

¿Lo vio a mí y a tu colega tomando un juicio formal?: un análisis estadístico de juicios de aceptabilidad

Direct object duplication in *rio de la plata* spanish: a statistical analysis of acceptability judgment

Alvarez, Federico; Gualchi, Santiago; Tallon, Laura; Zacchigna, Charo

 **Federico Alvarez**

Universidad de Buenos Aires, Argentina

 **Santiago Gualchi**

Universidad de Buenos Aires, Argentina

 **Laura Tallon**

Universidad de Buenos Aires, Argentina

 **Charo Zacchigna**

Universidad de Buenos Aires, Argentina

Rasal Lingüística

Sociedad Argentina de Estudios Lingüísticos, Argentina

ISSN: 2250-7353

ISSN-e: 2618-3455

Periodicidad: Semestral

vol. 2022, núm. 1, 2022

revistarasaral@gmail.com

Recepción: 01 Mayo 2021

Aprobación: 11 Agosto 2021

URL: <http://portal.amelica.org/amelia/journal/494/4942923002/>

DOI: <https://doi.org/10.56683/rs221056>

Algunos derechos reservados.



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial 4.0 Internacional.

Resumen: En el presente artículo se expone un procedimiento de recolección de juicios de aceptabilidad con el objetivo de contribuir al debate sobre la utilización de metodología cuantitativa en lingüística. Para esto se evalúan las hipótesis de Gualchi (2019) sobre la gramaticalidad de oraciones con duplicación del objeto directo del español rioplatense. Este sostiene que existen tres posibilidades de concordancia para el clítico cuando la posición de objeto directo está ocupada por una coordinación: total, con el primer coordinado o con el último coordinado; y dos posibles lecturas: colectiva y distributiva. Dicho trabajo se basa en un juicio informal y no considera todas las posibles combinaciones de las variables observadas. La presente investigación postula que someter el fenómeno a un estudio controlado cuasi-experimental de base cuantitativa alcanzará resultados en consonancia con los de Gualchi (2019), pero permitirá un mayor nivel de precisión en el análisis. Se realizaron dos experimentos de puntuación de la aceptabilidad de diferentes oraciones en una escala de Likert. Los resultados indican que cinco de las predicciones fueron respaldadas por los datos, pero el caso de la concordancia total y lectura distributiva mostró altos puntajes de aceptabilidad, de manera opuesta a las predicciones.

Palabras clave: metodología, juicios de aceptabilidad, sintaxis, español, clíticos.

Abstract: This article presents a procedure for collecting judgments of acceptability in order to contribute to the debate on the use of quantitative methodology in linguistics. For this, the authors evaluated the hypotheses of Gualchi (2019) on the grammaticality of sentences with duplication of the direct object in Rioplatense Spanish. Gualchi (2019) argues that there are three possibilities for the clitic's agreement when the position of direct object is occupied by a coordination: agreement with the whole construction, agreement with the first coordinated or with the last coordinated; and two possible readings: collective and distributive. The mentioned work is based on informal judgments and does not consider all the possible combinations of the observed variables. This research postulates that submitting the phenomenon to a quasi-experimental study on a quantitative basis will result in the same conclusions in terms of acceptability, but will provide a greater level of detail in the analysis. Two

experiments were conducted where the acceptability of different sentences was evaluated through a Likert scale. The results indicate that five of the predictions were supported by the data, but the case of total agreement and distributive reading showed high acceptability scores, opposite to the predictions.

Keywords: methodology, judgments of acceptability, syntax, Spanish, clitics.

1. Introducción

Los juicios de aceptabilidad como metodología de recolección de datos en lingüística son objeto de polémica. Hay investigadores que defienden el uso de juicios informales como práctica tradicional que ha contribuido al avance en lingüística, y hay quienes sostienen que los juicios de los propios investigadores no son representativos de la totalidad de los hablantes de la lengua estudiada y además están sujetos a una diversa cantidad de sesgos. El presente artículo se postula como una contribución a este debate que, sin desestimar el valor de los juicios informales como punto de partida, fomenta la utilización de métodos cuantitativos de investigación. Nuestro objetivo es aplicar un abordaje experimental de un fenómeno lingüístico en pos de un mayor nivel de precisión en el análisis en comparación al de la introspección del investigador. Para esto, partimos del trabajo de Gualchi (2019) y presentamos dos procedimientos formales de recolección de juicios de aceptabilidad.

A continuación, en la sección 1.1 desarrollaremos las distintas posturas en torno a los juicios de aceptabilidad, para dar pie en 1.2 al marco teórico postulado por Gualchi (2019). En la sección 2 se enuncian objetivos e hipótesis del trabajo. En las secciones 3 y 4 presentaremos la recolección de datos junto con sus respectivos análisis y resultados. Finalmente, en 5 se presentan las conclusiones y líneas de investigación futuras.

1.1 Los juicios de aceptabilidad

Se llama juicios de aceptabilidad al conjunto de tareas empleadas para evaluar las intuiciones de los hablantes sobre el estatus de distintas oraciones en la lengua que hablan, en términos de si estas oraciones les suenan “bien” o “mal” (Schütze y Sprouse, 2014). Implica solicitar a los hablantes que reporten si consideran que una secuencia de palabras determinada es o no una posible oración de su lengua; la interpretación de esta secuencia se halla implícita o se estipula explícitamente. El supuesto fundamental que subyace a este tipo de tareas es que la aceptabilidad se percibe de manera espontánea en reacción a estímulos lingüísticos que sean (o se asemejen a) oraciones. La aceptabilidad cobra sentido en comparación con el concepto de gramaticalidad: mientras que las oraciones gramaticales son aquellas que tienen una representación lícita producto de las computaciones mentales disponibles, la aceptabilidad de una oración se ve modulada por factores extralingüísticos (Sprouse, 2007). Una característica de la aceptabilidad es que, en tanto representación y/o computación cognitiva, no hay manera de medirla directamente, pudiendo recurrir a métodos tanto mediados

por procesos estratégicos, como en los reportes de los sujetos, o no, como en las neuroimágenes.

However, inferences —drawn from any dependent measure — about either the linguistic representations or computations are always indirect. And these inferences are no more indirect in reading times or event-related potentials, etc., than in acceptability judgements: across all dependent measures we take some observable (e.g., a participant's rating on an acceptability judgement task or the time it took a participant to read a sentence) and we try to infer something about the underlying cognitive representations/processes (Gibson y Fedorenko, 2013, p. 93).

En la literatura se ha sugerido que las conclusiones obtenidas a partir de datos de juicios no necesariamente aportan información respecto de cómo está constituida la facultad del lenguaje, en tanto su “realidad psicológica” no haya sido comprobada a través de algún procedimiento experimental que recurra a alguna otra variable dependiente como el tiempo, la tasa de errores o respuestas electrofisiológicas (Edelman y Christiansen, 2003). Esta concepción, sin embargo, se basa en un malentendido: siguiendo a Marantz (2005), los juicios de aceptabilidad constituyen en sí mismos un dato relevante sobre el comportamiento y la cognición humanos del cual una teoría debe dar cuenta.

Otra objeción a los juicios de aceptabilidad es que requieren de conciencia metalingüística, o sea, la capacidad de tratar al lenguaje como objeto de atención y evaluación. Esto, se dice, podría volverlos artificiales y reducir su validez externa (Bresnan *et al.*, 2007). Sin embargo, aunque los reportes de un sujeto sobre su propia percepción requieren que sea consciente de ella y pueda reportarla deliberadamente, estos reportes han sido empleados como un tipo de dato válido en la construcción de teorías sobre la cognición porque tienden a ser sistemáticos y permiten la construcción de teorías falsables.

Según Cowart (1997), la manera más sencilla de recolectar respuestas en la investigación sintáctica consiste en instancias de opciones binarias: el sujeto solo puede responder “sí” o “no”. Un creciente reconocimiento de la existencia de posibilidades intermedias (tanto en los juicios como en la propia gramática), y la frecuente incomodidad de los informantes enfrentados con elecciones forzadas, ha llevado a varios investigadores a ofrecer a sus sujetos una o más opciones de respuesta entre los extremos de la escala. En algunos casos, se ha tratado de ofrecer tantas posibilidades de respuesta como niveles de gramaticalidad se postulan en la gramática. Esto no es necesariamente un problema: si hay más respuestas posibles que categorías provistas por la gramática, no se pierde información al suministrar un exceso de opciones. Si se ofrece una cantidad de respuestas posibles reducida, como en el de la respuesta binaria, puede suponerse que las oraciones aceptables obtendrán una cantidad mayor de respuestas positivas y que la proporción de respuestas positivas va a variar de manera continua en la medida en la que la gramaticalidad también lo haga. Por lo tanto, con un número elevado de informantes, incluso un procedimiento con una baja resolución podría detectar diferencias pequeñas en la aceptabilidad de un estímulo.

