

## UNA NOTA ACERCA DEL USO DE JUICIOS EN TEORÍA GRAMATICAL

Carlos Muñoz Pérez

**RESUMEN.** Los juicios de aceptabilidad son la principal fuente de datos del gramático. Sin embargo, su utilización ha sido criticada de diversas maneras en la bibliografía. Esta nota presenta dos de los debates en torno al uso de juicios de aceptabilidad: (i) el que se centra en la “artificialidad” de los datos basados en juicios, crítica planteada por algunos representantes de la lingüística de *corpus* y (ii) el que se centra en torno a la “informalidad” de la tarea psicológica del juicio de aceptabilidad, crítica planteada por diversos psicolingüistas y psicólogos experimentales.

*Palabras clave:* juicios de aceptabilidad, metodología, gramática, *corpus*, experimentos.

**ABSTRACT.** Acceptability judgments are the main source of data for the grammarian. However, its use has been criticized in several ways in the literature. This note presents two of the debates regarding acceptability judgments: (i) the one that centers on the “artificiality” of data based on judgments, a critic raised by some proponents of *corpus* linguistics, and (ii) the one that centers on the “informality” of the psychological task of acceptability, a critic raised by several psycholinguists and experimental psychologists.

*Keywords:* acceptability judgments, methodology, grammar, *corpus*, experiments.

**RESUMO.** Os julgamentos de aceitabilidade são a principal fonte de dados do gramático. Entretanto, seu uso tem sido criticado de diferentes modos na literatura. Este trabalho apresenta dois dos debates em torno do uso dos julgamentos de aceitabilidade: (i) o que se centra na “artificialidade” dos dados baseados em julgamentos, crítica levantada por alguns representantes da lingüística de *corpus*, e (ii) o que se centra em torno a “informalidade” da tarefa psicológica do julgamento de aceitabilidade, crítica feita por diversos psicolinguistas e psicólogos experimentais.

*Palavras-chave:* julgamentos de aceitabilidade, metodologia, gramática, *corpus*, experimentos.

**1. INTRODUCCIÓN.** Una de las formas más utilizadas para obtener datos lingüísticos es a través de la “intuición” de los hablantes. Si bien esta clase de metodología es atribuida principalmente al paradigma generativo (dado que dicho marco teórico procuró revalorizar este tipo de evidencia al enfrentar la postura empirista del estructuralismo norteamericano), debe aclararse que prácticamente no hay enfoque que aborde el estudio



*Signo y Señal*, número 26, diciembre de 2014, pp. 107-120

Facultad de Filosofía y Letras (UBA)

<http://revistas.filo.uba.ar/index.php/sys/index>

ISSN 2314-2189

de las lenguas naturales sin recurrir a datos obtenidos a partir del conocimiento implícito de un hablante.

Si bien son muchas las clases de datos que se dice pueden obtenerse a partir del conocimiento lingüístico de los hablantes<sup>1</sup>, voy a restringir mis afirmaciones a aquel tipo de dato que reporta la buena o mala formación de una expresión lingüística a partir de la “intuición” de uno o más informantes, el llamado *juicio de aceptabilidad*. Al utilizar este término, y no el más (erróneamente) estandarizado de *juicio de gramaticalidad*, sigo a Chomsky (1965), quien considera que los hablantes sólo pueden acceder a la *aceptabilidad* de una expresión lingüística: un informante puede notar en mayor o menor medida si una expresión es admisible o no en su lengua, pero no puede determinar *a priori* si esto se debe a factores gramaticales (i.e., morfosintácticos), fonológicos, semánticos, pragmáticos, de procesamiento, etcétera<sup>2</sup>. En otras palabras, la noción de *aceptabilidad* tiene un carácter empírico y mensurable, mientras que la *gramaticalidad* de una expresión sólo puede ser deducida a partir del estudio analítico de los patrones aceptables.

El siguiente dato, tomado de un artículo de tipología (Keenan y Dryer 2007, 331), ilustra cómo se utilizan los juicios de aceptabilidad en gramática:

(1) *Letón*

Es tieku	macits	(*no mates).
Yo soy	enseñado	por madre
<i>Yo soy instruido</i> (lectura participial).		

