PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

FACULTAD DE LETRAS

Métodos de Investigación en Letras

César Antonio Aguilar

Facultad de Lenguas y Letras

07/09/2017

caguilara@uc.cl

Síntesis de la clase pasada (1)

En la clase pasada vimos cómo podíamos organizar una investigación a partir del uso de corpus lingüísticos como fuente de datos. Así, lo que nos propusimos fue:

1. Formular primero las que serán nuestras preguntas de investigación, las cuales nos ayudan a delimitar el fenómeno lingüístico que queremos analizar, además de que van perfilando las hipótesis que guiarán nuestro trabajo.

2. Plantear nuestro conjunto de hipótesis, las cuales son las respuestas posibles que, conforme a nuestro conocimiento teórico, consideramos que pueden responderlas.

Síntesis de la clase pasada (2)

Para fijar nuestras hipótesis, establecemos un contraste entre **dos supuestos opuestos: uno que plantea que nuestra hipótesis es eficaz para explicar el fenómeno que estamos estudiando, la cual llamamos *hipótesis alternativa* o *de investigación*; y otra que niega la anterior, llamada *hipótesis nula*.**

Estableciendo un marco teórico (1)

Un paso importante que hay que seguir en una investigación es el desarrollo de lo que se conoce como **marco teórico**, que en concreto se refiere al **conjunto de conocimientos previamente adquiridos que soportan nuestras hipótesis.** Esto lo podemos entender mejor viendo la siguiente gráfica:

Estableciendo un marco teórico (2)

Ahora bien, incluir un marco teórico en nuestra investigación es muy bueno, básicamente por las siguientes razones:

Estableciendo un marco teórico (3)

A las 6 funciones que hemos señalado antes, hay un factor más que deben tomar en cuenta: si su investigación requiere de un patrocinio para llevarse a cabo, hay distintas instituciones a las cuales pueden acudir, tanto públicas como privadas…, pero en todos los casos, **siempre hay un comité científico que evaluará su propuesta:**

National Science Foundation
WHERE DISCOVERIES BEGIN

CONICYT
Ministerio de Educación
Gobierno de Chile

CONICYT
Comisión Nacional de Investigación Científica y Tecnológica

CONACYT
Consejo Nacional de Ciencia y Tecnología

En América, países como Estados Unidos, Chile, México y otros más, cuentan con instituciones gubernamentales que, bajo la conducción de comités científicos (integrado por investigadores y profesores importantes para un área de conocimiento), quienes evalúan propuestas de investigación, ponderando sobre todo qué tan viable es la solución de un problema **conforme al marco teórico que el investigador propone.**

Estableciendo un marco teórico (4)

Por poner un ejemplo, en el caso de **Fondo Nacional de Desarrollo Científico y Tecnológico, FONDECYT** (www.conicyt.cl/fondecyt), si van a concursar por algún presupuesto, se les pide explícitamente en sus formatos de participación el que desarrollen un **marco teórico**, acompañado de una revisión bibliográfica**:**

Tenga presente lo estipulado en las Bases del Concurso FONDECYT de Iniciación en Investigación 2013 e Instrucciones para postular.
Este archivo debe contener un máximo de **17 páginas,** incluyendo secciones obligatorias y opcionales. Secciones obligatorias: Formulación del proyecto, Marco Teórico y Discusión Bibliográfica, Referencias Bibliográficas, Hipótesis, Objetivos, Metodología, Plan de Trabajo, Difusión del proyecto a la sociedad.
Se recomienda mantener el formato tamaño carta, fuente Verdana tamaño 10 y extensión sugerida para cada sección.

III. FORMULACIÓN DEL PROYECTO, MARCO TEÓRICO Y DISCUSIÓN BIBLIOGRÁFICA:

Hoy en día, debido al incremento exponencial de datos contenidos en enormes repositorios de documentos, el área de biomedicina requiere tiene un gran interés en el desarrollo de técnicas, métodos y herramientas que le ayuden a identificar, extraer, integrar y clasificar dichos datos desde tales repositorios, con miras a obtener conocimientos relevantes sobre hechos tan diversos como los factores que intervienen en una enfermedad, la cantidad de pacientes que muestran un conjunto de síntomas, el nombre de un virus identificado con diferentes nomenclaturas, etc.
Para resolver parte de estas necesidades, la Biblioteca Nacional de Medicina de Estados Unidos (ingl. National Library of Medicine, o NLM) desarrolló en 1977 la base de datos bibliográfica MedLine, la cual cuenta con un total aproximado de 21 millones de citas y referencias a artículos provenientes de 4.500 revistas médicas publicadas en los Estados Unidos y en más de 70 países (Ananiadou, Kell y Tsujii 2006; Sainani 2008). Ahora bien, para acceder a tal repositorio, el NLM ha desarrollado el motor de consulta PubMed, el cual brinda acceso gratuito a los datos contenidos en MedLine.