Asimismo, existe una controversia en cuanto a la utilización de juicios informales (reportes del propio investigador) y la necesidad de implementar más juicios formales (reportes de sujetos experimentales). En este ámbito, Muñoz Pérez (2014) señala que el hecho de que los estímulos presentados en juicios sean oraciones construidas deliberadamente para ese fin ha sido un motivo de

discusión respecto de su pertinencia en la investigación sobre el lenguaje. El hecho de que en ocasiones las oraciones presentadas sean evidentemente artificiales podría quitarle validez ecológica al estudio. Sin embargo, según Sprouse (2015), el uso de juicios informales permite obtener información sobre estructuras de baja frecuencia en el habla espontánea, pero propias de la gramática de la lengua estudiada. De acuerdo con Muñoz Pérez (2014), habría que demostrar que la calidad de los datos usualmente empleados en la construcción de teorías es insuficiente para considerarlos fiables y que el uso de juicios informales provoca errores sistemáticos para descartarlos como metodología en favor del uso de juicios formales. En este sentido, no considera que la evidencia proporcionada por los detractores de los juicios informales cumpla con estas condiciones. Por lo tanto, si bien no encuentra motivos para rechazar el uso de experimentos, sostiene que tampoco se justifica adoptarlos como única metodología válida para la recolección de juicios.

En discrepancia con la postura que considera innecesario el uso de juicios formales, Gibson y Fedorenko (2013) discuten los argumentos más utilizados en este debate. Para comenzar señalan que la recolección de datos, en general, no es difícil si no hay limitaciones respecto de la cantidad de hablantes o de la cantidad de juicios que pueden tomarse por hablante. Por lo cual, en la mayoría de los casos no hay dificultades que imposibiliten tomar un volumen de datos adecuado para realizar un análisis cuantitativo. Si bien en otras ciencias cognitivas a veces se acepta un número reducido de participantes en un experimento, eso suele estar asociado a un gran número de estímulos por sujeto, mientras que en los juicios informales de los lingüistas suele haber una cantidad muy escasa de ejemplos. Asimismo, plantean que es cierto que el uso de juicios formales y métodos cuantitativos no garantiza la ausencia de variables que intervengan en el fenómeno bajo estudio y no hayan sido consideradas. Tampoco asegura que los resultados vayan a ser correctamente interpretados. Sin embargo, eso no quiere decir que sean funcionalmente idénticos a los juicios informales. El uso de métodos cuantitativos ayuda a minimizar y evidenciar errores que imposibilitan la generalización de los resultados (una muestra limitada y/o no representativa de sujetos o estímulos, sesgos por parte de los investigadores o efectos del contexto en la toma de datos). También consideran que el trabajo de recolección de datos necesario para el uso de juicios formales no es ineficiente ni dificulta el avance teórico de la disciplina. Existen situaciones en las que la recolección de datos es efectivamente difícil, por ejemplo, el relevamiento de lenguas con muy pocos hablantes o cuando se buscan hablantes bilingües con niveles de competencia similares entre sí en ambas lenguas. Aunque en estos casos cualquier dato es mejor que ningún dato, la falta de rigor metodológico es más dañina para el avance de la teoría lingüística que el requisito de recolección formal (Gibson y Fedorenko, 2013, p. 96). Según los autores, si suponemos un artículo académico en el cual hay 50 enunciados empíricos (es decir, 50 juicios informales), de los cuales es correcto el 90%, hay 2.118.760 teorías posibles que son efectivamente consistentes con los datos. Sin un experimento cuantitativo es imposible determinar cuál de esas teorías es la que mejor se ajusta a la realidad. Más aún, esas combinaciones son solo válidas en el caso de que un experimento permita solo la aceptación o el rechazo totales de una hipótesis dada. Si consideramos la posibilidad de apoyo o rechazo

parciales, el número de teorías posibles es aún mayor. La evaluación cuantitativa de patrones empíricos puede servir como base para el desarrollo teórico.

Gibson y Fedorenko (2013) estiman que los lingüistas, principalmente los investigadores que realizan sus propios juicios, pueden estar sujetos al sesgo de confirmación o a sesgos teóricos, por lo cual es preferible el uso de sujetos ingenuos (*naïve*). Para ellos, presentar juicios en reuniones de investigadores, clases, o foros públicos como congresos no es equivalente a la realización de juicios formales en una muestra de varios sujetos. Finalmente, consideran que si bien es cierto que los lingüistas tienden a rechazar los resultados que no son generalizables, esta situación se limita a los casos en los que son hablantes nativos de la lengua sobre la cual se realizaron los juicios. Cuando no es así, los investigadores quedan a la merced de resultados imprecisos.

Además, Gibson y Fedorenko (2013) ponen como ejemplo casos en la literatura en los que el uso de juicios informales derivó en conclusiones equivocadas, según se determinó en instancias posteriores. Sprouse y Almeida (2013) señalan que la selección de Gibson y Fedorenko (2013) no fue aleatoria, sino que incluye fallos ya reconocidos en la literatura. Por esto, presentan dos estudios (Sprouse y Almeida, 2012; Sprouse, Schütze y Almeida, 2013) que retoman la totalidad de los juicios presentes en un libro de sintaxis y una muestra de juicios sobre el inglés estadounidense publicados en distintos números de la revista *Linguistic Inquiry*. Al someterlos a una serie de experimentos cuantitativos obtienen tasas de replicación de 98% y 95%, las cuales están dentro del marco aceptado por las ciencias cognitivas. En respuesta, Gibson, Piantadosi y Fedorenko (2013) señalan que los juicios evaluados en los trabajos de Sprouse, Schütze y Almeida confirman que la recolección de datos no es privativamente trabajosa y que sus tasas de replicación no son representativas del estándar de los datos en la teoría lingüística, dado que varios de los juicios que reportan son sobre datos no controversiales. En consecuencia, indican que, en primer lugar, los juicios cuya tasa de replicación es crucial son aquellos que permiten distinguir entre teorías y, en segundo lugar, que la tasa de errores que reportan es sobre cada juicio. El problema con esto es que, en el caso de que se desee plantear una teoría que pueda dar cuenta con exactitud de todos los datos que se usan para sostenerla (sin considerar su capacidad predictiva sobre otros datos), la cantidad de juicios que puede emplearse está fuertemente limitada. Más aún, si bien los trabajos de Sprouse, Schütze y Almeida encuentran que pudieron replicar 95-98% de los juicios evaluados, Gibson, Piantadosi y Fedorenko señalan que pudieron encontrar que había juicios mal reportados y que consiguieron determinar cuáles únicamente a través del uso de juicios formales. Además, indicaron que el uso de experimentos cuantitativos tiene la virtud adicional de permitir comparar múltiples condiciones entre sí en una misma escala, independientemente de la cantidad de variables en las que difieran.

Más allá de sus diferencias, estos autores (Gibson y Fedorenko, 2013; Gibson, Piantadosi y Fedorenko, 2013; Schütze y Sprouse, 2014; Sprouse, Schütze y Almeida, 2013; Sprouse, 2015; Sprouse y Almeida, 2013; Sprouse y Almeida, 2012) coinciden en que el uso de experimentos cuantitativos posibilita la discusión sobre el tamaño de los efectos observados en los juicios, y que esto es necesario para el progreso del desarrollo de teorías sobre el lenguaje.

Con el objetivo de identificar los casos donde el uso de juicios formales es estrictamente necesario, Marantz (2005) propone que hay tres clases de juicios de aceptabilidad posibles: en primer lugar, los casos en los que la cadena evaluada constituye una secuencia de palabras o morfemas a los que no se les puede asignar una representación fonológica apropiada, es decir, no es posible pronunciarlas como si se tratara de una frase u oración. Estos casos se emplean como ejemplos evidentes de opciones no disponibles en el sistema gramatical de una lengua dada y no ameritan el uso de experimentos para constatar su validez.

