que se inscribe esta investigación. En tercer lugar, he reseñado un conjunto de estudios, tanto en 2.5.1. como en 2.5.2., los cuales han contribuido en diferente medida a la investigación que he llevado a cabo; en todos ellos se aprecia el interés que este tema ha suscitado, y sigue haciéndolo, entre los lingüistas que se han propuesto caracterizar, con la mayor precisión posible, el sujeto en español, así como determinar los factores lingüísticos y extralingüísticos que condicionarían su expresión. Estas investigaciones podrían contribuir también a la enseñanza del español como segunda lengua. Por último, espero que los resúmenes de los estudios mencionados en este capítulo permitan observar la precisión metodológica alcanzada en los estudios de variación sintáctica en los últimos años, así como la creación y la aplicación de distintas variables lingüísticas para la explicación del fenómeno del sujeto variable.



### 3. METODOLOGÍA

#### 3.1. El corpus

El corpus de esta investigación está compuesto por treinta y seis (36) muestras de habla del corpus PRESEEA-Caracas 2004-2010 (Bentivoglio y Malaver 2006).<sup>41</sup> Cada una de las grabaciones del corpus tiene una duración mínima de cuarenta y cinco (45) minutos. Las grabaciones son entrevistas semidirigidas, cuyos temas son: la forma de vida de los caraqueños, las festividades, la vida familiar, el contraste entre la Caracas de ayer y la de la actualidad, entre otros. En las entrevistas participaron el hablante y dos entrevistadores, estos últimos se encargaron de que el hablante fuera lo más natural y espontáneo en su discurso (Bentivoglio y Sedano 1993; Silva-Corvalán 2001).

Los hablantes del corpus son caraqueños de padres caraqueños, distribuidos en igual proporción según los criterios extralingüísticos de sexo, edad y grado de instrucción, según se muestra en el cuadro 6.

---

<sup>41</sup> El corpus completo está compuesto por ciento ocho (108) muestras de habla, pertenecientes al Proyecto PRESEEA, el cual está financiado por el Consejo de Desarrollo Científico y Humanístico (CDCH) de la Universidad Central de Venezuela (UCV), identificado con el código PG.07.00.5760.2004, y fue coordinado por las profesoras Paola Bentivoglio, Irania Malaver y María José Gallucci de la UCV.

Cuadro 6: El corpus

20 a 34 años		35 a 54 años		55 en adelante		Total de hablantes
Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	
CARA_H13_073	CARA_M13_079	CARA_H23_085	CARA_M23_091	CARA_H33_097	CARA_M33_103	
CARA_H13_074	CARA_M13_080	CARA_H23_086	CARA_M23_092	CARA_H33_098	CARA_M33_104	
CARA_H13_075	CARA_M13_081	CARA_H23_087	CARA_M23_093	CARA_H33_099	CARA_M33_105	
CARA_H13_076	CARA_M13_082	CARA_H23_088	CARA_M23_094	CARA_H33_100	CARA_M33_106	
CARA_H13_077	CARA_M13_083	CARA_H23_089	CARA_M23_095	CARA_H33_101	CARA_M33_107	
CARA_H13_078	CARA_M13_084	CARA_H23_090	CARA_M23_096	CARA_H33_102	CARA_M33_108	
<b>12</b>		<b>12</b>		<b>12</b>		<b>36</b>

Como puede observarse en el cuadro 6, cada conversación transcrita del corpus está identificada con un código alfanumérico que debe leerse de izquierda a derecha de la siguiente forma: i) **CARA** = Caracas; ii) sexo: **h** = hombre, **m** = mujer; iii) grupo generacional: **1** = de 20 a 34 años, **2** = de 35 a 54 años, **3** = de 55 años en adelante; iv) grado de instrucción del hablante: **3** = más de quince años de escolaridad, universitario; y v) ubicación del hablante en su casilla (Bentivoglio y Malaver 2006).

En el presente estudio solo he tomado en consideración los datos de los hablantes con grado de instrucción 3, ya que cuando empecé la investigación solo las grabaciones de este nivel estaban completamente transcritas y corregidas.

### 3.2. Las exclusiones

Se excluyeron del análisis los casos cuyos sujetos gramaticales son:

i) pronombres indefinidos (*uno, alguien, algunos*, etc.), según se ilustra en (4):

(4) *algunos* se robaban las prendas y todo eso (CARA\_M23\_095)

ii) pronombres modificados por el adjetivo *mismo(a)*, como en (5):

(5) me indujo a mí a seguir estudiando más/ a/ a/ retroalimentarme de los cursos que *ellos mismos* daban en el <siglas = [CIED]> </siglas> (CARA\_H13\_074)

iii) pronombres relativos, como en (6):

(6) yo tenía un vecino *que* trabajaba con un amigo mío en una empresa publicitaria (CARA\_H13\_074)

iv) oraciones impersonales: las construidas con *se* impersonal, como en (7), y las construidas con *sujeto plural indeterminado*,<sup>47</sup> como en (8):

- (7) y también *se empezó a tomar* los liceos (CARA\_M23\_095)  
 (8) Fue la época en que *tomaban* a la Universidad (CARA\_M23\_095)

v) no referenciales como se aprecia en (9):

- (9) eso es el sacerdocio porque *mucha gente* piensa que el sacerdote solamente es el que reza (CARA\_H33\_098)

vi) obligatorios, porque hay contraste como en (10) o contraste paralelo<sup>48</sup> como en (11), respectivamente:

- (10) o que de repente *mis amigos* no podían cuando *yo* podía/ o de repente *yo* no podía y *ellos* sí podían/ (CARA\_H13\_074)  
 (11) y *yo* lo veía mal y *él* me veía (CARA\_M32\_091)

vii) mencionados dentro de una cita textual, como en (12):

- (12) <estilo directo> no/no le hagas caso a mi papá/ *mi papa tiene un carácter* </estilo directo>/ (CARA\_H33\_097)

viii) y también los sujetos de cláusulas hendidas, pseudohendidas, pseudohendidas inversas, y construcciones con el verbo *ser* focalizador, como se ilustra en los ejemplos del (13) al (16):

- (13) *fue la que* la crió a ella (CARA\_M23\_089)  
 (14) a los veinte años *fue que* Ø tuvo televisión (CARA\_H13\_075)  
 (15) él *lo que quería era* que lo reengancharan (CARA\_M33\_101)  
 (16) entonces ya *piensa es* en trabajar (CARA\_M13\_082)

---

<sup>47</sup> Esta clasificación está basada en Sedano 2011.

<sup>48</sup> Se entiende por contraste paralelo “la oposición entre dos pares de elementos de dos oraciones distintas” (Sedano 1990:39).

### 3.3. Las variables de análisis

La variable dependiente es la alternancia entre la presencia y la ausencia del sujeto de 3ª persona, tanto plural como singular, en las cláusulas con verbo conjugado. Esta variable tiene dos variantes: *presencia* y *ausencia*, según puede observarse en (17):

- (17) a. /presencia/  
 ÉL *bajaba* con sus flores (CARA\_M33\_103)  
 b. /ausencia/  
 y Ø las *vendía* aquí (CARA\_M33\_103)

En la presente investigación utilicé ocho variables independientes, seis lingüísticas y dos extralingüísticas; estas últimas son las inherentes a los hablantes del corpus: sexo y edad (ver cuadro 6 en §3.1)

Las variables lingüísticas independientes de este estudio son: i) *número gramatical*; ii) *ambigüedad*; iii) *conectividad discursiva*; iv) *tipología verbal*; v) *distancia clausular*; y vi) *turno de la conversación*.

La descripción, la explicación y los ejemplos de cada una de las variables lingüísticas independiente están en la sección 4 dedicada al análisis de los resultados, con el fin de brindar una interpretación integral de los resultados de cada una de las variables independientes explicativas.

### 3.4. Procedimientos de análisis de los datos del corpus

#### 3.4.1. Selección y codificación de los casos

En esta investigación cada texto del corpus fue segmentado en cláusulas finitas, unidades lingüísticas con un verbo conjugado en 3ª persona singular o plural. Cada uno de los sujetos de estas cláusulas es un *caso*, y el conjunto de todos ellos es el *universo de casos* del estudio. He analizado los primeros cincuenta (50) casos de sujetos cuyos referentes poseen los rasgos semánticos [+ animado] y [+ específico] y han sido mencionados al menos una vez en el discurso previo.

Extraje y enumeré manualmente todas las cláusulas con sujetos de 3ª persona del corpus, tanto los analizados (las formas variables), como los no analizados (las formas obligatorias). En los ejemplos, los sujetos explícitos aparecen en letras versales, mientras que los verbos que concuerdan con los sujetos tanto explícitos como implícitos están en cursivas. La segmentación de los textos se hizo en cláusulas, según puede apreciarse en el ejemplo (18):

- (18)
1. *MI TÍO<sub>1</sub> es más que mi tío,*
  2. *Ø<sub>2</sub> es mi amigo,*
  3. *Ø<sub>3</sub> es una persona open mind*
  4. *que<sub>4</sub> no se enrolla por tonterías*
  5. *Ø<sub>5</sub> ni anda con conflictos ¿sabes?*
  6. *una persona que<sub>6</sub> escucha.*
  7. *Ø<sub>7</sub> habla ¿sabes?*
  8. *Cuando Ø<sub>8</sub> tiene que escuchar*
  9. *Ø<sub>9</sub> escucha,*
  10. *y cuando Ø<sub>10</sub> tiene que hablar*
  11. *Ø<sub>11</sub> habla.*
  12. *Ø<sub>12</sub> es muy prudente en sus comentarios,*
  13. *cosa que todavía me falta un poco aprender,*
  14. *a veces ... no siempre lo pongo en práctica.*
  15. *E.. MI TÍO<sub>13</sub> no tuvo hijos. (CARA\_M13\_079)*

Como puede observarse en (18), la segmentación se hace a partir de la primera mención del referente sujeto, la cual se identifica con el subíndice 1 (línea 1); luego se enumera también con subíndices cada una de las siguientes menciones del mismo referente.

La forma en que puede aparecer el referente sujeto puede ser: i) una FNL como en las líneas 1 y 15; ii) un pronombre relativo, como en las líneas 4 y 6; y/o iii) anáfora cero, como en las líneas 2, 3 y 5, 7-12. Es importante señalar que los hablantes, al introducir por primera vez un nuevo referente con la función de sujeto, tienden a codificarlo explícitamente,<sup>49</sup> por lo que dicho contexto no es variable, como sí pueden serlo los siguientes.

---

<sup>49</sup> Aunque en el discurso natural son raros los casos de referentes sujetos mencionados por primera vez que sean no explícitos, sí es posible encontrarlos en los textos literarios.

Codifiqué en hojas de Excel cada uno de los casos que cumplían con los criterios establecidos en esta sección. Se entiende por codificación el procedimiento mediante el cual se rotulan las características –lingüísticas y sociales– de cada caso bajo análisis en una secuencia alfanumérica, ordenada y biunívoca, de tantos caracteres como variables de estudios haya. Este proceso es necesario para el análisis de los datos por medio del programa estadístico empleado, GOLDVARB X. Para la codificación se asignó a cada una de las variantes de las variables de estudios un código, según puede observarse en el cuadro 7. Cada uno de los casos bajo análisis está identificado con una secuencia alfanumérica constituida por los códigos de cada una de las variables. En el cuadro 8 se presenta un ejemplo de la codificación realizada.

Cuadro 7: Código de los casos según las variables de estudio

Tipo	Variables	Variantes	Código
Sociales	1. Sexo del hablante	Hombre	h
		Mujer	m
	2. Grupo generacional	20-34 años	1
35-54 años		2	
55 en adelante		3	
Lingüísticas	3. Dependiente	presencia del sujeto de 3 <sup>a</sup> persona	p
		ausencia del sujeto de 3 <sup>a</sup> persona	a
	4. Turno en la conversación	igual turno	i
		cambio de turno	c
	5. Ambigüedad	morfológicamente no ambiguo	n
		morfológicamente ambiguo	a
	6. Conectividad discursiva	grado 1	1
		grado 2	2
		grado 3	3
	7. Tipología verbal	Verbos de proceso mental	1
		Verbos de proceso relacional	2
		Verbos de proceso material	3
		Verbos de proceso verbal	4
		Verbos de proceso existencial	5
		Verbos de proceso de modulación	6
8. Distancia	Cláusulas adyacentes	1	
	Una cláusula de separación	2	
	De dos a cuatro cláusulas de separación	3	
	De cinco o más cláusulas de separación	4	
	No aplica	0	
9. Número gramatical	Singular	s	
	Plural	p	

Cuadro 8: Codificación de un caso

Código	Caso
<i>h3pin333s</i>	entonces ÉL <i>se regresó</i> pa' Boconó

El código que aparece en el cuadro 8 se lee de izquierda a derecha en la misma

secuencia en la que aparecen las variables de la columna 2 del cuadro 7: el código h3pin33p3s del caso “ÉL se regresó pa’ Boconó” indica que: i) el hablante es un hombre (h); ii) su edad está entre los 55 y más años (3); iii) el sujeto de 3ª persona es explícito (p); iv) está en el mismo turno de la conversación (i); v) el verbo es morfológicamente no ambiguo (n); vi) tercer grado de la conectividad discursiva (3); vii) el verbo es de proceso material (3); viii) la distancia referencial entre los dos sujetos es de dos a cuatro cláusulas (3) y; ix) el número del sujeto es singular (s).

#### 3.4.2. Análisis estadísticos: descriptivo e inferencial

##### 3.4.2.1. Estadística descriptiva: frecuencias absolutas y relativas y prueba del $\chi^2$

Calculé las frecuencias absolutas y relativas del universo de los casos. También, apliqué la prueba del chi cuadrado ( $\chi^2$ ), la cual es una prueba estadística que se utiliza para “determinar si existe asociación en un cuadro de contingencia (el cruce de dos variables) obtenido con datos de una muestra” (Briones 2000:118), sin calcular, no obstante, la magnitud de la relación. En las ciencias sociales, se considera que dos variables tienen una relación estadística cuando la prueba del  $\chi^2$  indica que el valor de  $p$  es menor que 0,05.