Estableciendo un marco teórico (5)

Otro ejemplo más cercano: muchas convocatorias que lanza nuestra universidad, p. e., los llamados **concursos puentes:**

Para más detalles, pueden consultar la página: http://investigacion.uc.cl/Fondos-concursables/concursos-vri.html

Estableciendo un marco teórico (6)

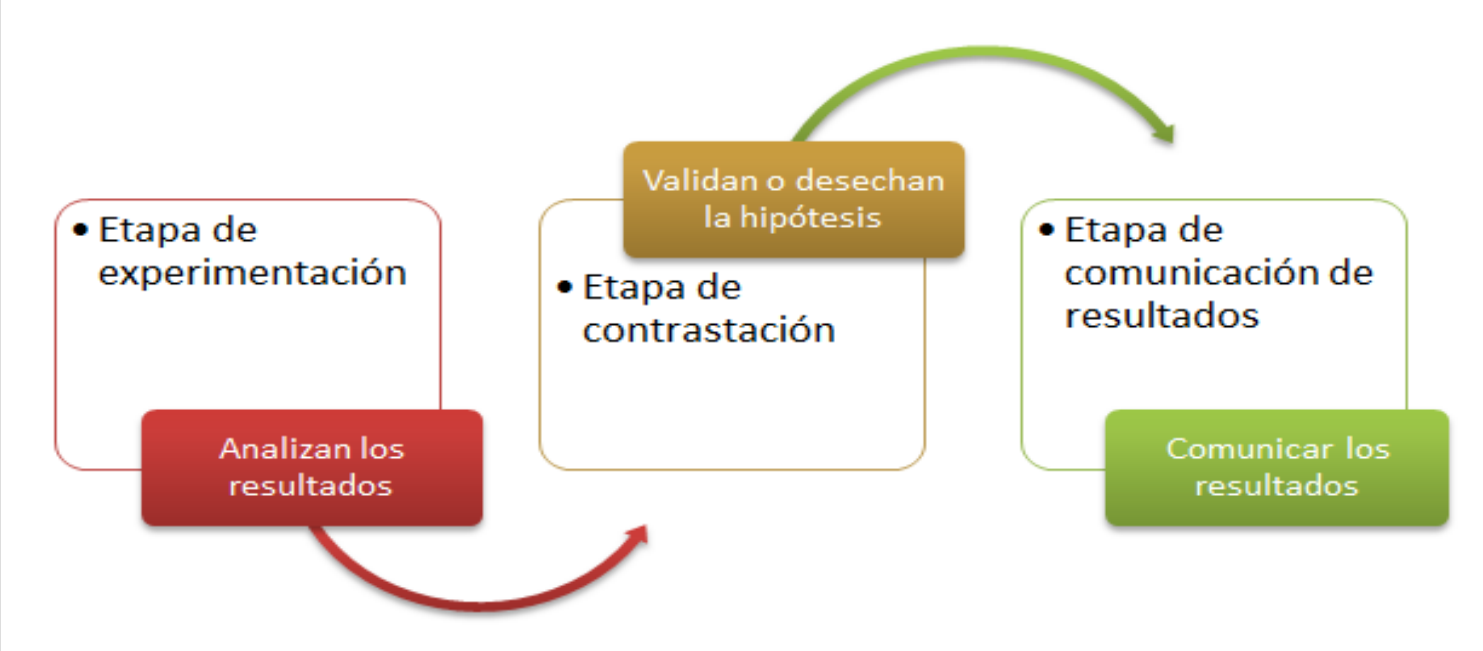
La elaboración de un marco teórico considera el cumplimiento de los siguientes pasos:

Como podrán darse cuenta, la base más importante para configurar un marco teórico es la **lectura** de todo el material bibliográfico que puedan revisar, relacionado con su problema de investigación.

