

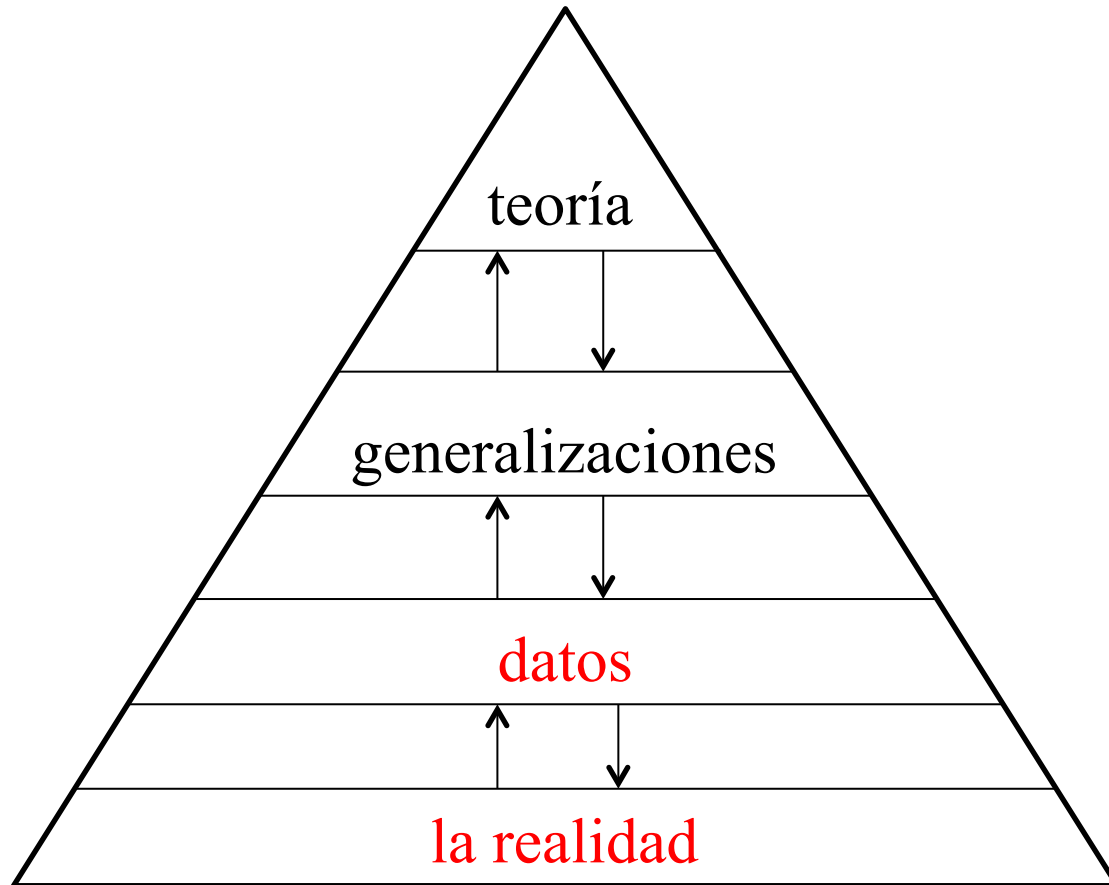
Metodología de la investigación en lingüística: el análisis gramatical

Octubre de 2012

UNIDAD 4: ORIGEN Y TRATAMIENTO DE LOS DATOS

¿De dónde vienen los datos del gramático?

Carlos Muñoz Pérez
UBA



El lenguaje presenta dificultades prácticamente únicas para su estudio: desde un punto de vista estrictamente platónico se puede decir que ***no existe***, que es una mera abstracción. ¿Cómo se estudia aquello que no existe, que no podemos ver o tocar?

El problema en términos más teóricos es:

¿Cómo accedemos a la Lengua-I que queremos investigar y definir?

Una manera de hacerlo es a través de la *Lengua-E*. El estudio de la Lengua-E podría permitirnos inferir algunas de las propiedades de la Lengua-I.

Una idea más osada es intentar acceder a la Lengua-I *a partir de métodos y técnicas especialmente diseñados para ello* con el fin de obtener información más precisa sobre la naturaleza de la Lengua-I.

En este sentido, hay tres tipos de datos que utiliza un lingüista para elaborar teorías sobre la Lengua-I.