##### 3.4.2.2. Estadística inferencial

###### 3.4.2.2.1. Análisis de regla variable

El análisis de regla variable se hizo mediante el programa GOLDVARB X que

está diseñado especialmente para estudiar la variación lingüística (Sankoff, Tagliamonte y Smith 2005). Tagliamonte (2006:129) señala que: “The variable rule program [...] is a one of the most appropriate method available for conducting statistical analysis on natural speech”.<sup>50</sup> Guirado (2011:49) señala que “El procedimiento de regresión múltiple permite considerar simultáneamente todas las variables y la posible combinación de las mismas, lo que equivale a las virtuales hipótesis que podrían explicar la variación”. GOLDVARB le permite al investigador analizar tres o más variables independientes de forma simultánea a través de uno o más procedimientos de estudios estadísticos. El primer procedimiento o método directo se denomina *binomial varbrul* o *one-level*; este ofrece los porcentajes y los pesos probabilísticos de todas las variantes del estudio, así como también los errores de las celdas analizadas. Vale señalar que: “El valor de los errores da cuenta de la adecuación de los datos a un modelo teórico. Se considera que las celdas cuyo valor de error es más alto de 3,84 pueden ser interpretadas como la expresión de una pobre adecuación al modelo” (Hernández Campoy y Almeida 2005:274). Si el modelo de análisis propuesto no da cuenta del fenómeno estudiado, una solución que puede tomar el investigador es eliminar las celdas cuyos valores de error están por encima de 3,84. El segundo procedimiento del análisis que ofrece GoldVarb es el método por paso o *step-up and down*, el cual ofrece: i) un conjunto de modelos que podrían

---

50 MT: El programa de regla variable [...] es uno de los métodos más apropiados para la realización de análisis estadísticos del habla natural.

explicar los datos estudiados y ii) la selección del modelo más adecuado (Hernández Campoy y Almeida 2005:277). GOLDVARB X es una nueva versión del programa de regla variable.

Los pesos probabilísticos superiores a 0,50 indican que una o más variantes de una variable independiente favorecen significativamente una de las dos variantes de la variable dependiente, que, en este estudio, es la *presencia del sujeto de 3ª persona*. Por el contrario, las variantes cuyos pesos son menores a 0,50 desfavorecen la presencia explícita de un sujeto y las que tienen pesos iguales o cercanos a 0,50 ni favorecen ni desfavorecen la alternancia (Hernández Campoy y Almeida 2005:273; Tagliamonte 2006:252; Guirado 2011:49).

El programa también ofrece los valores de las siguientes pruebas de bondad: el *input*, el  $\chi^2$ , el logaritmo de verosimilitud y la significación del modelo:

- El *input* representa la probabilidad de que una de las variantes de la variable dependiente –en este caso un sujeto explícito de 3ª persona– aparezca en los datos analizados sin tomar en cuenta las demás variables, lingüísticas y/o extralingüísticas del estudio (Hernández Campoy y Almeida 2005:273).
- El  $\chi^2$  indica “la importancia de las diferencias entre las probabilidades obtenidas por el conjunto de factores” (Hernández Campoy y Almeida 2005:277).
- El logaritmo de verosimilitud es una medida de probabilidades, cuyo valor ideal es cero (0). Esta prueba permite que el investigador conozca hasta qué punto los datos analizados y el modelo teórico propuesto son significativos:

“la verosimilitud sirve tanto para indicar cuáles son las probabilidades más adecuadas para los factores de un grupo, como para determinar cuál es la combinación de factores que mejor se corresponde con los datos” (Moreno Fernández 1994:111).

- La *significación* es “la probabilidad de que los resultados del análisis se deban al azar. La significación nunca debe superar el umbral del .05” (Moreno Fernandez 1994:127).

#### 3.4.2.2.2. Diseño de los modelos de análisis

Hay que señalar que las variables *conectividad discursiva* y *distancia clausular* tienen algunas similitudes, puesto que ambas surgen de las reflexiones teóricas y metodológicas acerca de la variable *correferencialidad*. Por esta razón, he llevado a cabo dos análisis estadísticos separados, con el fin de que no se produjera ningún tipo de solapamiento entre las variables en el procedimiento de regresión múltiple.

El primer modelo de análisis, compuesto por todas las variables lingüísticas y extralingüísticas a excepción de la variable *distancia clausular*, lo he denominado *modelo con la variable conectividad discursiva*. Del segundo modelo se ha excluido la variable *conectividad discursiva*, y lo he denominado *modelo con la variable distancia clausular*.



#### 4. RESULTADOS Y EJEMPLOS

En el cuadro 9 se presenta la distribución en términos absolutos y relativos de los casos de sujeto variable de 3ª persona, singular y plural:

Cuadro 9: Distribución del sujeto de 3ª persona

	N	%
Presencia	853	48,3
Ausencia	913	51,7
Total	1.766	100

Como puede observarse en el cuadro 9, de las 1.766 cláusulas<sup>51</sup> con sujeto de 3ª persona –singular y plural– que he analizado, 913 tienen sujeto implícito, lo que representa el 51,7%, y 853 tienen un sujeto explícito, lo que representa el 48,3% del universo de casos analizados, con una diferencia de 3,4 puntos porcentuales. Estos resultados llaman mucho la atención puesto que el porcentaje de sujetos explícitos es bastante alto, si lo comparamos con los resultados de Quijada 2006, quien analiza también el dialecto caraqueño (ver §2.5.1.8. y el comentario después del cuadro 10). En el cuadro 10 y en el gráfico 1, presento la distribución de la presencia~ausencia del sujeto de 3ª persona en esta y en otras investigaciones:

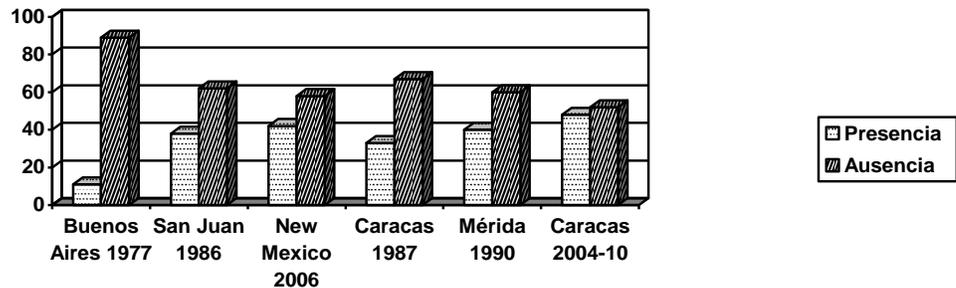
---

<sup>51</sup> Según lo propuesto en el apartado 3.2, habría debido analizar 1.800 casos (50 en cada grabación). Sin embargo, hay dos muestras de habla (CARA\_H23\_088 y CARA\_H33\_098) en las que había menos de cincuenta casos analizables. Esto se debe, quizá, a la misma naturaleza de los estudios de variación morfosintácticos, como bien lo señala Silva-Corvalán 2001.

Cuadro 10: Distribución de la presencia y ausencia del sujeto de 3ª persona en esta y otras investigaciones

Ciudad y autor		Presencia	Ausencia	Total
Buenos Aires (Barrenechea y Alonso 1977)	<i>n</i> %	55 11,31	431 88,69	486
San Juan de Puerto Rico (Morales 1986)	<i>n</i> %	237 38,28	382 61,72	619
New Mexico (Dumont 2006)	<i>n</i> %	430 42	598 58	1.028
Caracas (Quijada 2006)	<i>n</i> %	623 33	1.253 67	1.876
Mérida (Balasch 2008)	<i>n</i> %	210 40	313 60	523
Caracas (Martínez Lara 2012)	<i>n</i> %	853 48	913 52	1.766

Gráfico 1: Distribución de los sujetos explícitos e implícitos según el dialecto



En el cuadro 10<sup>27</sup> y en el gráfico 1 se muestra que, en el estudio sobre el español hablado en Buenos Aires, el porcentaje de sujetos explícitos es muy bajo (11,31%), si se compara con el del resto de los estudios, en los que esa frecuencia sobrepasa ampliamente el 30%; los dialectos de Caracas 2004-2010, New Mexico y Mérida (Venezuela) son los que tienen los porcentajes más altos de sujetos explícitos: 48%,

<sup>27</sup> Tanto los datos de Buenos Aires así como los de San Juan de Puerto Rico corresponden a los de 3ª persona singular de hablantes monolingües.

42% y 40%, respectivamente. Es importante hacer una acotación sobre los porcentajes de sujetos explícitos del presente estudio (48%) y los de Quijada 2006 (33%) en la muestra del CSC'87, cuya diferencia es de 15 puntos porcentuales. Esto podría indicar un incremento significativo de uso de sujetos explícitos en el dialecto caraqueño en el lapso de tiempo transcurrido entre las dos muestras, aun cuando creo que no se trata de un cambio lingüístico, como parecería, sino más bien de las diferencias entre las muestras del CSC'87 y la de PRESEEA-CARACAS 2004-2010. Este es un problema que vale la pena estudiar con más detenimiento en futuras investigaciones.

#### 4.1. Frecuencias, prueba de $\chi^2$ y ejemplos

En este apartado presento las variables lingüísticas y extralingüísticas usadas en esta investigación. Describo: i) la variable y el objetivo que se persigue con ellas, y presento los ejemplos ilustrativos; ii) el cuadro con las frecuencias absolutas y relativas, los resultados de la prueba del  $\chi^2$  (chi cuadrado) y un gráfico ilustrativo, en los casos que así lo ameriten; y iii) los comentarios acerca de los resultados obtenidos; los problemas hallados y sus alcances.

##### 4.1.1. Número gramatical

Con esta variable me he propuesto verificar si en el habla actual de Caracas el número gramatical (singular/plural) puede condicionar la expresión variable del sujeto de 3ª persona. Las variantes de esta variable son: singular, en (19), y plural, en (20):

- (19) Singular
- a. /presencia/  
MI MAMÁ *atravesaba* por una situación bastante difícil en su vida (CARA\_M23\_084)
- b. /ausencia/  
y Ø *cayó* en una depresión (CARA\_M23\_084)
- (20) Plural
- a. /presencia/  
MIS SOBRINOS me adoran/ muy cómico/ (CARA\_H13\_078)
- b. /ausencia/  
o sea, digamos que logro como que ponerme a su nivel y<alargamiento/> Ø [mis sobrinos] *se divierten*/ (CARA\_H13\_078)

En el cuadro 11 presento los resultados obtenidos:

Cuadro 11: Presencia vs. ausencia del sujeto de 3ª persona según NÚMERO GRAMATICAL

		Presencia	Ausencia	Total	%
Singular	<i>n</i>	<b>748</b>	718	1.466	83
	%	<b>51</b>	49		
Plural	<i>n</i>	105	195	300	17
	%	35	65		
Total	<i>n</i>	853	913	1.766	100
	%	48,3	51,7		
$\chi^2 = 25,60 > 3,841$		$p = 0,0000 < 0,05$			

En el cuadro 11 se aprecia que no hay una gran diferencia porcentual entre la frecuencia de sujetos explícitos e implícitos cuando estos están en singular (51% vs. 49%, respectivamente). Por el contrario, cuando el sujeto es plural, el porcentaje de implícitos alcanza el 65%. La prueba del  $\chi^2$  indica que la variable dependiente y la variable *número gramatical* están relacionadas estadísticamente.

En el cuadro 12 presento los resultados de la variable *número gramatical* en los estudios hechos sobre el habla de dos dialectos de Venezuela (Caracas y Mérida):

Cuadro 12: Presencia y ausencia del sujeto de 3<sup>a</sup> persona en Venezuela según la variable NÚMERO GRAMATICAL

Número	Estudios		Presencia %	Ausencia %
Singular	Quijada 2006	Caracas 1987	30	70
	Balash 2008	Mérida 1995	40	60
	Martínez Lara 2012	Caracas 2004-2010	51	49
Plural	Quijada 2006	Caracas 1987	35	65
	Balash 2008	Mérida 1995	43	57
	Martínez Lara 2012	Caracas 2004-2010	35	65

El cuadro 12 muestra las semejanzas y las diferencias de uso de los sujetos explícito de 3<sup>a</sup> persona según el número gramatical en los tres estudios hechos sobre dos dialectos de Venezuela. En cuanto a los sujetos singulares, puede observarse que hay: i) hay un aumento paulatino de presencia explícita de sujetos desde el primer estudio hecho sobre los dos dialectos venezolanos hasta el presente (30%, 40% y 51%, respectivamente); y, ii) hay un mayor porcentaje de sujetos explícitos en la muestra actual de Caracas (51%) que en la de hace veinte años (30%). Los sujetos plurales explícitos, por su parte, tanto en el estudio de Quijada 2006 como en el de Martínez Lara 2012, tienen porcentajes iguales (35%); en cambio, en el estudio de Balash 2008, sobre el dialecto de Mérida, el porcentaje de sujetos plurales explícitos es mayor (43%).

Resumiendo, en los dos corpus del habla de Caracas, los sujetos explícitos plurales tienen la misma frecuencia de uso (35% / 65%).

#### 4.1.2. Ambigüedad

El objetivo de esta variable es determinar si el factor *ambigüedad* condiciona la alternancia entre la presencia y la ausencia del sujeto de 3ª persona.<sup>28</sup> Parto del supuesto de que el sujeto tiende a ser implícito cuando el verbo de la cláusula bajo análisis no es morfológicamente ambiguo y explícito cuando el verbo es morfológicamente ambiguo. Esta variable tiene dos variantes, no ambiguo y morfológicamente ambiguo, como se ilustra en (21) y (22):

- (21) No ambiguo
- a. /presencia/  
ELLA *salió* corriendo (CARA\_M33\_105)
  - b. /ausencia/  
y Ø *fue* a buscar a su papá y Ø lo *encontró* entre los que habían declarado muertos (CARA\_M33\_105)
- (22) Morfológicamente ambiguo
- a. /presencia/  
MI VECINO *era* mi partero en el campo petrolero (CARA\_M33\_108)
  - b. /ausencia/  
y Ø *había recibido* una herencia importante (CARA\_M33\_105)

En el cuadro 13 aparecen los resultados obtenidos:

---

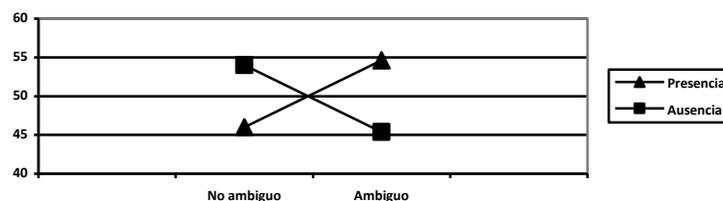
<sup>28</sup> Hay que acotar que cuando en una misma cláusula o en cláusulas adyacentes o cercanas hay dos referentes que pueden funcionar como sujetos de un mismo verbo es posible que se cree un contexto ambiguo. En el presente estudio no he codificado ningún caso *morfológica y contextualmente ambiguo*, porque estos son muy escasos en las muestras analizadas, lo que cuantitativa y estadísticamente constituye un problema para el estudio.