Experimentación en lingüística (1)

En una investigación, el punto en donde se concilian el marco teórico junto con la metodología es el proceso de **experimentación**.

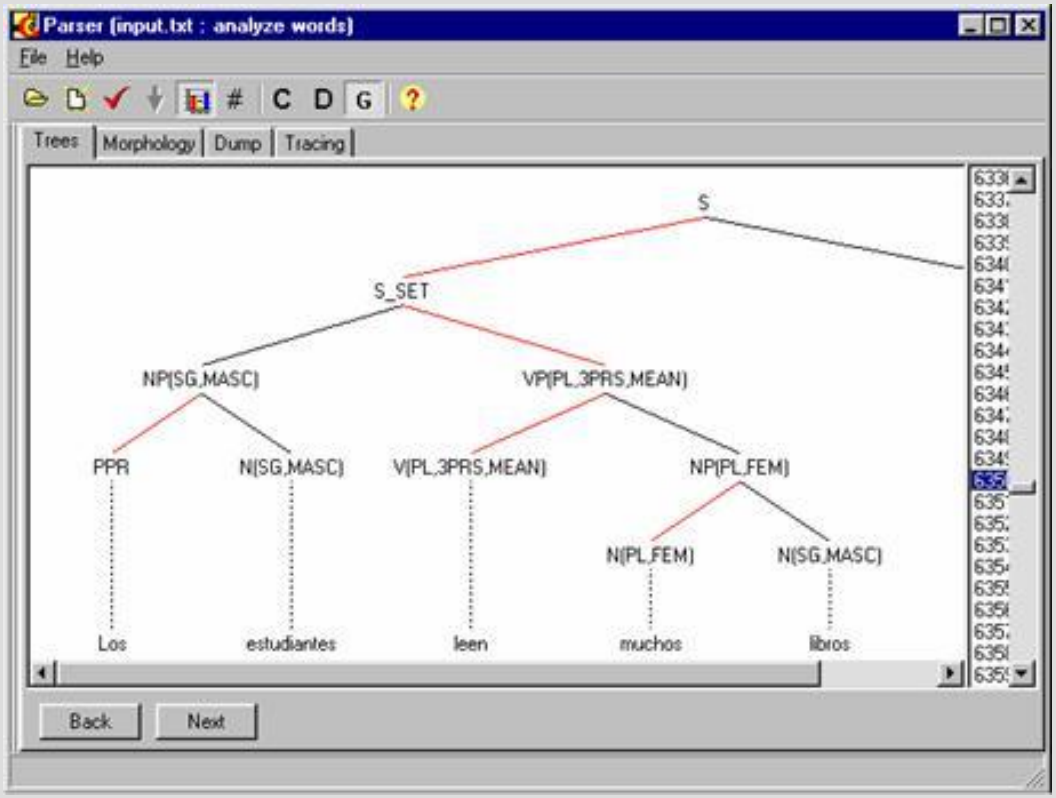
En lingüística, como en otras ciencias, la experimentación puede ser vista como un conjunto de acciones que nos permiten poner en relieve los supuestos de los que partimos con los datos que generamos al analizar el fenómeno que estamos estudiando.



Experimentación en lingüística (2)

La mayoría de los lingüistas aplican 3 tipos de criterios para desarrollar sus experimentos: (i) uno basado en la colecta y contraste de **ejemplos reales**; (ii) otro basado en la colecta de información a partir del uso de encuestas y estudios de tipo cualitativo; y finalmente (iii), los que consideran la aplicación de métodos cuantitativos, sobre todo basados en la investigación de corpus.

Hoy en día, de hecho, estos criterios se pueden combinar, lo que da lugar al desarrollo de **métodos híbridos**. Tales métodos son muy usados en lingüística computacional.