- Datos que existen independientemente del «deseo» del lingüista por estudiar el lenguaje (i.e., *corpus*).
- Conducta verbal inducida con el fin específico de estudiar la Lengua-I (*introspección*).
- Observación de los procesos mentales y fisiológicos que se dan durante la utilización inconsciente de la Lengua-I (e.g., *un experimento psicolingüístico*).

A su vez, con respecto a la conducta verbal inducida, podemos hacer una nueva distinción a nivel práctico.

- Conducta verbal inducida en un marco casual. Experimento psicológico no formal (i.e., *juicio de gramaticalidad*).
- Conducta verbal inducida en un marco estrictamente controlado según los parámetros de la psicología experimental.

Nuestra discusión sobre las fuentes de datos del gramático consistirá en:

- Un análisis superficial sobre la utilización de corpus en gramática.
- Una discusión profunda acerca de la fuente de datos más corriente en los estudios gramaticales: la introspección.
- Un breve justificativo a la última unidad del seminario (falta todavía): la gramática experimental.

corpus como fuente de datos

Empecemos por una definición de *corpus*. La tomo del diccionario de lingüística de David Crystal.

Corpus:

Una colección de datos lingüísticos, ya sean textos escritos o transcripciones de habla, que puede ser usada (i) *como punto de partida de la descripción lingüística* o (ii) *como medio de verificar hipótesis sobre el lenguaje*.

La noción de corpus nos lleva a hablar de la llamada *lingüística de corpus*.

corpus como fuente de datos

La lingüística de corpus no es el estudio de ningún componente particular del sistema lingüístico.

Lingüística de corpus:

«Área de la lingüística» que se aboca al desarrollo de procedimientos y métodos para estudiar el lenguaje.

(si bien muchos lingüistas de corpus no parecen estar muy contentos con la caracterización de su área como «mera metodología»).

Existe una dicotomía entre los enfoques que se denominan *corpus-based* y *corpus-driven*.

corpus como fuente de datos

Los estudios *corpus-based* típicamente usan datos de corpus para explorar una teoría o hipótesis establecida en la bibliografía con el fin de validarla, refutarla o refinarla. La definición de la lingüística de corpus en tanto *método* subyace a este uso de los datos de corpus en lingüística.

Los estudios *corpus-driven* rechazan la caracterización de la lingüística de corpus como método, y asumen, en cambio, que el corpus en sí mismo debería ser la única fuente de hipótesis sobre el lenguaje. Involucran, entonces, que *el corpus encarna una teoría del lenguaje*.

ventajas del uso de corpus

Ventaja N° 1:

Los corpus contienen muestras de lengua que fueron producidas cuando el hablante no estaba concentrado en cómo se expresaba (a nivel de forma).

Muchas veces, los hablantes consideran inaceptables expresiones *que utilizan sin darse cuenta*. Este tipo de fenómeno no se capta a partir del uso exclusivo de juicios gramaticales.

Uno de los primeros en observar esto fue Labov a mediados de los 70s con respecto a la utilización de *anymore*.

ventajas del uso de corpus

En la mayoría de los dialectos del inglés, *anymore* se utiliza negativamente.

(1) Trains don't stop here anymore.

Sin embargo, en algunos dialectos se lo utiliza en contextos gramaticales positivos para dar cuenta de cosas que son negativas en su valor.

(2) John is smoking a lot anymore.

'John empezó a fumar mucho, lamentablemente'.

Labov y sus colaboradores encontraron que la gente que suele producir este tipo de expresiones lo hacía sin darse cuenta.

ventajas del uso de corpus

Cuando se les presentaba una oración como *John is smoking a lot anymore* decían que nunca habían escuchado algo parecido, no lo reconocían como inglés, pensaban que significaba 'John no está fumando' y mostraban los mismos signos de desconcierto que cualquiera tiene ante hablantes de otras variedades de la propia lengua. Esto describe la conducta de Jack Greenberg, un constructor de 58 años. Sus juicios de introspección eran tan convincentes que tuvimos que aceptarlos por descripciones válidas de su gramática. Pero dos semanas más tarde se escuchó que le decía a un plomero: *Do you know what's a lousy show anymore?* Johnny Carson'.

(Labov 1975)

ventajas del uso de corpus

Ventaja N° 2:

Hay fenómenos gramaticales que se dan en contextos particulares, sólo detectables en un corpus.