El dato corresponde a las llamadas *pasivas sin agente* del letón. El nombre viene dado, justamente, porque se trata de construcciones de carácter inacusativo aspectualmente perfectivas formadas a partir de un verbo típicamente transitivo que no admiten la presencia de un complemento agente. ¿Cómo se sabe que este tipo de pasiva no admite complemento agente? Porque los hablantes del letón consideran inaceptables estas for-

1 No haré referencia, por ejemplo, al llamado *paradigma introspeccionista* (Wundt 1896), que asume que es posible acceder conscientemente a los procesos y representaciones mentales responsables, por ejemplo, de las formas lingüísticas. Para una defensa de esta metodología particular, véase Talmy (2000).

2 Otro modo de plantear la cuestión es en términos de implicación: si una oración es agramatical, debe entonces ser inaceptable; en cambio, si una oración es inaceptable, esto no implica necesariamente que se deba a factores gramaticales (i.e., que sea agramatical). En definitiva, las oraciones agramaticales son un subconjunto de las oraciones inaceptables.

mas si se les adjunta un agente. El juicio de los hablantes del letón, entonces, permite caracterizar ciertos aspectos de la gramática de su lengua.

Casos como el de (1) u otros más complejos son los que vienen a la mente cuando se piensa en usar datos basados en juicios como base empírica para la teoría gramatical. Sin embargo, no hace falta consultar una obra especializada para encontrar patrones de aceptabilidad como fundamento para diversas cuestiones. Considérese algo tan elemental como reconocer las funciones gramaticales básicas. ¿Cuáles son las pruebas, por ejemplo, para reconocer un objeto directo en español? Como dice la RAE (2010, 656): “en casi todos los análisis tradicionales se destaca la posibilidad de que el complemento directo pueda sustituirse por un pronombre átono acusativo de tercera persona”. Siguiendo este razonamiento, el encorchetamiento de (2a) es correcto porque la oración de (2b) es aceptable:

- (2) a. No pienso leer [<sub>OD</sub> este libro].  
 b. No pienso leerlo.

De forma análoga, el análisis de (3a) debe ser considerado incorrecto porque la oración de (3b) no es aceptable para los hablantes de español:

- (3) a. [<sub>OD</sub> Este libro] fue leído por los estudiantes.  
 b. \*Lo fue leído por los estudiantes.

El uso de juicios no solo se encuentra estandarizado para la definición de las funciones sintácticas. Son innumerables las distinciones taxonómicas de carácter empírico que se basan en la no aceptación de algún tipo de expresión lingüística. Por ejemplo, puede decirse que el verbo *ser* es un estado porque \**Yo soy bueno rápidamente* es inaceptable (con el significado relevante), o puede decirse que *arena* es un nombre de masa porque la oración \**Tengo dos arenas es inaceptable* (con el significado relevante).

En resumen, el uso de juicios de aceptabilidad es una pieza fundamental del razonamiento gramatical y constituye la base empírica más importante para las teorías de la gramática. Es por esta razón que los debates en torno a la “calidad” de los datos basados en patrones de aceptabilidad cobran vital importancia. El objetivo de esta nota es presentar críticamente algunos de estos debates. Por tanto, en la sección 2 haré referencia a

la crítica más extendida a los datos lingüísticos obtenidos a partir de juicios de aceptabilidad: su naturaleza “artificial” frente a datos extraídos de *corpus* (i.e. datos “reales”). Luego, en la sección 3, presentaré el debate de mayor novedad con respecto a la base empírica de la teoría gramatical: su carácter informal frente a, por ejemplo, los métodos de obtención de datos en psicolingüística. La sección 4, por último, contiene las conclusiones del trabajo.

**2. “DATOS REALES” VS. “DATOS INVENTADOS”.** La crítica más conocida al uso de juicios de aceptabilidad es introducida por quienes distinguen entre “datos reales”, emisiones lingüísticas que pueden ser documentadas a partir de la recolección de *corpus*, y “datos inventados”, patrones lingüísticos que no son observados directamente en la actuación lingüística y son diseñados por el investigador para ser contrastados con el conocimiento intuitivo de un hablante (e.g., Meyer 2004; Sampson 2001, 2007; Sinclair 2004). Algunos de los autores que sostienen esta distinción consideran que los estudios basados en juicios de aceptabilidad se desvían de la práctica científica normal:

*Es importante basar el estudio del lenguaje en datos reales —instancias verdaderas de habla o escritura— más que en datos artificiales o “inventados” (Meyer 2004, xiii; mi traducción).*

*Los lingüistas que utilizan las intuiciones de los hablantes y no evidencia observable de modo interpersonal como base empírica para la descripción lingüística están eligiendo dar la espalda a la ciencia (Sampson 2007, 7; mi traducción).*

No parece adecuado basar en tal diferencia metodológica una distinción entre “buena” y “mala” lingüística. En filosofía de la ciencia se reconocen dos maneras de acceder a los datos de un determinado dominio empírico: la *observación* y la *experimentación* (e.g. Herschel 1966). En este sentido, la llamada *lingüística de corpus* podría ser considerada una instanciación pura del método observacional. Los juicios, en cambio, tienen un carácter experimental, ya que buscan corroborar una relación causal establecida teóricamente entre dos variables. Para ejemplificar esto, tomemos una oración como (4):

(4) Había un chico en la plaza.