La segunda clase la componen los casos en los que se evalúan ejemplos de generalizaciones no controversiales sobre una lengua dada. Aquí las consideraciones metodológicas son útiles a fin de determinar cuál es la cobertura de un ejemplo dado. Este tipo de juicios suele dar cuenta de postulados sobre orden de constituyentes o configuraciones representativas de caso y concordancia. En tercer lugar, presenta los juicios contrastivos sobre ejemplos que permiten decidir entre teorías en competencia. El tipo de fenómenos determinados a partir de estos juicios incluye (pero no se limita a) dominios de localidad para dependencias a larga distancia, juicios sobre correferencia y juicios sobre el alcance semántico de elementos como cuantificadores. En estos casos, las cadenas pueden recibir una representación fonológica y una representación semántica total o parcial. Estos juicios se relacionan con la interpretación semántica o con la correcta formación de la asociación entre sonido y significado. Es en estas circunstancias en las que un juicio formal es más necesario.

Estas tres clases ofrecen un marco útil para plantear los casos en los que es importante realizar juicios formales. Sin embargo, la discusión desarrollada previamente considera únicamente juicios sobre el inglés. Si bien hay una gran cantidad de investigadores dedicados a la construcción de teorías sobre la gramática de otras lenguas, el tamaño de la comunidad de investigadores de habla inglesa es mucho mayor, lo cual permite un control más riguroso de los trabajos publicados. En este marco, Linzen y Oseki (2018) realizaron una serie de experimentos sobre hebreo y japonés, tomando ejemplos correspondientes a las clases 2 y 3 de Marantz, a partir de los cuales concluyen que, si bien los lingüistas entrenados pueden reconocer qué juicios pertenecen a las clases 2 y 3 y los juicios de la clase 2 se replican sin mayor dificultad, la utilización de juicios formales es crítica para la correcta determinación de las diferencias en los juicios de clase 3.

1.2 Clíticos y concordancia en el español rioplatense

Para constatar la sensibilidad de los juicios formales a contrastes teóricamente relevantes, tomamos entonces un fenómeno que, en tanto caso de dependencia de larga distancia, constituye un ejemplo de la clase 3: la concordancia en oraciones con doblado de clítico acusativo y objetos directos que contienen una coordinación de sintagmas de determinante (SD).

El doblado de clíticos ha sido observado en diversas lenguas pertenecientes a distintas familias lingüísticas a partir de distintas propuestas (véase Anagnostopoulou, 2017 para un estado de la cuestión). En el español rioplatense es un fenómeno ampliamente estudiado (véase, por ejemplo, Di Tullio y Zdrojewski, 2006; Estigarribia, 2013; Seco, 2013; Correa, 2004; Mazzuchino, 2013; Sánchez y Zdrojewski, 2013). En la literatura se lo suele relacionar con

rasgos formales como la animacidad, definitud o especificidad del nombre o SN (Di Tullio y Zdrojewski, 2006; Mazzuchino, 2013), así como también se estudia su función en el discurso, enfatizando, introduciendo participantes, marcando la estructura de la información (Estigarribia, 2013; Sánchez y Zdrojewski, 2013). En estudios sobre concordancia y doblado de clíticos en caso dativo encontramos gran cantidad de ejemplos en los que el SN presenta diversas configuraciones: SN en singular, en plural, sustantivos colectivos, gentilicios, construcciones coordinadas de sustantivos y nombres propios, entre otros (Barbeito, 2017, 2018, 2019; Barbeito, Murata y Peri, 2018; Barbeito y Peri, 2019; Mojedano Batel, 2014).

Cuando ahondamos en los corpus o ejemplos de doblado de clíticos acusativo que se brindan en la literatura encontramos que estos incluyen oraciones donde el SN de OD está en singular o en plural. Sin embargo, el eje del presente trabajo son las estructuras coordinadas:

(1)

- a. Nos vio a mí y al doctor.
- b. Me vio a mí y al doctor.
- c. *Lo vio a mí y al doctor

Como antecedentes encontramos solo dos estudios de este fenómeno particular. El primero de ellos es Camacho (2003), quien postula una división de lenguas de acuerdo con si la concordancia parcial implica un cambio de significado o no. Para explicar esto, presenta dos variedades del español (lengua en la que la concordancia no afecta el significado) con características disímiles, tomando ejemplos con doblado de clíticos.

(2)

- a. *Lo vi a Juan y Pedro.
- b. Lo vi a Juan, y a Pedro (tomado de Camacho, 2003, 99)

De acuerdo con Camacho (2003), para ambas variedades, chilena y peruana, respectivamente, la concordancia parcial presenta características de *gapping* o “vaciado”, es decir, elipsis del verbo o de una proyección verbal (Gallego, 2011). El autor consideró que este tipo de oraciones son el resultado de una coordinación de oraciones y un posterior borrado del contenido sintáctico repetido, esto es, el clítico y el verbo:

(3)

Me vio a mí y lo vio al doctor.

En estas variedades la concordancia parcial solo está permitida cuando los coordinados están antecedidos por la preposición a y el segundo está separado por una pausa. Una característica que aboga por la noción de que son construcciones de *gapping* es el hecho de que casos con concordancia parcial como (2b) permiten adverbios como *ayer*, *hoy* y *también*, mientras que esto no es posible si el clítico es plural.

En cuanto a la concordancia total, las variedades se diferencian. La variedad chilena no acepta la coordinación de dos SN cuando cada uno está antecedido por a (4b), mientras que la variedad peruana permite que tanto la coordinación como los SN estén antecedidos por a (5a y b). No obstante, según sea el caso, cada tipo de coordinación se interpreta de manera diferente: la coordinación del tipo de (5a) se interpreta de manera colectiva (un evento), mientras que la coordinación del tipo de (5b) se interpreta de manera distributiva (dos eventos).

(4)

(Dialecto A)

- a. Lo-s vi a Juan y Pedro.
- b. *Lo-s vi a Juan y a Pedro.

(5)

(Dialecto B)

- a. ?Lo-s vi a Juan y Marta.
- b. Lo-s vi a Juan y a Marta (tomado de Camacho, 2003, p. 100)

En este sentido, Camacho (2003, p. 100) señala que este contraste “se seguiría si la coordinación de dos SP (como en 4b) siempre involucrara la coordinación de los nodos más altos con el *gapping* [...] esto explicaría por qué no es posible tener concordancia total en esos casos: no hay un nodo nominal plural con el cual concordar”.

En cuanto a la variedad rioplatense, encontramos la propuesta de Gualchi (2019), quien señala que, a diferencia de lo que ocurre en el dialecto peruano estudiado por Camacho (2003), en español rioplatense oraciones como la de (6a) son posibles. Asumiendo, en el marco del Minimalismo (Chomsky, 1995, 2000, 2001, 2008), una estructura de la facultad del lenguaje en la que el componente semántico opera sobre las estructuras generadas por el componente sintáctico, el análisis en términos de elipsis empleado en (3) resulta problemático para este tipo de construcción, dado que *juntos* requiere de la unidad sintáctica de la secuencia “a mí y al doctor”, como se marca en (6b). Así, la derivación como resultado de un proceso de borrado del clítico y el verbo (6c) no es aplicable, ya que *juntos* debe proyectarse sobre individuos y no sobre proposiciones (obsérvese el contraste en (7)).

(6)

- a. Me vio a mí y al doctor juntos.
- b. Me vio [a mí y al doctor] juntos.
- c. *[Me vio a mí y lo vio al doctor] juntos.

(7)

- a. (Llegaron) Juan y María juntos.
- b. *Canta y baila juntos.

A partir de estas observaciones, Gualchi (2019) deriva los casos de concordancia con el primer coordinado y lectura distributiva en términos similares a los propuestos por Camacho (2003):

(8)

- a. Me vio a mí ayer y al doctor hoy.
- b. [[ST Me vio a mí ayer] y [ST lo vio al doctor hoy]]

En cambio, para derivar las construcciones del tipo de (6a), propone una estructura más simple: *a mí y al doctor* es un sintagma en el que la conjunción y es núcleo, el doctor ocupa la posición de complemento y yo ocupa la posición de especificador. Siguiendo a Zhang (2010), las coordinaciones proyectan la categoría de su especificador, en este caso, del determinante. Así, la estructura de la coordinación queda recogida en (9a) y la del SV en (9b):

(9)

- a. [SD [SD yo] [y [SD el doctor]]]
- b. [SV vio [SD/OD yo y el doctor]].