Considérense los siguientes datos:

- (3) Saw no one.
- (4) Hurt myself when trying to cut the roses.
- (5) Left the party exhausted.

Haegeman (1990) observa que oraciones con sujeto nulo como (3-5) son recurrentes en un tipo de texto particular: el diario íntimo.

ventajas del uso de corpus

El fenómeno no se restringe a la primera persona:

(6) Saw no one after we had left the party.

(7) Had hurt himself while trying to cut the roses.

Haegeman observa que estos sujetos nulos no pueden aparecer en subordinadas (8), con objetos topicalizados (9), o con movimiento-wh (10). Sin embargo, no es necesario que aparezca en primera posición (11).

(8) * I wonder when will see her again.

(9) * This book, did not approve of .

(10) * When will be able to meet him?

(11) So shall stop writing for a day.

ventajas del uso de corpus

Ventaja N° 3:

Sólo a través de corpus podemos estudiar lenguas ya extintas o en peligro de desaparecer.

¿Se acuerdan los datos del rumano sobre copias múltiples que vimos la clase pasada?

Cine ce precede?

Quién qué precede

*Cine precede ce?

Quién precede qué

‘¿Quién precede qué?’

*Ce ce precede?

Qué qué precede

Ce precede ce?

What precedes what

‘¿Qué precede qué?’

ventajas del uso de corpus

Parece que en griego antiguo se daba un fenómeno muy similar. Para empezar, es normalmente observado que en esta lengua un modificador genitivo podía aparecer a la izquierda o a la derecha del núcleo nominal al que modifica.

[[hee tólma] [tóon legóntoon]]

EL-NOM VALOR-NOM ELLOS-GEN HABLANTES-GEN

‘la valentía de los que hablan’

[[hee [tóu himatíoon] ergasías]

EL-NOM EL-GEN MULTITUD-GEN LEY-NOM

‘La ley de la multitud’

ventajas del uso de corpus

Dada la recurrencia de esta alternancia, sería esperable que existiera también el siguiente par.

[[[tóon	oikeíoon]	tinàs]	[tóon	ekeínoon]]
los-GEN	esclavos-GEN	algunos-GEN	los-GEN	aquellos-GEN

*[[tóon	[tóon	ekeínoon]	oikeíoon]	tinàs]
los-GEN	los-GEN	aquellos-GEN	esclavos-GEN	algunos-GEN
<i>‘algunos de los esclavos de aquellos’</i>				

Golston (1995) realizó una búsqueda en un corpus que reunía más de 500 años de literatura griega antigua, pero no encontró ningún caso en el que se diera un patrón sintáctico como el segundo.

ventajas del uso de corpus

Ventaja N° 4:

Los corpus permiten el análisis de fenómenos no observables a partir del análisis introspectivo de oraciones aisladas (e.g., identificación de referentes a través del discurso).

Ventaja N° 5:

En los corpus electrónicos, es posible automatizar la búsqueda de datos y la generación de estadísticas.

ejemplo de uso de un corpus

Veamos un ejemplo puntual de la utilización de datos de corpus para apoyar un análisis gramatical.

En su estudio sobre la sintaxis de las construcciones aditivo-sustractivas del español, Brucart analiza patrones como el siguiente:

- (12) Leyó **más** libros
- (13) a. Leyó **dos** libros **más**
b. *Leyó **más** dos libros.
- (14) a. Leyó **algunos** libros **más**.
b. *Leyó **más** algunos libros.
- (15) a. Leyó {**muchos/bastantes/pocos**} libros **más**.
b. *Leyó **más** {**muchos/bastantes/pocos**} libros.
c. Leyó {**muchos/bastantes/pocos**} **más** libros.

ejemplo de uso de un corpus

Las proyecciones involucradas en esta estructura serían un *Sintagma de Grado* (SG) y un *sintagma de cuantificador* (SQ).

[_{SG} ____ [_{G'} *más* [_{SQ} ____ [_{Q'} *dos* [_{SN} *libros*]]]]

Reconsideremos los ejemplos bajo esta estructura.

- (12) Leyó **más** libros
- (13) a. Leyó dos libros **más**
b. *Leyó **más** dos libros.
- (14) a. Leyó algunos libros **más**.
b. *Leyó **más** algunos libros.
- (15) a. Leyó {muchos/bastantes/pocos} libros **más**.
b. *Leyó **más** {muchos/bastantes/pocos} libros.
c. Leyó {muchos/bastantes/pocos} **más** libros.

ejemplo de uso de un corpus

Brucart utiliza datos de corpus para mostrar que la asimetría entre cuantificadores como *algunos* y *bastantes* tiene un carácter diacrónico.