Con el fin de determinar cuáles son las propiedades de esta clase de construcciones, un gramático podría elaborar diversas *hipótesis* que expliciten los tipos de constituyentes que pueden formar parte de ellas. Consideremos el caso puntual de (5):

- (5) Las construcciones existenciales con *haber* requieren obligatoriamente un argumento nominal de tipo indefinido.

La hipótesis de (5) predice que no es posible formar una construcción existencial con *haber* con un argumento definido. ¿Cómo podemos saber si esta hipótesis es adecuada? El primer paso para poner a prueba (5) es construir un estímulo idéntico a la oración (4) en el que únicamente varíe la definitud del argumento verbal (i.e. una oración en la que se reemplace *un* por *el*). En este caso se dice que la definitud es una *variable independiente* cuya manipulación debería producir un cambio en la aceptabilidad de la expresión, la *variable dependiente*:

- (6) Había *el* chico en la plaza.

El segundo paso consiste en presentar este estímulo a hablantes de español a modo de *experimento psicológico*<sup>3</sup>: se espera que los hablantes informen si perciben o no dicha forma como una expresión de su propia lengua. Si los hablantes rechazan (6) (y admiten (4)), se corrobora la existencia de una relación (causal) entre la definitud del argumento de una construcción existencial con *haber* y su aceptabilidad, por lo que la hipótesis (5) puede seguir siendo sostenida.

Un posible corolario de considerar que *corpus* y juicios son las “versiones lingüísticas” de *observación* y *experimentación* es que ambas deberían ser consideradas formas de obtener datos complementarias e igualmente válidas. Esto es especialmente cierto si se consideran las restricciones y limitaciones inherentes a cada uno de estos métodos. Así, del mismo modo en que los juicios de aceptabilidad son totalmente inútiles para obtener datos referidos a las frecuencias de uso o para estudiar una lengua muerta, los *corpus* carecen de utilidad al estudiar, por ejemplo, un fenómeno lingüístico tan característico como la ambigüedad, ya que sólo

3 Suele asumirse que la tarea de juicio de aceptabilidad es análoga a los experimentos propios de la *psicofísica*, área de la psicología que estudia la relación entre la magnitud de los estímulos sensoriales y la intensidad con que son percibidos (e.g. Bard, Robertson y Sorace 1996). Para un argumento contrario a esta asimilación, véase Schütze (1996).

a través de la “intuición” de un hablante se puede saber que las expresiones de (7) tienen más de un significado<sup>4</sup>:

- (7) a. La matanza de los cazadores.
- b. Juan vio a la niña con el telescopio.
- c. Todos los estudiantes tienen dos profesores.

Otra limitación de los *corpus* que debe ser suplida a partir de la utilización de juicios es que no aportan *evidencia negativa*, datos con respecto al tipo de expresiones que no existen en una lengua (e.g. Chomsky 1957, McEnery y Wilson 2001). Esto es de gran importancia si se aspira a describir explícita y satisfactoriamente el funcionamiento de una gramática particular. Por ejemplo, supongamos que, en vez de la hipótesis (5), se intenta defender (8):

- (8) Las construcciones existenciales con *haber* requieren obligatoriamente un argumento nominal.

A partir del experimento informal en el que se evaluó la inaceptabilidad de (6) es posible afirmar que (8) es una hipótesis falsa (o, al menos, demasiado amplia), dado que no cualquier argumento nominal puede acompañar a haber en construcciones existenciales. Sin embargo, el tipo de dato necesario para poder realizar esta afirmación (e.g., *\*Había el chico en la plaza*) no puede obtenerse de modo no inductivo de ningún *corpus*<sup>5</sup>, ya que estos consisten en una colección de emisiones que efectivamente han sido utilizadas (i.e., datos positivos).

Para mayor complejidad, lo que en cambio sí puede encontrarse en un *corpus* son ciertos casos de combinaciones de *había + el+ N*: en el CREA (Corpus de Referencia del Español Actual), por ejemplo, existen al menos 148 secuencias de este tipo<sup>6</sup>:

- (9) [...] cuando era joven *había el franquismo*, que veía con malos ojos los dos frentes en los que siempre se ha movido mi obra.

