

Una nota acerca del uso de juicios en teoría gramatical¹

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 a los juicios de aceptabilidad: (i) el que se centra en la “artificialidad” de los datos basados en juicios, crítica planteada por los 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 the 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.

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 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 hablantes², voy a restringir mis afirmaciones a aquel

¹ Agradecimientos.

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*³. El siguiente dato, tomado de un artículo de tipología (Keenan y Dryer 2007, 331), ilustra el funcionamiento de este método.

- (1) *Letón*
Es tieku macits (*no mates).
Yo soy enseñado por madre
'Yo soy instruido' (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 formas 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 GRAE (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

² 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).

³ Prefiero no utilizar el término más estandarizado de *juicio de gramaticalidad* ya que sigo a Chomsky (1965) al considerar que los hablantes sólo pueden acceder a la *aceptabilidad* de una expresión lingüística. En pocas palabras, un hablante puede saber si una expresión está bien o mal formada en su lengua, pero no puede *a priori* determinar si esto se debe a factores gramaticales (i.e., morfo-sintácticos), fonológicos, semánticos, pragmáticos, de procesamiento, etc. Es labor del gramático definir esto a partir del estudio analítico de los patrones aceptables.

persona”. Siguiendo este razonamiento, el encorchetamiento de (2a) es correcto porque la oración de (2b) es aceptable.

- (2) a. No pienso leer [_{OD} 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. [_{OD} 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. Algunos de los autores que sostienen esta distinción consideran que los estudios basados en juicios de aceptabilidad se desvían de la práctica científica normal:

En décadas recientes, la lingüística ha dejado de ser una ciencia empírica en la práctica. Las "gramáticas" (término utilizado para designar las descripciones formalizadas y predictivas de las lenguas) no han tenido en cuenta las observaciones acerca del comportamiento lingüístico concreto. (Sampson 2002, 1).

El hecho es que 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).

No parece adecuado basar en esta diferencia metodológica la distinción entre “buena” y “mala” lingüística. Generalmente, 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*⁴: 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 corpora carecen de utilidad al estudiar, por ejemplo, un fenómeno lingüístico tan característico como la ambigüedad: sólo a través de la “intuición” de un hablante se puede saber que las expresiones de (7) tienen más de un significado⁵.

(7) a. La matanza de los cazadores.

⁴ 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).

⁵ 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?”.

- 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, dado que no cualquier argumento nominal puede acompañar a *haber* en construcciones existenciales. Sin embargo, el tipo de dato necesario para realizar esta afirmación de modo no inductivo (e.g. **había el chico en la plaza*) no existe en ningún corpus.

Uno 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 real⁶. En este sentido, hay autores que asumen 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" (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).

⁶ 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)

Si bien es cierto que muchas de las expresiones sometidas a juicios pueden ser difíciles de utilizar en una conversación, este no es un parámetro para evaluar la validez global de esta metodología. En primer lugar, los criterios de adecuación como la validez ecológica se aplican a investigaciones 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 asignarse 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. En el contexto de los estudios gramaticales, la validez interna se relaciona en forma directa con la construcción de *pares mínimos* (i.e. dos o más expresiones que son idénticas salvo por un rasgo o valor en contraste⁷) y con la evaluación de la hipótesis con respecto a dicho par. 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

⁷ El término *par mínimo* tiene su origen y formulación estricta en el campo de la fonología (e.g. Harris 1951). Su extensión a otros niveles de análisis lingüístico (e.g. sintaxis, semántica) no implicó, hasta el momento, una reformulación precisa de dicha definición.

trata de un criterio que busca evaluar si es posible establecer la misma relación de causa-efecto a través de diferentes informantes/participantes, contextos y tareas experimentales⁸. Existe, en definitiva, una importante diferencia entre los tres tipos de criterios presentados: tanto la validez interna y 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 corpora (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, etc. 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 (e.g. Edelman y Christiansen 2003, Ferreira 2005, Wasow y Arnold 2005, Gibson y Fedorenko 2010).

⁸ Para una tipología y descripción de las tareas de juicios más comúnmente utilizadas, véase Schütze y Sprouse (2014).