Ciertos cuantificadores léxicos pueden ascender al especificador de SG sin ir acompañados del SN. Se trata de *muchos*, *bastantes* y, en menor medida, de *pocos*. Parece que esta clase ha experimentado cambios en la historia del español. Así, el CORDE y el DCRLC dan noticia de construcciones de ascenso del cuantificador en solitario con *algún* y *otro*.

CORDE = Corpus Diacrónico del Español

DCRLC = Diccionario de construcción y régimen de la lengua castellana

ejemplo de uso de un corpus



Real Academia Española - Corpus Diacrónico del Español (CORDE)

Consulta:	<input type="text" value="alguna más"/>		
Criterios de selección:			
Autor:	<input type="text"/>	Obra:	<input type="text"/>
Cronológico:	<input type="text"/> <input type="text"/>	Medio:	Geográfico:
		(Todos) Libros Periódicos Revistas Miscelánea Orales	(Todos) Argentina Bolivia Chile Colombia Costa Rica
Tema:	(Todos) 11.- Lfrica 12.- Narrativa 121.- Breve 1211.- Relato breve tradicional 1212.- Relato breve culto		

ejemplo de uso de un corpus

Nº	CONCORDANCIA		AÑO
1	iento á lo que, segun ellos, no envuelve realidad	alguna, más en el fondo piensan como aquel filósofo.	** 1877
2	el futuro de reparación se alargue a veces, para	alguna, más de lo que se quisiera. Quizá la carretera	** 1958
3	isión del Protonotario no hay hasta agora Novedad	alguna, más de que, haviéndole dado vn Criado que le	** 1644
4	rta cantidad de Dinero sobre que le hablaron con	alguna más libertad de la que pide la distancia de la	** 1643
5	Un correo que ha venido de Italia no trae novedad	alguna más de que se continuaba en Turín la tregua, b	** 1639
6	sin más carga de la con que se fué, y el Rey con	alguna más salud de la que llevó. El Presidente que f	** 1659
7	rso, en el que salvo las esculturas mencionadas y	alguna más daba una triste impresión de impotencia, d	** 1935
8	n que en su egerçio tengan util ni conveniençia	alguna más que desear el vien y aumento de sus repúbl	** 1684
9	las tales casas ni labrar ni hedificar en él cosa	alguna, más que finque para siempre fecho muladar en	** 1478
10	verbo ad verbum, non añadiendo ni menguando cossa	alguna más de quanto está en la dicha apuntadura oreg	** 1504
11	sas se han aumentado, en que parece pudiera haver	alguna más. dificultad. 10. Antonio de Herrera en su	** 1648
12	s que trae consigo el acto, y sólo sirve de darle	alguna más fuerza y autoridad, como yá queda dicho, y	** 1648
13	o puedan echar ni echen en lana tynta (ni en otra	alguna) más de medio açunbre de agua en cada quartill	** 1500
14	E. en la primera ocasión, y sin admitirle excusa	alguna más de venir a dicho don Antonio Ternero a viv	** 1783
15	s indios de la provincia de los condes se reparta	alguna más cantidad que antes, por me constar que des	** 1578
16	, los dichos naturales deban ni puedan pagar cosa	alguna más de lo que buenamente podían y debían pagar	** 1574
17	pagando los derechos de la entrada; y si hallare	alguna más adelante, se entienda ser la suya, y quede	** 1574
18	ho cabildo sin el corregidor ni tratar otra cosa	alguna más de lo que a él toca, so pena que si se tra	** 1572
19	ldo no puedan tomar más que los otros, ni persona	alguna más de aquello que verosimilmente hubieren men	** 1572
20	como está ordenado, los impida, ni ocupe en cosa	alguna más de en las sobredichas. titulo xxiv De la	** 1572
21	ue los indios cuando vuelvan a sus tierras lleven	alguna más ganancia, proveyéndolo como más convenga a	** 1569
22	a del cacique principal que sea con más anchura e	alguna más autoridad que la de los indios particulare	** 1569
23	nsa de alguna cosa que se les dio, de llevar cosa	alguna, más de lo que fuere tasado, so pena que por c	** 1544
24	to trabajo; y por esto, porque parezca que tienen	alguna más libertad no los repartí como los otros; y	** 1524
25	icios insignes que en ella haya, describiendo con	alguna más extensión los que por su magnificiencia, a	** 1765

ejemplo de uso de un corpus

La idea es, entonces, que el español antes permitía el ascenso de cuantificadores «solos» de manera generalizada, pero que luego esta propiedad se restringió a elementos como *muchos*, *bastantes* y *pocos*.