En un paso posterior de la derivación, se ensambla la proyección verbal con un núcleo *v* que establece *Agree* (Chomsky, 2000, 2001) con un SD dentro de su dominio de localidad. Entendiendo que un SD es local respecto de *v* si se encuentra en una Configuración Mínima (Rizzi, 2001) con este núcleo, serán locales aquellos SD que son mandados-c por *v*, pero que no sean mandados-c por otro SD:

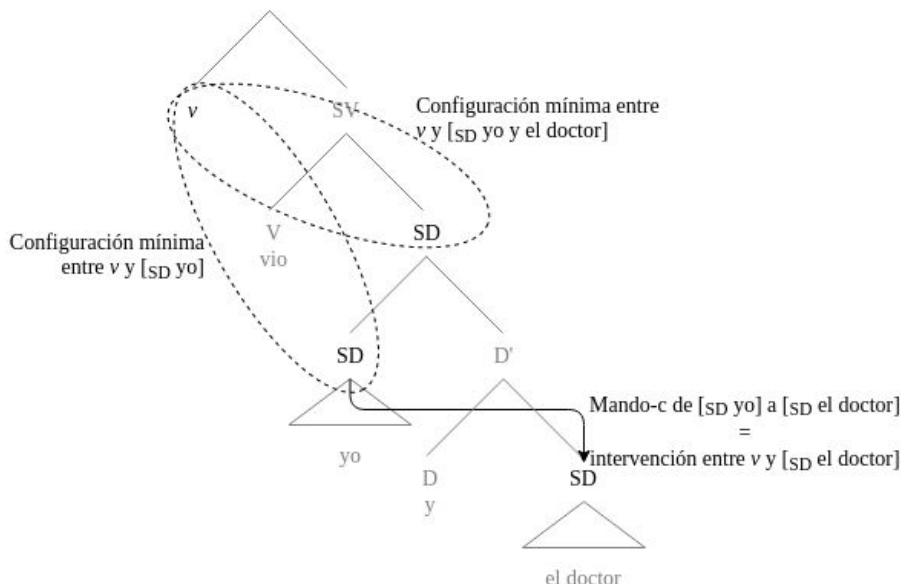


Figura 1
Configuraciones mínimas entre *v* y los SD

Figura 1. Configuraciones mínimas entre *v* y los SD

Así, *v* puede establecer *Agree* con *yo* y *el doctor* y con *yo*, pero no puede hacerlo con *el doctor*, dado que *yo* interviene entre *v* y *el doctor*. En el primer caso, en el que *v* concuerda con *yo* y *el doctor*, la concordancia será total, mientras que en el caso en el que *v* concuerda con *yo* será parcial con el primer coordinado. Ya sea que el clítico es la materialización de la concordancia de objeto entre *v* y el objeto directo o producto de movimiento/copiado desencadenado por dicha operación, los rasgos de persona, género y número de este elemento quedan determinados por los elementos intervenientes en la operación de *Agree* entre *v* y el SD-meta en su dominio de mando-c.

2. Objetivos e hipótesis

En su trabajo Gualchi (2019) considera que en construcciones con doblado de clíticos y objetos directos coordinados existen tres posibilidades de concordar: (i) concordancia total (CT) con el OD; (ii) concordancia con el primer coordinado (CPC), y (iii) concordancia con el último coordinado (CUC). Únicamente para el caso de las construcciones con CPC, estudió dos posibles interpretaciones: (i) una lectura distributiva (LD) de dos eventos; y (ii) una lectura colectiva (LC) de un único evento. A continuación, evaluaremos empíricamente las predicciones derivadas de Gualchi (2019), con el objetivo de demostrar la utilidad de los juicios formales para contrastar condiciones que no forman un par mínimo, y verificar la presencia de diferencias graduadas entre condiciones, que escapan a los

juicios informales binarios. Para ello, hemos conducido dos etapas de recolección de juicios de aceptabilidad bajo los siguientes supuestos metodológicos: (i) los hablantes no acceden conscientemente a su gramática mental y, por lo tanto, no pueden juzgar la gramaticalidad de las oraciones (Chomsky, 1965; Schütze y Sprouse, 2013), (ii) los hablantes pueden emitir juicios de aceptabilidad, que dependen de la gramaticalidad de las oraciones y de otras variables extralingüísticas (Haider, 2007; Sprouse y Almeida, 2012; Sprouse y Schütze, 2017), y (iii) a partir del control de las variables extralingüísticas, podemos emplear los juicios de aceptabilidad como indicador de la gramaticalidad de las oraciones juzgadas (Featherston, 2007).

En el presente trabajo hemos retomado las dos variables propuestas por Gualchi (concordancia y lectura), completando el diseño factorial, que resulta en los siguientes tipos de construcciones:

(10)

- a. CTxLC: Nos conoció a mí y al doctor juntos.
- b. CTxLD: Nos conoció a mí ayer y al doctor hoy.
- c. CPCxLC: Me conoció a mí y al doctor juntos.
- d. CPCxLD: Me conoció a mí ayer y al doctor hoy.
- e. CUCxLC: Lo conoció a mí y al doctor juntos.
- f. CUCxLD: Lo conoció a mí ayer y al doctor hoy.

De acuerdo con el marco teórico y los análisis propuestos por Gualchi (2019) esperamos encontrar los siguientes resultados. En primer lugar, que para las oraciones con CT solo la lectura colectiva sea aceptable dado que, en estas construcciones el clítico *nos* debe concordar con un objeto directo que coordina “a mí y al doctor”.² En segundo lugar, para las oraciones con CPC, esperamos que ambas lecturas sean aceptables. Finalmente, las oraciones con CUC deberían juzgarse con el menor valor de aceptabilidad ya que no serían posibles en español.

3. Experimento piloto

Se realizó un experimento piloto donde cada sujeto puntuó, en una escala de Likert de 1 a 5, un conjunto de oraciones que abarcaban las seis combinaciones posibles entre los tres valores de concordancia y los dos valores de lectura considerados.

3.1 Diseño experimental

3.1.1 Participantes

Sobre un total de 387 personas que respondieron en forma voluntaria un formulario de Google, descartamos a aquellos sujetos que reportaron no ser de nacionalidad argentina, residir fuera de la provincia de Buenos Aires, tener nivel avanzado en alguna lengua diferente del español, o tener una primera lengua distinta del español. La muestra final consistió en 260 personas de entre 16 y 65 años (Mediana 26, IQR 10).³

3.1.2 Materiales

Se diseñaron oraciones estímulo combinando los niveles propuestos de concordancia (CPC, CT y CUC) y lectura (LC y LD), de manera tal que aparecieran las seis combinaciones posibles entre las dos variables. Para recolectar las respuestas se empleó una escala de Likert de 1 a 5, presentada a través de un formulario de Google. Se les presentó a los sujetos una única lista de 58 estímulos⁴. Los primeros 4 eran de práctica y fueron descartados en el análisis, 18 eran *targets* y 36 eran *fillers*. Los 18 estímulos fueron elaborados de modo tal que la lista estaba integrada por seis tipos de estímulos, con tres ítems por condición:

(11)

- a. CTxLC: Nos conoció a mí y al ingeniero juntos.
- b. CPCxLC: Me conoció a mí y al ingeniero juntos.
- c. CUCxLC: Lo conoció a mí y al ingeniero juntos.
- d. CTxLD: Nos conoció a mí el lunes y al ingeniero el viernes.
- e. CPCxLD: Me conoció a mí el lunes y al ingeniero el viernes.
- f. CUCxLD: Lo conoció a mí el lunes y al ingeniero el viernes.

En todos los casos había un ítem con un primer coordinado pronominal singular y un segundo coordinado nominal singular. Para no dar lugar a ambigüedades, la lectura fue forzada, ya fuera colectiva o distributiva, añadiendo un adjunto que especificara la existencia de un evento o de dos. En el caso de la lectura colectiva se añadió el adjunto *juntos* al final de las 9 oraciones, y en el caso de la lectura distributiva se añadió un adjunto a cada coordinado, temporal en cinco casos y locativo en los cuatro restantes.

En todos los casos, el segundo coordinado está en masculino debido a un problema encontrado en una prueba preliminar: durante la generación de los estímulos, las investigadoras del presente trabajo consideraron inaceptables los casos en los que el primer coordinado era una primera persona (*a mí*) y el segundo estaba en femenino. Esto se debe a que, al posicionarse como esa primera persona, el *juntos* no concordaría con un “nosotras” (por ejemplo, *Me conoció a mí — mujer — y a mi hermana juntos*). Por este motivo, se tomó la decisión de que el primer coordinado sea siempre femenino y el segundo sea siempre masculino. De este modo, también se evitó la ambigüedad de tener un clítico masculino singular que pudiera concordar con ambos coordinados (*Lo halló a ella y a su profesor juntos* vs. *Lo atendió a él y al señor juntos*). Los 36 *fillers* estaban compuestos por 18 oraciones consideradas inaceptables por violar la estructura argumental (*Los nadó a los alumnos*) y 18 aceptables (*Lo mató al gasista*). Todos los *fillers* de ambos tipos comenzaban con un clítico para asemejarse a las oraciones *target*.

