

Metodología de la investigación en lingüística: el análisis gramatical

noviembre de 2012

UNIDAD 6: GRAMÁTICA EXPERIMENTAL

Tareas para la recolección de juicios y
efectos de saciedad sintáctica

Carlos Muñoz Pérez

UBA

recordando algunas cosas

Rescatemos algunos conceptos antes de empezar con los contenidos de la unidad.

- Los juicios de gramaticalidad son la única fuente de evidencia negativa para los lingüistas.
- Muchos críticos de la GG observan una crisis en el marco teórico por la inadecuación de sus métodos de recolección de datos (introspección, intuición).
- Sin embargo, hay estudios que muestran que la aplicación de métodos de experimentación formal tendría poco impacto en la base empírica de la teoría lingüística.

recordando algunas cosas

A pesar de no modificar sustancialmente los resultados de una investigación, los datos experimentales tienen ciertas ventajas.

- Son «mejores» para datos en los que es inherente cierta ambigüedad o gradiencia (juicios difíciles).
- Son datos «menos debatibles», «más defendibles».
- Son el tipo de herramienta que hay que adoptar si queremos convencer a gente fuera de la lingüística de que nuestro trabajo es serio.

tareas de recolección de datos

Se puede hacer una mínima tipología de las tareas de recolección de datos en base a dos criterios.

- **Tareas no numéricas**
- **Tareas numéricas**

Las tareas no numéricas están pensadas para detectar diferencias cualitativas (e.g, esta oración es buena, esta oración es mala), pero pierden de vista la magnitud de la diferencia.

Las tareas numéricas, por el contrario, miden la magnitud de la diferencia, pero pierden de vista la diferencia cualitativa.

tarea de elección forzosa (EF)

La EF es una tarea no numérica que está explícitamente diseñada para comparar cualitativamente dos formas gramaticales.

Se les presentan a los participantes dos (o más) oraciones y se les pide que elijan cual es la más aceptable.

- Todos los ladrones fueron atrapados por la policía.
- Fueron todos los ladrones atrapados por la policía.



tarea de elección forzada (EF)

Ventajas de la EF

- Es una tarea muy fácil y rápida de presentar: no requiere entrenamiento de ningún tipo.
- Nos presenta de manera directa diferencias entre formas gramaticales.

Desventajas de la EF

- Proveen información cualitativa de forma muy indirecta (e.g., el 75% eligió la forma X).
- No sabemos en qué punto de la escala de aceptabilidad se encuentran las formas gramaticales estudiadas.

tarea de YES/NO

La prueba YN es una tarea no numérica que está diseñada para probar la relación entre una oración y dos categorías: *acceptable* e *inacceptable*. Nótese el contraste con la EF, en donde la relación se daba entre dos (o más) oraciones.

Se le pide a los participantes que evalúen si una determinada forma gramatical es acceptable o no.

- ¿Qué dijo Juan que quién compró? ☐ Si ☒ No

Recuerden que no todos los hablantes comparten las nociones de acceptable/inacceptable, ni es seguro que estos valores tengan una correlación gramatical.

Ventajas de la YN

- Es una tarea muy fácil y rápida de presentar: no requiere entrenamiento de ningún tipo.

Desventajas de la YN

- Es menos sensible que la EF para detectar diferencias/similitudes entre formas gramaticales (es necesario comparar la cantidad de respuestas positivas por cada oración presentada).
- No sabemos en qué punto de la escala de aceptabilidad se encuentran las formas gramaticales estudiadas.

tarea de Escala de Likert (EL)

En la tarea de EL a los participantes se les presenta una escala numérica cuyos extremos se definen como aceptable e inaceptable y se les pide que ubiquen cada oración en un punto de la escala.

Generalmente las escalas se diseñan con extremos impares (e.g., 1-5 o 1-7) para que haya un punto intermedio (3 o 4).

- Juan le parece a Luís ser un buen candidato.

☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5

tarea de Escala de Likert (EL)

Ventajas de la EL

- Es una tarea muy intuitiva.
- Permite ver las diferencias en la aceptabilidad de diversas formas gramaticales

Desventajas de la EL

- La escala sugiere que los intervalos son uniformes, pero no es posible saber si todos los hablantes consideran del mismo modo la diferencia entre 4 y 5 que la diferencia entre 2 y 3 (piénsenlo en términos de notas del 1 al 10). Esto puede reducirse si al principio del experimento se presentan ejemplos de oraciones con su valoración en la escala.

tarea de estimación de magnitudes (EM)

En la tarea de EM a los participantes se les brinda una oración de referencia y se les dice que la aceptabilidad de dicha oración tiene un valor numérico particular (e.g., 100). La oración de referencia recibe el nombre de *Estándar* y su valor numérico se denomina *Módulo*.

Los participantes deben asignar valores numéricos a nuevas oraciones en proporción al Módulo. Así, una oración que es "el doble de aceptable" que la Estándar recibe una puntuación de 200.

- Algunos niños juegan en el parque. 100
- Algunas trampas, los jugadores las conocen. 30

tarea de Estimación de Magnitudes (EM)

La EM fue especialmente diseñada para subsanar el problema de los intervalos no uniformes que plantean las EL.

Dado que la Estándar es la única unidad de medida (en tanto proporción) se supone que los intervalos resultantes son mucho más detallados que los de las EL.

Los grados de gramaticalidad son potencialmente infinitos.

OJO: el valor de la Estándar determina la «longitud» de la escala a ser utilizada.

ESTANDÁR = La pelota fue pateada por Juan. **100**

Vs.

ESTANDÁR = La pelota fue por Juan tenida. **100**

tarea de Estimación de Magnitudes (EM)

Desventajas de la EM

- A pesar de la ilimitada cantidad de posibles valores, los participantes se manejan con pocas variables.
- Tienden a elegir múltiplos de 5 y de 10, y eso no tiene nada que ver con el estímulo.
- Puede que toda la idea de las magnitudes variables esté, lisa y llanamente, mal...

¿lo qué?

La EM fue originalmente utilizada en *psicofísica*¹, y parece que ahí funciona bastante bien.

¹ Estudio de la relación entre la magnitud de un estímulo físico y la intensidad con que es percibido.

tarea de Estimación de Magnitudes (EM)

En psicofísica, también existe la *Producción de Magnitudes*, que sería la contrapartida de la EM.

Estimación de Magnitudes

- Algunos niños juegan en el parque. 100
- Algunas trampas, los jugadores las conocen. —

Producción de Magnitudes

- Algunos niños juegan en el parque. 100

Instrucción: ahora produzca una oración con módulo = 50

Las proporciones no parecen funcionar con la gramática.

Los participantes pueden estar tratando las EM como EL «tuneadas», de escala abierta.

la naturaleza común de las tareas

A pesar de diferir en el tipo de dato que arrojan, las cuatro tareas presentadas se basan en la misma actividad cognitiva: reportar una percepción de aceptabilidad gramatical.

Si bien existen variaciones en cuanto a las escalas que se obtienen para una misma forma gramatical, se ha demostrado que las cuatro pruebas arrojan resultados similares, sobre todo cuando el objetivo es rastrear diferencias entre estímulos (e.g., aceptabilidad).

diseño experimental

A continuación, les voy a presentar algunas pautas (muy) generales para la confección de experimentos.

Con *experimento* no me refiero solamente al formato propio de la psicología experimental, sino a cualquier marco controlado en el que se intenten obtener juicios de gramaticalidad (e.g., encuesta).

En particular, no voy a decir nada sobre cuestiones de estadística (básicamente porque no tengo idea).

Dos opciones: aprender estadística o hacerse amigo de alguien que sepa estadística.

No hay una manera estándar de confeccionar las instrucciones para un experimento con juicios de gramaticalidad. Sin embargo, siempre hay que aclararle a los participantes que hay ciertos aspectos de las oraciones que **NO** nos interesan, por lo que no deben afectar sus respuestas.

- Prescripción gramatical.
- Si la oración sería o no utilizada en el mundo real.
- Valores de verdad de la oración.
- Si es posible «entender» la oración por más que suene mal.