Como observa Sáez del Alamo, estos cuantificadores son los mismos que pueden funcionar como atributos.

(16) Tres libros son {muchos / bastantes / pocos}.

(17) *Tres libros son {algunos / numerosos / varios / tres}.

(Para más datos sobre el tema, vayan al artículo de Brucart. Les dejo el link al final del PPT).

desventajas del uso de corpus

Desventaja N° 1:

¿Cómo buscar datos que sean relevantes para la investigación que estamos llevando a cabo?

Supongan que quieren estudiar... eh... no sé... casos de Control no obligatorio. Para eso pueden necesitar oraciones del siguiente tipo:

(18) PRO evadir impuestos perjudica a la clase trabajadora.

¿Cómo se hace para encontrar este tipo de estructura en un corpus como el CREA?

desventajas del uso de corpus



Real Academia Española - Corpus de Referencia del Español Actual (CREA)

Consulta:	<input type="text" value="evadir impuestos perjudica a la clase trabajadora"/>		
Criterios de selección:			
Autor:	<input type="text"/>	Obra:	<input type="text"/>
Cronológico:	<input type="text"/> <input type="text"/>	Medio:	Geográfico:
		(Todos) Libros Periódicos Revistas Miscelánea Oral	(Todos) Argentina Bolivia Chile Colombia Costa Rica
Tema:	(Todos) 1.- Ciencias y Tecnología. 101.- Biología. 102.- Veterinaria. 103.- Ecología. 104.- Tecnología.		

desventajas del uso de corpus

Real Academia Española.

Banco de datos del Español

No existen casos para esta consulta.

Pulse el botón de *Retroceso* de su navegador

Bueno, en realidad es una crítica bastante injusta hacia el CREA. Este tipo de cosas hay que buscarlas en un corpus en el que los datos estén etiquetados sintácticamente.

desventajas del uso de corpus

Base de Datos Sintácticos del Español Actual

Verbos

Verbos en un subesquema

Funciones que componen el esquema:

- ☒ Sujeto
- ☒ Complemento directo
- ☐ Complemento indirecto
- ☐ Complemento agente

- ☐ Suplemento
- ☐ Complemento adverbial
- ☐ Complemento modal
- ☐ Otros complementos preposicionales

desventajas del uso de corpus

Base de Datos Sintácticos
del Español Actual

Verbos

Verbos en un subesquema

Datos acerca del subesquema:

Datos para Sujeto (S)

Cláusula de infinitivo ▼

Datos para Complemento directo (D)

[+anim] ▼

Enviar

Borrar

desventajas del uso de corpus

Base de Datos Sintácticos del Español Actual

Verbos

Esquemas

Subesquemas

Ejemplos

Pre

FAVORECER

[1 ejemplo]

1/1

Activa

SD

Scf Dan

BAIRES: 123, 15

(123, 13)

(123, 14)

(123, 15)

(123, 16)

(123, 17)

mucho tiempo, ¿no?

Inf. B.- Sí, por un lado... eh... copiar pautas- - -

culturales de la- - - patrona la favorece por ese aspecto,

pero por otro lado la embroman cuando yo- - - estoy segura

que no hay ninguna conciencia de clase.

Nos fue bastante bien: encontramos un caso. La pregunta, claro, es si sólo un caso alcanza para un estudio del fenómeno en cuestión.

desventajas del uso de corpus

El problema es que los lingüistas suelen tratar con datos mucho más complejos que ese.

(19) El chico que tiene la gorra que te mostré ayer no la usa.

El ejemplo de (19) es, todavía, bastante simple. Sin embargo, un buscador como el recién visto no permite buscar datos con ese nivel de anidamiento.

Ahora bien, supongamos que no encontramos ningún dato que siga el esquema o tenga las propiedades que estamos buscando (e.g., el corpus es muy chico, no tiene la diversidad necesaria). Esto plantea otro problema...

desventajas del uso de corpus

Desventaja N° 2:

¿Cómo se interpreta la ausencia de un dato en el corpus?