De los cuatro ítems de prueba, dos eran aceptables y dos eran inaceptables. Se proveyó una solución para todos ellos y estaban acompañados de un texto explicativo sobre su resolución, a fin de que los participantes entendieran bien la consigna. Por ejemplo, para el ítem de práctica *Los chicos están cansada*, se proporcionó el siguiente texto:

A diferencia de las anteriores, esta oración no parece estar bien formada. Sería difícil imaginar que un hablante de español emita una oración como esta y, en cambio, sí podríamos imaginar que un extranjero que está aprendiendo español la use. Una buena puntuación para esta oración es 1-para nada natural.

3.1.3 Procedimiento

Al comienzo del formulario se pidieron los datos personales de cada sujeto: edad, género, nivel de estudios, carrera, profesión, residencia en sus primeros 10 años de vida, idioma en el que le hablaron los padres y finalmente nivel de otro/s idioma/s. La recolección de datos se realizó de forma totalmente anónima. Se pidió a los sujetos que evalúen las oraciones de acuerdo con qué tan naturales les sonaban, siendo 1 “para nada natural” y 5 “totalmente natural”. Al inicio se presentó la siguiente consigna, junto a los cuatro ítems de práctica:

En este cuestionario vas a encontrar una serie de oraciones. Tu tarea es calificarlas con un número entre 1 y 5 según qué tan naturales te suenan. Es importante remarcar que no se espera que contestes según un criterio de corrección como el de la escuela o el de la RAE. Por lo general, un buen criterio es imaginar si podrías escuchar esta oración en un contexto cotidiano. No pienses mucho tus respuestas. En esta ocasión, la primera impresión suele ser correcta.

A continuación se incluyeron los estímulos ordenados de forma pseudoaleatoria, de modo que no aparecieran dos ítems experimentales consecutivos.

3.2. Análisis y resultados

Las respuestas se resumen en el gráfico 1: en el gráfico de la izquierda podemos ver que en los casos de LC (lectura colectiva) tanto para la condición de CPC (concordancia con el primer coordinado) como para CT (concordancia total) la mediana es 5 (el mayor puntaje de aceptabilidad posible). Es decir, la mitad o más de los juicios fueron puntuados con ese valor, mientras que en el caso de CUC (concordancia con el último coordinado) la mediana es 2 y casi la mitad de los juicios se puntuaron con un 1 (el menor puntaje posible). En el caso de LD (lectura distributiva), vemos que la CPC tiene mediana 5, la CT 4, y la CUC 1.

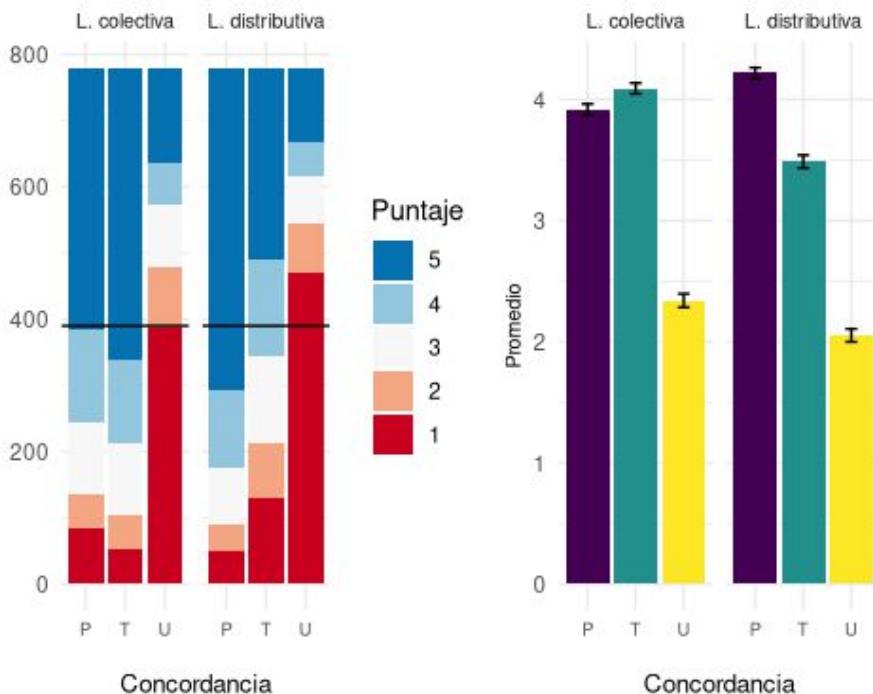


Gráfico 1

Puntuaciones obtenidas en el primer experimento

En el gráfico de la izquierda, el tamaño de la sección de cada barra indica la frecuencia absoluta de los puntajes, el color corresponde al puntaje recibido, y la línea negra que atraviesa las barras indica la mediana. El gráfico de barras de la derecha presenta la media de cada condición. 5 representa el valor más alto de aceptabilidad “totalmente natural”, mientras 1 es el valor más bajo “para nada natural”.

Gráfico 1. Puntuaciones obtenidas en el primer experimento

A la derecha, en el gráfico 1, vemos que las condiciones CPCxLC, CTxLC y CPCxLD poseen todas un promedio similar, alrededor de 4 puntos (CPCxLC: $M = 3.91$, $ES = 0.049$; CTxLC: $M = 4.09$, $ES = 0.044$; CPCxLD: $M = 4.22$, $ES = 0.043$). Del mismo modo, los casos de concordancia con el último coordinado en ambas condiciones de lectura tienen un promedio similar entre sí (CUCxLC: $M = 2.34$, $ES = 0.056$; CUCxLD: $M = 2.05$, $ES = 0.054$). Finalmente, la condición CTxLD tiene un promedio bastante superior al de los casos de CUC, pero es menor al de los otros casos (CTxLD: $M = 3.48$, $ES = 0.053$).

Parametrizamos un modelo lineal con efectos mixtos usando la librería lme4 (Bates, Maechler, Bolker y Walker, 2015) para R (R Core Team 2020), tomando lectura, concordancia y sus interacciones como efectos fijos y a los sujetos como efectos aleatorios (con interceptos y pendientes aleatorias por cada uno de los efectos fijos) junto con la combinación verbo-sustantivo (solamente interceptos). Usando el método de Satterthwaite implementado en el paquete lmerTest (Kuznetsova, Brockhoff y Christensen, 2017) obtuvimos valores-p para las diferentes condiciones. La concordancia mostró un efecto significativo ($F = 63.5894$, $p < 0.001$), mientras que el efecto de lectura no. La interacción entre estas dos variables tampoco alcanzó el umbral de significación ($F = 3.8323$, $p = 0.051$).

Cuadro 1
Efectos principales e interacciones significativas del experimento piloto

Efectos principales	
Lectura	$F = 2.034, p = 0.178$
Concordancia	$F = 63.5894, p < 0.001$
Interacciones	
Lectura-concordancia $F = 3.8323, p = 0.051$	

Cuadro 1. Efectos principales e interacciones significativas del experimento piloto

Sin embargo, un análisis de comparaciones múltiples con corrección de Holm reveló que las condiciones de CUC mostraron diferencias significativas respecto de todas las demás condiciones, pero no entre sí: para los casos de LC, los puntajes de las oraciones con CUC fueron mucho más bajos que los de las oraciones con CPC ($= -1.57$, ES = 0.243, $z = -6.479$, $p < 0.001$), o que tuvieran CT ($= -1.75$, ES = 0.244, $z = -7.185$, $p < 0.001$). Algo similar ocurrió en los casos de LD, donde hubo puntajes mucho más bajos para las oraciones con CUC que con CPC ($= -2.17$, ES = 0.242, $z = -8.951$, $p < 0.001$) o de CT ($= -1.43$, ES = 0.242, $z = -5.932$, $p < 0.001$). En cambio, no hubo diferencias significativas entre los puntajes asignados a CPC para ambas lecturas ($= -0.29$, ES = 0.236, $z = -1.218$, $p = 0.828$). Al comparar las oraciones CTxLD con aquellas predichas como aceptables (CPCxLC, CTxLC, y CPCxLD), encontramos que solo presentan diferencias significativas con las oraciones CPCxLD, teniendo CTxLD una aceptabilidad levemente menor ($= -0.731$, ES = 0.238, $z = -3.074$, $p = 0.026$). Las condiciones predichas como aceptables no mostraron diferencias significativas entre sí.

