

EL EFECTO DEL TURNO DE HABLA
Y DEL DISCURSO DIRECTO
EN LA EXPRESIÓN DE SUJETOS PRONOMINALES
DE PRIMERA PERSONA SINGULAR

EFFECTS OF TURN-TAKING AND DIRECT SPEECH
IN THE EXPRESSION OF FIRST PERSON SINGULAR
PRONOMINAL SUBJECTS

LEONOR OROZCO

Universidad Nacional Autónoma de México

leonor.orozco@comunidad.unam.mx

orcid: 0000-0002-9516-7940

RESUMEN: El propósito de esta investigación es estudiar la expresión de los sujetos pronominales de primera persona singular considerando dos factores relacionados con la estructura de la interacción: el turno de habla y el discurso directo. En el análisis se incluyen además otros factores lingüísticos que han resultado significativos en estudios anteriores. Los resultados de la investigación indican que tanto el inicio de turno de habla como los contextos en discurso directo favorecen la expresión de *yo* en entrevistas sociolingüísticas. Esto muestra la necesidad de analizar el efecto de elementos relacionados con la estructura de la interacción en los estudios de los sujetos pronominales.

Palabras clave: sujetos pronominales; primera persona singular; estructura de la interacción; turno de habla; discurso directo.

ABSTRACT: The aim of this article is to study the expression of singular first-person pronominal subjects by analyzing two factors related to the structure of interaction: turn-taking and direct speech. The analysis also takes into account other linguistic factors that have proved to be significant in previous studies. The results indicate that both turn-taking and direct speech are statistically significant for our understanding of the expression of *yo/self* in sociolinguistic interviews. This demonstrates the need to incorporate factors related to the structure of the interaction in studies of pronominal subjects.

Keywords: pronominal subjects; singular first-person; structure of interaction; turn-taking; direct speech.

Recepción: 1º de diciembre de 2020; aceptación: 25 de noviembre de 2021.

INTRODUCCIÓN

El español se caracteriza por ser una lengua en la que la marcación de sujeto no es obligatoria (parámetro de sujeto nulo o *pro drop*) y en la que la información sobre la persona gramatical puede recuperarse gracias a la flexión verbal; por tanto, la expresión de sujeto constituye un fenómeno variable, como se muestra en los ejemplos (1*a*) y (1*b*). En (1*a*) no hay sujeto pronominal explícito, pero la flexión del verbo *ver* concuerda con una primera persona singular. En ese contexto también es posible emplear un sujeto explícito, como se muestra en (1*b*). Por ello, la sociolingüística hispánica se ha concentrado en una pregunta de investigación que procura entender cuáles son los factores lingüísticos y extralingüísticos que motivan la expresión de sujetos pronominales¹.

(1) *a.* \emptyset *vi* la habilidad que tienen ustedes en la guitarra;

b. *yo vi* la habilidad que tienen ustedes en la guitarra.

(Entrevista 17, *El habla popular*.)

Gracias a los estudios sociolingüísticos, se sabe que entre los factores más importantes que explican la variabilidad en la expresión de los sujetos pronominales se encuentran la persona gramatical, el cambio de referencia y la categoría semántica del verbo. Asimismo, aunque con algunas divergencias entre estudios, la ambigüedad morfológica ha sido significativa. También se ha analizado la influencia del tiempo, el modo, la modalidad, el tipo de cláusula, la perseverancia o *priming* (por ej., Travis 2007) y la frecuencia léxica (Erker y Guy 2012), entre otros factores².

En la mayoría de las investigaciones —realizadas en diferentes variedades dialectales— se ha observado que la primera persona singular suele tener porcentajes de expresión más altos que las otras (cf. Orozco y Guy 2008; Sánchez Arroba 2013; Michnowicz 2015; Manjón Cabeza *et al.* 2016; Limmerick 2019)³. Esta característica llama

¹ Otros aspectos ligados al parámetro de sujeto nulo, como el orden sujeto-verbo, también han sido objeto de estudio de la sociolingüística hispánica. Véase, por ejemplo, DAUPHINOIS & ORTIZ LÓPEZ 2016.

² Puesto que los estudios sobre el tema son realmente abundantes, y lo que me interesa en este trabajo es centrarme de manera exclusiva en el sujeto pronominal de primera persona singular, me limito a señalar algunos textos que ofrecen un excelente panorama de la investigación sobre sujetos pronominales en el mundo hispánico: FLORES FERRÁN 2007, POSIO 2018 y MARTÍN BUTRAGUEÑO 2020.

³ En el español de la Ciudad de México, CANTERO SANDOVAL (1976) reporta que los sujetos pronominales de primera persona son los más frecuentes, aunque este autor sólo contabiliza sujetos explícitos.

la atención, pues, más allá de las diferencias estructurales con otros pronombres —como su carácter deíctico y la imposibilidad de que la primera persona alterne con frases nominales plenas—, se ha señalado que su expresión se relaciona con diversas estrategias que tienen que ver con el involucramiento del hablante (Davidson 1996). Sin embargo, se precisa de más estudios sociolingüísticos que analicen en detalle la incidencia de otro tipo de componentes en la variación de sujetos pronominales.

El propósito de esta investigación es analizar dos factores vinculados con la estructura de la interacción en entrevistas sociolingüísticas y comprender cómo se interrelacionan con otros factores previamente analizados en la bibliografía. La hipótesis de trabajo plantea que la expresión del pronombre de primera persona singular está motivada, en primer lugar, por el turno de habla, como en (2), en que el pronombre *yo* ocurre al principio, cuando el informante rompe a hablar en respuesta a una intervención apelativa del encuestador:

- (2) Enc.— Y ¿cómo trabajan ustedes en ese sentido? Digo, ¿cuál es el plan, digamos?
 Inf.— *Yo* no he trabajado así; pero en el sentido de la rehabilitación, la tendencia es hacer que el niño actúe como cualquier chico normal, sin sobreprotegerlo, es decir, sin cuidarlo demasiado (Entrevista 5, *La norma culta*).

En segundo lugar, se espera que la reproducción de estructuras dialógicas y la recreación de situaciones de enunciación en los fragmentos en discurso directo motive también la expresión del *yo*, tal y como se observa en (3), en que la entrevistada da voz a una tercera persona, su pretendiente, al narrar que éste buscaba pretextos para ir a verla, pidiendo permiso a Lolita —la propietaria de la casa en donde ella vivía— de visitarla y hacerle obsequios. En el fragmento que trata de recrear, pone en boca de su pretendiente formas de tratamiento explícitas (*Lolita, usted*), así como formas autorreferenciales (*me, yo*), con lo cual se realiza a los participantes de ese acto de enunciación:

- (3) Ya. Empezó allí. Y luego ya todos los días, ya no sabía cómo... hablar a... “Lolita... este... ¿me permitirá usted que la vaya... que le obsequie *yo* a usted una... un busto de Beethoven?” (Entrevista 11, *La norma culta*).

ANTECEDENTES

El primero de los estudios que se centraron en el estudio de los sujetos de primera persona en español corresponde a Paola Bentivoglio (1987), quien analiza tanto la primera persona singular como plural y

concluye que solamente los sujetos singulares favorecen la expresión. En esta investigación, el cambio de referencia, la ambigüedad morfológica, los verbos de percepción y cognición, así como el cambio de turno, favorecen la expresión del *yo*, mientras que el énfasis —factor al que algunas gramáticas (Alarcos Llorach 1969; Gili Gaya 1970; RAE 1973) atribuían la presencia del sujeto— no resultó estadísticamente significativo. Más aún, la autora muestra que los contextos enfáticos o contrastivos son escasos en los datos analizados y expone el problema que implica la interpretación divergente del término *énfasis* según diversos autores.

Investigaciones posteriores a la de Bentivoglio que también se abocaron a los pronombres de primera persona confirman, en términos generales, las tendencias de su estudio. El cambio de referencia constituye un factor cuya incidencia es clara y constante en varios estudios realizados en comunidades monolingües muy diversas, así como en comunidades bilingües (Lowther 2004), con hablantes no nativos (Geeslin & Gudmestad 2016) y en conversaciones en las que hay *code switching* (Torres Cacoullous & Travis 2010).

La ambigüedad morfológica es un factor con resultados e interpretaciones divergentes. En la mayoría de los estudios, éste sí es estadísticamente significativo (Travis 2005 y 2007; Pérez Brabandere 2010; Torres Cacoullous & Travis 2010; Manjón Cabeza *et al.* 2016), con excepción de Lowther (2004). Sin embargo, Silva Corvalán (1997) sugiere que la expresión de sujetos pronominales es mayor en tiempos verbales donde hay ambigüedad entre primera y tercera persona singular a causa de las funciones pragmáticas que éstos desempeñan y no a la ambigüedad morfológica. La autora indica que el copretérito, pospretérito y los tiempos en modo subjuntivo suelen desempeñar funciones de *background* en el discurso narrativo; en ellas se da relevancia al referente, lo que propicia la expresión de sujetos pronominales, mientras que en tiempos como el pretérito la expresión de sujetos pronominales es menor porque éstos desempeñan funciones de *foreground* en las que el evento tiene la relevancia.