Sería raro encontrar en un corpus oraciones que tengan varios niveles de anidamiento (los hablantes suelen no emitirlas). Estas estructuras, sin embargo, son perfectamente posibles en español.

Otras lenguas, en cambio, restringen el anidamiento sintáctico (i.e., no se puede tener una subordinada dentro de otra subordinada).

¿Se puede captar esta diferencia a partir del corpus?

desventajas del uso de corpus

Desventaja N° 3:

Los *treebanks* (corpus sintácticamente anotados) dependen de análisis sintácticos particulares.

Desventaja N° 4:

El acceso a muchos corpus es pago.

Si, acceder el *Penn Treebank* (un corpus particularmente rico) nos costaría U\$3150.

Desventaja N° 5:

Ningún corpus posee evidencia negativa.

desventajas del uso de corpus

Consideren las siguientes oraciones:

(20) [La tortuga ninja]_i bailaba mientras *pro*_i comía pizza.

(21) Mientras *pro*_i comía pizza, [la tortuga ninja]_i bailaba.

(22) **pro*_i bailaba mientras [la tortuga ninja]_i comía pizza.

Principio C:

Una Expresión Referencial no puede estar coindizada con un elemento que la mande-c.

Los chicos no tienen acceso a evidencia negativa como (22). ¿Cómo hacen para adquirir este principio?

Respuesta de Chomsky: ese principio es innato.

desventajas del uso de corpus

Los lingüistas no tenemos acceso consciente a las propiedades de nuestra Lengua-I. Y, sin embargo, tenemos que construir una teoría sobre la Lengua-I.

Si la evidencia con la que tenemos que manejarnos se restringe al corpus (i.e., a las oraciones bien formadas), ¿cómo podríamos llegar a postular algo así como el *Principio C*?

Para poder definir las propiedades de la gramática, es necesario contar con evidencia negativa. Y los corpus no la proveen.

desventajas del uso de corpus

Desventaja N° 6:

En un corpus no se tiene ningún tipo de control sobre las muchas variables que pueden condicionar una emisión lingüística.

Este es, básicamente, el motivo por el que la ciencia no se basta con datos provenientes de la observación, sino que postula la necesidad de experimentación.

Consideremos el siguiente estudio observacional.

desventajas del uso de corpus

Un estudio con mujeres de 14 a 21 años comenzó en 1979. En 1988 estas mujeres tenían 1239 hijos entre 6 y 9 años. Las mujeres y los niños fueron entrevistados en 1988 y, nuevamente, en 1990. Dos de las cosas que se medían era el grado de conducta antisocial de los chicos y la frecuencia con que se les pegaba. Los resultados mostraron que los chicos que fueron golpeados más frecuentemente en 1988 mostraron más conductas antisociales en 1990 que los chicos que fueron menos golpeados.

desventajas del uso de corpus

¿Son los golpes el motivo de la conducta antisocial? Tal vez sí, pero hay otras explicaciones posibles. Tal vez los chicos más problemáticos en 1988 eran justamente los más golpeados, y quienes eran menos problemáticos eran golpeados menos.

El problema con los estudios observacionales es que las variables involucradas en un fenómeno no están controladas (y muchas veces son desconocidas) por el investigador.

Adaptado de Dehlert, G. (2010).
A First Course in Design and Analysis of Experiments.

en definitiva...

Los datos que encontramos en un corpus son, sin duda, valiosos...

...pero no es posible hacer gramática únicamente con este tipo de datos.

Recreo.

En cuanto volvamos, una introducción a los problemas que plantea la utilización de juicios de gramaticalidad.

repartiendo lecturas para la próxima

Alguien si o si tiene que ofrecerse para presentar los siguientes artículos.

Son artículos cortos y sin complicaciones técnicas.

Wasow y Arnold (2005). «Intuition in linguistic argumentation».

Phillips (2009). «Should we impeach armchair linguistics?».

la introspección como fuente de datos

Simultáneamente, la mayor virtud y falencia de la teoría lingüística es que los juicios de aceptabilidad sean utilizados como datos primarios.

Thomas Bever

la introspección como fuente de datos

La «intuición de hablante nativo» es por lejos la más utilizada fuente de datos en los estudios lingüísticos.