Una manera de controlar este último factor es estableciendo como consigna que lo que se entiende por *acceptable* es una oración que podría ser emitida por un hablante nativo.

Algo llamativo es que, a diferencias de otros tipos de experimentos, existe el consenso de que las instrucciones para este tipo de experimento no cambian alteran sustancialmente los resultados.

El cuidado en el diseño experimental debe caer en otros aspectos (materiales, estímulos, ítems de relleno, etc.).

Lamentablemente, estas mismas consideraciones hacen que no haya bibliografía específica para el diseño de instrucciones.

práctica previa al experimento

Las tareas de juicio gramatical son generalmente consideradas intuitivas para la mayor parte de los hablantes. La práctica del protocolo experimental sólo suele darse con las tareas numéricas (e.g., EL o EM).

En las tareas de *Escala de Likert* es común encontrar anclajes de ciertas formas gramaticales a ciertos puntos de la escala: un ejemplo del tipo de patrón sintáctico con menos o más valor en la escala (o ambos).

También suele darse una pequeña práctica sin dar aviso al participante en la que aparezcan construcciones distintas a las estudiadas y que lleven a ubicar valores a lo largo de toda la escala.

práctica previa al experimento

En tareas de *Estimación de Magnitudes* es normal incluir una fase de práctica en la cual los participantes entiendan el la lógica del experimento.

La idea de que se trabaja con proporciones se introduce a través de ayudas gráficas (e.g., una barra larga para 100, la mitad de la barra para 50, etc.).

Dado que la EM puede no ser otra cosa que una EL, no parece inadecuado incluir las pruebas de EL a las de EM.

diseño factorial

Si estamos llevando a cabo un experimento (y no simplemente pidiendo un juicio), lo más seguro es que nos estemos preguntando algo más complejo que simplemente evaluar la aceptabilidad o no aceptabilidad de un par mínimo.

En esos casos, es recomendable hacer un *diseño experimental factorial*.

Un *experimento factorial* mide dos (o más) factores de comportamiento de las unidades experimentales a partir de un conjunto de estímulos que cubre todas las posibles combinaciones de dichos factores.

Imaginemos que queremos estudiar la interacción del llamado *D-Linking* y las restricciones sobre la extracción de nominales complejos.

1. a. ¿Qué hiciste el comentario de que compró Juan ?
b. ¿Qué libro hiciste el comentario de que compró Juan ?

¿Podemos a partir de esto decir que el *D-Linking* mejora la aceptabilidad de las extracciones desde nominales complejos? **NO.**

Aun podría ser el caso de que el *D-Linking* mejore la aceptabilidad de ***todas*** las oraciones...

Para probar esto, necesitamos comparar oraciones que no contengan violaciones de nominales complejos.

- 2. a. ¿Qué comentaste que compró Juan?
- b. ¿Qué libro comentaste que compró Juan?

La pregunta ahora es si la diferencia entre (1a) y (1b) es igual a la diferencia entre (2a) y (2b). Esto nos dirá si el D-Linking mejora (i) sólo extracciones desde un nominal complejo (1b) o (ii) construcciones con extracción a nivel general (2b).

Las oraciones de (1) y (2) forman un experimento con diseño factorial porque hay dos factores en juego (tipo de extracción y tipo de elemento-wh) que tienen cada uno dos posibles valores (\pm isla y \pm D-Linking).

1. a. ¿Qué hiciste el comentario de que compró Juan ?
b. ¿Qué libro hiciste el comentario de que compró Juan ?
2. a. ¿Qué comentaste que compró Juan?
b. ¿Qué libro comentaste que compró Juan?

múltiples realizaciones léxicas

Cada forma gramatical que estudiemos debe estar representada por más de una realización a nivel léxico.

3. a. La pelota será pateada por el jugador.
- b. La policía fue arrestada por los ladrones.
- c. La computadora es reparada por un técnico.
- d. ...

A la vez, estas formas deben estar distribuidas entre los participantes, para que cuestiones como la plausibilidad pragmática o la contribución semántica de ciertas palabras afecten los resultados.

En la mayor parte de los experimentos es útil introducir ítems de relleno (i.e., oraciones que no se relacionan con la investigación). Estos cumplen tres objetivos.