Con respecto a la categoría semántica de los verbos, en estudios recientes se sugiere que, si bien los verbos de cognición favorecen la presencia de sujetos pronominales —y más concretamente los de primera persona singular—, *creer* y *saber*, al ser los verbos cognitivos más frecuentes, pueden crear un efecto de clase. De hecho, Ramos (2016, p. 121) apunta que la emergencia de la categoría de verbos de cognición como factor favorecedor de la expresión de sujetos pronominales no se había consolidado en español durante los siglos XIV a XVI; en su análisis de este período, la clase semántica del verbo no resultó estadísticamente significativa, y sí lo fueron los factores cambio de referencia, distancia y el tiempo verbal.

Además, las construcciones con el verbo *creer* en presente de indicativo (*yo creo*) favorecen claramente la expresión de *yo*, tendencia que sugiere un alto grado de gramaticalización (Posio 2013, 2014) y su procesamiento como unidad (Travis & Torres Cacoulllos 2012). Estas particularidades han llamado la atención de varios estudiosos dedicados a analizar las diferencias semánticas, discursivas y pragmáticas relacionadas con su expresión u omisión, así como con su anteposición o posposición al verbo (véanse, por ej., Aijón Oliva y Serrano 2010; Xie 2019; Hennemann 2016; Soler Bonafont 2016).

En cuanto al cambio de turno se ha señalado que éste no opera de igual manera para todas las personas que conforman el paradigma pronominal, pero que sí es un factor pertinente para la primera persona singular (Alfaraz 2015; Manjón Cabeza *et al.* 2016). En estudios sobre la primera persona hay resultados divergentes a los de Bentivoglio (1987). En un trabajo más reciente llevado a cabo también con datos de Caracas, el cual retoma de Bentivoglio la oposición mismo turno/ turno diferente, Pérez Brabandere (2010) encuentra que este factor no es significativo. De igual manera, en su investigación sobre el efecto del *priming* en datos de narrativas en Nuevo México y de conversaciones espontáneas en Colombia, Travis (2007) indica que la posición en el turno no es pertinente. Tampoco lo es en el estudio de Torres Cacoulllos y Travis (2010) con datos de hablantes de español en Nuevo México, en el cual se favorece la expresión siempre y cuando la realización previa del sujeto pronominal se exprese, el sujeto sea diferente, haya verbo psicológico y ambigüedad morfológica. Sin embargo, en otra investigación de Travis y Torres Cacoulllos (2012), el factor “unidad entonacional” en correlación con los verbos cognitivos incrementa las frecuencias de explicitud de manera notable (hasta 83%, comparado con los contextos de no inicio de unidad entonacional). Por último, en un estudio en que se analizan datos de español y de portugués, Posio (2008) advierte que hay una tendencia más sistemática en español —en comparación con el portugués— a la expresión de sujetos pronominales cuando hay más turnos de habla.

El efecto del discurso directo en la expresión de los sujetos pronominales de primera persona no se ha analizado en otros estudios en español. Sin embargo, se ha mencionado que en el diálogo reportado los sujetos pronominales explícitos permiten aclarar quién es el enunciatario (Lastra y Martín Butragueño 2015), y en las investigaciones sobre discurso directo se dice que los discursos citados requieren “necesariamente la reconstrucción de su situación de enunciación correspondiente” (Maldonado González 1999, pp. 3555 ss.). En este sentido me parece que, si se reproduce un acto de enunciación, la necesidad de referir al enunciatario y al destinatario podría motivar la presencia de sujetos pronominales de primera y segunda persona

singular, como indican estudios anteriores. En Orozco 2020 se mostró que la expresión del pronombre *usted* se ve favorecida en fragmentos elaborados a partir del discurso directo. En un estudio sobre este tipo de discurso en javanés, lengua en que menudea la elisión de pronombres, Ewing (2014) destaca que el diálogo construido motiva la presencia de la primera persona singular.

En suma, los resultados de las investigaciones previas muestran que hay restricciones claras en cuanto a la expresión de sujetos pronominales en español, pero que algunas no operan de igual manera con todos los pronombres. Además, se ha observado que, si se consideran nuevos factores en el estudio de los sujetos pronominales, emerge información que contribuye al entendimiento de este proceso variable, por lo cual, en consecuencia, los modelos estadísticos varían en función de los factores considerados. Por esta razón, estimo necesario hacer más estudios sociolingüísticos basados en pronombres específicos, como lo plantean Geeslin y Gudmestad (2016). Asimismo, me parece pertinente analizar de qué manera los factores relacionados con la estructura de la interacción inciden en la expresión pronominal de primera persona singular.

METODOLOGÍA

Los datos que aquí se analizan provienen de los corpus *La norma culta* (1971) y *El habla popular de la Ciudad de México* (1976), ambos coordinados por Lope Blanch y grabados entre 1967 y 1974. Estos materiales son valiosos en varios sentidos. En primer lugar, dan cuenta de una sincronía en el español de la Ciudad de México que, siguiendo la metodología variacionista, puede emplearse tanto para realizar estudios en tiempo aparente como en tiempo real (Labov 1994). En segundo lugar, estos corpus se conforman por tres tipos de eventos: entrevistas sociolingüísticas de un participante; grabaciones en las que participan dos hablantes que se conocen y tienen una relación de cercanía, y conferencias. Ello posibilita analizar los efectos de las diferentes dinámicas de interacción en la variación. Para este estudio se seleccionaron 24 entrevistas, de las cuales 12 corresponden a *La norma culta* y 12 a *El habla popular*. En cada corpus se seleccionaron 7 entrevistas de un participante y 5 grabaciones de dos participantes, por lo que se cuenta con datos correspondientes a 17 participantes de cada corpus, lo que suma un total de 34 hablantes, hombres y mujeres de diferentes edades y niveles de escolaridad. Es importante mencionar que en las 14 entrevistas la búsqueda de contextos no se limitó a las participaciones de los entrevistados; sin embargo, se observó que sólo hubo 4 contextos de ocurrencia de primera persona singular correspondientes a los entrevistadores. Asimismo, se documentaron

17 ocurrencias de sujetos de primera persona singular provenientes de otros participantes que interactuaron de manera esporádica: todas ellas forman parte del análisis.

La selección de los datos se basó en el principio de exhaustividad sociolingüística y se realizó a partir de la delimitación del contexto variable (cf. Tagliamonte 2012), esto es, se seleccionaron todos los contextos en que la alternancia entre omisión y expresión del sujeto pronominal de primera persona es posible. En (4) se muestra un ejemplo de alternancia posible: el verbo *agarré* se acompaña de un sujeto pronominal explícito pospuesto (*agarré yo*), mientras que *me dediqué a bolear* no tiene sujeto explícito. En ambos casos la alternancia entre omisión y expresión se hace posible (*agarré yo ~ agarré/ me dediqué a bolear ~ yo me dediqué a bolear*):

- (4) fue como entonces agarré *yo* el cajón, y \emptyset me dediqué a boliar (Entrevista 7, *El habla popular*).

Por lo demás, se excluyeron los contextos en que dicha alternancia no es viable, ya sea porque la omisión o la expresión de sujeto es obligatoria, como ocurre, por ejemplo, en los focos contrastivos en que no es posible omitir el pronombre *yo*, en los contextos con el adjetivo *mismo* (*yo mismo*) o cuando se presenta información focal nueva, como se muestra en (5) (ejemplo tomado de Silva Corvalán y Enrique Arias 2017, p. 173), en el que, mediante la pregunta, se solicita información sobre el sujeto que realizó la acción de traer el diario; por tanto, se espera que la respuesta explicita el sujeto, de modo que en tal contexto sería agramatical omitirlo⁴.

- (5) A. ¿Quién trajo este diario?
B. Lo traje *yo* / * \emptyset lo traje.