- Juicios de gramaticalidad.
- Intuiciones sobre el orden de las operaciones (e.g., en la formación de palabras).
- Intuiciones acerca de la correspondencia entre formas lingüísticas (e.g., felicidad o infelicidad de un tipo de respuesta ante una determinada pregunta).
- Identificación de ambigüedad estructural (e.g., *la matanza de los cazadores*).

el juicio de gramaticalidad

Vamos a centrarnos, sobre todo, en los juicios de gramaticalidad, por más que muchas cosas que digamos van a aplicarse también a los otros casos de introspección.

Hay cuatro motivos bien definidos para utilizar juicios de gramaticalidad (sin contar el clásico «es rápido y fácil»).

Motivo 1:

Elicitar juicios de gramaticalidad permite analizar reacciones a emisiones lingüísticas que raramente aparecen en corpus.

el juicio de gramaticalidad

Motivo 2:

Los juicios de gramaticalidad proveen información negativa.

Motivo 3:

A partir de los juicios de gramaticalidad es posible descartar cuestiones como falsos comienzos, errores de habla, oraciones sin terminar, etc.

Motivo 4:

La utilización de juicios permite controlar el conjunto de variables que constituyen la *actuación lingüística*.

el juicio de gramaticalidad

*Y así, los lingüistas «modernos» pasan su tiempo sacando y poniendo estrellitas sobre oraciones improbables del tipo **John, Bill y Tom se mataron entre sí** (por tomar un ejemplo al azar de una publicación), las cuales elaboran con reprimida frustración ya que son muy difíciles de ubicar en una conversación. Estos ejemplos no guardan ningún parecido con las oraciones que componen el texto que pretende explicarlas.*

(Becker 1975)

el juicio de gramaticalidad

Más allá de críticas como la de Becker, los juicios de gramaticalidad son indudablemente problemáticos:

- *Carecen de validez ecológica*: a los sujetos se les pide realizar una tarea que es totalmente distinta al uso normal del lenguaje. (Aunque, como señala Cowart, no es del todo desatinado proponer que los juicios de gramaticalidad guardan una función lingüística).
- *No siguen los procedimientos de la psicología experimental*: no se sigue un estándar de experimentación; es una técnica informal.

el juicio como actuación

*La noción de **aceptabilidad** no debe ser confundida con la de **gramaticalidad**. La aceptabilidad es algo que pertenece al terreno de la actuación, mientras que la gramaticalidad pertenece al campo de la competencia [...]. Como la aceptabilidad, la gramaticalidad también tiene grados, pero ambas escalas no coinciden. La gramaticalidad es sólo uno de varios factores involucrados en la aceptabilidad. Es improbable que sólo un criterio operacional pueda llegar a ser suficiente para definir la noción de gramaticalidad.*

(Cáp. I de *Aspectos*)

el juicio como actuación

Paradoja de la lingüística (Householder 1973):

La única manera de corroborar si una gramática es adecuada es a partir de los juicios de gramaticalidad de los hablantes, pero estos juicios son inaccesibles.

La gramaticalidad debe ser considerada, de momento, una noción teórica que debe ser construida a partir de varios tipos de evidencia.

Esto complica muchísimo las cosas. Si la noción importante para determinar los datos, la gramaticalidad, es una construcción: ¿cómo se espera que haya consenso con respecto a qué es un dato y que no?

el juicio como actuación

Hay pocos procedimientos experimentales para obtener información sobre la intuición de los hablantes nativos. [...] Cuando un procedimiento es propuesto, debe ser analizado a partir de un estándar provisto por el conocimiento tácito que intenta especificar y describir. Si un procedimiento lograra satisfacer este requerimiento, podríamos confiar en él para decidir entre juicios complejos. Esto, sin embargo, es más una esperanza a futuro que una realidad presente.

*Sin embargo, el problema crítico hoy no es la escases de evidencia, sino la inadecuación de las teorías actuales para explicar grandes cantidades de evidencia. Me parece que mejorar la recolección de datos es de poca importancia para la problemática actual. **Tal vez llegue el día en que el tipo de dato que ahora tenemos en abundancia sea insuficiente para responder preguntas más profundas sobre la estructura del lenguaje.*** (Cáp. I de Aspectos)

el juicio como técnica informal

La recolección de datos informal lleva inevitablemente a ciertos problemas. Considérense los siguientes datos de correferencia hacia atrás:

(23) ?I knew him_i when Harvey_i was a little boy.