Cuadro 13: Presencia vs. ausencia del sujeto de 3ª persona según AMBIGÜEDAD

		Presencia	Ausencia	Total	%
Ambiguo	<i>n</i>	<b>263</b>	219	482	27,3
	%	<b>54,6</b>	45,4		
No ambiguo	<i>n</i>	590	694	1.284	72,7
	%	46	54		
Total	<i>n</i>	853	913	Total	100
	%	48,3	51,7		
$\chi^2 = 10,41 > 3,841$		$p = 0,0015 < 0,05$			

El cuadro 13 muestra que hay una mayor tendencia a que un sujeto de 3ª persona se mencione explícitamente cuando el verbo de la cláusula es morfológicamente ambiguo; por el contrario, hay más sujetos implícitos cuando la morfología del verbo no es ambigua. La prueba de  $\chi^2$  indica que la relación entre la variable dependiente y la *ambigüedad* no se debe al azar en esta muestra. Los resultados corroboran la hipótesis de que un sujeto tiende a aparecer de forma explícita cuando el verbo tiene morfología ambigua; además, confirman lo postulado por las gramáticas tradicionales, así como por las investigaciones previas en las que se ha comprobado esta relación (ver cuadro 1). Estos resultados se ilustran en el gráfico 2:

Gráfico 2: Presencia~ausencia del sujeto de 3ª persona según AMBIGÜEDAD



En el cuadro 14 se comparan los resultados de esta investigación con otras dos llevadas a cabo en el español de Venezuela:

Cuadro 14: Comparación de los porcentajes de uso de la presencia y la ausencia del sujeto de 3ª persona en Venezuela según AMBIGÜEDAD

Ambigüedad	Estudios		Presencia %	Ausencia %
Morfológicamente ambiguo	Quijada 2006	Caracas 1987	37	53
	Balash 2008	Mérida 1995	36	64
	Martínez Lara 2012	Caracas 2004-2010	54,6	45,4
Morfológicamente no ambiguo	Quijada 2006	Caracas 1987	33	67
	Balash 2008	Mérida 1995	42	58
	Martínez Lara 2012	Caracas 2004-2010	46	54

Según puede verse en el cuadro 14, tanto en el estudio de Quijada 2006 como en el de Balash 2008, los porcentajes de sujetos explícitos son similares (37% y 36%, respectivamente) cuando la morfología del verbo es ambigua. Por el contrario, en la presente investigación, la frecuencia de sujetos explícitos en contextos con verbos morfológicamente ambiguos es más alta (54,6%). Vale acotar, además, que en este estudio la presencia de un sujeto es directamente proporcional a la ambigüedad de la morfología del verbo (54,6% vs. 45,4%), lo que no se da en Quijada 2006 (37% vs. 53%).

#### 4.1.3. Conectividad discursiva

He usado la variable con las tres variantes propuestas por Quijada 2006, ya que tuve que tomar las mismas decisiones metodológicas de la investigadora. Parto de la hipótesis de que debería haber una mayor frecuencia de sujetos explícitos

cuando la conexión entre cláusulas con sujetos correferenciales es interrumpida y más sujetos implícitos o tácitos cuando la conexión es óptima. Las tres variantes de la *Conectividad discursiva* son:

- a) *Conectividad 1* u *óptima* (este grado reúne los grados 1 y 2 de la propuesta de Paredes Silva 1993): el sujeto de la cláusula bajo análisis es correferencial con el sujeto de la cláusula precedente, independientemente de su forma así como del modo, tiempo y aspecto del verbo, según se muestra en (23):

(23) *Conectividad 1*

- a. /presencia/

mi mamá es comerciante/ este<alargamiento/> ELLA fue presidenta de Comprebar (CARA\_H13\_077)

- b. /ausencia/

Ø aprendió a manejar ahí mismo y<alargamiento/>/ y Ø decía que le había costado ochocientos bolos una cuestión así (CARA\_H23\_089)

- b) *Conectividad 2* (corresponde al grado 4 de Paredes Silva 1993): el sujeto de la cláusula no es correferencial con el sujeto de la cláusula precedente, pero sí lo es con otro elemento lingüístico que cumple una función sintáctica distinta a la de sujeto (objeto directo, indirecto, complemento circunstancial, atributo), según se observa en (24):

(24) *Conectividad 2*

- a. /presencia/

entonces me abracé con ella y ELLA se puso a llorar conmigo (CARA\_H13\_073)

- b. /ausencia/

y bueno, el hecho es que conocí al tipo, al chamito y Ø es igualito al abuelo (CARA\_H13\_077)

- c) *Conectividad 3* (está compuesta por los grados 3, 5 y 6 de Paredes Silva 1993): el sujeto de la cláusula no es correferencial con el sujeto de la cláusula precedente ni está presente en alguna estructura lingüística que cumpla otra función sintáctica. Este grado de conectividad puede darse cuando en la interacción confluyen factores como la introducción de una aclaración, el cambio de tema de conversación o un cambio de turno de habla, según se ilustra en el ejemplo (25):

## (25) Conectividad 3

## a. /presencia/

mi mamá estaba desesperada de que ya uno terminara porque definitivamente uno cuando está estudiando es u<palabra\_cortada/>/ cuando no tiene trabajo es una carga/ por eso MI MAMÁ en cada vacación nos buscaba trabajo/ (CARA\_M23\_092)

## b. /ausencia/

entonces Ø me dijo/ <cita> bueno/ pero/ ¿por qué no propones eso en Radio Caracas? porque ahorita están un poco<alargamiento/>/ buscando ese tipo de cosas por la Ley Resorte y todo lo demás </cita> entonces/ bueno/ e<alargamiento/> sin/ sin/ reflexionar mucho Ø me consiguió una cita con/con/ con la/ la productora de Radio Caracas (CARA\_M23\_093)

En el cuadro 15 se presentan los resultados obtenidos al cruzar la variable dependiente con la *Conectividad discursiva*:

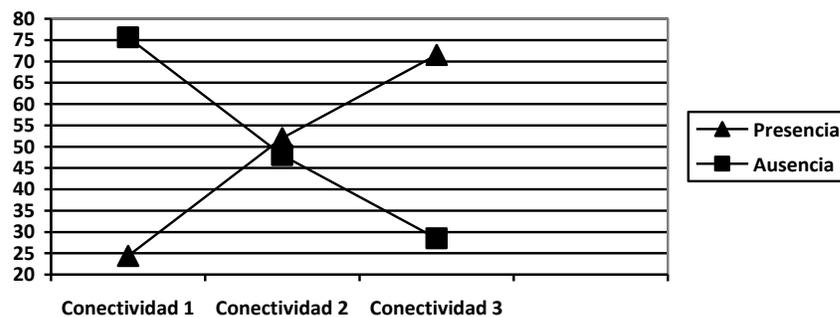
Cuadro 15: Presencia vs. ausencia del sujeto de 3ª persona según CONECTIVIDAD DISCURSIVA

		Presencia	Ausencia	Total	%
Conectividad 3	<i>n</i>	<b>557</b>	222	779	44,1
	%	<b>71,5</b>	28,5		
Conectividad 2	<i>n</i>	103	97	200	11,3
	%	51,5	48,5		
Conectividad 1	<i>n</i>	193	594	787	44,6
	%	24,4	75,6		
Total	<i>n</i>	853	913	Total	100
	%	48,3	51,7		
$\chi^2 = 346,93 > 5,991$		2 g.d.1	$p = 0,0000 < 0,05$		

En el cuadro 15 se observa que la presencia del sujeto de 3ª persona es inversamente proporcional al grado de conectividad, puesto que la presencia de un sujeto explícito aumenta en la medida en que disminuye el grado de conectividad entre las cláusulas (de 24,4% a 71,5%); lo contrario, obviamente, ocurre cuando el sujeto es implícito (de 75,6% a 28,5). La prueba del  $\chi^2$  muestra que hay una relación estadística entre la

variable dependiente y la *Conectividad discursiva*, ya que el valor del  $\chi^2$  es mayor al valor esperado ( $347,93 > 5,991$ ) y  $p$ , menor que 0,05. Estos resultados comprueban la hipótesis de que un sujeto explícito es más frecuente cuando hay una interrupción entre las cláusulas, y uno implícito lo es cuando hay una mayor conectividad. El gráfico 4 ilustra mejor estos resultados:

Gráfico 3: Presencia~ausencia del sujeto de 3ª persona según CONECTIVIDAD DISCURSIVA



En el cuadro 16 se comparan los resultados de la ausencia~presencia del sujeto de 3ª persona según la variable *Conectividad discursiva* en las dos investigaciones sobre el habla de Caracas:

Cuadro 16: Porcentajes de la presencia y ausencia del sujeto de 3ª persona en Caracas según CONECTIVIDAD DISCURSIVA

Conectividad discursiva	Estudios		Presencia %	Ausencia %
	Quijada 2006 Martínez Lara 2012	Caracas 1987 Caracas 2004-2010		
Grado 1	Quijada 2006	Caracas 1987	26	74
	Martínez Lara 2012	Caracas 2004-2010	24,4	75,6
Grado 2	Quijada 2006	Caracas 1987	44	56
	Martínez Lara 2012	Caracas 2004-2010	51,5	48,5
Grado 3	Quijada 2006	Caracas 1987	48	52
	Martínez Lara 2012	Caracas 2004-2010	71,5	28,5

En el cuadro 16 se observa que, en la investigación de Quijada 2006, la frecuencia de sujetos explícitos no supera el 50% en ninguno de los tres grados de conexión (26%, 44% y 48%). Estos resultados se asemejan mucho a los de la presente investigación en lo que atañe al grado 1 de *Conectividad discursiva*. Esta semejanza no se mantiene, sin embargo, en los resultados relacionados con el grado 3 de la conexión, ya que en el corpus más reciente (2004-2010) la frecuencia de sujetos explícitos alcanza 71,5%, un porcentaje más alto (23,5 puntos porcentuales) que el encontrado por Quijada en el corpus de 1987 (48%). Esto quizá se deba a que en el corpus PRESEEA-Caracas 2004-10 puede haber más cambios de tópico y de turno, así como la introducción de nuevos referentes sujetos y cláusulas impersonales que en el CSC'87. Juzgo inconveniente, sin embargo, adelantar conclusiones no sustentadas por un amplio estudio acerca de las posibles diferencias discursivas en ambas muestras.

En el gráfico 4 se ilustran los resultados de los dos estudios:

Gráfico 4: Porcentaje de la presencia explícita de un sujeto de 3ª persona según *Conectividad discursiva* en Quijada 2006 y Martínez Lara 2012



#### 4.1.4. Tipología verbal

El objetivo que me propongo al utilizar esta variable es determinar si la semántica de los verbos puede condicionar la manifestación variable del sujeto de 3ª persona, singular y plural. Las variantes son los seis tipos de procesos propuestos por el grupo de investigación ADESSE (§2.4.4.): mental, relacional, material, verbal, existencial y de modulación, ilustrados en los ejemplos del (26) al (31):

##### (26) Verbos de procesos mentales

- a. /presencia/  
ELLA *estudia* canto lírico (CARA\_H33\_101)
- b. /ausencia/  
no te rías/ Ø *creían* que las íbamos a secuestrar (CARA\_M13\_084)

##### (27) Verbos de procesos relacionales

- a. /presencia/  
ELLA *tiene* diez hijos (CARA\_M33\_103)
- b. /ausencia/  
Ø *tiene* nueve hembras (CARA\_M33\_103)

- (28) Verbos de procesos materiales
- a. /presencia/  
ÉL *toca* la ventana de mi cuarto todas las mañanas (CARA\_M13\_084)
- b. /ausencia/  
y Ø *toca* la ventana (CARA\_M13\_084)
- (29) Verbos de procesos verbales
- a. /presencia/  
EL TIPO me *dice* <cita>pana/ mira o sea/ no podemos reconstruirle el rostro por si lo quieren ver ¿no?/ (CARA\_H13\_073)
- b. /ausencia/  
y Ø me *dice* <cita> bueno/ Juancho/ si quieres te acompaño </cita> (CARA\_H13\_073)
- (30) Verbos de procesos existenciales
- a. /presencia/  
I: sí/ MI PAPÁ *nació* aquí en Caracas igual que mi mamá/ (CARA\_H13\_077)
- b. /ausencia/  
Ø *murió* hace<alargamiento/> ocho años/ por cáncer de pulmón creo/ sí/ (CARA\_H13\_077)
- (31) Verbos de modulación
- a. /presencia/  
ÉL *trata* de conseguir la ropa que yo tengo (CARA\_H13\_074)
- b. /ausencia/  
Ø *se deja* influenciar conscientemente de/ de/ de su entorno/ (CARA\_H13\_074)

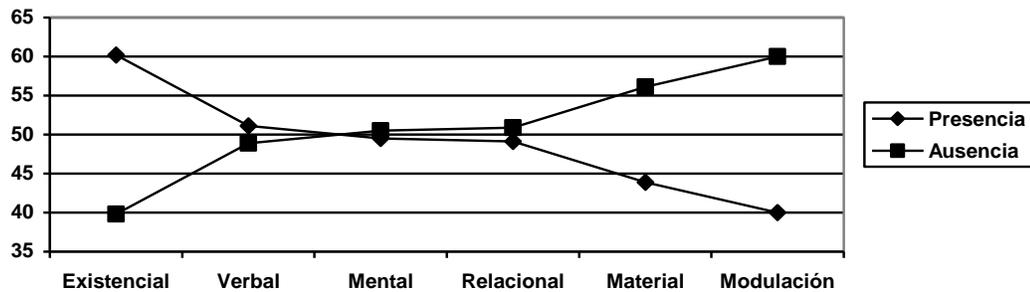
En el cuadro 17 se presentan los resultados del cruce entre la variable dependiente y la *Tipología verbal*:

Cuadro 17: Presencia vs. ausencia del sujeto de 3ª persona según TIPOLOGÍA VERBAL

Verbos		Presencia	Ausencia	Total	%
De procesos existenciales	<i>n</i> %	<b>77</b> <b>60,2</b>	51 39,8	128	7,2
De procesos verbales	<i>n</i> %	113 51,1	108 48,9	221	12,5
De procesos mentales	<i>n</i> %	104 49,5	106 50,5	210	11,9
De procesos relacionales	<i>n</i> %	295 49,1	306 50,9	601	34
De procesos materiales	<i>n</i> %	244 43,9	312 56,1	556	31,5
De modulación	<i>n</i> %	20 40	30 60	50	2,8
Total	<i>n</i> %	853 48,3	913 51,7	1.766	100
$\chi^2 = 13,91 > 11,070$		5 g.d.l		$p = 0,0161 < 0,05$	

En el cuadro 17 puede apreciarse claramente que los verbos de procesos existenciales son los que tienden a aparecer con un sujeto explícito ( $77/128 = 60,2\%$ ); los verbos de procesos verbales, mentales y relacionales tienen sujetos explícitos e implícitos casi con la misma proporción (51,1%, 49,5% y 49,1%, respectivamente), mientras que los verbos de procesos materiales y de modulación favorecen los sujetos implícitos (43,9% y 40%). La prueba del  $\chi^2$  indica que hay una relación estadística entre la variable dependiente y la *tipología verbal*. En el gráfico 5 se ilustran estos resultados:

Gráfico 5: Presencia~ausencia del sujeto de 3ª persona según TIPOLOGÍA VERBAL



#### 4.1.5. Distancia clausular

He empleado esta variable con el fin de determinar si la distancia clausular entre los sujetos correferenciales podría condicionar la manifestación variable del sujeto. Parto del supuesto de que una mayor distancia entre cláusulas con sujetos correferenciales favorece la presencia de un sujeto explícito. Las variantes de esta variable son cuatro, según los sujetos bajo análisis estén en cláusulas: i) adyacentes; ii) separadas por una cláusula; iii) separadas por dos, tres o cuatro cláusulas; y, iv) separadas por cinco o más cláusulas, lo que se ilustra en los ejemplos del (32) al (35):