Además de los contextos en que la alternancia no es posible, quedaron fuera del análisis todas las ocurrencias de los verbos *creery decir*

⁴ La exclusión de contextos sigue los criterios de la *Guía de codificación para el estudio de los sujetos pronominales* de PRESEEA (Bentivoglio et al. 2011), que cito a continuación: “se excluyen de la codificación contextos obligatorios: a) frases idiomáticas (*Yo qué sé, Tú sabes...*) si no presentan variación...; b) usos enfáticos con mismo/ misma (*Yo mismo lo hice, Lo hiciste tú misma*); c) sujetos focales: i) Sujeto es información nueva que responde a una pregunta: *¿Quién lo trajo? Lo traje yo*; ii) Sujeto es indiscutiblemente necesario por ser foco de contraste: *Pepe estudia y yo trabajo*; d) Excluir contextos en donde NO es posible la presencia de un pronombre personal. Por ejemplo: i) verbos impersonales: *hacer, haber* (\emptyset *Hay una persona en la calle, Hace frío*); ii) cláusulas relativas con sujeto relativizado (*Un estudiante que era de Alcalá*). El sujeto explícito es posible en estilo coloquial en algunos dialectos (*Un estudiante que él era de Alcalá*), pero no los consideramos en la codificación básica; iii) cláusulas con *se* impersonal (*Se come bien en Alcalá*)”.

en presente de indicativo cuando funcionan como marcadores discursivos y, por tanto, no presentan alternancia pronominal. A manera de ejemplo, menciono una entrevista de dos participantes en la que el verbo *decir*, conjugado en primera persona en presente de indicativo, se documentó un total de 116 veces en las que cumplía con tres funciones distintas, a saber, introductor de discurso directo, marcador reformulador y parentético. Este verbo suele emplearse en las entrevistas con sujeto pronominal omitido, como se muestra en (6). Por su parte, el verbo *creer* en presente de indicativo también presentó una alta ocurrencia en las entrevistas, sobre todo acompañado de un sujeto pronominal explícito, como en el ejemplo (7), en que se documenta un caso de *creer* parentético:

- (6) Inf. B.— A mí me gusta la guitarra. Te oigo a ti tocar la guitarra tan bien, que... Sí... *Digo...* hasta... hasta me da envidia. Luego me... me miro los dedos, y no sé ni para dónde menearlos, [235] ¿no?
 Inf. A.— *Digo*, con...
 Enc.— Es pura agilidad, ¿no?
 Inf. A.— Es pura agilidad. *Digo*, sí.
 Inf. B.— Es práctica, ¿verdad?
 Inf. A.— Es práctica. *Digo...*
 Inf. B.— Y...
 Inf. A.— Un poco de aburrimiento... *Digo...* Pero sí se llega a aprender lo que quieres, ¿no? (Entrevista 16, *El habla popular*).
- (7) pues <~ps> se parece a su papá [*yo creo*] (Entrevista 1, *La norma culta*).

La determinación de excluir estos contextos responde a una serie de decisiones relacionadas con la delimitación del contexto variable (cf. Tagliamonte 2012). En el caso de los sujetos pronominales, la bibliografía ha mostrado que los verbos *creer* y *decir* se encuentran entre los cinco más frecuentes, y que éstos suelen emplearse como marcadores discursivos o se presentan en secuencias formulaicas (Posio 2013 y 2014). Travis y Torres Cacoullós (2012) concluyen que la secuencia *yo + creo* se procesa como unidad. Además, los estudios sobre el papel de la frecuencia léxica en la expresión de sujetos pronominales (Erker y Guy 2012; Bayley *et al.* 2013) han confirmado que la frecuencia interactúa con otros factores de manera que amplía o elimina restricciones lingüísticas que operan en la expresión de sujetos pronominales. En el corpus que analizan Erker y Guy (2012), las formas verbales más frecuentes fueron *creo*, *sé*, *digo* y *tengo*; de ellas sólo *creo* mostró un porcentaje alto de expresión pronominal, mientras que las otras tres no. Asimismo, el verbo *creer* presenta las frecuencias más altas de expresión de sujetos pronominales en

datos de diferentes dialectos del español y en la Ciudad de México (Martín Butragueño 2020).

La selección de los factores lingüísticos que se analizan en este trabajo toma como punto de partida los que han resultado significativos en la expresión de sujetos pronominales en otros estudios. Asimismo, y en función de los objetivos de la investigación, se analizan dos factores vinculados con la estructura de la interacción. La decisión de incluir factores que han resultado sobresalientes para entender cómo actúa la variación en la expresión de sujetos pronominales, además de dos factores relacionados con la estructura de la interacción, permitirá analizar cómo opera cada uno de ellos y cuál es la importancia relativa de cada grupo de factores en cuanto a la expresión de la primera persona singular.

Los factores sometidos a análisis son correferencia, modo, tiempo, ambigüedad y modalidad. Los dos que se relacionan con la estructura de la interacción son turno de habla y discurso directo. Para la clasificación de los factores examinados se retomaron las propuestas de Bentivoglio *et al.* (2011) y de Otheguy y Zentella (2012), lo que permite la comparabilidad entre diversos estudios sociolingüísticos. En la Tabla 1 se describe cada uno de los factores y sus diferentes niveles:

TABLA 1

Factores analizados

<i>Factor</i>	<i>Niveles de cada factor</i>
Cambio de referencia	a) Cambio total. b) Cambio parcial. c) Correferencia.
Tiempo verbal	Se distinguieron todos los tiempos documentados; la información detallada se presenta más adelante en Tabla 4.
Modo	a) Indicativo. b) Subjuntivo.
Modalidad	a) Declarativa. b) Interrogativa.
Ambigüedad morfológica	a) Sí hay ambigüedad morfológica. b) No hay ambigüedad morfológica.
Posición en el turno de habla	a) Inicio de turno. b) Otra posición en el turno.
Discurso directo	a) Se ubica en un fragmento en discurso directo. b) No se ubica un fragmento en discurso directo.
Verbo	Como factor aleatorio.

Con el fin de mostrar cómo se clasificaron los factores que se enlistan en Tabla 1, presento un ejemplo en (8), en el que hay tres verbos concordantes con la primera persona singular. El seguimiento de la referencia se codifica teniendo en cuenta el sujeto de la cláusula previa al contexto analizado, de modo que, sí hay un cambio total de referencia en el primer verbo (*me quedo yo*), pues el sujeto de la cláusula previa es una tercera persona expresada con la frase nominal “mi hija”. En cambio, los dos contextos verbales que siguen (*voy a arreglar*; *me pongo a arreglar*) se clasifican como “correferencia” porque en ambos se mantiene el mismo referente. En cuanto a los demás factores, se observa que las tres ocurrencias están en presente de indicativo, que no hay ambigüedad morfológica, que no se ubican en un fragmento en discurso directo y que ninguna de ellas está en inicio de turno. En (2) se muestra un contexto en el que el pronombre *yo* se documenta a inicio de turno (*yo no he trabajado así*), mientras que (3) corresponde a un contexto en el que la primera persona ocurre en un fragmento en discurso directo.

- (8) Ya luego se va m'hija, y *me quedo yo*. *Voy arreglar*... este... mis camas, arreglar mis cuartos, y luego ya... este... se... ya *me pongo a arreglar* mi cocina (Entrevista 9, *El habla popular*).

Los datos fueron sistematizados en Excel. El análisis estadístico se realizó con el apoyo de Goldvarb (Sankoff *et al.* 2015), a partir del cual se pudo hacer conteos de frecuencias y tabulaciones cruzadas. En el programa Rbrul (Johnson 2009) se efectuó una regresión logística mediante un modelo de efectos mixtos en el que se considera el verbo como factor aleatorio. Por último, se obtuvo un análisis no paramétrico mediante árboles de inferencia condicional y bosques aleatorios en Language Variation Suite (Scrivner & Díaz Campos 2016).

ANÁLISIS DE LOS DATOS

En este apartado se presenta la distribución general de la alternancia pronominal y se describe la colinealidad entre los factores de tiempo, ambigüedad y modo. Posteriormente, se ofrecen los resultados estadísticos del modelo de efectos mixtos, del árbol de inferencia condicional y de los bosques aleatorios. El apartado cierra con la discusión de los resultados.

Resultados globales

Los datos analizados suman un total de 2 706 contextos de ocurrencia; de ellos, en 1 818 (67.18%) no hay sujeto pronominal, 685 (25.31%)

tienen un sujeto antepuesto, y 203 (7.5%), un sujeto pospuesto. A fin de realizar el análisis de regresión, se agruparon los sujetos antepuestos y pospuestos para contrastar la expresión de sujetos pronominales con la omisión, de modo que ambos suman 888 ocurrencias (32.82%)⁵. Estos datos se muestran en la Tabla 2.

TABLA 2

Distribución de los sujetos pronominales de primera persona singular

<i>Sin sujeto pronominal</i>	<i>Sujeto antepuesto</i>	<i>Sujeto pospuesto</i>
1818 (67.18%)	685 (25.31%)	203 (7.5%)
	<i>Con sujeto pronominal</i>	
	888 (32.82%)	
<i>Total de datos analizados</i>	2 706 (100%)	

Los resultados porcentuales para cada uno de los factores analizados se presentan en las dos siguientes tablas. En Tabla 3 se ubican los factores de cambio de referencia, modalidad, posición en el turno de habla y discurso directo. Las frecuencias de expresión para el factor cambio de referencia concuerdan con las tendencias de estudios anteriores, es decir, se hace evidente el incremento en la expresión de *yo* cuando ocurre un cambio total en la correferencia, en que se alcanza un 46.7%, casi 20 puntos porcentuales arriba de aquellos contextos en que la correferencia es parcial (27.5%). Con menor porcentaje de expresión (23.8%), se ubican los casos que mantienen la correferencia; en este sentido, los factores de cambio total y cambio parcial tienen comportamiento semejante al compararlos con el de cambio de referencia.

También es notable la diferencia que puede advertirse en el factor modalidad; es precisamente su especie declarativa la más frecuente en el corpus y la que presenta porcentajes de expresión más altos (33.2%), que casi duplican los de la modalidad interrogativa (17.9%).