(24) ?We'll just have to fire him_i, wheter McIntosh_i likes ir or not.

(25) The boy who loves her_i claims that Mary_i is a genius.

Langacker (1969) dice que (23) y (24) son malas. Las compara con una oración buena como (25).

Reinhart (1972) dice que (23) y (24) son buenas. Las compara con una oración claramente mala.

el juicio como técnica informal

El hecho de que el método de recolección de evidencia no permita que los investigadores se pongan de acuerdo con respecto a la naturaleza de los datos es sumamente problemático.

Por cierto, Carden y Dietreich (1981) realizan un experimento y demuestran que oraciones como (23) y (24) son aceptadas en la misma medida que una oración completamente inaceptable (1 de cada 30 veces).

Langacker tenía razón.

el juicio como técnica informal

En otros casos, lo que se ignora es la gran variación entre los sujetos en los juicios de gramaticalidad.

Lasnik y Saito (1984) realizan grandes modificaciones sobre el ECP y proponen operaciones de borrado de complementante en FL con el fin de predecir la ambigüedad de oraciones como esta:

(26) Why do you think that he left?

Los autores asumen que este juicio es sistemático entre los hablantes de inglés.

el juicio como técnica informal

Sin embargo, Aoun et al. (1987) proponen un análisis alternativo en el mismo dominio prediciendo la agramaticalidad de muchos casos similares a (26).

Algunos hablantes alegan tener una interpretación de *why* dentro de la subordinada en (51a) [***Why did she say that there are men outside?***] incluso si el complementante está presente. Sin embargo, encontramos que cuando se les pide que repitan la oración, lo hacen omitiendo el *that*, como si no lo hubieran percibido. [...]. La *aceptabilidad* de (51a) no nos parece que se siga de su *gramaticalidad* a menos que adoptemos también como buenas (26)-(28) [e.g., ****Who remembers what we bought why?***] y similares; así, creemos que un proceso analítico extra está involucrado en este caso.

el juicio como técnica informal

En concreto, ¿cuáles son los problemas sobre esta discusión sobre el ECP?

- Lasnik y Saito no muestran señal de saber que sus juicios no son compartidos por toda la comunidad lingüística. Su explicación no hace ninguna referencia a por qué tantos hablantes encuentran incorrecta la extracción de *why*.
- Aoun et al. hace una aseveración muy fuerte a partir de un estudio muy poco riguroso basado en un número ínfimo de hablantes.

(Dato de Schütze (1996): según sus mediciones, 2/3 de los hablantes acepta la extracción de *why*).

resumiendo lo visto hoy

- Introdujimos los tipos de fuente de evidencia relevantes para el estudio de la gramática.
- Señalamos las ventajas y desventajas de utilizar corpus como fuente de datos.
- Introdujimos brevemente la problemática inherente a trabajar con juicios introspectivos en gramática.

Sólo unas palabras más...

un problema externo a la teoría

Como señala Schütze, cualquier comentario sobre el método de recolección de evidencias por parte de los lingüistas debe mencionar también el hecho de que muchas lenguas están desapareciendo.

- El 82% de las casi 7000 lenguas conocidas es hablado por poblaciones de menos de 100.000 habitantes. Y el 39% por poblaciones de menos de 10.000 hablantes (Evans y Levinson 2009).

un problema externo a la teoría

- El 8% de las lenguas del mundo ya está prácticamente extinto (Evans y Levinson 2009).
- Una lengua muere cada dos semanas (Evans y Levinson 2009).
- Entre el 20% y 50% de las lenguas que se hablan en el mundo hoy no están siendo aprendidas por una nueva generación (Krauss 1996).

Como señalan Levinson y Evans, esta pérdida de diversidad restringe drásticamente nuestra capacidad de entender científicamente qué hace posible el lenguaje humano.

algunas referencias

- Lo aquí presentado se basa en las ideas de un artículo de Schütze titulado *Linguistic Evidence and Grammatical Theory* (campus).
- La exposición sobre la introspección y los juicios de gramaticalidad se sigue de los capítulos 1 y 2 del libro de Schütze *The Empirical Base of Linguistics* (campus).
- El mencionado artículo de Brucart se encuentra en: <http://filcat.uab.cat/clt/publicacions/reports/pdf/GGT-02-4.pdf>