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 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 estas críticas afirman que el método tradicional de recolección de datos debe ser totalmente abandonado y reemplazado por un conjunto de estándares metodológicos estrictos que en conjunto reciben el nombre de *syntaxis experimental* (e.g. Cowart 1997). 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 de modo general, 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), algo similar a intentar mostrar la poca fiabilidad de un tipo de diagnóstico médico a partir de la existencia de un único falso positivo. Por supuesto, el problema es complejo y no puede zanjarse únicamente mostrando que existen datos erróneos en la bibliografía, sino que es necesario demostrar (i) que el uso de datos erróneos es moneda corriente y (ii) que el método tradicional realmente es poco fiable (i.e. 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 espurios ni

a realizar elecciones teóricas de peso a partir de datos controversiales⁹. 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 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*¹⁰ 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 dichos en un contexto experimental formal con 40 participantes. Resultó que los datos fueron coincidentes en un 98% de los casos, es decir que el número de falsos positivos y falsos negativos del método tradicional no superó el 2%. En un segundo experimento (Sprouse, Schütze y Almeida 2012), se evaluó una selección aleatoria de datos del período 2001-2010 del journal *Linguistic Inquiry* con resultados similares: los datos resultaron replicados en un 95% de los casos. Dado que en ambos experimentos la proporción de datos erróneos obtenidos a partir de la metodología tradicional fue muy baja (entre el 2% y el 5%, respectivamente), la idea de que dichos datos resultan poco fiables tuvo que ser descartada.

⁹ Entre otros factores, esto puede deberse al estricto control que realizan los gramáticos con respecto a los datos: por lo general, existe un consenso comunitario acerca de los patrones que es necesario explicar.

¹⁰ 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.

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 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.

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. Mis cursivas).

En este contexto, 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. Conclusiones

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.

Referencias

Bard, E. G., Robertson, D., y Sorace, A. 1996. "Magnitude estimation of linguistic acceptability". Language: 32-68.

- Becker, J. 1975. The phrasal lexicon. En *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing*. Association for Computational Linguistics. 60-63.
- Brewer, M. 2000. "Research Design and Issues of Validity". En Reis, H. y Judd, C. (eds) *Handbook of Research Methods in Social and Personality Psychology*. Cambridge: CUP.
- Brunswick, E. 1956. *Perception and representative design of psychological experiments*. Berkeley: University of California Press.
- Button, K., Ioannidis, J., Mokrysz, C., Nosek, B., Flint, J., Robinson, E. y Munafò, M. 2013. "Power failure: why small sample size undermines the reliability of neuroscience". *Nature Reviews Neuroscience* 14: 365-376.
- Campbell, D. 1957. "Factors Relevant to the Validity of Experiments in Social Settings". *Psychological Bulletin* 54: 297-312.
- Chomsky, N. 1955. *The Logical Structure of Linguistic Theory*. Chicago: University of Chicago Press (1975).
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Boston: MIT Press.
- Cowart, W. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks: Sage Publications.
- Edelman, S., y Christiansen, M. H. 2003. "How seriously should we take Minimalist syntax?". *Trends in Cognitive Sciences* 7: 60-61.
- Ferreira, F. 2005. "Psycholinguistics, formal grammars, and cognitive science". *The Linguistic Review* 22: 365-380.
- Gibson, E. y Fedorenko, E. 2010. "Weak quantitative standards in linguistics research". *Trends in Cognitive Sciences* 14: 233-234.
- NGRAE. 2010. *Nueva Gramática de la Lengua Española*. Buenos Aires: Espasa.
- Harris, Z. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Herschel, J. F. W. 1966. *Preliminary Discourse on the Study of Natural Philosophy*. New York: Johnson Reprint Corporation

Keenan, E. y Dryer, M. 2007. "Passive in the world's languages". En Timothy Shopen (ed.), *Language Typology and Syntactic Description Volume I: Clause Structure*. Cambridge: CUP.

McEnery, T. y Wilson, A. 2001. *Corpus Linguistics. An Introduction*. Edinburgh: Edinburgh University Press,

Phillips, C. 2009. "Should we impeach armchair linguists". *Japanese/Korean Linguistics* 17: 49-64.

Sampson, G. 2002. *Empirical linguistics*. Continuum.

Sampson, G. 2007. "Grammar without grammaticality". *Corpus linguistics and linguistic theory* 3: 1-32.

Schütze, C. .1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Schütze, C. y Sprouse, J. 2014. "Judgment data". *Research Methods in Linguistics* 27.

Sprouse, J. y Almeida, D. 2012. "Assessing the reliability of textbook data in syntax: Adger's Core Syntax". *Journal of Linguistics* 48: 609-652.

Sprouse, J., Schütze, C. y Almeida D. 2012. "Assessing the reliability of journal data in syntax: Linguistic Inquiry 2001-2010". Ms, University of California y Michigan State University.

Talmy, L. (2000). *Toward a cognitive semantics, Vol. 1: Concept structuring systems*. Boston: The MIT Press.

Wasow, T. y Arnold, J. 2005. "Intuitions in linguistic argumentation". *Lingua* 115: 1481-1496.

Wundt, Wilhelm. 1896. *Grundriss der Psychologie*. Leipzig: Engelmann.