#### (32) Cláusulas adyacentes

##### a. /presencia/

MI MAMÁ *se separó* de mi papá/ MI MAMÁ *se fue*<alargamiento/> a vivir a<alargamiento/> Valencia (CARA\_H13\_077)

##### b. /ausencia/

Ø *compra* dos cajas de cerveza/ Ø *compra* parrilla/ (CARA\_H13\_073)

- (33) Una cláusula de separación
- a. /presencia/  
y ÉL *está* por publicar en enero/ en Colombia le van a publicar su/ un primer libro de poemas de él<sub>1</sub> porque ÉL siempre *ha escrito* narrativa/ (CARA\_H13\_076)
- b. /ausencia/  
y Ø *está haciendo* un postgrado/ de paso la tesis la trae ¡uf!/ ¡loca!/ ¡pobrecita!<sub>1</sub> ¡Ø no *duerme!* (CARA\_M33\_104)
- (34) De dos a cuatro cláusulas de separación
- a. /presencia/  
ÉL *estaba preocupado* porque nosotras estábamos ahí<sub>1</sub> y no nos pasara nada<sub>2</sub> pues al mismo tiempo/ yo estaba preocupada<sub>3</sub> porque ÉL *estuviera* fuera (CARA\_M13\_081)
- b. /ausencia/  
entonces Ø *tenía* un revolver/ entonces e<(:)>/ <estilo directo> mire<sub>1</sub> <(:)> usted </estilo directo>/ tenía ya como/ dos años de/ de/ <palabra cortada> ma <palabra cortada>/ de amores con mi novia/ con mi esposa actual<sub>2</sub> y entonces Ø me *dijo* e<(:)> <estilo directo> si usted no se casa con mi hija lo mato </estilo directo>/ (CARA\_H33\_097)
- (35) De cinco o más cláusulas de separación
- a. /presencia/  
y *dice* MI MAMÁ que estaba muy chiquita<sub>1</sub> todavía no hablaba muy claro<sub>2</sub> pero estábamos sentados en la piscina frente al bar<sub>3</sub> y yo pasé todo el día pidiéndole guindas al señor del bar<sub>4</sub> él me las metía en un vasito<sub>5</sub> y MI MAMÁ me *decía* <cita>pero ¡ya! </cita> (CARA\_M13\_079)
- b. /ausencia/  
y Ø me *empezó a hacer* como una especie de muestreo <cita> ¿hablas italiano?<sub>1</sub> ¿tienes pasaporte?<sub>2</sub> </cita>/ yo no sé qué más cosas/<sub>3</sub> a la postre yo tuve que/que intervenir en la conversación<sub>4</sub> y le dije<sub>5</sub> <cita> bueno/ me está preguntando mucho<sub>6</sub> ¿qué quiere conmigo?<sub>7</sub> </cita>/ <risas =E> entonces Ø me *dijo* que era que quería mandarme a<(:)> Roma a estudiar/ (CARA\_H33\_098)

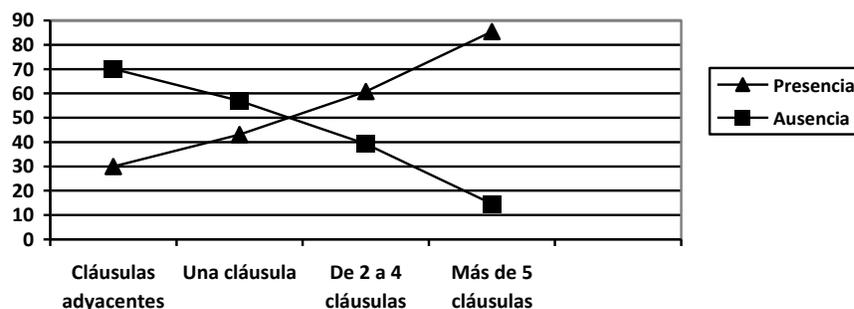
En el cuadro 18 se aprecian los resultados de la variable *Distancia clausular*:

Cuadro 18: Presencia vs. ausencia del sujeto de 3ª persona según  
DISTANCIA CLAUSULAR

		Presencia	Ausencia	Total	
De cinco o más cláusulas de separación	<i>n</i> %	<b>345</b> <b>85,4</b>	59 14,6	404	22,9
De dos a cuatro cláusulas de separación	<i>n</i> %	138 60,8	89 39,2	227	12,9
Una cláusula de separación	<i>n</i> %	100 43,1	132 56,9	232	13,1
Cláusulas adyacentes	<i>n</i> %	270 29,9	633 70,1	903	51,1
Total	<i>n</i> %	853 48,3	913 51,7	1.766	100
$\chi^2 = 361,76 > 11,070$		4 g.d.l		$p = 0,0000 < 0,05$	

Los resultados del cuadro 18 confirman la hipótesis de que el porcentaje de presencia explícita de un sujeto crece en la medida en que aumenta la distancia entre los dos sujetos correferenciales (de 30% a 85%); e, inversamente, el porcentaje de ausencia de un sujeto explícito disminuye (de 70% a 15%), si la distancia entre los dos es menor. La prueba del  $\chi^2$  indica que hay una relación estadística entre la variable dependiente y la *distancia clausular* ( $p = 0,0000$ ), lo que permite rechazar la hipótesis nula. En el gráfico 6 se ilustran estos resultados:

Gráfico 6: Presencia~ausencia del sujeto de 3ª persona  
según DISTANCIA CLAUSULAR



#### 4.1.6. Turno de la conversación

El objetivo que me he propuesto con esta variable es determinar si el cambio de turno en la interacción es un factor que puede condicionar la manifestación variable del sujeto. Parto del supuesto de que los hablantes mencionan el sujeto explícitamente al retomar el turno de habla, como lo han propuesto algunos autores, entre ellos Bentivoglio (1987:39). Esta variable tiene dos variantes: igual turno vs. cambio de turno, como se ilustra en (36) y (37), respectivamente:<sup>29</sup>

(36) Igual turno

a. /presencia/

Mi primito tenía catorce años/ y ÉL *se puso* a manipular el arma de su papá (CARA\_H13\_073)

b. /ausencia/

entonces como son <énfasis> tan bonitas </énfasis> y Ø *perdonan* y tal (CARA\_H13\_075)

(37) Cambio de turno

a. /presencia/

*Hab.:* de esa señora que salvó a su papá/ tenía un baúl muy grande de doble fondo y en ese doble fondo guardaba unas <palabra cortada> toroco <palabra cortada>/¿cómo se llama? unas/ eso/ no/ unas ¿cómo es que se dice eso?/ una<(:)>s ¿las monedas esas de oro?

*E2:* morocotas/

*Hab.:* unas morocotas/ las morocotas/ y ELLA *sacaba* las morocotas/ (CARA\_M33\_105)

b. /ausencia/

*Habl.:* respirar <risas = "todos"/>

*E1:* no/ pero <risas= > ¿qué hace en la calle todo el día?

*Habl.:* ¿en la calle? este<alargamiento/> bueno/ si Ø *no está* con mi tía Ø *está* en la casa de la playa/ (CARA\_H13\_075)

---

<sup>29</sup> No he considerado como interrupciones del turno de habla del entrevistado las pequeñas interjecciones (*ujum, ajá, ok*, etc.) hechas por uno de los investigadores o por cualquier otra persona presente en la entrevista, puesto que tales expresiones son un tipo de estímulo, mas no interrupciones o turnos de habla (Bentivoglio 1987:40).

A continuación presento el cuadro 19 con los resultados de la variable *Turno de la conversación*:

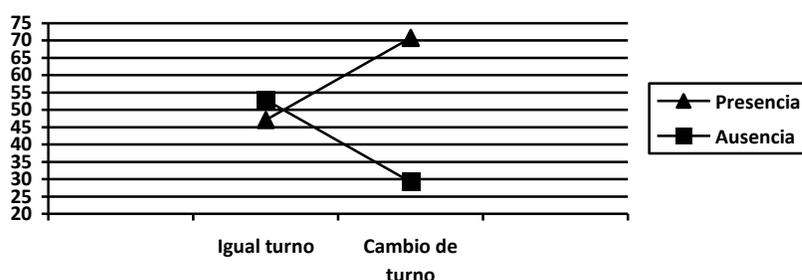
Cuadro 19: Presencia vs. ausencia del sujeto de 3ª persona según TURNO DE LA CONVERSACIÓN

		Presencia	Ausencia	Total	%
Cambio de turno	<i>n</i>	<b>63</b>	26	89	5
	%	<b>70,8</b>	29,2		
Igual turno	<i>n</i>	790	887	1.677	95
	%	47,1	52,9		
Total	<i>n</i>	853	913	1.766	100
	%	48,3	51,7		
$\chi^2 = 18,98 > 3,841$		$p = 0,0000 < 0,05$			

En el cuadro 19 se observa que la mayor parte de las cláusulas analizadas (1.677 = 95%) están dentro de un mismo turno de habla y que en este caso casi no hay diferencia entre la presencia vs. ausencia del sujeto: 47,1% vs. 52,9%, respectivamente. Por el contrario, la tendencia a mencionar explícitamente el sujeto alcanza casi el 71% cuando hay un cambio de turno en la conversación; vale acotar que los casos de cambio de turno representan el 5% de la muestra analizada. Según la prueba del  $\chi^2$  hay una relación entre la variable dependiente y el *turno de la conversación*, lo que permite afirmar que los resultados no se deben al azar. Se valida, entonces, la hipótesis propuesta por Bentivoglio (1987:39), de que el emisor, al retomar el turno de habla, tiende a mencionar el sujeto bajo la forma de un pronombre o de una FNL.

En el gráfico 7 se ilustran estos resultados:

Gráfico 7: Presencia~ausencia del sujeto de 3ª persona según TURNO DE LA CONVERSACIÓN



En el cuadro 20 se comparan los resultados de los dos estudios sobre el español de Caracas (Quijada 2006 y el presente):

Cuadro 20: Porcentaje de la presencia y ausencia del sujeto de 3ª persona en Caracas según el CAMBIO DE TURNO

Turno de la conversación	Estudios		Presencia	Ausencia
			%	%
Igual turno	Quijada 2006	Caracas 1987	32	67
	Martínez Lara 2012	Caracas 2004-2010	47,1	52,9
Cambio de turno	Quijada 2006	Caracas 1987	51	49
	Martínez Lara 2012	Caracas 2004-2010	70,8	29,2

Según puede apreciarse en el cuadro 20, en la muestra de Caracas de 1987 como en la actual, los hablantes mencionan explícitamente el sujeto con más frecuencia cuando hay un cambio de turno en la conversación; sin embargo, la frecuencia de sujetos explícitos en el presente estudio es mayor que en el de Quijada 2006 (70,8% y 51%, respectivamente). Las razones para este aumento no son explicables fácilmente; el mismo podría deberse al hecho de que en el corpus más reciente hay un mayor número de turnos de habla que en el anterior, lo que podría dar lugar a una mayor

presencia de sujetos explícitos. Esta hipótesis solo puede confirmarse haciendo un recuento de los turnos de habla en ambas muestras, asunto al que dedicaré mi atención en un futuro próximo.

#### 4.1.7. Variables extralingüísticas: Sexo y Edad

En los cuadros 21 y 23 puede observarse la distribución de los porcentajes de la ausencia~presencia del sujeto según las variables extralingüísticas *Sexo* y *Edad*.

Cuadro 21: Presencia vs. ausencia del sujeto de 3ª persona según SEXO

		Presencia	Ausencia	Total	%
Hombres	<i>n</i>	404	462	866	49
	%	46,7	53,3		
Mujeres	<i>n</i>	449	451	900	51
	%	49,9	50,1		
Total	<i>n</i>	853	913	1.766	100
	%	48,3	51,7		
$\chi^2 = 1,85 < 3,841$		$p = 0,1734 > 0,05$			

En el cuadro 21 se observa que no hay una gran diferencia porcentual entre la presencia y la ausencia del sujeto según la variable *sexo*, pues solo los hombres muestran 6,6 puntos porcentuales de diferencia entre ambas variantes (46,7% y 53,3%), mientras que el porcentaje obtenido por las mujeres es prácticamente igual. La prueba del  $\chi^2$  confirma que no hay una relación estadística entre las dos variables y que debe aceptarse la hipótesis nula.

A continuación se presenta un cuadro comparativo entre los resultados de Quijada 2006 y los de este estudio:

Cuadro 22: Distribución de la presencia y ausencia del sujeto de 3ª persona en la investigación de Quijada 2006 y en esta, según la variable SEXO

Sexo	Estudios		Presencia %	Ausencia %
Hombres	Quijada 2006	Caracas 1987	30	70
	Martínez Lara 2011	Caracas 2004-2010	46,7	53,3
Mujeres	Quijada 2006	Caracas 1987	35	65
	Martínez Lara 2011	Caracas 2004-2010	49,9	50,1

En el cuadro 22 se muestra que, en ambos estudios, tanto los hombres como las mujeres mencionan un sujeto explícito de 3ª persona casi con la misma frecuencia, aun cuando las mujeres son las que los usan más. En el estudio de Quijada 2006, los porcentajes de uso de sujetos implícitos son bastante altos que los sujetos implícitos, en ambos grupos. En cambio, en el presente estudio, los porcentajes de sujetos implícitos y explícitos de hombres y mujeres son muy similares.

En el cuadro 23 pueden observarse los resultados obtenidos de acuerdo con la variable *edad*:

Cuadro 23: Presencia vs. ausencia del sujeto de 3ª persona según EDAD

		Presencia	Ausencia	Total	%
De 55 en adelante	<i>n</i>	303	280	583	33
	%	52	48		
De 35 a 54 años	<i>n</i>	277	306	583	33
	%	47,5	52,5		
De 20 a 34 años	<i>n</i>	273	327	600	34
	%	45,5	54,5		
Total	<i>n</i>	853	913	1.766	
	%	48,3	51,7		
$\chi^2 = 5,18 < 5,991$		2 g.d.l		$p = 0,75118 > 0,05$	

En el cuadro 23 puede apreciarse que no hay una gran diferencia entre la presencia~ausencia de un sujeto según la edad de los hablantes. La prueba del  $\chi^2$  indica que no hay una relación estadística entre la variable dependiente y la *Edad*.

#### 4.1.8. Resumen de los resultados del análisis descriptivo

En los apartados del 4.1.1. al 4.1.7. he presentado cada una de las variables independientes –lingüísticas y extralingüísticas–, los ejemplos respectivos, las frecuencias absolutas y relativas, así como los resultados de la prueba del  $\chi^2$ .

Según los resultados obtenidos en esta primera parte, puede decirse que en la muestra analizada los sujetos explícitos tienen una frecuencia de uso de 48% y los implícitos, 52%.

Si se toma en cuenta cada una de las variables independientes por separado, las frecuencias relativas más altas de sujetos explícitos las tienen las variables lingüísticas *distancia clausular* (85,4%), *conectividad discursiva* (71,5%), y *turno de la conversación* (70,8%).

