

Juicios de aceptabilidad del español: cuándo y cómo llevar a cabo un experimento formal

Santiago Gualchi
(Universidad de Buenos Aires)

Federico Alvarez
(Universidad de Buenos Aires)

Laura Tallon
(Universidad de Buenos Aires)

Charo Zacchigna
(Universidad de Buenos Aires)

Los juicios de aceptabilidad (a veces llamados juicios de gramaticalidad) se han convertido en una práctica común en distintas áreas de la lingüística, especialmente en lingüística formal. No obstante, esta metodología no ha permanecido libre de controversias. En los últimos 20 años, se ha alzado un debate entre quienes defienden los métodos tradicionales (en adelante, métodos informales) de recolección de juicios y quienes demandan una migración hacia métodos más formales.

Los procedimientos informales emplean muestras de conveniencia, esto es, la recolección de observaciones se realiza en función de la facilidad de su obtención y no se lleva a cabo un control de las variables que puedan sesgarlas como ocurriría en un muestreo aleatorio o estratificado (Diez, Barr y Çetinkaya-Rundel, 2012). Particularmente, en la mayor parte de los casos, estos métodos reflejan los juicios del mismo investigador. Si, en cambio, el experimentador decide solicitar información a terceros, con frecuencia, los entrevistados resultan ser los colegas del investigador interrogados en condiciones experimentales pobres (por ejemplo, conversaciones casuales o mensajería instantánea). Los detractores de estos métodos han señalado que podrían estar provocando una pérdida de exactitud en nuestros modelos del lenguaje y han sugerido que la lingüística debería adoptar procedimientos más rigurosos, en línea con las prácticas comunes en las ciencias cognitivas. Por el contrario, sus defensores consideran que el enfoque informal conduce a un desarrollo más veloz de la lingüística sin perder exactitud de una manera significativa.

En este contexto, Sprouse y Almeida (2012), y Sprouse, Schütze y Almeida (2013) han estudiado la validez de métodos informales y métodos formales. Estos autores llevaron a cabo experimentos en los que analizaron formalmente juicios de aceptabilidad del inglés reportados previamente en la literatura científica que habían sido recolectados informalmente. Como resultado, consiguieron replicar los resultados en un 98% y un 95% de los casos, respectivamente. Estos experimentos, junto con resultados que señalan la crisis metodológica

prevalente en las ciencias cognitivas, son empleados por Muñoz Pérez (2014) para sostener que los juicios informales son suficientes para la construcción de teorías lingüísticas.

Sin embargo, Linzen y Oseki (2018) han apuntado que estos estudios se basaron en muestras de juicios seleccionados aleatoriamente y, en consecuencia, han incluido numerosos juicios evidentes (por ejemplo, **the bear snuffleds*). En este sentido, Marantz (2005) distingue entre distintos tipos de juicios. El ejemplo mencionado se corresponde con un juicio de tipo 2, dado que se juzga una generalización no controversial acerca de una lengua, particularmente, que el morfema de tercera persona singular y el de pretérito simple en inglés se encuentran en distribución complementaria. Como resultado, la tasa de replicabilidad de estas muestras no solo queda determinada por juicios cuestionables o sensibles, sino también por juicios cuyo valor es indiscutido.

De este modo, la interpretación de esta tasa se vuelve difusa. Más aún, Linzen y Oseki (2018) señalaron que los juicios del inglés se ubican en una posición especial en la literatura, ya que, la comunidad sintáctica ostenta, en general, una competencia nativa, o cercana a nativa, de esta lengua, lo que conduce a que los juicios incorrectos en esta lengua sean sujetos a mayores controles y acaben siendo filtrados en el proceso de revisión por pares. A partir de estas observaciones, estos autores diseñaron un experimento en el que evaluaron la tasa de replicabilidad de contrastes cuestionables del japonés y del hebreo reportados en la literatura en inglés. Los juicios que analizaron pertenecían al tercer grupo en la clasificación de Marantz (2005) (por ejemplo, juicios acerca de dominios de localidad para dependencias a larga distancia), que pueden ser poco concluyentes mediante la recolección informal. Como resultado, la mitad de los contrastes del hebreo y un tercio de los contrastes del japonés no se repitieron o, si se dieron, fue en la dirección opuesta (la oración reportada como agramatical fue más aceptada que la reportada como gramatical).

Motivados por estos experimentos, en este workshop presentaremos un primer acercamiento a un procedimiento formal de recolección de juicios de aceptabilidad para el español. En nuestro experimento solicitamos a los participantes que puntuaran una serie de oraciones en una escala de 1 a 5. Las oraciones se usaron para evaluar las hipótesis de Gualchi (pendiente de aprobación) respecto a la concordancia entre objetos directos y clíticos. Para facilitar la obtención de una muestra representativa, recolectamos datos mediante el formato de encuesta online. Si bien este mecanismo dificulta el control de variables contextuales, que, a su vez, aumentan el ruido, el tamaño de la muestra eleva sustancialmente el poder estadístico (esto es, la probabilidad de encontrar un contraste verdadero). Asimismo, tomamos como variable de control la formación en lingüística de los sujetos que contestaron la encuesta con el fin de prevenir un potencial efecto de sesgo teórico. Además, hemos incluido una fase de práctica para asegurarnos que los sujetos hubieran entendido correctamente la consigna. Es necesario hacer

notar que, en concordancia con Linzen y Oseki (2018), no sugerimos abandonar la recolección informal de juicios de aceptabilidad, sino ofrecer una herramienta dirigida a asistir en la recolección de datos para construcciones particularmente desafiantes. Esperamos que la disponibilidad de nuevos procedimientos formales para la recolección y análisis de datos lingüísticos contribuya hacia el desarrollo de una ciencia del lenguaje más rigurosa.

Referencias

- Diez, D., Barr, C. y Çetinkaya-Rundel, M. (2012). *OpenIntro Statistics*. CreateSpace.
- Gualchi, S. (pendiente de aprobación). La concordancia parcial en los clíticos acusativos del español rioplatense. *Exlibris*.
- Linzen, T. y Oseki, Y. (2018). The reliability of acceptability judgments across languages. *Glossa*, 3(1), 1-25.
- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *Linguistic Review*, 22(2-4), 429-445.
- Muñoz Pérez, C. (2014). Una nota acerca del uso de juicios en teoría gramatical. *Signo y seña*, 26, 107-120.
- 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, y. Schütze, C. T. & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry 2001-2010*. *Lingua*, 134, 219-248.