⁵ La agrupación de sujetos antepuestos y pospuestos no implica la asunción de equivalencia de significado pragmadiscursivo en ambas posiciones. (ARRIETA COYOTL [2021] analiza actualmente los sujetos pospuestos en este corpus.) Otros autores han mostrado diferencias pragmáticas relacionadas con la posición. Por ejemplo, SILVA CORVALÁN y ENRIQUE ARIAS (2017, p. 183) sugieren que la posposición del sujeto funciona como estrategia desfocalizadora. SERRANO (2014) reconoce que la posición del sujeto otorga a los enunciados diferentes enfoques cognitivos: “El sujeto preverbal ocupa la posición lineal canónica SV, queda destacada su agentividad y es probable que ya haya sido mencionado en el discurso precedente, por lo que resulta más prominente que la variante pospuesta. Esta última, en cambio, adopta una colocación más informativa, entendida como novedosa o imprevista, aportando una focalización pragmática. Por ello, su prominencia cognitiva y su agentividad disminuyen” (p. 327).

Con respecto a las dos variables relacionadas con la estructura de la interacción vemos que, en efecto, el inicio del turno de habla es un contexto que incrementa la expresión de *yo* (39.4% vs. 32% cuando no hay inicio de turno) y que los contextos en discurso directo tienen un porcentaje de expresión ligeramente superior a aquellos en que no hay discurso directo (34.2% vs. 32.7%).

TABLA 3

Distribución de los factores cambio de referencia, modalidad, posición en el turno de habla y discurso directo

<i>Factor</i>	<i>Expresión</i>	<i>Omisión</i>
Cambio de referencia		
Cambio total	46.7% (481)	53.3% (549)
Cambio parcial	27.5% (57)	72.5% (150)
Correferencia	23.8% (350)	76.2 (1 119)
Modalidad		
Declarativa	33.2% (876)	66.8% (1 763)
Interrogativa	17.9% (12)	82.1% (55)
Posición en el turno de habla		
Inicio de turno	39.4% (115)	60.6% (177)
Otra posición en el turno	32.02% (773)	67.98% (1 641)
Discurso directo		
Sí	34.2% (93)	65.8% (179)
No	32.7% (795)	67.3% (1 639)

En Tabla 4 figuran los resultados porcentuales de la expresión de sujetos pronominales para los diferentes tiempos verbales documentados, los cuales se ligan a las variables modo y ambigüedad. Los tiempos se separan en dos categorías: ambiguos y no ambiguos. Esta división revela una tendencia en la que todos los tiempos ambiguos presentan frecuencias de expresión mayores al 40%, mientras que los tiempos no ambiguos no superan en ningún caso el 35.1%. Los contextos de ambigüedad morfológica promedian 56.7% de expresión, mientras que los no ambiguos llegan apenas a 26.4%. Además, en esta tabla se advierte la correlación entre ambigüedad y modo; como se sabe, todos los subjuntivos son ambiguos, y casi todos los indicativos —con excepción del copretérito, pospretérito y antecopretérito— son no ambiguos, de manera que las diferencias observadas en el factor ambigüedad son similares a las del factor modo: el modo indicativo alcanza un porcentaje de expresión de 32.3%, mientras que el subjuntivo llega a 51.2%.

TABLA 4

Distribución de los factores tiempo, modo y ambigüedad morfológica

TAM	Expresión	Omisión
<i>Ambiguos</i>		
Copretérito indicativo	56.4% (257)	43.6% (199)
Pospretérito indicativo	69.6% (16)	30.4% (7)
Presente subjuntivo	41.5% (17)	58.5% (24)
Pretérito subjuntivo	61.5% (24)	38.5% (15)
Antepresente subjuntivo	100% (1)	0% (0)
Antecopretérito indicativo	70.8% (17)	29.2% (7)
Antepretérito subjuntivo	50% (2)	50% (2)
<i>No ambiguos</i>		
Presente indicativo	30% (342)	70% (797)
Pretérito indicativo	21.2% (161)	78.8% (598)
Futuro perifrástico	13.9% (16)	86.1% (99)
Futuro indicativo	18.2% (2)	81.8% (9)
Antepresente indicativo	35.1% (33)	64.9% (61)
<i>Total</i>	32.82% (888)	67.2% (1 818)

Esta distribución sugiere que hay colinealidad estadística entre los tres factores mencionados. Esto es, no existen combinaciones posibles en cada nivel del factor tiempo para cada nivel del factor ambigüedad. Por ejemplo, todas las ocurrencias en presente de indicativo se clasifican también como casos no ambiguos. Lo mismo ocurre en relación con el modo, pues, para la primera persona, el modo subjuntivo siempre es ambiguo, de manera que todos los tiempos en modo subjuntivo no covarían con el factor ambigüedad. La distribución de estos tres factores sugiere la necesidad de probar diferentes modelos estadísticos para corroborar cómo se articula cada uno de ellos, tema que se discute en el siguiente apartado.

Modelo de efectos mixtos

Para el análisis estadístico se decidió usar un modelo de efectos mixtos con el verbo como factor aleatorio; éste se procesó mediante el programa Rbrul (Johnson 2009). Primero se hizo un análisis con todos los factores (análisis a), y, posteriormente, tres análisis más en los que se omitieron dos de los tres factores que muestran colinealidad, es decir, en el análisis (b) se incluyó el tiempo, pero no la ambigüedad ni el modo; en (c), el modo, pero no la ambigüedad ni

el tiempo, y en (d), la ambigüedad, pero no el modo ni el tiempo. De esta manera, se realizaron en total cuatro modelos estadísticos, tres de los cuales fueron descartados por el programa porque falló la convergencia, esto es, el modelo indicó que había colinealidad entre factores⁶. El análisis (d) fue el único en que no hubo problemas de convergencia. Esto coincide con las observaciones de la Tabla 4, en cuyo modelo se muestra que el factor ambigüedad resultó más adecuado para explicar la variación que los factores tiempo y modo.

TABLA 5

Resultados del modelo de efectos mixtos

<i>Factor</i>	<i>Log-odds</i>	<i>N</i>	<i>Proporción</i>	<i>Peso probabilístico</i>
Ambigüedad morfológica (p = 1.06e-49)				
Ambiguo	0.811	571	0.567	0.692
No ambiguo	-0.811	2135	0.264	0.308
Correferencia (p = 2.12e-40)				
Cambio completo de referencia	0.810	1030	0.467	0.692
Cambio parcial	-0.337	207	0.275	0.417
Sin cambio	-0.473	1469	0.238	0.384
Discurso directo (p = 1.92e-04)				
Sí	0.293	272	0.342	0.572
No	-0.293	2434	0.327	0.427
Modalidad (p = 2.91e-04)				
Declarativa	0.59	2639	0.332	0.643
Interrogativa	-0.59	67	0.179	0.357
Posición en el turno de habla (p = 2.13e-03)				
Inicio de turno	0.223	292	0.394	0.556
Otra posición en el turno	-0.223	2414	0.320	0.444
Verbo como factor aleatorio	Intercepta 0.571	2706	0.328	
<i>Input: 0.33; Total: N 2706; Desviación: 3004.104; Df: 8; Intercepta: -0.707; Gran media: 0.328</i>				

⁶ GUY (1988, p. 128, *apud* TAGLIAMONTE 2012, p. 132) comenta que el hecho de que no se dé la convergencia en un modelo estadístico supone una advertencia de que hay factores no ortogonales, es decir, colinealidad entre factores. Enseguida se muestran las advertencias de no convergencia en Rbrul para los modelos (a), (b) y (c): (a) *Model failed to converge with max|grad| = 0.023201 (tol = 0.002, component 1)*; (b) *Model failed to converge with max|grad| = 0.01526 (tol = 0.002, component 1)*; (c) *Model failed to converge with max|grad| = 0.00251626 (tol = 0.002, component 1)*.

Los resultados del modelo de efectos mixtos del análisis (d) se muestran en Tabla 5. La primera columna enlista los factores⁷ y sus diferentes niveles; en la segunda se presentan los *log-odds* o probabilidades: aquellos con valor positivo favorecen la expresión de sujetos de primera persona, pero no aquellos con valor negativo⁸. En las columnas subsecuentes se indica la frecuencia absoluta (N), la frecuencia relativa o proporción y, por último, el peso probabilístico tal como se expresa en el programa Goldvarb, en el cual pesos superiores a 0.500 se interpretan como favorecedores de la expresión de los sujetos pronominales de primera persona singular.