La prueba del  $\chi^2$  muestra que todas las variables lingüísticas independientes tienen una relación estadística con la variable dependiente, contrariamente a lo que ocurre con las variables extralingüísticas, cuyos valores de  $\chi^2$  son superiores a 0,05, por lo que su distribución debe considerarse debida al azar.

#### 4.2. Análisis con GoldVarb X

El análisis estadístico de los datos ha sido realizado mediante el programa GOLDVARB X, el cual indica la magnitud de la relación o el peso probabilístico entre la variable dependiente y todas las variables independientes en conjunto (Paolillo 2002; Hernández Campoy y Almeida 2005; Tagliamonte 2006). Los cálculos hechos con GOLDVARB X me han permitido establecer la contribución que las variables independientes tienen, en conjunto, sobre la variante de la variable dependiente *presencia del sujeto de 3ª persona*.

Como se ha indicado antes, con dos de las variables lingüísticas (*conectividad discursiva* y *distancia clausular*) se busca explicar el fenómeno estudiado como parte de una estrategia de cohesión textual usada por el hablante. Estas dos variables tienen similitudes en su naturaleza discursiva, por lo cual no es posible considerarlas juntas en el análisis multivariante con el programa GOLDVARB X (ver §3.4.2.2.1.). Por tal motivo, fue necesario crear dos modelos para determinar la relación en conjunto de las variables propuestas en esta investigación (ver §3.4.2.2.2.). Un primer modelo incluye, dentro del conjunto de todos los factores que contribuyen al fenómeno, solo la variable *conectividad discursiva*. En un segundo modelo, junto al grupo de factores estudiados, se incluye solo la variable *distancia clausular*. A continuación, presento los resultados de estos dos análisis.

#### 4.2.1. Modelo con la variable *conectividad discursiva*

En el cuadro 24 se presentan las variables que el programa GOLDVARB X ha seleccionado como significativas para la *presencia del sujeto de 3ª persona*. Estas variables aparecen ordenadas de forma jerárquica de mayor a menor según el valor de su respectivo rango. En el cuadro se resaltan en negritas las variantes con un peso probabilístico mayor a 0,50.

Cuadro 24: Resultados del análisis con GLODVARB X en el modelo con la variable CONECTIVIDAD DISCURSIVA

VARIABLES	Nº DE CASOS	% DE PRESENCIA	PESO PROBABILÍSTICO
CONECTIVIDAD			
Grado 3	557/779	72	<b>0,73</b>
Grado 2	103/200	52	<b>0,53</b>
Grado 1	193/787	25	0,25
<i>Rango</i>			48
TIPOLOGÍA VERBAL			
Existenciales	77/128	60	<b>0,61</b>
Mentales	104/210	50	<b>0,54</b>
Relacionales	295/601	49	<b>0,51</b>
Verbales	113/221	51	0,47
Materiales	244/556	44	0,45
Modulación	20/50	40	0,43
<i>Rango</i>			18
NÚMERO GRAMATICAL			
Singular	748/1466	51	<b>0,58</b>
Plural	105/300	35	0,42
<i>Rango</i>			16
AMBIGÜEDAD			
Morfológicamente ambiguo	263/482	55	<b>0,54</b>
Morfológicamente no ambiguo	590/1284	46	0,46
<i>Rango</i>			8
$\chi^2$ celdas = 1,0266	N = 1.766		Verosimilitud (-987,784) = -1019,261 Significación = 0,044
Total $\chi^2$ = (45) 62,953	Input 0,458		

Como puede apreciarse en el cuadro 24, las variables CONECTIVIDAD DISCURSIVA, TIPOLOGÍA VERBAL, NÚMERO GRAMATICAL y AMBIGÜEDAD contribuyen significativamente a la *presencia del sujeto de 3ª persona*.

Las variantes con pesos probabilístico más altos son: i) el *grado 3 de conectividad discursiva* (0,73); ii) los *verbos de procesos existenciales* (0,61); y iii) el *número singular* (0,58). Las variantes *verbos de procesos mentales* (0,54); *morfológicamente ambiguos* (0,54) y *verbos relacionales* (0,51) no favorecen ni frenan la presencia de un sujeto explícito, puesto que sus pesos probabilísticos están cercanos a 0,50 (ver §3.4.2.1)

Los valores presentados en el cuadro 24 indican que el *input* es 0,458; el valor de  $\chi^2$  (62,953) es mucho mayor al valor esperado (45); y la significación es 0,044 (< 0,05). Según todos estos valores, puede decirse que este primer modelo de análisis da cuenta satisfactoriamente de la alternancia presencia *vs.* ausencia del sujeto de 3ª persona en el español hablado en Caracas en 2004-2010.

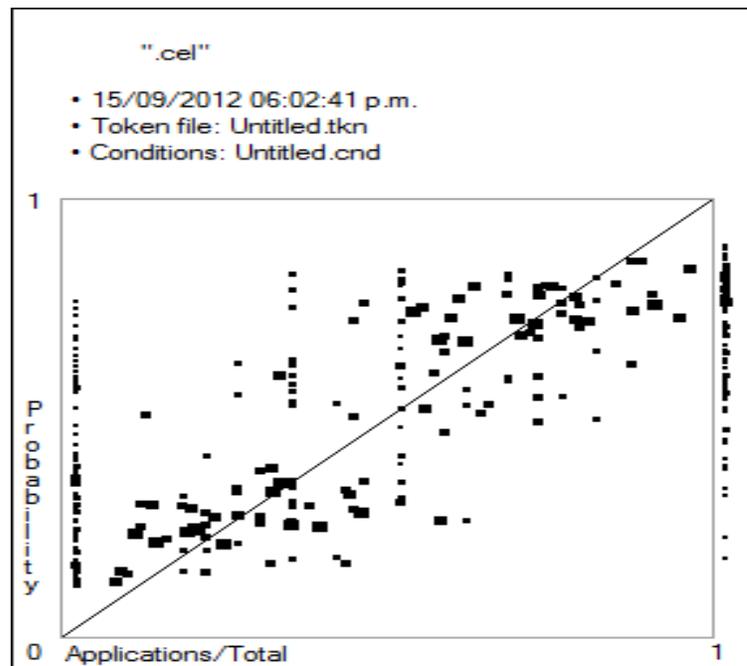
El programa GOLDVARB X no escogió como significativas las variables *turno de la conversación*, el *sexo* y la *edad*, ya que en este primer modelo estas tres variables no establecen una relación simultánea con el resto de las variables, de modo que no contribuyen en conjunto al fenómeno. En el cuadro 25 se presentan los pesos que obtuvieron estas variables individualmente:

Cuadro 25: Pesos probabilístico de las variables no significativas según GOLDVARB X en el modelo de CONECTIVIDAD DISCURSIVA

VARIABLES	N° DE CASOS	% DE PRESENCIA	PESO PROBABILÍSTICO
TURNO DE LA CONVERSACIÓN			
Cambio de turno	63/89	70,8	0,53
Igual turno	790/1.677	47,1	0,48
SEXO			
Mujer	449/900	49,9	0,52
Hombre	404/866	46,7	0,48
EDAD			
De 55 años en adelante	303/583	52	0,54
De 20 a 34 años	273/600	45,5	0,48
De 35 a 54 años	277/583	47,5	0,46

En el gráfico 8 se presenta el diagrama de dispersión:

Gráfico 8: Diagrama de dispersión de la presencia explícita de un sujeto de 3ª persona con la variable CONECTIVIDAD DISCURSIVA



En el gráfico 8 puede verse que los cuadros negros, que representan cada uno de los casos del universo analizado, están cerca del eje, lo que refleja que hay una correspondencia entre los valores esperados y los reales (Moreno Fernández 1994:126).

#### 4.2.2. Modelo con la variable *distancia clausular*

En un primer análisis con este modelo, el programa GOLDVARL X señaló la hipótesis nula como *aceptada*, en virtud de que los valores de significación de las pruebas de bondad, tanto del logaritmo de verosimilitud (0,059) como del  $\chi^2$  (0,781), eran superiores a lo esperado ( $>0,05$ ); de modo que no es posible afirmar la veracidad de los resultados, ya que los datos obtenidos podrían ser producto del azar (ver anexo 7) Debido a esto, tal como aconsejan Hernández Campoy y Almeida (2005:274), decidí adecuar el modelo teórico propuesto eliminando todos los casos cuyo margen de error estuviera por encima de 3,84 (ver §3.4.2.2.1). Una vez eliminadas las celdas con error, procedí al análisis del segundo modelo, esta vez con 1.645 casos.

En el cuadro 26 se presentan los resultados obtenidos para el segundo modelo teórico propuesto, en el que he incluido la variable *distancia clausular* y excluido la *conectividad discursiva*, tal como se indica en el subapartado 3.4.2.2.2.

Cuadro 26: Resultados del análisis GoldVarb X del segundo modelo con la variable DISTANCIA CLAUSULAR

VARIABLES	Nº DE CASOS	% DE PRESENCIA	PESO PROBABILÍSTICO
DISTANCIA CLAUSULAR			
Más de 5 cláusulas de separación	333/375	89	<b>0,85</b>
De 2 a 4 cláusula de separación	138/218	63	<b>0,55</b>
Una cláusula de separación	84/209	40	0,32
Cláusulas adyacentes	249/840	30	0,23
<i>Rango</i>			62
NÚMERO GRAMATICAL			
Singular	712/1373	52	<b>0,60</b>
Plural	92/269	34	0,40
<i>Rango</i>			20
TURNO DE LA CONVERSACIÓN			
Cambio de turno	63/85	74	<b>0,59</b>
Igual turno	741/1557	48	0,41
<i>Rango</i>			18
EDAD			
De 55 años en adelante	290/522	56	<b>0,56</b>
De 20 a 34 años	252/568	44	0,49
De 35 a 54 años	262/552	48	0,46
<i>Rango</i>			10
AMBIGÜEDAD			
Morfológicamente ambiguo	248/437	57	<b>0,54</b>
Morfológicamente no ambiguo	556/1205	46	0,46
<i>Rango</i>			8
$\chi^2$ celdas = 0,8917 Total $\chi^2$ (56) = 52,825	N = 1.642 Input 0,616		Verosimilitud (-873,084) = -899,497 Significación = 0,014

En el cuadro 26 se observa que las cinco variables que contribuyen significativamente a la presencia de un sujeto explícito son, en orden jerárquico, de mayor a menor según el valor del rango: DISTANCIA CLAUSULAR, NÚMERO GRAMATICAL, TURNO DE LA CONVERSACIÓN, EDAD y AMBIGÜEDAD.

Las variantes que contribuyen significativamente a la presencia de un sujeto son: *Más de 5 cláusulas de separación* (0,85), *número singular* (0,60), *cambio de*

*turno* (0,59) y *De 55 años en adelante* (0,56). Las variantes *De 2 a 4 cláusula de separación* (0,55) y *morfológicamente ambiguo* (0,54) ni condicionan ni frenan la presencia de un sujeto explícito, puesto que sus valores están cercanos a 0,50.

Según puede observarse en el cuadro 26, los valores obtenidos con GOLDVARB X indican que el *input* de la variante presencia de un sujeto de 3ª persona es 0,616; el valor total del  $\chi^2$  es 52,825, menor al esperado 56; el logaritmo de verosimilitud es -899,497, mayor al máximo posible -873,084; y, la significación es 0,014 (<0,05). Estos valores indican que este segundo modelo de análisis, cuyos datos con valores de error superior a 3,84 han sido eliminados, da cuenta del fenómeno estudiado.

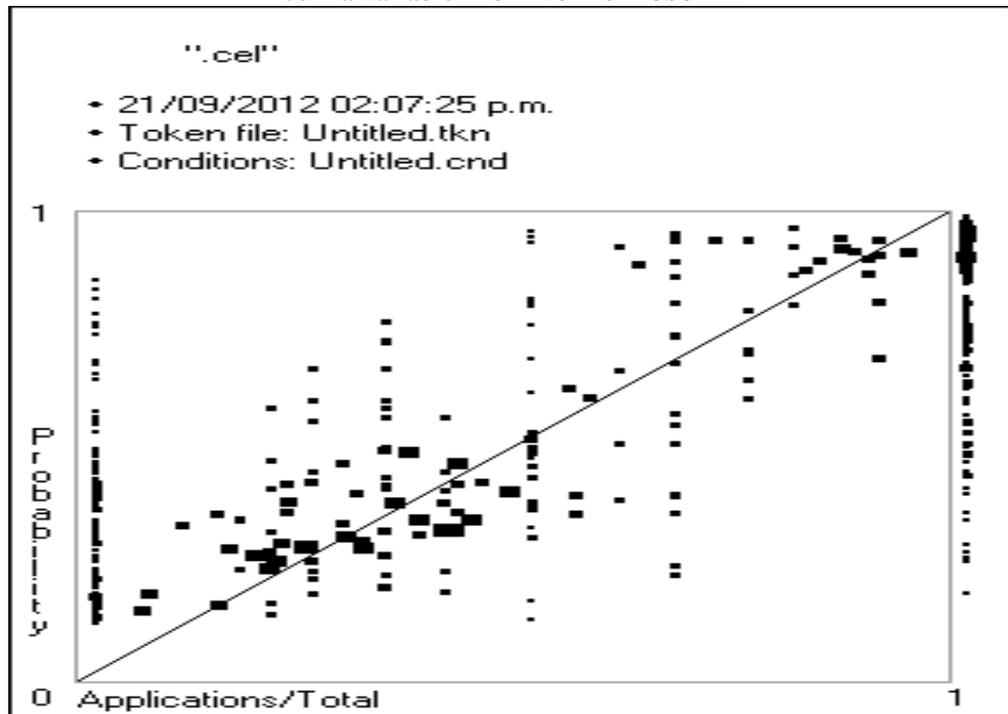
En el cuadro 27 se presentan los valores de las variables que no resultaron significativas por GOLDVAR X en el segundo modelo de análisis:

Cuadro 27: Pesos probabilístico de las variables no significativas según GOLDVARB X en el modelo de DISTANCIA CLAUSULAR

VARIABLES	Nº DE CASOS	% DE PRESENCIA	PESO PROBABILÍSTICO
TIPOLOGÍA VERBAL			
Existenciales	77/127	61	0,61
Relacionales	283/570	50	0,53
Mentales	95/185	51	0,51
Materiales	225/507	44	0,46
Verbales	104/204	51	0,46
Modulación	20/49	41	0,43
SEXO			
Mujer	427/847	50	0,51
Hombre	377/795	47	0,49

En el gráfico 9 se presenta el diagrama de dispersión del modelo de análisis con la variable *distancia clausular*.

Gráfico 9: Diagrama de dispersión de la presencia explícita de sujetos de 3ª persona con la variable DISTANCIA CLAUSULAR



En el gráfico 9 puede observarse que la mayoría de los datos analizados (representados por los cuadros negros) están cerca del eje, lo que significa que los casos se ajustan al modelo del análisis.