Estos contrastes sugieren que, de las seis condiciones estudiadas, las tres que se predijeron como aceptables fueron confirmadas, mientras que de las que se predijeron como inaceptables hubo dos (CUCxLC y CUCxLD) que fueron coherentes con las predicciones teóricas. Sin embargo, CTxLD, si bien mostró diferencias significativas con una de las condiciones aceptables, no lo hizo con las otras dos, y en cambio sí tuvo diferencias significativas con los otros dos casos inaceptables. Esto indica que, en primer lugar, el uso de operacionalizaciones no binarias de la aceptabilidad permite identificar diferencias significativas entre grupos que trascienden la distinción aceptable/inaceptable y, en segundo lugar, muestra que la predicción teórica no es consistente con las observaciones del experimento.

En este primer estudio piloto hay limitaciones metodológicas que nos llevaron a realizar un segundo estudio, a saber: (i) consideramos la posibilidad de que una escala de cinco puntos no ofreciera suficiente resolución como para distinguir apropiadamente diferencias entre condiciones; (ii) no se controlaron de manera rigurosa los adjuntos de los ítems; y (iii) al trabajar únicamente con una lista, cada verbo solo fue observado en una condición. Este último punto implica que no es posible determinar si hay algún efecto producido por el verbo y no por la condición. A partir de estas observaciones, diseñamos el segundo estudio, que presentamos a continuación.

4. Segundo experimento

Con el objetivo de confirmar las diferencias observadas en el primer experimento se realizó un segundo juicio de aceptabilidad. En este caso se hizo una revisión de los estímulos con el objetivo de volverlos más consistentes (con relación a los adjuntos) y se buscó una mayor resolución a través del uso de una escala de Likert de 0 a 10.

4.1 Diseño experimental

4.1.1 Participantes

Sobre un total de 292 personas que respondieron en forma voluntaria un formulario de Google, descartamos a aquellos sujetos que reportaron no ser de nacionalidad argentina, residir fuera de la provincia de Buenos Aires, tener nivel avanzado en alguna lengua diferente del español, o tener una primera lengua distinta del español. Además, solicitamos información sobre formación académica, en base a las cuales filtramos a los sujetos que poseyeran estudios de nivel terciario o universitario en carreras con instrucción explícita sobre gramática, como Letras, Traducción o Fonoaudiología. La muestra resultante consistió en 88 personas de entre 17 y 81 años (Mediana 31, IQR 25).

4.1.2 Materiales

Para llevar a cabo este experimento se amplió la escala de Likert a 11 puntos para obtener una mayor amplitud, que permita identificar diferencias entre condiciones que no se vieron en el experimento piloto. Se confeccionaron 6 listas de 58 estímulos cada una. A cada sujeto se le presentó una única lista a través de un formulario de Google. Todas las listas tenían el mismo formato: los primeros 4 ítems eran de práctica, y de los 54 restantes eran 18 oraciones *target* y 36 *fillers*⁵. Se eligió trabajar con 6 listas para que los verbos pudieran ser observados en todas las condiciones, pero que, a su vez, cada sujeto vea cada verbo solo una vez. Los *fillers* fueron los mismos para las 6 listas.

Los 18 estímulos críticos de cada lista fueron elaborados según su lectura y la concordancia del clítico, cruzando las variables de manera idéntica al experimento anterior. Sumando los estímulos de las 6 listas, la cantidad total de ítems para cada condición fue de 18:

(12)

- a. CTxLC: Nos abandonó a mí y al músico juntos el lunes.
- b. CPCxLC: Me abandonó a mí y al músico juntos el lunes.
- c. CUCxLC: Lo abandonó a mí y al músico juntos el lunes.
- d. CTxLD: Nos abandonó a mí el viernes y al músico el lunes.
- e. CPCxLD: Me abandonó a mí el viernes y al músico el lunes.
- f. CUCxLD: Lo abandonó a mí el viernes y al músico el lunes.

En todos los casos las oraciones contaban con un primer coordinado pronominal singular y un segundo coordinado nominal singular. Para no dar lugar a ambigüedades la lectura fue forzada añadiendo un adjunto que especificara

la existencia de un evento o de dos. En el caso de la lectura distributiva se añadió un adjunto temporal a cada coordinado, mientras que en el caso de la lectura colectiva se añadieron el adjunto *juntos* para inducir la lectura colectiva y luego un adjunto temporal para asemejar el estímulo a la otra condición. Como se puede observar, en este experimento para todas las condiciones se utilizaron adjuntos temporales; esto se debe a que se buscaba controlar algún posible efecto por el tipo de adjunto, y por la cantidad de adjuntos, ya que de este modo las oraciones de LC también cuentan con dos adjuntos. Al igual que en el primer experimento, en todos los casos el segundo coordinado fue presentado en masculino singular, mientras que todos los primeros coordinados de tercera persona eran femeninos.

La utilización de 6 listas en este experimento se hizo para completar el diseño de cuadrado latino. Por cada verbo de los estímulos *target* se generaba un *token set* con ese verbo en cada una de las seis condiciones a evaluar. Como se mencionó anteriormente, en el experimento piloto solo se incluyó una oración de cada *token set* en una única lista, por lo que cada verbo era presentado solo con una de las seis condiciones. En esta ocasión se confeccionaron seis listas para que todos los verbos fueran observados en todas las condiciones, pero solo aparecieran una vez en cada lista. Las listas se asignaron a los sujetos aleatoriamente. También se controló que cada una de las condiciones apareciera la misma cantidad de veces con el primer coordinado en primera, segunda y tercera persona.

Finalmente, los *fillers* estaban compuestos por 18 casos inaceptables por violar la estructura argumental (*La llegó a la portera el jueves*) y 18 aceptables (*Los convenció a los cirujanos el lunes*). Todos los *fillers* de ambos tipos comenzaban con un clítico y terminaban con un adjunto temporal para asemejarlos a los estímulos críticos. Los ítems de prueba fueron los mismos del experimento anterior.

4.1.3 Procedimiento

Al comienzo del formulario se pidieron los datos personales de cada sujeto: edad, género, nivel de estudios, carrera, profesión, residencia en sus primeros 10 años, idioma en el que le hablaron los padres y finalmente nivel de otro/s idioma/s. La recolección de datos se realizó de forma totalmente anónima. Se les pidió a los sujetos que evaluaran las oraciones según qué tan natural les sonaban, siendo 0 “para nada natural” y 10 “totalmente natural”. Al inicio se incluyeron los ítems de práctica y la siguiente consigna:

En este cuestionario vas a encontrar una serie de oraciones. Tu tarea es calificarlas con un número entre 0 y 10 según qué tan naturales te suenan. No contestes según un criterio de corrección como el de la escuela o el de la RAE. Un buen criterio es imaginar si podrías escuchar esta oración en un contexto cotidiano. No pienses mucho tus respuestas, la primera impresión es la correcta.

A continuación se incluyeron los estímulos ordenados de forma pseudoaleatoria, de modo de que no aparecieran dos ítems experimentales consecutivos. En la última página se les agradeció la participación y se les dio la opción de dejar un correo electrónico para participar de futuras investigaciones. La participación fue totalmente anónima.

4.2 Análisis y resultados

Las respuestas se resumen en el gráfico 2. Podemos ver que en los casos de lectura colectiva, tanto para la condición CPC como para CT, la mediana es 8, mientras que en el caso de CUC la mediana es 0, y todos los demás puntajes representan menos de la mitad de las respuestas en esta condición. En el caso de la lectura distributiva, vemos que CPC tiene mediana 10, la de CT es 8, y la de CUC, 3.