- Sirven para evitar los efectos de *priming*, y para evitar que el participante note qué se está investigando.
- Sirven para garantizar una mejor distribución de las respuestas a lo largo de la escala.
- Sirven para meter estímulos con otros objetivos investigativos (experimento 2x1).

Seguimos en 15.

una mini encuesta

A ninguno de ellos los exámenes le angustian.

Levante la mano quien considere aceptable esta
oración.

Piénsenlo 15 segundos.

una mini encuesta

A nadie sus defectos le avergüenzan.

Levante la mano quien considere aceptable esta
oración.

Piénsenlo 15 segundos.

una mini encuesta

A ninguno de ellos el dinero le obsesiona.

Levante la mano quien considere aceptable esta
oración.

Piénsenlo 15 segundos.

una mini encuesta

A nadie la música le gustó.

Levante la mano quien considere aceptable esta
oración.

Piénsenlo 15 segundos.

una mini encuesta

*A nadie la música le gustó.

La música **no** le gustó **a nadie**.

A nadie le gustó la música.

el mal del lingüista

Todo gramático ha notado que luego de trabajar mucho con algún tipo de construcción se vuelve insensible a los juicios sobre ella.

El efecto se manifiesta como la progresiva aceptación de formas que suelen ser consideradas agramaticales.

Este efecto ha recibido varios nombres informales (e.g., *enfermedad del lingüista*), pero sólo recientemente (Snyder 2000) ha sido acuñado el término actualmente más difundido: *saciedad sintáctica* (syntactic satiation).

el mal del lingüista

Mientras los lingüistas suelen considerar este efecto como un *gaje del oficio*, su existencia ha sido utilizada por los críticos como una prueba de que los juicios de gramaticalidad no son una fuente fiable de datos.

Snyder (2000) responde: si la inestabilidad de los juicios *es una propiedad de los juicios gramaticales*, y no sólo un síntoma de la labor del lingüista, entonces los no lingüistas deberían manifestarla también.

Es más. Si mostramos que sólo una parte de las formas inaceptables produce saciedad, entonces la inestabilidad de los juicios *podría ser una propiedad de ciertas estructuras o restricciones*.

el mal del lingüista

Snyder (2000) llevó a cabo un experimento en el que que estudiantes sin formación en lingüística mostraron efectos de saciedad en *tres de siete tipos de violaciones sintácticas*.

Esto no sólo avaló sus hipótesis, sino que hizo de los efectos de saciedad sintáctica un área de estudio dentro de la teoría lingüística.

De manera interesante, Snyder sugirió también que las violaciones que producen *saciedad* y las que no se deben a factores distintos dentro de la Facultad del Lenguaje.

el mal del lingüista

En particular, Snyder sugirió que las violaciones *saciables* se deben a ***factores extragramaticales***, como restricciones de procesamiento o límite de recursos computacionales.

En definitiva, el hecho de que ciertas violaciones puedan ser «superadas con práctica» lleva a pensar en la distinción competencia y actuación.

Algunas violaciones de aceptabilidad tendrían un origen gramatical (y no pueden ser saciadas), mientras que otras construcciones tendrían una estructura gramatical bien formada pero serían inaceptables por limitaciones propias de los sistemas de actuación lingüísticos.

Los siete tipos de violación testeados por Snyder son los siguientes:

Table 1

Violations tested in Snyder 2000

Violation	Example sentence
Adjunct	Who did John talk with Mary after seeing?
Complex-NP	Who does Mary believe the claim that John likes?
Left-branch	How many did John buy books?
Subject	What does John know that a bottle of fell on the floor?
<i>That</i> -trace	Who does Mary think that likes John?
<i>Want-for</i>	Who does John want for Mary to meet?
<i>Whether</i>	Who does John wonder whether Mary likes?

Un problema que observa Sprouse (2009) es que los resultados de Snyder (2000) son muy difíciles de replicar.