#### 4.2.3. Comparación de los resultados de los dos modelos de análisis propuestos

En 4.2.1. y 4.2.2., he presentado los resultados obtenidos a través del programa de regresión múltiple GOLDVARB X, el cual muestra los pesos probabilísticos de las variantes usadas en conjunto.

En el primer modelo, los resultados indican que la *conectividad discursiva*, la *tipología verbal*, el *número gramatical* y la *ambigüedad* condicionan en conjunto la

presencia de un sujeto explícito, mientras que, en el segundo modelo, la *distancia clausular*, el *número gramatical*, la *edad* y la *ambigüedad* son las variables que contribuyen simultáneamente a la presencia de un sujeto explícito de 3ª persona. Las variables *número gramatical* y *ambigüedad* resultaron significativas en ambos modelos. La variable extralingüística *sexo* no resultó significativa en ninguno de los dos modelos. Estos resultados concuerdan, en grandes rasgos, con el análisis de frecuencias absolutas y relativas (ver 4.1.1. al 4.1.7.), ya que la prueba del  $\chi^2$  indica que solo las variables lingüísticas tienen una relación estadística con la variable dependiente, pero no las sociales. Sin embargo, como ya he indicado, GOLDVARB X escogió como significativa la variable *edad* en el segundo modelo de análisis propuesto.

Los valores proporcionados por GOLDVARB X referentes a las distintas pruebas de bondad aplicadas a los dos análisis multivariados (Moreno Fernández 1994:137 y 141 y Tagliamonte 2006:133) permiten afirmar que ambos modelos se ajustan a esta investigación y que dan cuenta del fenómeno estudiado. Los resultados obtenidos tanto en el modelo con la variable *conectividad discursiva*, como en el de *distancia clausular*, indican que la presencia~ausencia del sujeto de 3ª persona es un fenómeno de variación lingüística más que sociolingüística, puesto que esta alternancia se debe a razones intrínsecas a la lengua, principalmente a la correferencialidad de dos sujetos gramaticales en el texto.

#### 4.2.4. Comparación de los resultados de GOLDVARB en los dos estudios del dialecto caraqueño (Quijada 2006 y Martínez Lara 2012)

La expresión variable del sujeto de 3ª persona en el habla de Caracas ha sido estudiada por Quijada 2006 y Martínez Lara 2012. El primer estudio está basado en el CSC'87 y el segundo, en el corpus PRESEEA-Caracas 2004-2010; ambos son una muestra del habla caraqueña, pero la estratificación social de los hablantes es distinta.<sup>30</sup> En ambas investigaciones se usaron los mismos criterios metodológicos para las variables lingüísticas *turno de la conversación* y *conectividad discursiva*; la variable lingüística *tipología verbal* y la extralingüística *edad* no son iguales. En el cuadro 28 se comparan los resultados de los estudios de Quijada 2006 y Martínez Lara 2012 con relación a las dos variables –*conectividad discursiva* y *turno de la conversación*– basadas en los mismo criterios:

Cuadro 28: Resultados de GOLDVARB en los dos estudios del sujeto variable de 3ª persona en el habla de Caracas

Variables	Quijada 2006	Martínez Lara 2012
CONECTIVIDAD DISCURSIVA		
Grado 3	<b>0,654</b>	<b>0,73</b>
Grado 2	<b>0,631</b>	<b>0,52</b>
Grado 1	0,415	0,25
TURNOS DE LA CONVERSACIÓN		
Cambio de turno	<b>0,649</b>	<b>0,59</b>
Igual turno	0,495	0,41

<sup>30</sup> Para más detalles sobre el corpus PRESEEA-CARACAS 2004-2010 ver §3.1., y el anexo 6 sobre el CSC'87.

En el cuadro 28 puede apreciarse que en ambos estudio el análisis de regla variable señala como significativas las mismas variantes *grado 3* y *2*, de CONECTIVIDAD DISCURSIVA, y *cambio de turno*, de TURNO DE LA CONVERSACIÓN.

Los pesos probabilístico obtenidos con la variable *conectividad discursiva* llaman la atención, puesto que el *grado 3* tiene un peso mayor en la muestra de habla actual que en la muestra del año 1987 (0,73 y 0,65, respectivamente); por el contrario, la variante *grado 2* tiene un peso menor en el presente (0,52 < 0,631).

En la variable *turno de la conversación*, los pesos probabilístico de *cambio de turno* están por encima de 0,50: en ambas investigaciones, entonces, esta variante condiciona la presencia de un sujeto explícito. En el presente estudio el valor es menor (0,59) que en el de Quijada 2006 (0,64).

Los resultados obtenidos con estas dos variantes en ambos estudios, me permiten hacer la siguiente pregunta:

¿Hubo un mayor número de módulos temáticos en las entrevistas del corpus PRESEEA-CARACAS 2004-2010 con respecto al CSC'87? Esto podría explicar, tal vez, la alta frecuencia de sujetos explícitos en el presente estudio, puesto que cada tema o tópico de conversación puede tener referentes distintos.

El análisis con GOLDVARB X de Quijada 2006 indica que la variable *sexo* actúa en conjunto con el resto de las variables, mientras que en la muestra analizada en la presente investigación esta variable no contribuye de manera conjunta con las otras variables a la presencia de un sujeto explícito de 3ª persona.



## 5. CONCLUSIONES

En el presente estudio he tratado de describir la alternancia entre la presencia y la ausencia del sujeto de 3ª persona, singular y plural, en el habla actual de la ciudad de Caracas, con el fin de conocer qué factores lingüísticos y extralingüísticos, en conjunto, pueden condicionar y/o favorecer la presencia explícita del sujeto en aquellos contextos en los que el sistema de la lengua permite la omisión. Para cumplir este objetivo, he analizado cuantitativamente, desde los postulados teóricos de la sociolingüística variacionista, una muestra del habla actual de Caracas proveniente del corpus del Proyecto PRESEEA Caracas 2004-2010.

El primer objetivo específico planteado ha sido determinar la frecuencia de uso del sujeto de 3ª persona, singular y plural. Los resultados muestran que en el habla actual de Caracas la ausencia de sujeto es un poco más frecuente que la presencia, con una diferencia de apenas 3,4 puntos porcentuales. Esa tendencia es común a todos los dialectos del español –tanto de Venezuela como de otros países hispanohablantes– en los que se ha estudiado el mismo fenómeno, pese a que en la presente investigación la frecuencia de sujetos explícitos es la mayor (48%) encontrada hasta ahora.

El segundo objetivo ha sido identificar los factores, lingüísticos y extralingüísticos, que pueden influir, simultáneamente, en el uso expreso del sujeto. Para ello, las variables independientes fueron sometidas, por separado, a un análisis descriptivo y, posteriormente, las mismas se agruparon con el fin de construir dos modelos teóricos para el análisis inferencial. Debido a que dos de las variables

utilizadas en estudios anteriores (*Conectividad discursiva* y *Distancia clausular*), ambas relacionadas con la correferencialidad de dos sujetos adyacentes, podrían causar un solapamiento en el análisis, se propusieron dos modelos de análisis de inferencias. El primer modelo –denominado *Modelo con la variable conectividad discursiva* (MCD)– incluye todas las cinco variables propuestas excepto la *distancia clausular*, mientras que en el segundo – *Modelo con la variable distancia clausular* (MDC)– se contemplan todas las variables menos la *conectividad discursiva*. Tanto del análisis descriptivo como el de inferencias permiten concluir que son los factores lingüísticos los que condicionan el uso de un sujeto explícito, ya que el programa multivariado GOLDVARB X no ha seleccionado como significativo ninguno de los factores sociales en el primer modelo, y solamente el de la *Edad* sí fue seleccionada en el segundo modelo, pero con un peso probabilístico poco relevante (0,56). El presente estudio confirma una vez más que la variación presencia~ausencia del sujeto –tanto de 1<sup>a</sup> como de 3<sup>a</sup> persona– depende en gran medida, aunque no exclusivamente, de los factores internos de la lengua, encabezados por las variables relacionadas con la correferencialidad (*Conectividad discursiva* en el MCD y *Distancia clausular* en el MDC), cuyos respectivos rangos (48 y 62) superan con mucho los rangos de todas las variables. Esta reflexión confirma los análisis y los hallazgos de las primeras investigaciones sobre la presencia~ausencia del sujeto en el español hablado, tanto el español en contacto con el inglés (Silva-Corvalán 1977 y 1982), como el de los monolingües caraqueños (Bentivoglio 1980 y 1987).

Basándome en los resultados obtenidos, debería escoger uno de los dos modelos propuestos, asunto muy difícil dilucidar. El único elemento que haría preferir el primer modelo (MCD) es de índole estadística, pues este se ajusta más a los datos, ya que el  $\chi^2$  total es mejor que el del segundo modelo (MDC):  $62,953 > 45$  Y  $52,825 < 56$ , respectivamente. Sin embargo, las dos variables consideradas en los dos modelos son muy similares –de ahí la necesidad de no emplearlas conjuntamente–, pues ambas se basan en la distancia que separa el sujeto bajo análisis del sujeto correferencial previo. Lo que diferencia una variable de la otra es, por una parte, la mayor precisión en determinar la correferencialidad existente entre los dos sujetos que ofrece la *Conectividad discursiva*, pues esta refleja las funciones gramaticales distintas que desempeña un mismo referente, pero no tiene la misma precisión respecto del número de cláusulas interpuestas. Por otra parte, la *Distancia clausular* indica con exactitud cuántas cláusulas hay entre los dos sujetos, pero no toma en cuenta el hecho de que un mismo referente puede cumplir otra función gramatical que no sea la de sujeto. Resumiendo, ambas variables tienen ventajas y desventajas si las comparamos; me parece prematuro inclinarme por una de ellas, y prefiero concluir que el investigador que analice esta variación escoja el modelo que mejor se ajuste a sus datos. Las investigaciones futuras, seguramente, ayudarán a resolver el problema que, por el momento, queda como un posible e interesante motivo de discusión.

Las otras variables lingüísticas, en ambos modelos, tienen rangos muy bajos respecto de los obtenidos tanto por la *Conectividad discursiva* en el MCD, como por la *Distancia clausular* en el MDC. El programa Goldvarb X ha seleccionado como significativas las variables *Número gramatical* –con la variante *singular*– y *Ambigüedad* –con la variante *morfológicamente ambigua*– cuyos rangos son 16 y 8, respectivamente. La *Tipología verbal* solo ha sido seleccionada en el MCD, y la *Edad* en el MDC.

Según los resultados obtenidos, tanto los del análisis descriptivo como del inferencial, puede concluirse que: i) las relaciones discursivas interclausulares condicionan la presencia~ausencia de un sujeto, puesto que este tiende a ser explícito en la medida en que disminuye la conectividad discursiva o aumenta la distancia entre dos menciones de un mismo referente; y ii) la coincidencia de las marcas verbales de persona y número gramatical entre la 1ª y la 3ª persona singular favorecen la presencia explícita del sujeto de 3ª persona en el habla actual de la ciudad de Caracas, por lo que mencionar explícitamente un sujeto puede entenderse como una estrategia lingüística que usan los hablantes para desambiguar los contextos que podrían causar ambigüedad.

El tercer y último objetivo de esta investigación era determinar si la variación entre la presencia y la ausencia de un sujeto de 3ª persona es estable o si hay indicios de un cambio lingüístico. Para este fin, comparé los resultados obtenidos de Quijada 2006, sobre el habla de Caracas de 1987, con los del presente estudio. La

comparación, basada solamente en los porcentajes, revela un incremento en el uso de sujetos explícitos en el dialecto caraqueño actual (48%) respecto del de hace veinte años (33%). Si se toman en cuenta otros criterios, es difícil afirmar que haya un cambio lingüístico en curso, puesto que: i) el lapso entre las dos muestras analizadas es muy breve; y ii) la muestra del presente estudio está compuesta por hablantes de un solo nivel de instrucción (los que tienen un nivel superior al bachillerato), mientras que los del corpus de 1987 están estratificados en tres niveles socioeconómicos: alto, medio y bajo. Por el momento, no es posible decir nada seguro respecto de si la variación presencia~ausencia de un sujeto de 3ª persona en el habla actual de Caracas está en variación estable o en proceso de cambio.

Los resultados de esta investigación permiten concluir que el sujeto variable de 3ª persona, singular y plural, en el habla actual de Caracas, es un fenómeno de la lengua que está condicionado por factores lingüísticos sensibles al discurso.

En futuras investigaciones, recomiendo que se tomen en cuenta los siguientes señalamientos metodológicos: i) ahondar más sobre variables lingüísticas cuya naturaleza sea discursiva; ii) analizar una muestra de habla más amplia en la que haya hablantes de más de un grado de instrucción; y, iii) hacer una comparación más detallada con datos de corpus recientes de hablantes de otras ciudades hispanohablantes, por ejemplo, los obtenidos gracias a PRESEEA.

Por último, debo acotar que los estudios lingüísticos basados en corpus orales son valiosos por la riqueza de datos que aportan sobre la realidad lingüística de una comunidad de habla, por lo que sería recomendable seguir haciendo investigaciones

basadas en corpus así como e incentivar la elaboración de una muestra de conversaciones.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- Alarcos Llorach, Emilio. 1999. *Gramática de la lengua española*. Madrid: Espasa.
- Albelda, Marta. 2004. Cortesía en diferentes situaciones comunicativas. La conversación coloquial y la entrevista sociológica semiformal. En Diana Bravo y Antonio Briz (eds.), *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*, 109-134. Barcelona: Ariel.
- Alcina, Juan y José Manuel Blecua. 1975. *Gramática española*. Barcelona: Ariel.
- Alonso, Amado y Pedro Henríquez Ureña. 1969. *Gramática castellana*. Buenos Aires: Losada S.A.
- Balash, Sonia. 2008. La conectividad discursiva en el discurso interactivo. En Joyce Bruhn y Elena Valenzuela (eds.), *Selected proceedings of the 10th Hispanic Linguistics Symposium*, 301-311. Somerville, MA: Cascadilla Proceedings Project.
- Barrenechea, Ana María y Alicia Alonso. 1977. Los pronombres personales sujetos en el español hablado en Buenos Aires. En Juan Manuel Lope Blanch (ed.), *Estudios sobre el español hablado en las principales ciudades de América*, 333-349. México, DF: Universidad Nacional Autónoma de México.
- Bayley, Robert y Lucinda Pease-Alvarez. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9.349-371.
- Bello, Andrés. 1986 [1847]. *Gramática de la lengua castellana*. Madrid: Arco/Libros.
- Bentivoglio, Paola. 1980. *Why "canto" and not "yo canto"? The problem of first person subject pronoun in spoken Venezuelan Spanish*. Tesis de maestría inédita, University of California, Los Angeles.
- Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Bentivoglio, Paola. 1999. Recursos estadísticos aplicados al análisis sintáctico: una propuesta metodológica. En Joseph H. Matluck y Carlos A. Solé (eds.), *Actas del Simposio internacional de la lengua española: pasado, presente y futuro*. 225-236. Austin, TX: Texas University Press.
- Bentivoglio, Paola e Irania Malaver. 2006. La lingüística de corpus en Venezuela: un nuevo proyecto. *Lingua Americana* 19.39-46.
- Bentivoglio, Paola y Mercedes Sedano. 1993. Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana. *Boletín de Lingüística* 8.3-35.