El modelo estadístico seleccionó todos los factores incluidos, lo cual quiere decir que todos son estadísticamente significativos. Para cada factor hay un nivel que favorece la expresión de *yo*. Para el factor de ambigüedad morfológica, los tiempos ambiguos favorecen la expresión del sujeto pronominal (*log-odds*: 0.811; peso probabilístico: 0.692); enseguida, en el factor de correferencia, el cambio completo de referencia es el único nivel que favorece la expresión (*log-odds*: 0.810; peso probabilístico: 0.692); después, se observa que el factor de discurso directo es, en efecto, favorecedor de la expresión, tal y como se planteó en la hipótesis (*log-odds*: 0.293; peso probabilístico: 0.572); luego, se ubica la modalidad declarativa como favorecedora (*log-odds*: 0.59; peso probabilístico: 0.643); por último, el inicio de turno muestra diferencia ante otras posiciones en el turno, lo que apoya la hipótesis del inicio de turno como favorecedor de la expresión (*log-odds*: 0.223; peso probabilístico: 0.556). Con respecto a los verbos, en cuanto factor aleatorio, interesa destacar que el modelo sugiere que las diferencias entre ítems verbales son sobresalientes (*log-odds*: 0.571), lo cual sugiere un comportamiento diferenciado por verbos e indica que, a pesar de ello, todos los efectos fijos incluidos en el análisis son estadísticamente significativos.

Estos resultados permiten ver el orden de importancia de cada uno de los factores, por lo que la jerarquía variable se organiza como en (a). Esto muestra que los dos factores relacionados con la estructura de la interacción se interrelacionan con los otros factores analizados.

⁷ En el modelo proporcionado por Rbrul, los factores o variables independientes son llamados “predictores”; sin embargo, en todo el texto uso el término *factores* para ser congruente con la tradición sociolingüística que los ha llamado “variables independientes” o “factores”.

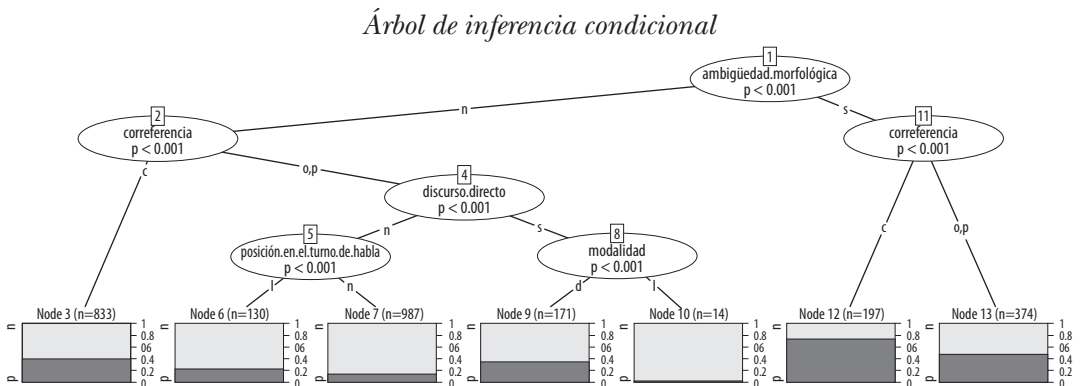
⁸ “Goldvarb factor weights range between 0 and 1, while log-odds can take on any positive or negative value from negative infinity to positive infinity and are anchored around zero. The coefficients are similar to the factor weights—they tell us the degree of contrast among factors and hierarchical organization of their effect. The intercept is similar to the «input» in Goldvarb—it provides a baseline from which the model predictions are built” (TAGLIAMONTE 2012, p. 141).

- (a) ambigüedad morfológica > cambio de referencia > discurso directo > modalidad declarativa > posición de inicio de turno de habla.

Árbol de inferencia condicional y bosque aleatorio

Los datos también se analizaron mediante modelos no paramétricos de regresión, concretamente, árboles de inferencia condicional y bosques aleatorios; éstos se elaboraron en el programa Language Variation Suite (Scrivner y Díaz Campos 2016). En la Figura 1 se ofrecen los resultados para el árbol de inferencia condicional. Éste nos muestra la estructura de los datos desde una perspectiva más icónica. La lectura de resultados se hace a partir del nodo ubicado más arriba, etiquetado con el número 1, que representa el factor con mayor peso en el modelo. A partir de ese nodo es posible observar cómo se van dando diversas bifurcaciones en los datos⁹.

FIGURA 1



El árbol de inferencia condicional confirma que la ambigüedad morfológica es el factor con mayor incidencia en la expresión de *yo*. A partir de este factor, los datos se separan en dos grupos, los contextos ambiguos a la derecha y los no ambiguos a la izquierda. A la derecha del árbol se ubican los contextos correspondientes a la ambigüedad morfológica, para los cuales el segundo factor de importancia es la correferencia (nodo 11), que a su vez se divide en dos ramas, la correspondiente al cambio de correferencia (nodo 12) y la que

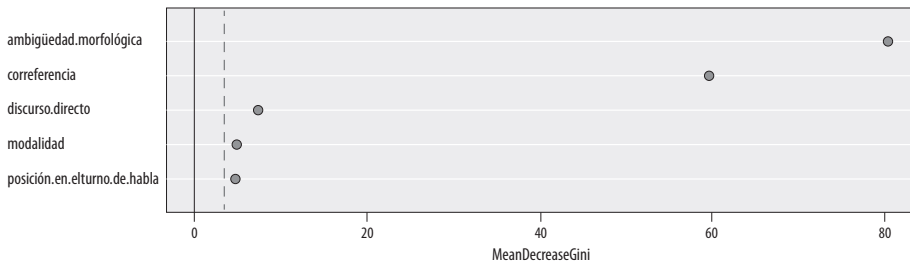
⁹ “Los árboles de inferencia condicional... establecen particiones recursivas binarias, de manera que se va eligiendo primero el predictor o variable independiente más vinculado con la variante dependiente de aplicación...; el predictor se escinde en dos valores o grupos de valores, y el procedimiento se va repitiendo de manera recursiva mientras haya variables significativas” (MARTÍN BUTRAGUEÑO 2020a, p. 651).

agrupa los casos de cambio parcial y total (nodo 13). El nodo 12 presenta la caja con mayor frecuencia de expresión en todo el árbol (75%); por tanto, si un contexto es ambiguo, y además hay cambio de referencia, es muy probable que se emplee un sujeto explícito.

En el lado izquierdo del árbol se ubican los casos de no ambigüedad morfológica; nuevamente, el factor de correferencia (nodo 2) es central y se divide de igual manera que en el nodo 11. Sin embargo, aquí se presentan más bifurcaciones, pues cuando no hay cambio en la referencia, o ésta es parcial, el siguiente factor que explica la presencia de *yo* es el discurso directo. En el nodo 4 vemos una separación que indica que el factor de posición en el turno de habla (nodo 5) destaca sólo en los casos en que no hay discurso directo, y que la modalidad de la oración (nodo 8) descuella únicamente en los contextos en discurso directo para la modalidad declarativa (37%), la cual difiere de la modalidad interrogativa, que no favorece la expresión de sujetos pronominales de primera persona (5%).

En suma, las variables mantienen la misma jerarquía en el árbol de inferencia condicional y en el modelo de efectos mixtos. Para terminar con la comparación de modelos estadísticos, los resultados del bosque aleatorio se muestran en Figura 2.

FIGURA 2

Bosque aleatorio

El bosque aleatorio¹⁰ sitúa los factores por orden de importancia de derecha a izquierda; la línea punteada marca el linde de la significatividad: si algún factor se ubicara a la izquierda de dicha línea, entonces no sería significativo, de modo que en este modelo se consideran como significativos todos los factores analizados, y la jerarquía se presenta como en (b). En el bosque aleatorio se muestra nuevamente que ambigüedad morfológica y correferencia son

¹⁰ Según TAGLIAMONTE (2012, p. 152), los bosques aleatorios “work through a data set by trial and error to establish whether a factor group is a useful predictor of variant choice or not. In so doing, they construct a large number of conditional inference trees, the random forest”.

los dos factores con más peso, pues ambos están cargados a la derecha, mientras que los tres restantes se encuentran hacia a la izquierda, más cerca del límite de la significatividad estadística. Es de notar que en los modelos se incluyen todos los factores analizados, y que se presentan según la misma jerarquía.

- (b) ambigüedad morfológica > cambio de referencia > discurso directo > modalidad declarativa > posición de inicio de turno de habla.

Discusión

En este apartado se comentarán los resultados a la luz de otras investigaciones sobre la expresión de sujetos pronominales en la Ciudad de México y se discutirán los patrones de los cinco factores seleccionados en los modelos estadísticos (ambigüedad morfológica, correferencia, discurso directo, modalidad y posición en el turno de habla), así como el orden en que fueron seleccionados.

Se registró un porcentaje de expresión para la primera persona singular de 32.82%. Este dato puede compararse con otros estudios de la Ciudad de México; al hacerlo, se destaca que los porcentajes de expresión son similares a los de este estudio. En primer lugar, hay que mencionar el trabajo de Cantero Sandoval (1978), quien analizó un corpus de habla mexicana de 25 horas de duración y encontró que la expresión del sujeto de primera persona ronda el 30%, y la no expresión, el 70%. A pesar de que el estudioso no indica con exactitud el corpus analizado ni los porcentajes, deduzco que trabajó con una muestra de las entrevistas de *La norma culta*. En segundo lugar, y dada la cercanía temporal, la comparación con la información que aportan Pozas Loyo y Martín Butragueño (2018) cobra relevancia. Ambos estudiosos examinan datos del corpus *Habla de la Ciudad de México* —grabado entre 1963 y 1969—, inscrito en el Corpus lingüístico oral “Juan M. Lope Blanch”, que ellos mismos coordinan desde 2014. Pozas Loyo y Martín Butragueño documentan para la primera persona singular un total de 34.7% de expresión. En datos más recientes del *Corpus sociolingüístico de la Ciudad de México* (CSCM), Lastra y Martín Butragueño (2015) encuentran un 24.7% de expresión. Como vemos, la cifra aquí registrada se acerca más a las de Cantero Sandoval (1978) y Pozas Loyo y Martín Butragueño (2018), lo cual era de esperar, si se considera que ambos corpus corresponden a la misma época; en el CSCM, sin embargo, parece haber un decremento en lo que toca a la expresión.