Table 2

Summary of satiation results and the replication problem (✓ = significant effect, (✓) = marginal effect, — = not tested)

	Snyder 2000	Hiramatsu 2000	Goodall 2005	Sprouse 1	Sprouse 2	Sprouse 3
Adjunct						
Complex-NP	✓		✓			
Left-branch						
Subject	(✓)	✓				
<i>That</i> -trace		✓				—
<i>Want-for</i>		✓	—			—
<i>Whether</i>	✓	✓	—			

Lógica de Sprouse: si algunas violaciones producen saciedad en algunos experimentos y otras violaciones los producen en otros, hay que empezar a mirar más de cerca los diseños experimentales. Hay gato encerrado.

Diseño experimental de Snyder (2000)

- 50 oraciones presentadas en 5 bloques de 10 oraciones.
- Cada bloque contenía una oración con una de las 7 violaciones (tres ítems de relleno).
- En resumen: cada participante estuvo expuesto a 35 oraciones malas y 15 buenas.

¿Cómo se calculó la *saciedad* en el experimento de Snyder?

- Se contó el número de respuestas positivas para cada violación en los dos primeros bloques.
- Se contó el número de respuestas positivas para cada violación en los dos últimos bloques.
- Si el número de respuestas positivas aumentaba, había saciedad; si el número de respuestas positivas disminuía, no había saciedad.

el mal del lingüista

Sprouse (2009): ¡ahaha! Si así evaluaste los datos, tengo una explicación alternativa a estos aparentes efectos de saciedad, *ecualización*.

- Supongamos que a los participantes les llama la atención estar respondiendo el 70% del tiempo con **NO** y preferirían estar respondiendo de forma más pareja (50% **SI**, 50% **NO**).
- Al final del tercer bloque (antes de los dos bloques cruciales del experimento), los participantes escucharon 21 oraciones que llevaban al **NO** y sólo 9 oraciones que llevaban al **SI**.
- Si adoptan una *estrategia de ecualización*, van a tender a responder con más **SI**s en los próximos bloques.

¿Cómo probar que la estrategia de ecualización (y no la *saciedad sintáctica* como la entiende Snyder) es la explicación para los datos del experimento?

Observemos que la explicación basada en la *hipótesis de estrategia de ecualización (HEE)* requiere:

- Un diseño desbalanceado (e.g., 70% vs. 30%).
- Que los participantes puedan llevar control de lo que van respondiendo.
- Que los participantes puedan cambiar sus respuestas en función al control que van llevando.

¿Y que requiere una explicación basada en la *hipótesis de la saciedad sintáctica (HSS)*?

Nada. Tal y como la postula Snyder, la saciedad es una manifestación de la Facultad del Lenguaje.

Consideremos, entonces, algunas predicciones basadas en ambas hipótesis.

Problema de replicación de resultados.

Las fallas para replicar los datos reflejan, según la HEE, que los participantes no están utilizando todos la misma estrategia para responder. La HSS no explica bien esto.

el mal del lingüista

Tareas con pocas opciones y diseño desbalanceado.

Con pocas respuestas es más fácil llevar control. Esto y el diseño no balanceado son las condiciones para que actúe la HEE. Resultados significativos.

Tareas con opciones ilimitadas y diseño balanceado.

La HEE no se puede aplicar bajo estas condiciones. Los resultados deberían no ser significativos.

Tareas con pocas opciones y diseño rebalanceado.

Rebalanceado significa que hay un mayor número de formas aceptables que inaceptables. La HEE predice que los resultados no deberían ser significativos.

Básicamente porque las otras opciones ya se han testado, Sprouse (2009) va a probar su hipótesis (**EHH**) bajo estas condiciones.

Tareas con opciones ilimitadas y diseño balanceado.

La HEE no se puede aplicar bajo estas condiciones. Los resultados deberían no ser significativos.

Tareas con pocas opciones y diseño rebalanceado.

Rebalanceado significa que hay un mayor número de formas aceptables que inaceptables. La HEE predice que los resultados no deberían ser significativos.

Vayamos primero con la tarea con «opciones ilimitadas y diseño balanceado».

Para «balancear» los datos basta con poner tantos estímulos aceptables como inaceptables. Pero... ¿qué tarea de juicio implica opciones ilimitadas?