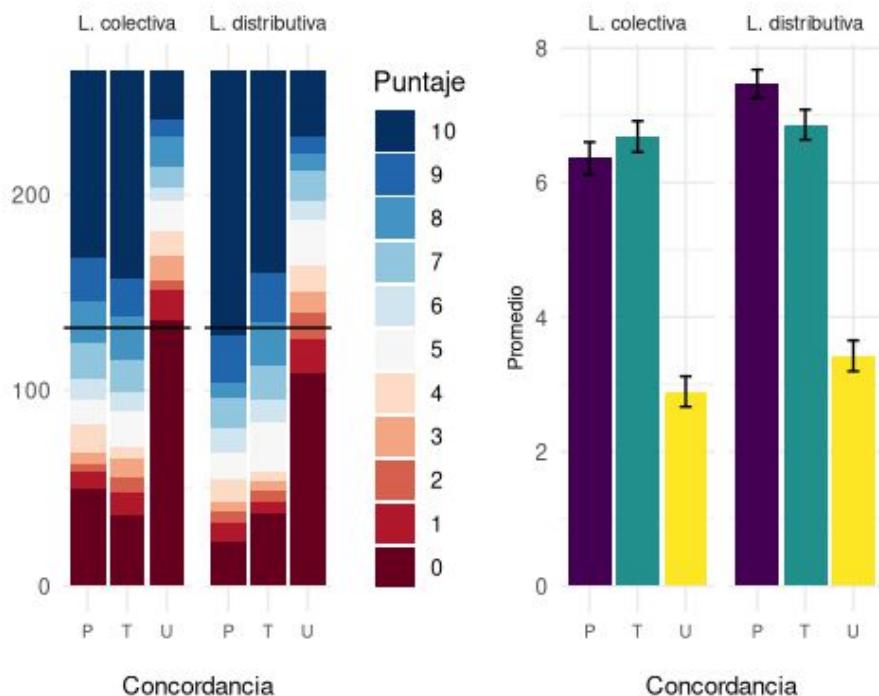


Gráfico 2
Puntajes obtenidos en el segundo experimento

En el gráfico de la izquierda, el tamaño de la sección de cada barra indica la frecuencia absoluta de los puntajes, el color corresponde al puntaje recibido, y la línea negra que atraviesa las barras indica la mediana. El gráfico de barras de la derecha presenta la media de cada condición. 10 representa el valor más alto de aceptabilidad “totalmente natural”, mientras 0 es el valor más bajo “para nada natural”.

Gráfico 2. Puntajes obtenidos en el segundo experimento

En el gráfico, a la derecha, vemos que las condiciones de CPC y CT, tanto en lectura colectiva como en lectura distributiva, poseen todas promedios elevados (CPCxLC: $M = 6.36$, $ES = 0.239$; CTxLC: $M = 6.68$, $ES = 2.31$; CPCxLD: $M = 7.47$, $ES = 0.212$; CTxLD: $M = 6.86$, $ES = 0.223$), mientras que los casos de CUC en ambos tipos de lectura tienen puntajes muy bajos (CUCxLC: $M = 2.89$, $ES = 0.226$; CUCxLD: $M = 3.42$; $ES = 0.229$).

Parametrizamos un modelo lineal con efectos mixtos tomando lectura, concordancia y sus interacciones como efectos fijos y como efectos aleatorios a los sujetos (con interceptos y pendientes aleatorias por cada uno de los efectos fijos) y la combinación verbo-sustantivo (solamente interceptos). Tanto la lectura como la concordancia mostraron efectos significativos (Lectura: $F = 14.4366$, $p < 0.001$; Concordancia: $F = 59.7635$, $p < 0.001$). Asimismo, la interacción entre estas dos variables también mostró un efecto significativo ($F = 3.9646$, $p = 0.0225$).

Cuadro 2
Efectos principales e interacciones significativas del experimento 2

Efectos principales	
Lectura	Lectura $F = 14.4366$, $p < 0.001$
Concordancia	$F = 59.7635$, $p < 0.001$
Interacciones	
Lectura-concordancia	$F = 3.9646$, $p = 0.0225$

Cuadro 2. Efectos principales e interacciones significativas del experimento 2

Un análisis de comparaciones múltiples con corrección de Holm reveló que las condiciones de CUC mostraron diferencias significativas con todas las demás condiciones, pero no entre sí: para los casos de lectura colectiva, los puntajes de las oraciones con CUC fueron mucho más bajos que los de CPC ($= -3.46$, ES = 0.340, $z = -8.890$, $p < 0.001$), o que tuvieran CT ($= -3.77$, ES = 0.414, $z = -9.115$, $p < 0.001$). Algo similar ocurrió en los casos de lectura distributiva, donde hubo puntajes mucho más bajos para CUC que con CPC ($= -4.04$, ES = 0.394, $z = -10.258$, $p < 0.001$) o de CT ($= -3.43$, ES = 0.427, $z = -8.040$, $p < 0.001$). En cambio, para ambas condiciones de lectura, no hubo diferencias significativas entre los puntajes asignados a ambos casos de CUC ($= 0.51$, ES = 0.252, $z = 2.039$, $p = 0.297$). El único par que presenta diferencias significativas corresponde a las diferencias entre CPCxLD y CPCxLC ($= 1.09$, ES = 0.242, $z = 4.525$, $p < 0.001$).

Estos contrastes sugieren que, de las seis condiciones previstas, las tres que se predijeron como aceptables fueron confirmadas, mientras que de las que se predijeron como inaceptables hubo dos (CUC en ambas condiciones de lectura) que fueron coherentes con las predicciones teóricas. Sin embargo, hubo una (CTxLD) que no mostró ninguna diferencia significativa con los casos aceptables.

5. Discusión

En cuanto a la teoría, los resultados obtenidos sugieren que puede ser necesario rever el marco teórico, al menos en lo concerniente a las predicciones sobre los casos en los que un clítico concuerda con ambos coordinados en oraciones que refieren a dos eventos disjuntos (CTxLD). Asimismo, cabe considerar la posibilidad de que los altos valores de aceptabilidad reportados para esta condición se deban a que la interpretación no es de lectura distributiva sino una lectura de lista, donde los coordinados se interpretan como parte de una enumeración. En este sentido, si bien la estructura que habilitaría la CT con LD no ofrece una derivación exitosa, podrían existir otras estructuras posibles que tengan una misma representación ortográfica. Esto permitiría a los participantes puntuar el ítem experimental recurriendo a una estructura sintáctica que no es la esperada. Para comprobar esto, resulta necesario proponer un nuevo diseño experimental que apunte a estudiar las diferencias entre distintas estructuras sintácticas posibles.

En términos metodológicos, las diferencias significativas entre condiciones que responden a una aceptabilidad graduada sugieren que un análisis cuantitativo es sensible a diferencias de grado que no son evidentes en la ponderación cualitativa de juicios informales. Además, el hecho de que algunas diferencias significativas se deban a interacciones entre variables implica que los métodos cuantitativos facilitan la comparación de condiciones más allá de los pares mínimos a los que suelen estar ligados los juicios informales.

Puntualmente en el caso del análisis estadístico, el uso de modelos de efectos mixtos posibilita el análisis unificado de las respuestas de múltiples sujetos, tomando en cuenta que cada uno posee una gramática propia y que las variables bajo estudio pueden afectarlos de diferente manera. Lo mismo sucede para los ítems experimentales: cada oración empleada puede interactuar diferencialmente con las variables de interés. Es posible, sin embargo, que el modelo utilizado no sea el más apropiado para los datos disponibles: el procedimiento empleado supone que la variable de respuesta es numérica, continua, y que la distancia entre valores consecutivos es idéntica para todos los casos, pero la escala de Likert puede corresponderse con una variable ordinal en la que las distancias entre un puntaje de 4 y un puntaje de 5 por un lado, y un puntaje de 5 y un puntaje de 6, por el otro, no sean idénticas. Esta situación obligaría a la adopción de técnicas de modelado diferentes.

Sobre la recolección de datos confirmamos lo sostenido por la bibliografía a favor de los juicios formales. Conseguir un número suficiente de sujetos es relativamente sencillo y una muestra abre la posibilidad de controlar múltiples variables, incluido el propio conocimiento de los sujetos sobre el objeto de estudio. También nos parece importante tener en cuenta que el diseño de un procedimiento experimental fuerza a tener en cuenta múltiples variables que pueden no surgir en un juicio informal, como ocurrió en nuestro caso con los adjuntos temporales o la ausencia de concordancia para sujetos experimentales mujeres.

Consideramos que estos resultados aportan argumentos a favor del uso de métodos estadísticos para el análisis de juicios. Esperamos que el uso de procedimientos formales para la recolección y análisis de datos lingüísticos contribuya al desarrollo de una ciencia del lenguaje más rigurosa.

Agradecimientos

Queremos agradecer especialmente al Grupo de Estadística para el Estudio del Lenguaje (GESEL) por su apoyo durante todo el proceso de investigación y publicación de este trabajo.

Referencias

- Anagnostopoulou, E. (2017). Clitic doubling. En M. Everaert y Henk van Riemsdijk (Eds.), *The Blackwell companion to Syntax* (2 ed., pp. 1-56). Blackwell Publishing.
- Barbeito, V. (2017). La ausencia de concordancia entre el pronombre dativo y su referente nominal en esquemas verbales ditransitivos. *Pragmalingüística*, 25, 50-61.