Con respecto a la ambigüedad, se observó que este factor se ubica en la cima de la jerarquía; ello se debe sin duda a que las flexiones de primera y tercera persona singular —a diferencia del resto de los

pronombres— son morfológicamente ambiguas en todo el modo subjuntivo, así como en copretérito, antecopretérito, pospretérito y antepospretérito en el modo indicativo. Estos resultados coinciden con estudios anteriores sobre la primera persona, en los cuales la ambigüedad suele ser un factor significativo para la expresión (Bentivoglio 1987; Blanco Canales 1999; Travis 2005 y 2007; Torres Cacoullós y Travis 2010). En cambio, en estudios que analizan todo el paradigma pronominal, el factor de ambigüedad no suele ubicarse entre los más altos en la jerarquía de restricciones (v.gr., Orozco y Guy 2008), lo cual quizá se deba a que, como no todos los sujetos pronominales presentan ambigüedad morfológica, el factor no sea relevante para el conjunto de los datos. Con respecto a los estudios anteriores en México, la ambigüedad morfológica ha resultado significativa en datos tanto de hablantes de la Ciudad de México (Lastra y Martín Butragueño 2015; Pozas Loyo y Martín Butragueño 2018) como en monolingües y bilingües en la Ciudad de Mérida (Michnowicz 2015).

Sobre el factor de correferencia, se comprobó que el cambio de referencia, ya sea parcial o total, motiva de manera sistemática el incremento en la frecuencia de *yo* en los niveles de todos los factores analizados. Asimismo, en el bosque aleatorio (Figura 2) es evidente que tanto la correferencia como la ambigüedad se separan del resto de los factores, lo cual sugiere que el peso de éstos es mayor que el de los tres factores restantes. Asimismo, en el árbol de inferencia condicional (Figura 1) se observa que el nodo 12, que ubica los contextos que reúnen las características <ambiguo> + <cambio de referencia>, alcanza mayor frecuencia de expresión (75%). El factor correferencia ha resultado significativo en todos los estudios sobre sujetos pronominales. En las investigaciones realizadas en el marco del PRESEEA, este factor es el más importante para la expresión pronominal (Martín Butragueño 2020); en los recién llegados a Nueva York, se sitúa en segundo lugar, sólo después de la persona gramatical (Otheguy, Zentella y Livert 2007).

El discurso directo fue factor significativo en los diferentes modelos estadísticos, lo cual prueba la hipótesis aquí planteada de que los contextos en discurso directo favorecen la expresión de *yo* porque el pronombre contribuye a recrear la estructura dialógica narrada colocando a los participantes en una situación de enunciación ajena a la de la entrevista. Lastra y Martín Butragueño (2015) sugieren que la expresión de sujetos pronominales podría contribuir a estructurar el diálogo reportado, y en Orozco (2020), éste fue relevante para la expresión del pronombre *usted*, pero no para la del pronombre *tú*. En el corpus analizado, los pronombres singulares (primera, segunda y tercera persona) suelen hacerse explícitos en los fragmentos en discurso directo.

La modalidad se situó jerárquicamente en cuarto lugar en el modelo de efectos mixtos, en el árbol de inferencia condicional y en el bosque aleatorio. Es de notar que la mayoría de los contextos del corpus corresponde a esta modalidad (97.5%), la cual alcanza el 33.2% de expresión. A partir del árbol de inferencia condicional, se advierte que la modalidad declarativa contribuye a la expresión de *yo* si hay fragmentos en discurso directo (véase nodo 9 en Figura 1) y en los contextos en que no hay ambigüedad ni cambio de referencia.

Los resultados que se relacionan con los factores de discurso directo y modalidad se ejemplifican en (9). En este ejemplo, el entrevistado explica mediante un diálogo reportado que, en su trabajo como zapatero, la decisión sobre el tipo de suelas que pone corresponde a sus clientes. En dicho ejemplo se observa que, al introducir las citas en discurso directo, el entrevistado emplea un sujeto explícito, ya sea gracias a una frase nominal (*el cliente, esa persona*) o a un sujeto pronominal (*él, yo*). Asimismo, los fragmentos en discurso directo en modalidad declarativa que sirven para dar voz a algún cliente se introducen con un sujeto explícito (*yo quiero hule, yo quiero suela*), no así los que dan voz al entrevistado (*te voy a cobrar, pos le voy a cobrar*): en ellos sólo destaca que el entrevistado da a sus interpelados el trato de *tú* o de *usted*, lo que sugiere cómo cambia su tratamiento pronominal ante diferentes clientes. En dicho ejemplo se muestra también que, aun cuando no hay ambigüedad morfológica ni cambio de referencia, se recurre a un sujeto explícito:

- (9) Enc.— ¿Y cómo escoge usted o suela o hule? ¿De qué manera?
 Inf.— Depende del cliente. *El cliente* me dice... este: “*Yo quiero hule*” o “*yo quiero suela*”. Y a la vez *él me dice*: “¿Cuánto me va a cobrar por un... par de suelas?” Entonces *yo le digo*: “Te voy a cobrar tanto”. *Esa persona me dice*: “No. Es muy caro”. “Bueno, pos le voy a cobrar tanto” (Entrevista 2, *El habla popular*).

Por último, el factor de posición en el turno de habla se ubica al final de la jerarquía. Sin embargo, es importante mencionar que en las diferentes tabulaciones cruzadas con otros factores se constató que el inicio de turno incrementa la expresión del sujeto de manera sistemática. En el árbol de inferencia condicional se vio que este factor contribuye a explicar la expresión en aquellos contextos en que no hay ambigüedad morfológica, ni cambio de referencia, ni discurso directo. Ello se debe a que el inicio de turno de habla es un lugar privilegiado, ya sea para apelar o para responder al interlocutor, como se muestra en (10), fragmento constituido por pares de pregunta-respuesta que se configuran en breves turnos de habla. En él, el encuestador está preguntando por los recuerdos de infancia del entrevistado, quien refiere que iba a la escuela y jugaba fútbol; al preguntarle por la posición en que jugaba, emplea un *tú* explícito con

claro carácter apelativo (y tú, ¿de qué jugabas?) y recibe una respuesta con un *yo* explícito. En este fragmento se observa además que no hay cambio de referencia:

- (10) Enc.— ¿Por dónde vivían?
 Inf.— Aquí, por... este... por la Clacotal... por donde está el Palacio de los Deportes.
 Enc.— ¡Ah! ¡Ya, ya! Y entonces, ¿ahí jugaban?
 Inf.— ¡Ahá!
 Enc.— tú, ¿de qué jugabas?
 Inf.— Este... *yo* jugaba de portero.
 Enc.— ¿Ah, te gustaba porterear?
 Inf.— Ahá (Entrevista 1, *El habla popular*).

Es verdad que la estructuración de turnos de habla en las entrevistas sociolingüísticas difiere en varios aspectos de la de una conversación. Por ejemplo, es posible que se presente una asimetría en el tipo de turnos de cada participante, pues, por la propia dinámica de este evento, se espera que el entrevistador haga preguntas y que el entrevistado responda; en este sentido, es probable que los turnos de habla de los entrevistadores sean más breves, y que se limiten a preguntar y a realizar turnos breves de tipo fático; en cambio, los turnos de los entrevistados suelen ser más complejos, pues involucran narraciones, descripciones o argumentaciones (cf. Calsamiglia y Tusón 2007). Sin embargo, al contrastar los datos de las entrevistas de un participante con los de las entrevistas de dos participantes, se observó la misma tendencia, esto es, los inicios de turno de habla que corresponden a un par adyacente de pregunta-respuesta sobre algo relacionado con el hablante motivan el empleo de un sujeto explícito. En (11) se muestra el ejemplo de una conversación entre dos participantes, en que también se emplea *yo* explícito a inicio de turno. Al igual que en el ejemplo (10), hay un turno de habla breve que corresponde a la pregunta del participante I y una respuesta del participante X que inicia con un *yo* explícito (*yo soy*); en esa respuesta hay tres verbos en primera persona y sólo el de inicio de turno se acompaña de sujeto explícito:

- (11) I: y ¿qué piensas del boicot que nos iban a levantar?
 X: pues *yo soy*/ partidario de la no discriminación y \emptyset considero que Sudáfrica/ era un país que no debió ser invitado puesto que practica la discriminación racial/ que va contra/ el reglamento olímpico/ y fue un/ gran acierto de/ México presentar esa moción para que no se admitiera Sudáfrica/ por las razones que/ \emptyset acabo de indicar (Entrevista 2, *La norma culta*).