- Blanco Canales, Ana. 1999. Presencia/ausencia de sujeto pronominal de primera persona en español. *Español Actual* 72.31-39.
- Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18.355-387.
- Bock, Kathryn y Zenzi Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology General* 129:2.177-192.
- Briones, Guillermo. 2002. *Metodología de la investigación cuantitativa en las ciencias sociales*. Bogotá: Instituto Colombiano para el Fomento de la Educación Superior.
- Briz, Antonio. 2004. Cortesía verbal codificada y cortesía verbal interpretada en la conversación. En Bravo, Diana y Antonio Briz (eds.), *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*, 67-93. Barcelona: Ariel.
- Briz Gómez, José. 2001. *El español coloquial en la conversación*. Barcelona: Ariel.
- Calsamiglia, Helena y Amparo Tusón. 1999. *Las cosas del decir*. Barcelona: Ariel.
- Cameron, Richard. 1992. *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Tesis doctoral inédita, University of Pennsylvania, Philadelphia.
- Cameron, Richard. 1993. Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change* 5.305-334.
- Cameron, Richard. 1995. The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6/7.1-27.
- Cameron, Richard y Nydia Flores-Ferrán. 2004. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1,1.41-65.
- Cantero, Gustavo. 1976. Peculiaridades en el empleo del pronombre personal *yo* en el habla culta de la ciudad de México. *Anuario de Lingüística* 14.233-237.
- Cifuentes, Hugo. 1980-81. Presencia y ausencia del pronombre personal en el habla culta de Santiago de Chile. *Boletín de Filología de la Universidad de Chile. Homenaje a Ambrosio Rabanales* 31.743-52.
- Dell, Gary. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93.283-321.
- Dixson, Robert y Julio Andújar. 1967. *Resumen práctico de la gramática inglesa*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

- Domínguez, Carmen Luisa y Elsa Mora. 1995. *Corpus sociolingüístico de la ciudad de Mérida (Venezuela)*. Mérida: Universidad de Los Andes.
- Dumont, Jenny. 2006. Full NPs as Subjects. En Nuria Sagarra y Jacqueline Almeida (eds.), *Selected Proceedings of the 9<sup>th</sup> Hispanic Linguistics Symposium*, 286-296. Somerville, MA: Cascadilla Proceedings Project.
- Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.
- Enríquez, Emilia V. 1986. La presencia de los pronombres personales sujeto en el mundo hispánico: un estudio comparativo. *Anuario de Lingüística Hispánica* XXIV.47-70.
- Fernández Ramírez, Salvador. 1951. *Gramática Española. El pronombre*. Madrid: Arco/Libros.
- Flores-Ferrán, Nydia. 2002. *Subject personal pronouns in Spanish narratives of Puerto Ricans in New York City: A sociolinguistic perspective*. München: Lincom Europa.
- Flores-Ferrán, Nydia. 2004. Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? *Language Variation and Change* 16.49-73.
- García, Erica. 1985. Shifting variation. *Lingua* 67.189-224.
- García-Miguel, José, Lourdes Costas y Susana Martínez. 2005. [Diátesis verbales y esquemas construccionales. Verbos, clases semánticas y esquemas sintáctico-semánticos en el proyecto ADESSE](#). En Gerd Wotjak y Juan Cuartero Otal (eds.), *Entre semántica léxica, teoría del léxico y sintaxis*, 373-384. Frankfurt am Main: Peter Lang.
- Gili y Gaya, Samuel. 1975. *Curso superior de sintaxis española*. La Habana: Editorial Pueblo y Educación.
- Guirado, Krístel. 2009. *(De)queísmo: uso deíctico y distribución social en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Gutiérrez Maté, Miguel. 2008. El estudio de la expresión del sujeto de tercera persona en documentos coloniales del Caribe. En Inés Olza Moreno, Manuel Casado Velarde y Ramón González Ruiz (eds.), *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística*, 285-296. Pamplona: Universidad de Navarra.

- Halliday, Michael A. K. 1967. Notes on transitivity and theme in English, Part II. *Journal of Linguistics* 3.199-244.
- Halliday, Michael A. K. 1994. *An introduction to functional grammar*. London: Edward Arnold.
- Hernández Campoy, Juan M. y Manuel Almeida. 2005. *Metodología de la investigación sociolingüística*. Málaga: Editorial Comares.
- Hochberg, Judith. 1986. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62,3.609-621.
- Hurtado, Luz Mercedes. 2001. *La variable expresión del sujeto en el español de los colombianos y colombo-americanos residentes en el condado de Miami-Dade*. Tesis doctoral inédita, University of Florida.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistic.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania.
- Labov, William. 1978. Where does the linguistic variable stop? A response to Beatriz Lavander. *Working Papers in Sociolinguistics*, vol. 44. Austin, TX: Southwest Educational Development Laboratory.
- Lamidi, Mufutau Temitayo. 2009. Parametric variation of the 3<sup>rd</sup> person singular in English and some Nigerian language. *The Linguistic Journal* 4,1.67-90.
- Lamíquiz, Vidal. 1972. *Morfosintaxis estructural del verbo español*. Sevilla: Publicaciones de la Universidad de Sevilla.
- Lavandera, Beatriz. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7.171-138.
- Li, Charles y Sandra Thompson. 1979. Third-person pronouns and zero anaphora in Chinese discourse. En Talmy Givón (ed.), *Discourse and syntax*, 311-335. New York: Academic Press.
- López Morales, Humberto. 2004. *Sociolingüística*. Madrid: Gredos.
- Lu, Hui-Chuan. 1997. La ausencia y la presencia del pronombre personal. *EPOS* XIII.117-133.
- Mayol, Laia. 2009. *Pronouns in Catalan: Information, discourse and strategy*. Tesis doctoral inédita, University of Pennsylvania, Philadelphia.
- Miró Vera, Ramona y Miguel Ángel de Pineda. 1990. Determinación sociolingüística de la presencia/ausencia del pronombre personal sujeto. En María Teresa Palet Plaja (ed.), *Sociolingüística andaluza 5: Habla de Sevilla y hablas americanas*, 37-44. Sevilla: Universidad de Sevilla.

- Montenegro, Liliana. 1982. Los pronombres personales sujetos en el habla estudiantil santiaguera. En Humberto López Morales y María Vaquero (eds.), *Actas del I Congreso internacional del Español de América*, 753-765. San Juan de Puerto Rico: Academia Puertorriqueña de la Lengua Española.
- Montes Miró, Rosa. 1986. Factores discursivos en el análisis de los pronombres sujetos en español. *Morphé* 1.45-71.
- Montolío, Estrella (Coord.) 2000. *Manual práctico de escritura académica*, Vol. III. Barcelona: Ariel.
- Morales, Amparo. 1982. La expresión de sujeto pronominal, primera persona, en el español de Puerto Rico. *Boletín de la Academia Puertorriqueña de la Lengua Española* 8.91-102.
- Morales, Amparo. 1986. Algunos aspectos de gramática en contacto: la expresión del sujeto en el español de Puerto Rico. *Anuario de Letras* 24.71-85.
- Morales, Amparo. 1997. La manifestación del sujeto en español: tendencias universales y variación dialectal. En Robert Hammond y Marguerite MacDonald (eds.), *Linguistic studies in honor of Bohdan Saciuk*, 87-103. West Lafayette, IN: Learning Systems, Inc.
- Moreno Fernández, Francisco. 1994. *Status quaestionis*: Sociolingüística, estadística e informática. *Lingüística* 6.95-154.
- Moreno Fernández, Francisco. 2005. *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Ariel.
- Orozco, Rafael y Gregory Guy. 2008. El uso variable de los pronombres sujetos: ¿qué pasa en la costa Caribe colombiana? En Maurice Westmoreland y Juan A. Thomas (eds.), *Selected Proceedings of the 4<sup>th</sup> Workshop on Spanish Sociolinguistics*, 60-70. Somerville, MA: Cascadilla Proceedings Project.
- Otheguy, Ricardo y Ana Cecilia Zentella. 2007. Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York. En Richard Cameron y Kim Powski (eds.), *Spanish in contact: Policy, social and linguistic inquiries*, 275-295. Amsterdam/Philadelphia: John Benjamins.
- Paredes Silva, Vera Lúcia. 1993. Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5.35-49.
- Paredes Silva, Vera Lúcia. 2003. Motivações funcionais no uso do sujeito pronominal: uma análise em tempo real. En Maria da Conceição de Paiva y Maria Eugênia Lamoglia Duarte (Organização), *Mudança lingüística em tempo real*, 97-114. Rio de Janeiro: Contra Capa.

- Pérez Brabandere, Vanessa. 2010. *Los sujetos pronominales de primera persona en el español de Caracas 2004-2010*. Trabajo de grado, Maestría en Lingüística, Universidad Central de Venezuela, Caracas.
- Poplack, Shana. 1981. Mortal phonemes as plural morphemes. En David Sankoff y Henrietta Cedergren(eds.), *Variation ómnibus*,59-71. Edmonton: Linguistic Research Inc.
- Pottier, Bernard. 1975. *Gramática del español*. Madrid: Ediciones Alcalá.
- Quijada Bringtown, Milagros. 2006. *Ø cantaba... ¿él, ella o yo? Sujetos de tercera persona en el habla de Caracas*. Trabajo de grado de la maestría en Lingüística, Universidad Central de Venezuela, Caracas.
- Quilis, Antonio, Manuel Esgueva, M.<sup>a</sup> Luz Gutierrez y Pilar Ruiz-Va. 1998. *Lengua española*. Madrid: Universidad Nacional a Distancia.
- Real Academia Española. 1973. *El esbozo para una nueva gramática de la lengua española*. Madrid: Espasa.
- Real Academia Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa.
- Real Academia Española. 2010. *Nueva gramática de la lengua española. Manual*. Madrid: Espasa.
- Romaine, Suzanne. 1981. The status of variable rules in sociolinguistic theory. *Journal of Linguistic* 17.93-119.
- Samper Padilla, José Antonio, Clara Eugenia Hernández Cabrera y Magnolia Troya Déniz. 2006. Presencia/ausencia del sujeto pronominal de primera persona singular en la norma culta de España. En Concepción Company Company (ed.), *El español en América. Diatopía, diacronía e historiografía*, 87-109. México, DF: Universidad Nacional Autónoma de México.
- Sankoff, David. 1978. *Linguistic variation: Models and methods*. New York: Academic Press.
- Sankoff, David y Pascale Rousseau 1978. Advances in variable rule methodology. En David Sankoff (ed.), *Linguistic variation. Models and methods*, 57-69. New York: Academic Press.
- Sankoff, David, Sali Tagliamonte y Eric Smith. 2005. *Goldvarb X: A variable rule application for Macintosh and Windows*. Toronto: Universidad de Toronto. En línea: [http://individual.utoronto.ca/tagliamonte/Goldvarb/GV\\_index.htm](http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm) [Consulta: 01 de octubre de 2010].
- Seco, Manuel. 2001. *Gramática esencial del español*. Madrid: Espasa Calpe.

- Sedano, Mercedes. 1990. *Hendidas y otras construcciones con ser en el habla de Caracas*. Caracas: Universidad Central Venezuela.
- Sedano, Mercedes. 2011. *Manual de gramática del español. Con especial referencia al español de Venezuela*. Caracas: Universidad Central de Venezuela.
- Serra, Miguel, Elisabet Serrat, Rosa Solé, Aurora Bel y Melina Aparici. 2000. *La adquisición del lenguaje*. Barcelona: Ariel.
- Silva-Corvalán, Carmen. 1977. *A discourse study of word order in the Spanish spoken by Mexican-Americans in West Los Angeles*. Tesis de maestría inédita, University of California, Los Angeles.
- Silva-Corvalán, Carmen. 1982. Subject expression and placement in Mexican-American Spanish. En Jon Amastae y Lucía Elías-Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, 93-120. New York: Cambridge University Press.
- Silva-Corvalán, Carmen. 2001. *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.
- Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Travis, Catherine E. 2005. *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin/ New York: Mouton de Gruyter.
- Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19,2.101-135.



# Anexos



Anexo 1: CLASIFICACIÓN DE LOS VERBOS SEGÚN LOS PROCESOS RECONOCIDOS POR EL GRUPO DE INVESTIGACIÓN ADESSE

<b>1 MENTAL</b>	Verbos-ejemplo
11 <u>Sensación</u>	<i>Sentir</i>
111 <u>Volición</u>	<i>gustar, temer</i>
12 <u>Percepción</u>	<i>querer, desear</i>
13 <u>Cognición</u>	<i>ver, escuchar</i>
131 <u>Conocimiento</u>	<i>pensar, entender</i>
132 <u>Creencia</u>	<i>saber1, recordar</i>
14 <u>Elección</u>	<i>creer, opinar</i>
	<i>decidir, elegir</i>
<b>2 RELACIONAL</b>	
21 <u>Atribución</u>	<i>ser, parecer</i>
211 <u>Relación</u>	<i>equivaler, depender</i>
212 <u>Propiedad</u>	<i>estar2, quedar2</i>
2121 <u>Medida</u>	<i>medir, costar</i>
2122 <u>Apariencia</u>	<i>oler2, saber2</i>
213 <u>Denominación</u>	<i>nombrar2, llamar2</i>
22 <u>Poseción</u>	<i>tener, necesitar</i>
221 <u>Adquisición</u>	<i>conseguir, recibir</i>
222 <u>Transferencia</u>	<i>dar, pagar</i>
<b>3 MATERIAL</b>	
31 <u>Espacio</u>	<i>mover, girar</i>
311 <u>Desplazamiento</u>	<i>ir, andar, llevar</i>
312 <u>Localización</u>	<i>poner, pender</i>
313 <u>Postura-posición</u>	<i>sentar, agachar</i>
314 <u>Orientación</u>	<i>señalar, apuntar</i>
315 <u>Manera movimiento</u>	<i>aletear, mecer</i>
316 <u>Unión</u>	<i>juntar, añadir</i>
32 <u>Cambio</u>	<i>pintar, cocinar</i>
321 <u>Creación</u>	<i>crear, fabricar</i>
322 <u>Modificación</u>	<i>romper, secar</i>
3221 <u>Cuidado corporal</u>	<i>lavar, cepillar</i>
323 <u>Destrucción</u>	<i>destruir, eliminar</i>
3231 <u>Consumo</u>	<i>comer, gastar</i>
33 <b>Otros hechos</b>	
331 <u>Meteorología</u>	<i>llover, atardecer</i>
332 <u>Emisión</u>	<i>brillar, emitir</i>
333 <u>Contacto</u>	<i>golpear, tocar</i>
334 <u>Control</u>	<i>agarrar, coger</i>
335 <u>Uso</u>	<i>emplear, usar</i>
336 <u>Sustitución</u>	<i>reemplazar, suplir</i>
337 <u>Actividad</u>	<i>trabajar, actuar</i>