Si sólo contásemos con datos positivos como los de (4) y (9), nos veríamos tentados a tomar (erróneamente) por cierta la hipótesis (8) y a asumir

4 Con respecto a los ejemplos de (7), su evaluación en términos de juicios de aceptabilidad se realiza a partir de consignas del tipo “¿Es aceptable la oración X con un significado Y?”.

5 Para un método inductivo de obtención de datos negativos a partir de *corpus*, véase Stefanowitsch (2006).

6 Agradezco a un revisor anónimo por esta valiosa observación.

que cualquier tipo de nominal puede acompañar a *haber*. Sólo a partir de considerar también evidencia negativa es posible reconstruir las sutiles restricciones que obedecen los argumentos de las construcciones existenciales<sup>7</sup>. En suma, queda demostrado lo necesario de combinar criterios observacionales y experimentales en la descripción lingüística, y por qué los gramáticos deben hacer uso tanto de datos provenientes de *corpus* como de juicios.

Otro de los argumentos clásicamente utilizados para afirmar que los datos obtenidos a partir de juicios no tienen la misma relevancia que los datos de *corpus* apela (muchas veces, intuitivamente) a la noción de *validez ecológica*, un criterio de adecuación según el cual los fenómenos o efectos observados en el transcurso de una investigación deben darse de forma más o menos idéntica en el mundo real<sup>8</sup>. La siguiente cita de Becker (1975) ilustra esta postura, donde se supone que los datos tomados de *corpus* son “superiores” ya que han sido efectivamente utilizados en una situación comunicativa, mientras que los datos basados en juicios son muchas veces “difíciles de usar”:

*El lingüista “moderno” pasa su tiempo poniendo y sacando asteriscos de oraciones poco probables como “John, Bill and Tom killed each other”<sup>9</sup> (por seleccionar una al azar de una publicación reciente), las cuales elaboran con frustración reprimida y son difíciles de usar en una conversación. Estas oraciones no guardan ningún parecido con las oraciones que componen los textos que se supone las explican (Becker 1975, 60).*

Si bien es cierto que muchas de las expresiones sometidas a juicios pueden ser difíciles de utilizar en una conversación, este no puede ser un parámetro para evaluar la validez global de esta metodología. En primer lugar, los criterios de adecuación como la validez ecológica se aplican a in-

7 Véase McNally (2011) para una concisa presentación del problema y de sus alternativas explicativas.

8 La definición original de *validez ecológica* fue propuesta por Brunswick (1956). Sin embargo, la acepción actual del término refiere a lo que Brunswick denominaba *diseño representativo*. La definición ofrecida aquí se basa en la de Brewer (2000, 12).

9 Resulta curiosa la elección del ejemplo utilizado por Becker, ya que se trata de un dato especialmente valioso para quienes trabajan con construcciones recíprocas. Como señala Nedjalkov (2007), si bien las construcciones recíprocas prototípicas de dos participantes (e.g., *Juan y María se golpearon entre sí*) admiten únicamente una lectura del tipo “A golpeó a B y B golpeó a A”, esta interpretación no se extiende a construcciones recíprocas con participantes plurales (e.g., *Juan, María y Pedro se golpearon entre sí*), ya que “es improbable, difícil o directamente imposible que cada uno de los participantes cumpla ambos roles (agente y paciente) con respecto a cada uno de los demás” (Nedjalkov 2007, 9; mi traducción), lo que en definitiva muestra que las formas plurales de estas construcciones no pueden reducirse a una suma de relaciones de reciprocidad.

vestigaciones particulares y no a métodos de experimentación en general. En este sentido, dos experimentos A y B pueden estar basados en protocolos similares e incluso así uno podría tener validez ecológica y el otro no. Del mismo modo, la aplicación de un juicio de aceptabilidad sobre una oración como *Hay un chico en la plaza* puede ser considerada ecológicamente válida (entendiendo intuitivamente que la validez ecológica en lingüística refiere a la posibilidad de usar una expresión en un diálogo), mientras que lo mismo puede no sostenerse para una oración como *Juan, Guillermo y Tomás se mataron entre sí*.