CONCLUSIONES

En este artículo se presentaron los resultados de la distribución del sujeto pronominal de primera persona singular en dos corpus de habla de la Ciudad de México en torno a la década de 1970. Los resultados apuntan que el factor de ambigüedad morfológica es más relevante que los factores de modo y tiempo para dar cuenta de la variación de los sujetos de primera persona singular. Otros estudios sobre sujetos de primera persona han mostrado la pertinencia del factor de ambigüedad (Bentivoglio 1987; Travis 2005 y 2007; Pérez Brabandere 2010; Torres Cacoullós & Travis 2010). En cuanto a la correferencia, este estudio corrobora que se trata del factor que impone una de las restricciones lingüísticas más sólidas para dar cuenta de la expresión de sujetos pronominales.

Se esperaba que la posición al inicio de turno de habla y el discurso directo incidieran en la expresión de *yo*, y los resultados cuantitativos muestran que sí hay diferencias en cuanto a la frecuencia de expresión de la primera persona singular en estos contextos y que éstas son estadísticamente significativas, de modo que las hipótesis planteadas se comprobaron. Con la inclusión de estos dos factores en el análisis se pretendía mostrar la manera en que ambos inciden en la expresión pronominal y llamar la atención sobre el papel que la estructura de la cesión y toma de turno (tanto en la interacción como en la narración de interacciones mediante diálogo directo) desempeña en la expresión de *yo* en los datos aquí analizados. Sin embargo, por las particularidades de la dinámica de entrevista sociolingüística, es evidente que estos resultados no pueden generalizarse a otros tipos de interacción, por lo que aún es necesario explorar diversos conjuntos de datos en relación con el factor turno de habla.

Asimismo, las entrevistas sociolingüísticas tienen un importante componente narrativo, y en las secuencias narrativas en que los hablantes hacen uso del discurso directo, las formas autorreferenciales y de tratamiento ayudan a reconstruir una escena enunciativa mediante la mención explícita de los participantes; puesto que la única forma que desempeña la referencia a una primera persona es el pronombre de primera persona singular, el discurso directo propicia su empleo.

En esta investigación se recurrió al uso de un modelo estadístico de efectos mixtos —con el propósito de incorporar el papel de los ítems verbales (Johnson 2009)— y de modelos no paramétricos (Scrivner y Díaz Campos 2016) que corroboraron los resultados del modelo de efectos mixtos. Por último, el árbol de inferencia condicional permitió una visión detallada y jerarquizada de la interacción de variables en diferentes subconjuntos de los datos.

En suma, los resultados de esta investigación muestran la importancia de hacer más estudios sobre sujetos pronominales específicos, como lo sugieren Geeslin y Gudmestad (2016), y la pertinencia de incorporar nuevos factores en el análisis de la variación de sujetos pronominales.

REFERENCIAS

- AIJÓN OLIVA, MIGUEL ÁNGEL y MARÍA JOSÉ SERRANO 2010. "El hablante en su discurso: expresión y omisión del sujeto de *creo*", *Oralia: Análisis del Discurso Oral*, 13, pp. 7-38; doi: 10.25115/oralia.v13i.8100.
- ALARCOS LLORACH, EMILIO 1969. *Gramática estructural*, Gredos, Madrid.
- ALFARAZ, GABRIELA 2015. "Variation of overt and null subject pronouns in the Spanish of Santo Domingo", en *Subject pronoun expression in Spanish. A cross-dialectal perspective*. Eds. Ana M. Carvalho, Rafael Orozco y Naomi Lapidus Shin, Georgetown University Press, Washington, D.C., pp. 3-16.
- ARRIETA COYOTL, ABRAHAM SAMUEL 2021. *Posposición del sujeto pronominal de primera persona singular en el habla de la Ciudad de México*, tesis, Universidad Nacional Autónoma de México, México.
- BAYLEY, ROBERT, KRISTEN GREER & CORY HOLLAND 2013. "Lexical frequency and syntactic variation: A test of a linguistic hypothesis", *University of Pennsylvania Working Papers in Linguistics*, 19, 2, art. 4.
- BENTIVOGLIO, PAOLA 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*, Universidad Central de Venezuela, Caracas.
- BENTIVOGLIO, PAOLA, LUIS ORTIZ y CARMEN SILVA CORVALÁN 2011. "La variable «expresión del sujeto pronominal». Guía de codificación", *Preseea*, Grupo de análisis de expresión de sujetos, en http://preseea.linguas.net/Portals/0/Metodologia/guia_codificacion_sujetos_julio_2011.pdf.
- BLANCO CANALES, ANA 1999. "Presencia/ ausencia de sujeto pronominal de primera persona en español", *Español Actual*, 72, pp. 31-40.
- CALSAMIGLIA BLANCAFORT, HELENA y AMPARO TUSÓN VALLS 2007. *Las cosas del decir. Manual de análisis del discurso*, 2ª ed. act. Pref. de Teun A. van Dijk, Ariel, Barcelona.
- CANTERO SANDOVAL, JORGE GUSTAVO 1976. "Peculiaridades en el empleo del pronombre personal *yo* en el habla culta de la ciudad de México", *Anuario de Letras*, 14, pp. 233-237; doi: 10.19130/iifl.adel.14.0.1976.368.
- CANTERO SANDOVAL, JORGE GUSTAVO 1978. "Observaciones sobre la expresión innecesaria de los pronombres personales sujeto del español de México", *Anuario de Letras*, 16, pp. 261-264; doi: 10.19130/iifl.adel.16.0.1978.1272.
- DAUPHINOIS, ASHLEE & LUIS ORTIZ LÓPEZ 2016. "Microvariation in the null subject parameter: Word order in Cuban Spanish", en *Inquiries in Hispanic linguistics: From theory to empirical evidence*. Eds. Alejandro Cuza, Lori Czerwionka & Daniel Olson, John Benjamins, Amsterdam, pp. 281-299.
- DAVIDSON, BRAD 1996. "«Pragmatic weight» and Spanish subject pronouns: The pragmatic and discourse uses of «tú» and «yo» in spoken Madrid Spanish", *Journal of Pragmatics*, 26, 4, pp. 543-565; doi: 10.1016/0378-2166(95)00063-1.
- ERKER, DANIEL & GREGORY GUY 2012. "The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish", *Language*, 88, 3, pp. 526-557; doi: 10.1353/lan.2012.0050.
- EWING, MICHAEL C. 2014. "Motivations for first and second person subject expression and ellipsis in Javanese conversation", *Journal of Pragmatics*, 63, pp. 48-62; doi: 10.1016/j.pragma.2013.09.023.