338 <u>Competición</u>	<i>luchar, vencer</i>
34 <u>Comportamiento</u>	<i>portar2, comportar2</i>
342 <u>Fisiología</u>	<i>llorar, orinar</i>
343 <u>Relaciones sociales</u>	<i>acoger, casar2</i>
<b>4 VERBAL</b>	
41 <u>Comunicación</u>	<i>decir, hablar</i>
411 <u>Mandato</u>	<i>ordenar1, pedir</i>
42 <u>Valoración</u>	<i>criticar, felicitar</i>
<b>5 EXISTENCIAL</b>	
51 <u>Existencia</u>	<i>haber, ocurrir</i>
52 <u>Fase-tiempo</u>	
521 <u>Fase</u>	<i>empezar, acabar</i>
522 <u>Tiempo</u>	<i>durar, retrasar</i>
53 <u>Vida</u>	<i>vivir, matar</i>
<b>6 MODULACIÓN</b>	
60 <u>Causación</u>	<i>hacer2</i>
601 <u>Inducción</u>	<i>incitar, ayudar</i>
602 <u>Permiso</u>	<i>permitir, dejar2</i>
603 <u>Obligación</u>	<i>obligar, forzar1</i>
61 <u>Disposición</u>	<i>tratar, atreverse</i>
62 <u>Aceptación</u>	<i>aceptar, rechazar</i>
64 <u>Verbos de apoyo</u>	<i>hacer-0, dar-0</i>

Anexo 2: Binomial VARBRUL, 1 step del Modelo teórico de análisis con la variable  
*conectividad discursiva*

BINOMIAL VARBRUL, 1 step

Run # 1, 317 cells:

Convergence at Iteration 8

Input 0,474

<i>Group</i>	<i>Factor</i>	<i>Weight</i>	<i>App/Total</i>	<i>Input&amp;Weight</i>
1:	i	0.475	0.47	0.45
	c	0.525	0.71	0.50
2:	n	0.466	0.46	0.44
	a	0.534	0.55	0.51
3:	3	0.727	0.71	0.71
	1	0.248	0.24	0.23
	2	0.532	0.52	0.51
4:	2	0.509	0.49	0.48
	3	0.451	0.44	0.43
	4	0.468	0.51	0.44
	1	0.537	0.50	0.51
	5	0.604	0.60	0.58
5:	6	0.430	0.40	0.41
	s	0.588	0.51	0.56
6:	p	0.412	0.35	0.39
	h	0.481	0.47	0.45
7:	m	0.519	0.49	0.49
	1	0.484	0.46	0.46
7:	2	0.477	0.48	0.45
	3	0.540	0.52	0.51

Total Chi-square = 323,9646

Chi-square/cell = 1,0220

Log likelihood = -1015,800

Maximum possible likelihood = -820.920

Fit: X-square (304) = 389.760, rejected, p = 0.0000

Anexo 3: Binomial VARBRUL, *step-up/down* del Modelo teórico de análisis con la variable *conectividad discursiva*

All remaining groups significant

Groups eliminated while stepping down: 1 6 7  
Best stepping up run: #21  
Best stepping down run: #45

Run # 21, 54 cells:

Convergence at Iteration 5

Input 0.458

Group # 2 -- n: 0.462, a: 0.538

Group # 3 -- 3: 0.729, 1: 0.249, 2: 0.529

Group # 4 -- 2: 0.507, 3: 0.449, 4: 0.467, 1:  
0.535, 5: 0.608, 6: 0.431

Group # 5 -- s: 0.582, p: 0.418

Log likelihood = -1019.261 Significance = 0.044

Maximum possible likelihood = -987.784

Fit: X-square(45) = 62.953, rejected, p = 0.0006

Anexo 4: Binomial Varbrul, 1 step del Modelo teórico de análisis con la variable *distancia clausular* sin las celdas cuyo valor de error es mayor a 3,84

BINOMIAL VARBRUL, 1 step

Run # 1, 374 cells:

Convergence at Iteration 8

Input 0.616

<i>Group</i>	<i>Factor</i>	<i>Weight</i>	<i>App/Total</i>	<i>Input&amp;Weight</i>
1:	i	0.415	0.48	0.53
	c	0.585	0.74	0.69
2:	n	0.458	0.46	0.58
	a	0.542	0.57	0.65
3:	2	0.528	0.50	0.64
	3	0.459	0.44	0.58
	4	0.459	0.51	0.58
	1	0.514	0.51	0.63
	5	0.610	0.61	0.71
	6	0.429	0.41	0.55
4:	2	0.318	0.40	0.43
	1	0.229	0.30	0.32
	4	0.852	0.89	0.90
	3	0.557	0.63	0.67
5:	s	0.596	0.52	0.70
	p	0.404	0.34	0.52
6:	h	0.489	0.47	0.60
	m	0.511	0.50	0.63
7:	1	0.491	0.44	0.61
	2	0.450	0.47	0.57
	3	0.559	0.56	0.67

Total Chi-square = 333,495

Chi-square/cell = 0,8917

Log likelihood = -894,036

Maximum possible likelihood = -702,962

Fit: X-square (360) = 482,148, rejected, p = 0.0000

Anexo 5: Binomial VARBRUL, *step-up/down* del Modelo teórico de análisis con la variable *distancia clausular* sin las celdas cuyo valor de error es mayor a 3,84

All remaining groups significant

Groups eliminated while stepping down: 6 3

Best stepping up run: #24

Best stepping down run: #39

Run # 24, 64 cells:

Convergence at Iteration 7

Input 0.616

Group # 1 -- i: 0.413, c: 0.587

Group # 2 -- n: 0.457, a: 0.543

Group # 4 -- 2: 0.320, 1: 0.231, 4: 0.852, 3:  
0.551

Group # 5 -- s: 0.597, p: 0.403

Group # 7 -- 1: 0.490, 2: 0.449, 3: 0.561

Log likelihood = -899.497 Significance = 0.014

Maximum possible likelihood = -873.084

Fit: X-square(56) = 52.825, rejected, p = 0.0059

Anexo 6: Corpus Sociolingüístico de Caracas de 1.987

NIVEL SOCIO-ECONÓMICO		SEXO	EDAD							
			A: 14-29 años		B: 30-45 años		C: 46-60 años		D: 61 años en adelante	
			HOMBRES	MUJERES	HOMBRES	MUJERES	HOMBRES	MUJERES	HOMBRES	MUJERES
ALTO			<a href="#"><u>CA1MA.87</u></a>	<a href="#"><u>CA1FA.87</u></a>	<a href="#"><u>CB1MA.87</u></a>	<a href="#"><u>CB1FA.87</u></a>	<a href="#"><u>CC1MA.87</u></a>	<a href="#"><u>CC1FA.87</u></a>	<a href="#"><u>CD1MA.87</u></a>	<a href="#"><u>CD1FA.87</u></a>
			<a href="#"><u>CA1MB.87</u></a>	<a href="#"><u>CA1FB.87</u></a>	<a href="#"><u>CB1MB.87</u></a>	<a href="#"><u>CB1FB.87</u></a>	<a href="#"><u>CC1MB.87</u></a>	<a href="#"><u>CC1FB.87</u></a>	<a href="#"><u>CD1MB.87</u></a>	<a href="#"><u>CD1FB.87</u></a>
			<a href="#"><u>CA1MC.87</u></a>	<a href="#"><u>CA1FC.87</u></a>	<a href="#"><u>CB1MC.87</u></a>	<a href="#"><u>CB1FC.87</u></a>	<a href="#"><u>CC1MC.87</u></a>	<a href="#"><u>CC1FC.87</u></a>	<a href="#"><u>CD1MC.87</u></a>	<a href="#"><u>CD1FC.87</u></a>
			<a href="#"><u>CA1MD.87</u></a>	<a href="#"><u>CA1FD.87</u></a>	<a href="#"><u>CB1MD.87</u></a>	<a href="#"><u>CB1FD.87</u></a>	<a href="#"><u>CC1MD.87</u></a>	<a href="#"><u>CC1FD.87</u></a>	<a href="#"><u>CD1MD.87</u></a>	<a href="#"><u>CD1FD.87</u></a>
MEDIO ALTO			<a href="#"><u>CA2MA.87</u></a>	<a href="#"><u>CA2FA.87</u></a>	<a href="#"><u>CB2MA.87</u></a>	<a href="#"><u>CB2FA.87</u></a>	<a href="#"><u>CC2MA.87</u></a>	<a href="#"><u>CC2FA.87</u></a>	<a href="#"><u>CD2MA.87</u></a>	<a href="#"><u>CD2FA.87</u></a>
			<a href="#"><u>CA2MB.87</u></a>	<a href="#"><u>CA2FB.87</u></a>	<a href="#"><u>CB2MB.87</u></a>	<a href="#"><u>CB2FB.87</u></a>	<a href="#"><u>CC2MB.87</u></a>	<a href="#"><u>CC2FB.87</u></a>	<a href="#"><u>CD2MB.87</u></a>	<a href="#"><u>CD2FB.87</u></a>
			<a href="#"><u>CA2MC.87</u></a>	<a href="#"><u>CA2FC.87</u></a>	<a href="#"><u>CB2MC.87</u></a>	<a href="#"><u>CB2FC.87</u></a>	<a href="#"><u>CC2MC.87</u></a>	<a href="#"><u>CC2FC.87</u></a>	<a href="#"><u>CD2MC.87</u></a>	<a href="#"><u>CD2FC.87</u></a>
			<a href="#"><u>CA2MD.87</u></a>	<a href="#"><u>CA2FD.87</u></a>	<a href="#"><u>CB2MD.87</u></a>	<a href="#"><u>CB2FD.87</u></a>	<a href="#"><u>CC2MD.87</u></a>	<a href="#"><u>CC2FD.87</u></a>	<a href="#"><u>CD2MD.87</u></a>	<a href="#"><u>CD2FD.87</u></a>
MEDIO			<a href="#"><u>CA3MA.87</u></a>	<a href="#"><u>CA3FA.87</u></a>	<a href="#"><u>CB3MA.87</u></a>	<a href="#"><u>CB3FA.87</u></a>	<a href="#"><u>CC3MA.87</u></a>	<a href="#"><u>CC3FA.87</u></a>	<a href="#"><u>CD3MA.87</u></a>	<a href="#"><u>CD3FA.87</u></a>
			<a href="#"><u>CA3MB.87</u></a>	<a href="#"><u>CA3FB.87</u></a>	<a href="#"><u>CB3MB.87</u></a>	<a href="#"><u>CB3FB.87</u></a>	<a href="#"><u>CC3MB.87</u></a>	<a href="#"><u>CC3FB.87</u></a>	<a href="#"><u>CD3MB.87</u></a>	<a href="#"><u>CD3FB.87</u></a>
			<a href="#"><u>CA3MC.87</u></a>	<a href="#"><u>CA3FC.87</u></a>	<a href="#"><u>CB3MC.87</u></a>	<a href="#"><u>CB3FC.87</u></a>	<a href="#"><u>CC3MC.87</u></a>	<a href="#"><u>CC3FC.87</u></a>	<a href="#"><u>CD3MC.87</u></a>	<a href="#"><u>CD3FC.87</u></a>
			<a href="#"><u>CA3MD.87</u></a>	<a href="#"><u>CA3FD.87</u></a>	<a href="#"><u>CB3MD.87</u></a>	<a href="#"><u>CB3FD.87</u></a>	<a href="#"><u>CC3MD.87</u></a>	<a href="#"><u>CC3FD.87</u></a>	<a href="#"><u>CD3MD.87</u></a>	<a href="#"><u>CD3FD.87</u></a>
MEDIO BAJO			<a href="#"><u>CA4MA.87</u></a>	<a href="#"><u>CA4FA.87</u></a>	<a href="#"><u>CB4MA.87</u></a>	<a href="#"><u>CB4FA.87</u></a>	<a href="#"><u>CC4MA.87</u></a>	<a href="#"><u>CC4FA.87</u></a>	<a href="#"><u>CD4MA.87</u></a>	<a href="#"><u>CD4FA.87</u></a>
			<a href="#"><u>CA4MB.87</u></a>	<a href="#"><u>CA4FB.87</u></a>	<a href="#"><u>CB4MB.87</u></a>	<a href="#"><u>CB4FB.87</u></a>	<a href="#"><u>CC4MB.87</u></a>	<a href="#"><u>CC4FB.87</u></a>	<a href="#"><u>CD4MB.87</u></a>	<a href="#"><u>CD4FB.87</u></a>
			<a href="#"><u>CA4MC.87</u></a>	<a href="#"><u>CA4FC.87</u></a>	<a href="#"><u>CB4MC.87</u></a>	<a href="#"><u>CB4FC.87</u></a>	<a href="#"><u>CC4MC.87</u></a>	<a href="#"><u>CC4FC.87</u></a>	<a href="#"><u>CD4MC.87</u></a>	<a href="#"><u>CD4FC.87</u></a>
			<a href="#"><u>CA4MD.87</u></a>	<a href="#"><u>CA4FD.87</u></a>	<a href="#"><u>CB4MD.87</u></a>	<a href="#"><u>CB4FD.87</u></a>	<a href="#"><u>CC4MD.87</u></a>	<a href="#"><u>CC4FD.87</u></a>	<a href="#"><u>CD4MD.87</u></a>	<a href="#"><u>CD4FD.87</u></a>
BAJO			<a href="#"><u>CA5MA.87</u></a>	<a href="#"><u>CA5FA.87</u></a>	<a href="#"><u>CB5MA.87</u></a>	<a href="#"><u>CB5FA.87</u></a>	<a href="#"><u>CC5MA.87</u></a>	<a href="#"><u>CC5FA.87</u></a>	<a href="#"><u>CD5MA.87</u></a>	<a href="#"><u>CD5FA.87</u></a>
			<a href="#"><u>CA5MB.87</u></a>	<a href="#"><u>CA5FB.87</u></a>	<a href="#"><u>CB5MB.87</u></a>	<a href="#"><u>CB5FB.87</u></a>	<a href="#"><u>CC5MB.87</u></a>	<a href="#"><u>CC5FB.87</u></a>	<a href="#"><u>CD5MB.87</u></a>	<a href="#"><u>CD5FB.87</u></a>
			<a href="#"><u>CA5MC.87</u></a>	<a href="#"><u>CA5FC.87</u></a>	<a href="#"><u>CB5MC.87</u></a>	<a href="#"><u>CB5FC.87</u></a>	<a href="#"><u>CC5MC.87</u></a>	<a href="#"><u>CC5FC.87</u></a>	<a href="#"><u>CD5MC.87</u></a>	<a href="#"><u>CD5FC.87</u></a>
			<a href="#"><u>CA5MD.87</u></a>	<a href="#"><u>CA5FD.87</u></a>	<a href="#"><u>CB5MD.87</u></a>	<a href="#"><u>CB5FD.87</u></a>	<a href="#"><u>CC5MD.87</u></a>	<a href="#"><u>CC5FD.87</u></a>	<a href="#"><u>CD5MD.87</u></a>	<a href="#"><u>CD5FD.87</u></a>