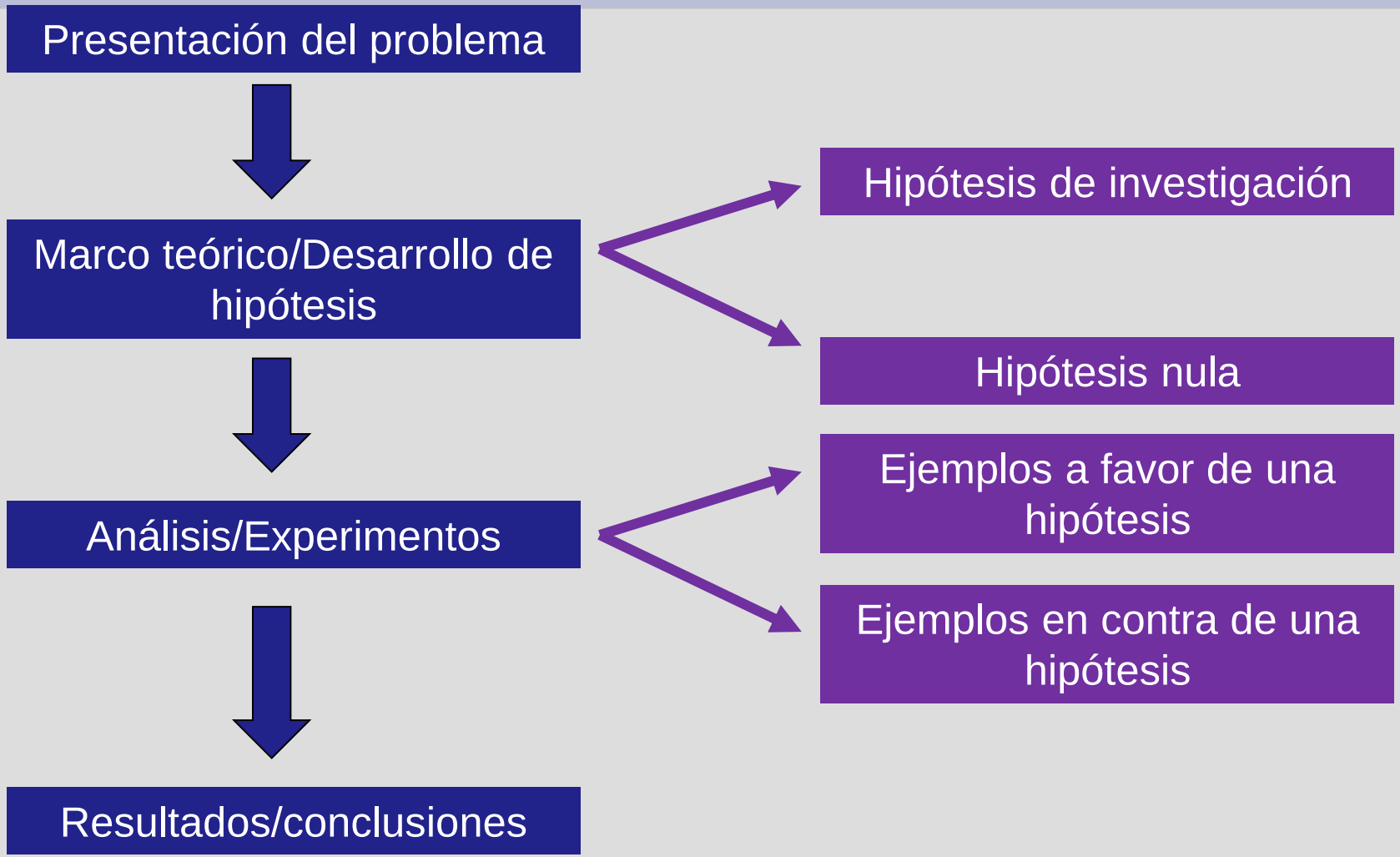


Experimentación en lingüística (3)

En el caso del primer criterio, se emplea mucho en trabajos derivados de la escuela generativista, y consiste en buscar, seleccionar y evaluar ejemplos concretos de secuencias lingüísticas (sean morfemas, palabras, frases u oraciones), cuya estructuración sea explicable a partir de una hipótesis determinada:

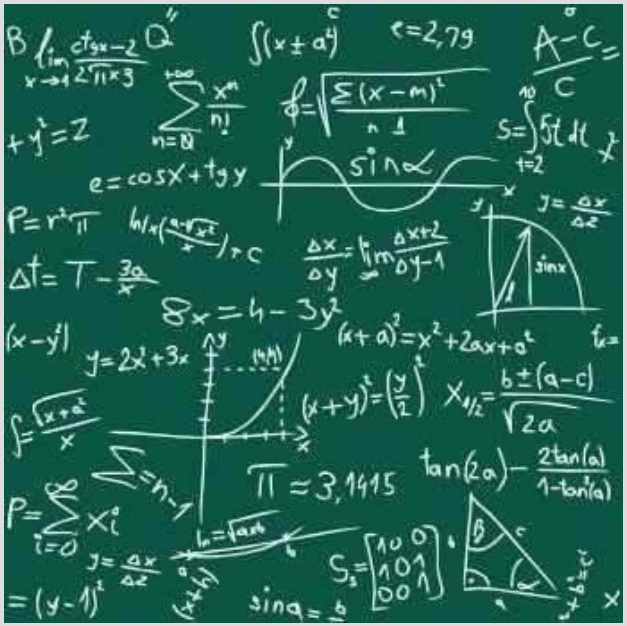
Experimentación en lingüística (4)

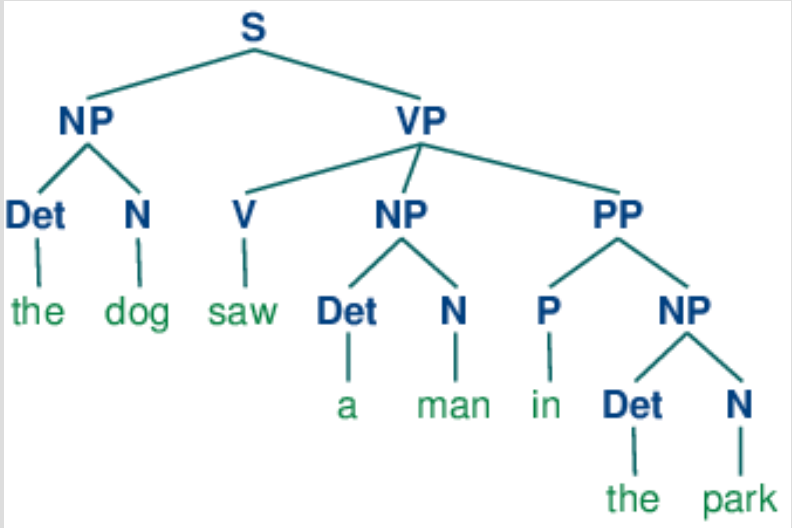
Por lo general, la forma de presentar esta clase de trabajos es la siguiente:



Experimentación en lingüística (5)

Lo que se pone en práctica con esta clase de ejemplos es un ejercicio de **introspección**, en donde el lingüista trabaja de un modo similar a lo que hacen los matemáticos o los físicos: se formulan una serie de cuestionamientos sobre un fenómeno determinado, desarrollan un modelo teórico que explica su generación, posteriormente desarrollan las reglas que subyacen en él, para terminar evaluando es su modelo considerando casos hipotéticos que ponen a prueba tales reglas (o **axiomas**).





Experimentación en lingüística (6)

Otro criterio diferente es el uso de **encuestas** para colectar datos. A grandes rasgos, una encuesta se puede definir del siguiente modo:

La **encuesta** es una búsqueda sistemática de información en la que el **investigador pregunta a los investigados** sobre los **datos** que desea obtener, y posteriormente reúne estos datos individuales para obtener durante la evaluación datos agregados.

La encuesta realiza a todos los entrevistados las mismas preguntas, en el mismo orden, y en una situación similar; de modo que **las diferencias son atribuibles a las diferencias entre las personas entrevistadas**.

Experimentación en lingüística (7)

Algunos tipos de encuestas son las siguientes:

Según la finalidad

Según su contenido

Según el procedimiento de encuesta

Según su dimensión temporal

Finalidad:

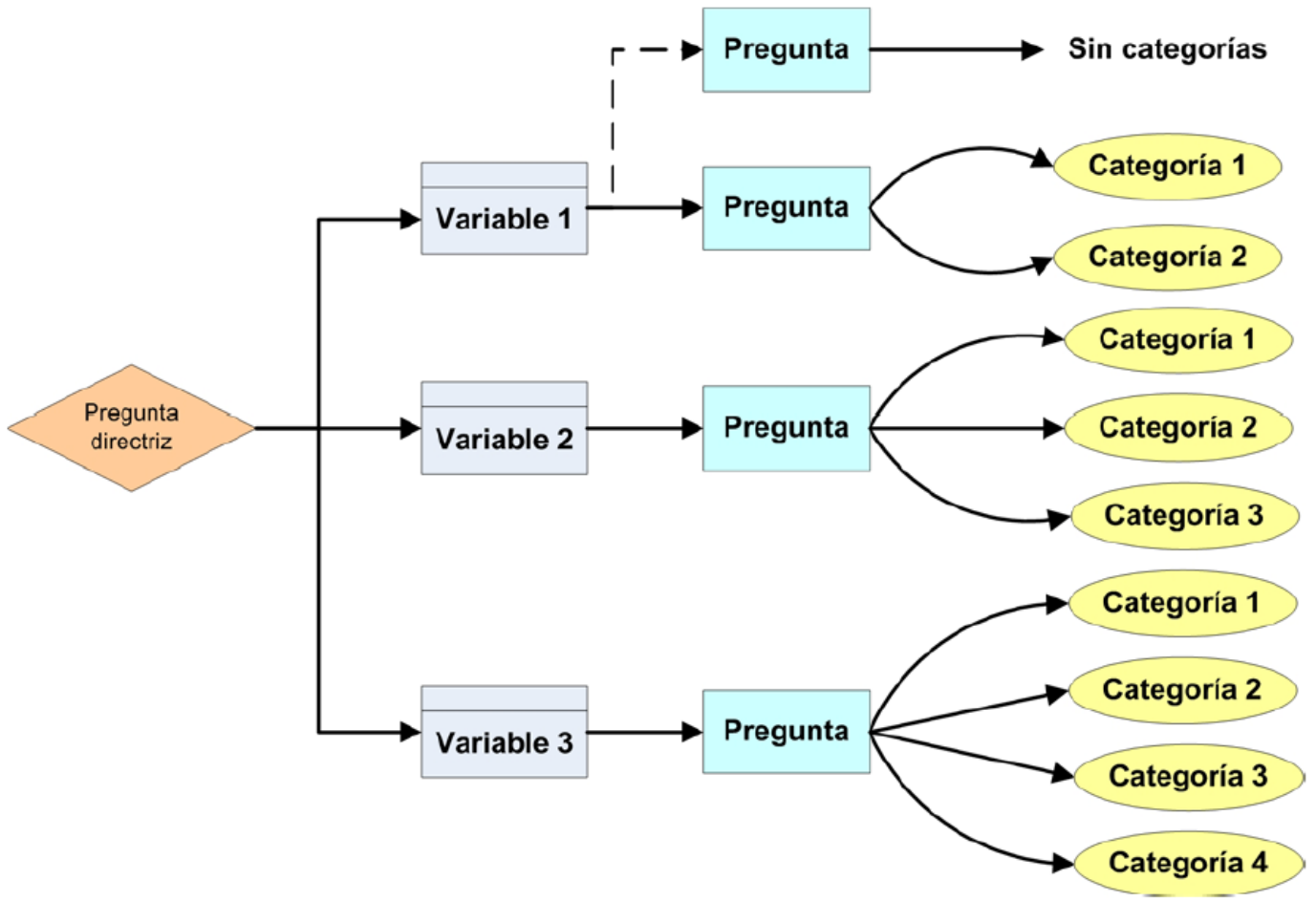
- Exploratoria
- Descriptiva
- Explicativa
- Predictiva
- Evaluativa

Contenido:

- Hechos
- Opiniones
- Actitudes o sentimientos

Experimentación en lingüística (8)

Para elaborar las preguntas que van a ser parte de nuestra encuesta, debemos de considerar las hipótesis que vamos a analizar, con miras a determinar cuáles serán nuestras **directrices**, así como sus posibles variantes, p. e.:



Experimentación en lingüística (9)

Dentro de un cuestionario, una pregunta puede contar con las siguientes características:

Estructura del cuestionario (la pregunta)

- Es el **elemento principal** de la entrevista
- Existen diversos **tipos**:
  - **Pregunta abierta o no estructurada**:
    - En la respuesta el **entrevistado puede expresarse libremente**
    - Tiene la desventaja de ser **difícil de codificar**.
    - Ejemplo: ¿Qué entiende por ecomodulación?
    - Son **costosas** de preguntar, codificar, tabular y analizar.
    - Deben ser **usadas con moderación**.
    - Los resultados dependen mucho de la calidad de las entrevistas y la codificación.