- FLORES FERRÁN, NIDIA 2007. "A bend in the road: Subject personal pronoun expression in Spanish after 30 years of sociolinguistic research", *Language and Linguistics Compass*, 1, 6, pp. 624-652; doi: 10.1111/j.1749-818X.2007.00031.x.
- GEESLIN, KIMBERLY L. & AARNES GUDMESTAD 2016. "Subject expression in Spanish. Contrasts between native and non-native speakers for first and second-person singular referents", *Spanish in Context*, 13, 1, pp. 53-79; doi: 10.1075/sic.13.1.03gee.
- GILI GAYA, SAMUEL 1970. *Curso superior de sintaxis española*, Vox, Barcelona.
- GUY, GREGORY R. 1988. "Advanced VARBRUL analysis", en *Linguistic change and contact. (Proceedings of the Sixteenth Annual Conference on New Ways of Analyzing Variation)*. Eds. Kathleen Ferrara, Becky Brown, Keith Walters & John Baugh, University of Texas, Austin, pp. 124-136.
- HENNEMANN, ANNE 2016. "A cognitive-constructionist approach to Spanish *creo* Ø and *creo yo* '[I] think'", *Folia Linguistica*, 50, 2, pp. 449-474; doi: 10.1515/flin-2016-0017.
- JOHNSON, DANIEL EZRA 2009. "Getting off the Goldvarb standard: Introducing Rbrul for mixed-effects variable rule analysis", *Language and Linguistics Compass*, 3, 1, pp. 359-383; doi: 10.1111/j.1749-818X.2008.00108.x.
- LABOV, WILLIAM 1994. *Principles of linguistic change*. T. 1: *Internal factors*, Blackwell, Oxford.
- LASTRA, YOLANDA y PEDRO MARTÍN BUTRAGUEÑO 2015. "Subject pronoun expression in oral Mexican Spanish", en *Subject pronoun expression in Spanish. A cross-dialectal perspective*. Eds. Ana María Carvalho, Rafael Orozco & Naomi Lapidus Shin, Georgetown University Press, Washington, D.C., pp. 39-57.
- LIMMERICK, PHILIP 2019. "The interaction effects on variable subject pronoun in Spanish", *Linguística y Literatura*, 77, pp. 300-320; 10.17533/udea.lyl.n77a13.
- LOPE BLANCH, JUAN M. (coord.) 1971. *El habla de la Ciudad de México. Materiales para su estudio*, Universidad Nacional Autónoma de México, México.
- LOPE BLANCH, JUAN M. (coord.) 1976. *El habla popular de la ciudad de México. Materiales para su estudio*, Universidad Nacional Autónoma de México, México.
- LOWTHER, KELLY A. 2004. "First person subject pronoun expression in the Spanish of Tucson", *Divergencias. Revista de Estudios Lingüísticos y Literarios*, 2, 2, pp. 41-50.
- MALDONADO GONZÁLEZ, CONCEPCIÓN 1999. "Discurso directo y discurso indirecto", en *Gramática descriptiva de la lengua española*. Dirs. Ignacio Bosque y Violeta Demonte, Espasa Calpe, Madrid, t. 3, pp. 3549-3595.
- MANJÓN CABEZA CRUZ, ANTONIO, FRANCISCA POSE FUREST y FRANCISCO JOSÉ SÁNCHEZ GARCÍA 2016. "Factores determinantes en la expresión del sujeto pronominal en el corpus PRESEEA de Granada", *Boletín de Filología*, 51, 2, pp. 181-207; doi: 10.4067/S0718-93032016000200007.
- MARTÍN BUTRAGUEÑO, PEDRO 2020. "An approach to subject pronoun expression patterns in data from the «Project for the Sociolinguistic Study of Spanish in Spain and America»", *Spanish in Context*, 17, 2, pp. 294-316; doi: 10.1075/sic.00060.mar.
- MARTÍN BUTRAGUEÑO, PEDRO 2020a. "La expresión del sujeto pronominal en la Ciudad de México: hacia un modelo de efectos mixtos", en *Propuestas metodológicas para la investigación lingüística. Aplicaciones teóricas y descriptivas*. Eds. Alan Pérez Barajas y Axel Hernández, Universidad de Colima-Universidad Nacional Autónoma de México, Colima-México, 639-684.
- MICHNOWICZ, JIM 2015. "Subject pronoun expression in Yucatan Spanish", en *Subject pronoun expression in Spanish. A cross-dialectal perspective*. Eds. A. M. Carvalho, R. O. & Naomi Lapidus Shin, Georgetown University Press, Washington, D.C., pp. 101-119.
- OROZCO, RAFAEL y GREGORY R. GUY 2008. "El uso variable de los pronombres sujetos: ¿qué pasa en la costa Caribe colombiana?", en *Selected proceedings of the 4th*

- workshop on Spanish sociolinguistics*. Eds. Maurice Westmoreland & Juan Antonio Thomas, Cascadilla Proceedings Project, Somerville, MA, pp. 70-80.
- OROZCO, LEONOR 2020. "El papel de la interacción en la expresión de sujetos pronominales de segunda persona", *Cuadernos de Lingüística de El Colegio de México*, 7, pp. 1-40; doi: 10.24201/clecm.v7i0.153.
- OTHEGUY, RICARDO, ANA CELIA ZENTELLA & DAVID LIVERT 2007. "Language contact in Spanish in New York toward the formation of a speech community", *Language*, 83, 4, pp. 770-802; doi: 10.1353/lan.2008.0019.
- OTHEGUY, RICARDO & ANA CELIA ZENTELLA (eds.) 2012. *Spanish in New York: Language contact, dialectal leveling, and structural continuity*, Oxford University Press, Oxford.
- PÉREZ BRABANDERE, VANESSA 2010. *Los sujetos pronominales de primera persona en el español de Caracas 2004-2010*, tesis, Universidad Central de Venezuela, Caracas.
- POSIO, PEKKA 2008. *Uso del pronombre personal sujeto de la primera persona del singular en español y portugués hablados: factores semánticos y pragmáticos*, tesis, Universidad de Helsinki, Helsinki.
- POSIO, PEKKA 2013. "The expression of first-person singular subjects in spoken peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences", *Folia Linguistica*, 47, 1, pp. 253-291; doi: 10.1515/flin.2013.010.
- POSIO, PEKKA 2014. "Subject expression in grammaticalizing constructions: The case of *creo* and *acho* «I think» in Spanish and Portuguese", *Journal of Pragmatics*, 63, pp. 5-18; doi: 10.1016/j.pragma.2013.07.001.
- POSIO, PEKKA 2018. "Properties of pronominal subjects", en *The Cambridge Handbook of Spanish linguistics*. Ed. Kimberly Geeslin, Cambridge University Press, Cambridge, pp. 286-306; doi: 10.1017/9781316779194.014.
- POZAS LOYO, JULIA y PEDRO MARTÍN BUTRAGUEÑO (coords.) 2014. "Proyecto para la preservación y estudio del Corpus lingüístico oral «Juan M. Lope Blanch»", El Colegio de México, México.
- POZAS LOYO, JULIA y PEDRO MARTÍN BUTRAGUEÑO 2018. "La expresión del sujeto pronominal en el Corpus Oral «Juan M. Lope Blanch»", ponencia presentada en el XI Congreso Internacional de Historia de la Lengua Española (Lima, Perú, agosto 6-10).
- RAE = Real Academia Española 1973. *Esbozo de una nueva gramática de la lengua española*, Espasa Calpe, Madrid.
- RAMOS, MIGUEL 2016. "Continuity and change. First person singular subject pronoun expression in earlier Spanish", *Spanish in Context*, 13, 1, pp. 103-127; doi: 10.1075/sic.13.1.05ram.
- SÁNCHEZ ARROBA, MARÍA ELENA 2013. "Compensación funcional y sujeto pronominal «tú» en el español de Cuba", *Revista Lengua y Sociedad*, 13, pp. 131-154; doi: 10.15381/lengsoc.v13i1.22631.
- SANKOFF, DAVID, SALI TAGLIAMONTE & ERIC SMITH 2015. *Goldvarb Yosemite: A multivariate analysis application for Macintosh*, en <http://individual.utoronto.ca/tagliamonte/goldvarb.html> [consultado el 25 de septiembre de 2020].
- SCRIVNER, OLGA & MANUEL DÍAZ CAMPOS 2016. "Language Variation Suite: A theoretical and methodological contribution for linguistic data analysis", *Proceedings of the Linguistic Society of America*, 1, 29, pp. 1-15; doi: 10.3765/plsa.v1i0.3734.
- SERRANO, MARÍA JOSÉ 2014. "El sujeto y la subjetividad: variación del pronombre *yo* en géneros conversacionales y de los medios de comunicación del español de Canarias", *Revista Signos. Estudios de Lingüística*, 47, 85, pp. 321-343; doi: 10.4067/S0718-09342014000200008.
- SILVA CORVALÁN, CARMEN 1997. "Variación sintáctica en el discurso oral: problemas metodológicos", en *Trabajos de sociolingüística hispánica*. Ed. Francisco Moreno Fernández, Universidad de Alcalá, Alcalá de Henares, pp. 115-136.

- SILVA CORVALÁN, CARMEN y ANDRÉS ENRIQUE ARIAS 2017. *Sociolingüística y pragmática del español*, Georgetown University Press, Washington, D.C.
- SOLER BONAFONT, M. AMPARO 2016. "Implicaciones pragmáticas del sujeto en el verbo «creo»", en *Nuevas investigaciones lingüísticas. XXX Congreso Internacional de la Asociación de Jóvenes Lingüistas*. Eds. Ángela Benito Ruiz, Pedro Pablo Espino Rodríguez y Bruno Revenga Saiz, Editorial de la Universidad de Cantabria, Santander, pp. 373-389.
- TAGLIAMONTE, SALI A. 2012. *Variationist sociolinguistics. Change, observation, interpretation*, Wiley-Blackwell, Oxford.
- TORRES CACOULLOS, RENA & CATHERINE TRAVIS 2010. "Variable *yo* expression in New Mexico: English influence?", en *Spanish of the U.S. Southwest: A language in transition*. Eds. Susana Rivera Mills & Daniel Villa, Iberoamericana-Vervuert, Madrid-Frankfurt/M., pp. 185-206; doi: 10.31819/9783865278692-012.
- TRAVIS, CATHERINE E. 2005. "The *yo-yo* effect: Priming in subject expression in Colombian Spanish", en *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages*. Eds. Randall Gess & Edward J. Rubin, John Benjamins, Amsterdam-Philadelphia, pp. 329-349; doi: 10.1075/cilt.272.20tra.
- TRAVIS, CATHERINE E. 2007. "Genre effects on subject expression in Spanish: Priming in narrative and conversation", *Language Variation and Change*, 19, 2, pp. 101-135; doi: 10.1017/S0954394507070081.
- TRAVIS, CATHERINE E. & RENA TORRES CACOULLOS 2012. "What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation", *Cognitive Linguistics*, 23, 4, pp. 711-748; doi: 10.1515/cog-2012-0022.
- XIE, YU 2019. "Dos interpretaciones de (*yo*) *creo* (*que*) y su uso atenuante-intensificador", *Estudios de Lingüística. Universidad de Alicante*, 33, pp. 211-231; doi: 10.14198/ELUA0000.33.11.