El segundo motivo por el cual la potencial falta de validez ecológica no invalida el uso de juicios se relaciona directamente con el carácter experimental de la tarea. Como ya se observó, un experimento no tiene como objetivo verificar la representatividad de un fenómeno con respecto al mundo cotidiano, sino de poner a prueba una hipotética relación causal entre dos variables. Por tanto, los criterios de validez relevantes para evaluar una tarea experimental deben estar dados a partir de ese objetivo. Fue Campbell (1957) quien introdujo los dos criterios de adecuación más utilizados hoy en día para tal fin: la *validez interna* y la *validez externa*. La *validez interna* refiere al valor de verdad que puede asignársele a la conclusión de una relación de causa-efecto observada en un experimento. Es decir, es un criterio orientado a comprobar si la variación observada en la variable dependiente se da inequívocamente por la manipulación de la variable independiente a partir de la eliminación de cualquier otro factor potencialmente incidente. Típicamente, los estudios gramaticales procuran alcanzar la validez interna a partir de la construcción de lo que los lingüistas estructuralistas denominaron *pares mínimos*, conjuntos de dos o más expresiones idénticas salvo por un rasgo o valor en contraste<sup>10</sup>, y de la evaluación de hipótesis con respecto a dichos pares. Por otra parte, la noción de *validez externa* refiere a qué tan generalizable es la relación observada entre la variable independiente y la variable dependiente en un determinado experimento. Es decir, se trata de un criterio que busca evaluar si es posible establecer la misma relación de causa-efecto a través de

10 El término *par mínimo* tiene su origen y formulación estricta en el campo de la fonología (e.g. Bloomfield 1933, Harris 1951). Lamentablemente (y hasta donde sé), su extensión a otros niveles de análisis lingüístico (e.g. sintaxis, morfología, semántica) no implicó, hasta el momento, una reformulación precisa de dicha definición.



diferentes informantes/participantes, contextos y tareas experimentales<sup>11</sup>. Existe, en definitiva, una importante diferencia entre los tres tipos de criterios presentados: tanto la validez interna como la externa se abocan a la relación entre un enunciado hipotético y un fenómeno empírico (que tiene lugar, en este caso, en un contexto experimental), mientras que la validez ecológica hace énfasis en la aplicabilidad de ese fenómeno empírico al mundo real (no experimental).

En resumen, entender que los juicios de aceptabilidad revisten un carácter experimental permite ubicarlos como una de las herramientas básicas del gramático. Descartar esta metodología parece errado dado que (i) los juicios suplen algunas de las falencias y restricciones de los *corpus* (y viceversa) y (ii) no es sostenible el criterio según el cual los datos de *corpus* son cualitativamente superiores. La conclusión a partir de estas consideraciones parece ser que el gramático no debe descartar apriorísticamente ninguna fuente de datos, sino utilizarlas complementariamente según los objetivos y necesidades de su investigación.

**3. EXPERIMENTOS FORMALES E INFORMALES.** Como se ha visto en la sección anterior, los juicios de aceptabilidad no son simples “intuiciones” acerca de cómo debe funcionar el lenguaje, sino que son verdaderos experimentos psicológicos en los que los participantes reportan cómo perciben un determinado estímulo lingüístico. Ahora bien, si este es el caso, entonces sería esperable que toda tarea de juicio de aceptabilidad cumpla con los estándares metodológicos de la psicología experimental e incorpore, por ejemplo, gran número de estímulos y participantes, uso de valores-p, evaluación de las magnitudes de los efectos, etcétera. Como es sabido, este tipo de pautas no suelen cumplirse. La manera “tradicional” de reportar juicios consiste en la evaluación de una oración o de un par de oraciones por el propio investigador, a veces con la opinión de algunos colegas. La “informalidad” de este método experimental ha llevado en años recientes a que se ponga en duda la fiabilidad de los datos en los que se basa la teoría gramatical, en particular la de orientación generativa:

*Desafortunadamente, los hallazgos de quienes hacen experimentos en lingüística rara vez juegan un papel en el trabajo de los gramáticos generativos. El desarrollo de la teoría parece seguir su propio curso, puesta a prueba solamente a través de las poco*

11 Para una tipología y descripción de las tareas de juicios más comúnmente utilizadas, véase Schütze y Sprouse (2014).

*confiables y a veces maleables intuiciones de los propios teóricos. Consecuentemente, las teorías son de cuestionable relevancia con respecto a los hechos del lenguaje* (Wasow y Arnold 2005, 495).

*Las teorías generativas parecen tener una base empírica muy débil debido a su informal recolección de juicios de gramaticalidad [...] Un conjunto de estándares [...] debería ser establecido. Si esta idea [...] se considerase, los desarrollos lingüísticos podrían una vez más ser relevantes a la empresa psicolingüística* (Ferreira 2005, 365).