- Barbeito, V. (2018). Uso no concordante del clítico dativo le. Ponencia presentada en *Coloquio de investigadores en Lingüística Cognitiva*. San Martín, UNSAM, julio de 2018.
- Barbeito, V. (2019). Estrategias léxicas en la manifestación de la concordancia: el uso de los cuantificadores. Ponencia presentada en *X Simposio de la Asociación Argentina de Lingüística Cognitiva*. San Juan, 4 y 5 de septiembre de 2019.
- Barbeito, V., Murata Missagh, J. y Peri, S. (2018). La ausencia de concordancia entre el clítico dativo y su referente nominal. *Textos en Proceso*, 3 (2), 128-143.
- Barbeito, V. y Peri, S. (2019). La conceptualización de la cantidad como estrategia creativa de los hablantes en el uso de la lengua: el problema de la concordancia. Ponencia presentada en *III Congreso de la ALFAL. Simposio “Configuraciones dinámicas, identidad y cognición”*. La Plata, abril de 2019.
- Bates, D., Maechler, M., Bolker, B. y Walker, S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Bresnan, J., Cueni, A., Nikitina, T. y Baayen, R. H. (2007). Predicting the dative alternation. En G. Bouma, I. Kraemer and J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*, pp. 69-94. Amsterdam: Royal Netherlands Academy of Science.
- Camacho, J. (2003). The structure of coordination. En J. Camacho, *The Structure of Coordination*, 33-90. Springer.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax* Cambridge. Multilingual Matters: MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MA. MIT Press.
- Chomsky, N. (2000). Minimalist inquiries: The framework (MITOPL 15). En H. Lasnik, R. Martin, D. Michaels, S. J. Keyser y J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89-155). Cambridge, MA: MIT Press.
- Chomsky, N. (2001). Derivation by phase. En M. Kenstowicz (Ed.) *Ken Hale: A life in language* (pp. 1-52). Cambridge, MA: MIT Press.
- Chomsky, N. (2008). On phases. En R. Freidin, C. P. Otero y M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory* (pp. 133-166). Cambridge, MA: MIT Press.
- Correa, M. (2004). *Sobre la ‘no opcionalidad’ del doblado de clíticos en español* [Tesis de doctorado, Universidad de Arizona].
- Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. SAGE Publications.
- Di Tullio, A. y Zdrojewski, P. (2006). Notas sobre el doblado de clíticos en el español rioplatense: asimetrías entre objetos humanos y no humanos. *Filología*, XXXVIII, 13-44.
- Edelman, S. y Christiansen, M. H. (2003). How seriously should we take minimalist syntax? A comment on Lasnik. *Trends in Cognitive Science*, 7(2), 60-61.
- Estigarribia, B. (2013). El modelo de las fuerzas discursivas y el doblado de clíticos rioplatense. *Signo y Seña*, 23, 119-142.
- Featherston, S. (2007). Data in generative grammar: the stick and the carrot. *Theoretical Linguistics*, 33 (3), 269-318.
- Gallego, A. J. (2011). Sobre la elipsis. *Cuadernos de la lengua española* 111. Arco libros.
- Gibson, E. y Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2), 88-124.

- Gibson, E., Piantadosi, S. T. y Fedorenko, E. (2013). Quantitative methods in syntax/semantics research: A response to Sprouse and Almeida (2013). *Language and Cognitive Processes*, 28(3), 229-240.
- Gualchi, S. (2019). La concordancia parcial en los clíticos acusativos del español rioplatense. *Exlibris*, 8, 152-163.
- Haider, H. (2007). As a matter of facts – comments on Featherston’s sticks and carrots. *Theoretical Linguistics*, 33, 381-394.
- Kuznetsova, A., Brockhoff, P. B. y Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26. doi:10.18637/jss.v082.i13.
- Linzen, T. y Oseki, Y. (2018). The reliability of acceptability judgments across languages. *Glossa: a journal of general linguistics*, 3(1), 1-25.
- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review*, 22(2-4), 429-445.
- Mazzuchino, M. G. (2013). El doblado de acusativo en español de Argentina: definitud, especificidad, presuposicionalidad y otras nociones conexas". *Lengua y habla*, 17, 118-152.
- Mojedano Batel, A. (2014). Variación de *le/les* en diferentes zonas hispanoparlantes: México, Colombia y España". *IULC Working Papers*, 14(2), 80-94.
- Muñoz Pérez, C. (2014). Una nota acerca del uso de juicios en teoría gramatical. *Signo y Seña*, 26, 107-120.
- R Core Team (2020). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rizzi, L. (2001). Relativized minimality effects. En M. Baltin y C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 89-110). Blackwell Publishing.
- Sánchez, L. y Zdrojewski, P. (2013). Restricciones semánticas y pragmáticas al doblado de clíticos en el español de Buenos Aires y de Lima. *Lingüística*, 29(2), 271-320.
- Seco, M. A. (2013). Neutralización de rasgos en clíticos de acusativo en el español hablado en Catamarca. En A. Martínez y A. Speranza (Eds.), *Rumbos sociolíngüísticos* (pp.121-129). Editorial FFyL-UNCuyo y SAL.
- Schütze, C. T. y Sprouse, J. (2013). Judgment data. En Podesva, R. J. y Sharma, D. (Ed.), *Research methods in linguistics* (pp. 7-50). Cambridge: Cambridge University Press.
- Sprouse, J. (2007). Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics*, 1, 123-134.
- Sprouse, J. (2015). Three open questions in experimental syntax. *Linguistics Vanguard*, 1(1), 89-100.
- Sprouse, J. y Almeida, D. (2012). Assessing the reliability of textbook data in syntax: Adger’s Core Syntax". *Journal of Linguistics*, 48(3), 609-652.
- Sprouse, J. y Almeida, D. (2013). The empirical status of data in syntax: A reply to Gibson and Fedorenko",. *Language and Cognitive Processes*, 28(3), 222-228.
- Sprouse, J. y Schütze, C. T. (2017). Grammar and the use of data. En B. Aarts, J. Bowie y G. Popova (Eds.), *The Oxford Handbook of English Grammar*. Oxford University Press. UCLA. <https://escholarship.org/uc/item/0n100842>
- Sprouse, J., Schütze, C. T. y Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua*, 134, 219-248.

Zhang, N. N. (2010). Cambridge Studies In Linguistics: Coordination in syntax. Vol. 123. Cambridge University Press.

Anexo

1. Estímulos del experimento piloto

Type	Condición	Text
Prueba experimental	CT	Algo murió en mi y el agente que lo mató es en el que pase. Los asesinó a él y a mí porque lo querían para la cena. Y el que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CPC	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CPL	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CLC	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CCL	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
Otros		Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.

2. Estímulos del segundo experimento

Type	Condición	Text
Prueba experimental	CT	Algo murió en mi y el agente que lo mató es en el que pase. Los asesinó a él y a mí porque lo querían para la cena. Y el que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CPC	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CPL	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CLC	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
	CCL	Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.
Otros		Algo murió en mi y el agente que lo mató es en el que pase. Me mataron a mí y el que lo mató lo querían para la cena.

Notas

- 1 Esta investigación ha sido financiada por el proyecto FiloCyT “Métodos cuantitativos en el estudio del lenguaje” (FC19-064), dirigido por la Dra. Mercedes Güemes, con sede en el Instituto de Filología y Literaturas Hispánicas “Dr. Amado Alonso”, Universidad de Buenos Aires.
- 2 Si bien Gualchi (2019) no se pronuncia respecto a la distinción de lecturas para las oraciones con CT, las derivaciones que propone para estas construcciones coinciden con las derivaciones de CPCxLC (excepto en la operación de Agree que desencadena el ensamblaje de v). En cambio, el análisis propuesto para CPCxLD no puede ser trasladado a las oraciones con CT, según los postulados teóricos desarrollados en la sección anterior.
*[Nos_i conoció [a mí ayer y lo conoció al doctor hoy]_j]
Como se observa en el ejemplo, la secuencia “a mí ayer y al doctor hoy” no puede constituir una unidad sintáctica bajo la hipótesis de la elipsis.
- 3 Nuestro marco teórico no realiza ninguna predicción sobre diferencias en los resultados a causa del grupo etario, pero en instancias futuras sería interesante investigar si hay variables demográficas que supongan diferencias en los juicios.
- 4 Ver Anexo “Experimento Piloto”.
- 5 Ver Anexo “Segundo experimento” para ver la lista completa de estímulos.