Estimación de Magnitudes

Reference sentence

What do you wonder whether Mary bought?

Acceptability: 100

Target sentence

What did Lisa meet the man that bought?

Acceptability: ____

el mal del lingüista

Table 3

Mean judgments (M) and standard deviations (SD) for each block in each experiment

	Subject		Adjunct		<i>Whether</i>		Complex-NP		Complex-NP with context	
	M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)
2 repetitions	−.14	(.24)	−.28	(.36)	.08	(.34)	−.01	(.28)	−.03	(.45)
4 repetitions	−.15	(.21)	−.32	(.37)	.12	(.35)	.01	(.28)	.08	(.44)
6 repetitions	−.07	(.30)	−.38	(.45)	.16	(.38)	−.04	(.27)	.01	(.41)
8 repetitions	−.11	(.21)	−.27	(.30)	.14	(.34)	−.01	(.36)	.05	(.40)
10 repetitions	−.09	(.22)	−.21	(.39)	.16	(.30)	−.07	(.36)	.08	(.38)
12 repetitions	−.09	(.22)	−.33	(.34)						
14 repetitions	−.10	(.27)	−.27	(.47)						

The lack of significant results indicates that there is no evidence of means or variances increasing over time, and thus no evidence of satiation with unlimited-response tasks and balanced designs. (Sprouse 2009: 336-337).

Vayamos ahora con el intento de probar las predicciones de la HEE para condiciones con *diseño rebalanceado* y opciones limitadas.

Dentro de este objetivo se llevaron a cabo dos experimentos.

- **Experimento 1:** un intento por replicar los datos de Snyder (2000). Ocho violaciones a lo largo de cinco bloques con dos ítems de relleno.
- **Experimento 2:** experimento con rebalanceo: 4 violaciones y seis ítems de relleno por bloque..

Vayamos ahora con el intento de probar las predicciones de la HEE para condiciones con *diseño rebalanceado* y opciones limitadas.

Dentro de este objetivo se llevaron a cabo dos experimentos.

- **Experimento 1:** un intento por replicar los datos de Snyder (2000). Ocho violaciones a lo largo de cinco bloques con dos ítems de relleno.
- **Experimento 2:** experimento con rebalanceo: 4 violaciones y seis ítems de relleno por bloque.

el mal del lingüista

Si bien en el *experimento 1* no se observó saciedad como la de Snyder, los datos resultan más *inestables* que los del *experimento 2*. En el *experimento 1*, 15 de 25 personas hicieron cambios de juicio; mientras que en el *experimento 2* sólo lo hicieron 2 de 19.

Table 6

Satiation results for experiments 1 and 2

	Experiment 1			Experiment 2		
	<i>No to yes</i>	<i>Yes to no</i>	<i>p value</i>	<i>No to yes</i>	<i>Yes to no</i>	<i>p value</i>
Adjunct	4	1	.38	0	0	1.0
Relative clause	3	4	1.0	2	0	.50
Complex-NP	1	1	1.0	0	0	1.0
<i>Whether</i>	2	3	1.0	0	2	.50
Coordinate structure	3	2	1.0	—	—	—
Sentential subject (infinitive)	3	0	.25	—	—	—
Sentential subject (finite)	3	1	.63	—	—	—
Left-branch	4	4	1.0	—	—	—

el mal del lingüista

El argumento de Sprouse parece ser bueno, pero nos deja sin explicación para explicar los efectos de saciedad que sí sufren los lingüísticas.

(y además tira por la borda las ideas de Snyder que, reconozcámoslo, estaban buenas).

El problema no está resuelto aún. En la actualidad se está discutiendo si el análisis de Sprouse es el adecuado, e incluso se utilizan los efectos de saciedad como diagnóstico para ciertas estructuras.

breve resumen

En la primera parte de la clase discutimos cuatro tareas para obtener juicios de gramaticalidad y algunas nociones básicas para el diseño de experimentos.

En esta segunda parte, se presentó y discutió la noción de *saciedad sintáctica*. Esto permitió que pongamos a prueba lo visto en la primera parte con respecto a la importancia del diseño experimental y las posibilidades de cada una de las tareas para obtener juicios.