Muchos de los autores que sostienen tales críticas (e.g. Edelman y Christiansen 2003, Ferreira 2005, Wasow y Arnold 2005, Gibson y Fedorenko 2010) afirman que el método tradicional de recolección de datos debe ser abandonado y reemplazado por un conjunto de estándares metodológicos estrictos que, a partir del trabajo de Cowart (1997), hoy en día son conocidos como *syntaxis experimental*. La idea es simple: los gramáticos deben basar sus teorías en juicios obtenidos según el modelo experimental de las ciencias cognitivas.

Por supuesto, no hay nada de malo en que los gramáticos realicen experimentos “formales” para corroborar sus hipótesis, pero la idea de que el método tradicional deba ser descartado por completo requiere un análisis más detenido. Por lo general, la bibliografía que aboga por el abandono de la metodología tradicional no ofrece ningún tipo de prueba o evidencia que permita mostrar por qué dicho procedimiento es poco fiable, sino que se limita a mostrar algunos casos en los que se han reportado juicios errados en la bibliografía. Este es el caso, por ejemplo, de Wasow y Arnold (2005), quienes como toda prueba para su argumentación presentan un estudio de caso sobre un juicio erróneo de Chomsky (1955). Esta clase de demostración no es (y no puede ser) razón suficiente para que los gramáticos abandonen el uso de experimentos “informales”. Ello equivaldría a descartar cierto tipo de diagnóstico médico a partir de la existencia de algunos pocos falsos positivos. De hecho, para que se justifique la renuncia a la metodología clásica antes debería demostrarse (i) que usualmente se utilizan datos espurios para dirimir cuestiones teóricas y que, por tanto, los gramáticos no tienen manera de identificarlos, y (ii) que el método tradicional realmente produce datos erróneos de modo sistemático. Con respecto a lo primero, Phillips (2009) observa que el uso de la metodología clásica en *syntaxis* no ha llevado a postular generalizaciones empíricas basadas en datos erróneos ni a realizar elecciones teóricas

importantes a partir de datos controvertidos<sup>12</sup>. Esto resulta evidente si se observa en qué aspectos se centran las discusiones entre modelos teóricos en competencia: los debates giran en torno a cómo explicar los datos o a cómo entenderlos (e.g. “el fenómeno X es sintáctico” vs. “el fenómeno X es discursivo”), pero no existe mayor controversia con respecto a la calidad de los datos ni a la existencia de los fenómenos.

El segundo punto, el que atañe a la fiabilidad general de la metodología tradicional, ha sido objeto en años recientes de una serie de experimentos reveladores. Sprouse y Almeida (2012) y Sprouse, Schütze y Almeida (2012) han intentado definir qué significa que el método clásico sea “poco fiable”. Su idea es realmente muy simple: el método es poco fiable si conlleva un número significativamente más alto de *falsos positivos* y de *falsos negativos*<sup>13</sup> con respecto a los datos que pueden ser obtenidos a partir de los procedimientos de la sintaxis experimental. A partir de esta definición, en un primer experimento (Sprouse y Almeida 2012) se evaluaron los datos basados en juicios presentes en un manual de teoría sintáctica (Adger 2003) y se procuró replicarlos en un contexto experimental formal con 40 participantes. Resultó que los datos fueron coincidentes en un 98 por ciento de los casos, es decir que el número de falsos positivos y falsos negativos del método tradicional no superó el 2 por ciento. En un segundo experimento (Sprouse, Schütze y Almeida 2012), se evaluó una selección aleatoria de datos del período 2001-2010 de la revista *Linguistic Inquiry* con resultados similares: los datos resultaron replicados en un 95 por ciento de los casos. Dado que en ambos experimentos la proporción de datos erróneos obtenidos a partir de la metodología

12 Entre otros factores, esto puede deberse al estricto control que realizan los gramáticos con respecto a los datos: existe un consenso a veces tácito entre pares de diferentes lenguas y variedades acerca de los patrones gramaticales que es necesario explicar. Esta actitud “comunitaria” con respecto a cuáles son los fenómenos de interés y los datos que mejor los representan supliría, en cierto modo, el poder estadístico que otorgan los estudios a larga escala (e.g. lingüistas de diversas variedades hispánicas coincidiendo en que un determinado efecto o contraste es real “equivaldría” a un estudio realizado utilizando gran número de hablantes de dichas variedades). Esta observación, por supuesto, es de orden puramente especulativo.

13 Un *falso positivo* se da cuando un experimento reporta una diferencia entre dos o más condiciones, pero tal diferencia no existe. Este sería el caso, por ejemplo, de un par de oraciones donde se reporte que una de ellas es aceptable y la otra no, siendo que realmente ambas son aceptables. Por otra parte, un *falso negativo* se da cuando un experimento no reporta diferencias entre dos o más condiciones. Esto es lo que ocurre, por ejemplo, con un par de oraciones donde se dice que ambas son aceptables y resulta que una de ellas realmente no lo es.

tradicional fue muy baja (entre el 2 y el 5 por ciento, respectivamente), la idea de que dichos datos resultan poco fiables tuvo que ser descartada.

Una segunda conclusión que puede derivarse de los experimentos recién mencionados es que el método tradicional ofrece datos de muy alta replicabilidad y generalidad: se trata de patrones que pueden volver a ser obtenidos en diferentes experimentos y con diferentes participantes. Esto no sólo revalida la base empírica más importante de la teoría gramatical, sino que le otorga un estatuto privilegiado entre otras subdisciplinas de la psicología cognitiva y de las ciencias de la mente en general. Por ejemplo, en un estudio reciente (Button et al. 2013) se mostró que los estándares estadísticos en neurociencia son tan bajos que una parte importante de los descubrimientos realizados en el campo serían falsos positivos<sup>14</sup>:

*Nuestros resultados indican que el poder estadístico medio en el campo de la neurociencia está probablemente entre alrededor del 8% y el 31% [...] Si el bajo poder estadístico observado en estos estudios es el típico de la literatura neurocientífica, esto tiene profundas implicaciones para el campo. Una consecuencia importante sería que la probabilidad de que cualquier hallazgo nominal refleje un efecto real es baja (Button et al. 2013, 7; mi traducción y mis cursivas).*

En este contexto general, simplemente no parece adecuado abandonar un método de obtención de datos que se ha probado confiable y que, además, resulta económico, rápido y fácil de aplicar.

**4. CONCLUSIÓN.** Se han presentado dos debates de vital importancia para la utilización de juicios de aceptabilidad como base empírica de la teoría gramatical. La conclusión general es que este método de recolección de datos se encuentra en plena vigencia a pesar de las diversas críticas que pueden hallarse en la bibliografía. Se trata, en definitiva, de un resultado positivo para el gramático de cualquier orientación teórica, dado que este ve garantizada la aptitud de una de las principales fuentes de datos a las que puede apelar.

14 Aunque esta situación está lejos de ser exclusiva de las ciencias de la mente. Para un estudio con afirmaciones similares con respecto a los estudios científicos en general, véase Ioannidis (2005).

## AGRADECIMIENTOS

Muchas de las ideas aquí presentes han sido expuestas en los seminarios *Metodología de la investigación en lingüística: el análisis gramatical* (2012) y *La ciencia de la gramática: método y práctica* (2014) de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires. Agradezco a los estudiantes de dichos cursos por sus comentarios y observaciones. Un agradecimiento especial corresponde a Andrés Saab y Augusto Trombetta, con quienes tuve la oportunidad de discutir (a veces acaloradamente) estos temas, y a un revisor anónimo que contribuyó a mejorar sustantivamente varios puntos del escrito original. Este artículo se benefició de mi participación como becario doctoral en el proyecto UBACyT 20020100100014 “Léxico y gramática: una perspectiva teórico-aplicada y su extensión a otras áreas” dirigido por Mabel Giammatteo. Cualquier error u omisión es de mi entera responsabilidad.

## BIBLIOGRAFÍA

- Bard, Ellen Gurman, Dan Robertson y Antonella Sorace. 1996. “Magnitude estimation of linguistic acceptability”. *Language* 72.1: 32-68.
- Becker, Joseph D. 1975. “The phrasal lexicon”. En *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing*, 60-63. Association for Computational Linguistics.
- Bever, Thomas G. 1970. “The influence of speech performance on linguistic structure”. En *Advances in psycholinguistics*, editado por Giovanni B. Flores D'Arcais y Willem J. M. Levelt, 21-50. Amsterdam: North-Holland.
- Brewer, Marilyn B. 2000. “Research design and issues of validity”. En *Handbook of research methods in social and personality psychology*, editado por Harry T. Reis y Charles M. Judd, 3-16. Cambridge: Cambridge University Press.
- Brunswick, Egon. 1956. *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson y Marcus R. Munafò. 2013. “Power failure: Why small sample size undermines the reliability of neuroscience”. *Nature Reviews Neuroscience* 14: 365-376.
- Campbell, Donald T. 1957. “Factors relevant to the validity of experiments in social settings”. *Psychological Bulletin* 54.4: 297-312.
- Chomsky, Noam. 1955 [1975]. *The logical structure of linguistic theory*. Chicago: University of Chicago Press.
- . 1957. *Syntactic structures*. The Hague: Mouton.
- . 1965. *Aspects of the theory of syntax*. Boston: MIT Press.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks: Sage Publications.
- Edelman, Shimon y Morten H. Christiansen. 2003. “How seriously should we take Minimalist syntax?”. *Trends in Cognitive Sciences* 7.2: 60-61.
- Ferreira, Fernanda. 2005. “Psycholinguistics, formal grammars, and cognitive science”. *The Linguistic Review* 22: 365-380.
- Gibson, Edward y Evelina Fedorenko. 2010. “Weak quantitative standards in linguistics research”. *Trends in Cognitive Sciences* 14.6: 233-234.
- Harris, Zellig S. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.
- Herschel, John Frederick William. 1966. *Preliminary discourse on the study of natural philosophy*. New York: Johnson Reprint Corporation.

- Ioannidis, John P. A. 2005. "Why most published research findings are false". *PLoS Medicine* 2.8: 696-701. doi:10.1371/journal.pmed.0020124.
- Keenan, Edward L. y Matthew S. Dryer. 2007. "Passive in the world's languages". En *Language typology and syntactic description*, volumen 1: *Clause structure*, editado por Timothy Shopen, 325-361. Cambridge: Cambridge University Press.
- McEnery, Tony y Andrew Wilson. 2001. *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McNally, Louise. 2011. "Existential sentences". En *Semantics: An international handbook of natural language meaning*, volumen 2, editado por Klaus von Heusinger, Claudia Maienborn y Paul Portner, 1.829-1.848. Berlin: de Gruyter.
- Meyer, Charles F. 2004. *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Nedjalkov, Vladimir P. 2007. "Overview of the research: Definitions of terms, framework, and related issues". En *Reciprocal constructions*, volumen 1, editado por Vladimir P. Nedjalkov, 3-114. Amsterdam: John Benjamins.
- Phillips, Colin. 2009. "Should we impeach armchair linguists?". En *Japanese/Korean Linguistics*, volumen 17, editado por Shoishi Iwasaki, Hajime Hoji, Patricia M. Clancy y Sung-Ock Sohn, 49-64. Stanford: CSLI Publications.
- Real Academia Española. 2010. *Nueva gramática de la lengua española*. Buenos Aires: Espasa.
- . 2014. *Corpus de referencia del español actual*. <http://corpus.rae.es/creanet.html>.
- Sampson, Geoffrey R. 2001. *Empirical linguistics*. Londres: Continuum.
- . 2007. "Grammar without grammaticality". *Corpus linguistics and linguistic theory* 3.1: 1-32.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Schütze, Carson T. y Jon Sprouse. 2013. "Judgment data". En *Research methods in linguistics*, editado por Robert J. Podesva y Devyani Sharma, 27-50. Cambridge: Cambridge University Press.
- Sinclair, John. 2004. *Trust the text: Language, corpus and discourse*. Londres: Routledge.
- Sprouse, Jon y Diogo Almeida. 2012. "Assessing the reliability of textbook data in syntax: Adger's *Core Syntax*". *Journal of Linguistics* 48.3: 609-652.
- Sprouse, Jon, Carson T. Schütze y Diogo Almeida. 2012. "Assessing the reliability of journal data in syntax: *Linguistic Inquiry* 2001-2010". Ms, University of California y Michigan State University.
- Stefanowitsch, Anatol. 2006. "Negative evidence and the raw frequency fallacy". *Corpus Linguistics and Linguistic Theory* 2.1: 61-77.
- Talmy, Leonard. 2000. *Toward a cognitive semantics*, volumen 1: *Concept structuring systems*. Boston: The MIT Press.
- Wasow, Thomas y Jennifer Arnold. 2005. "Intuitions in linguistic argumentation". *Lingua* 115: 1.481-1.496.
- Wundt, Wilhelm. 1896. *Grundriss der Psychologie*. Leipzig: Engelmann.

### Carlos Muñoz Pérez

Universidad de Buenos Aires

[munozperezc@gmail.com](mailto:munozperezc@gmail.com)

Trabajo recibido el 30 de abril de 2014 y aprobado el 11 de noviembre de 2014.