Experimentación en lingüística (10)

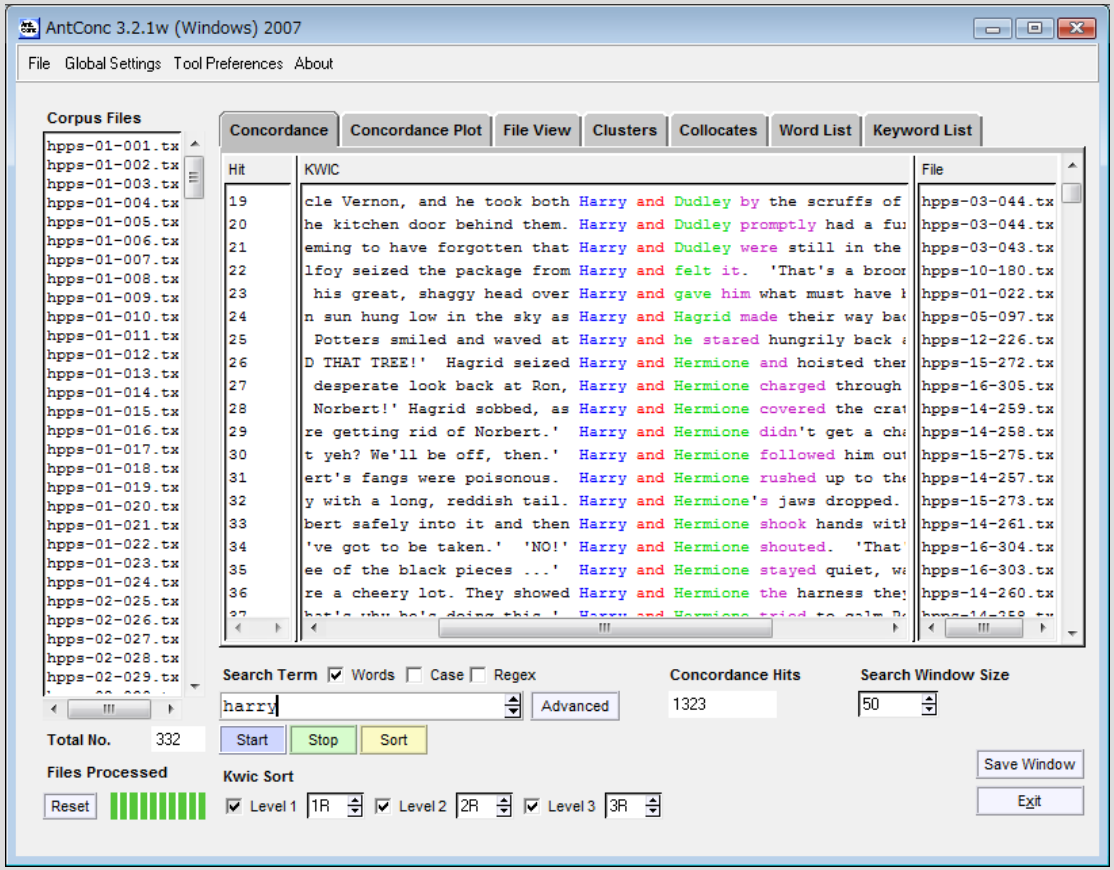
No todas las preguntas tienen que tan abiertas o permitir más de una sola respuesta. Se pueden plantear preguntas más estructuradas que orienten al entrevistado a dar una respuesta única, o en todo caso, se le pueden ofrecer un conjunto de **opciones múltiples** que le brinden un margen de libertad para elegir su respuesta:

- Se **mencionan** al entrevistado **las posibles opciones** de respuesta prefijada.
- Pueden ser **dicotómicas** (2 opciones) o en **abanico** (3 o más opciones)
  - Ejemplo Dicotómica: ¿Recibe usted ayudas PAC? 1)Si, 2)No
  - Asegurarse de que la pregunta en realidad tiene solo dos respuestas.
  - "No se" o "no responde" son también respuestas legítimas.
  - Si hay más de dos posibles respuestas, inclúyalas como alternativas, si esto facilita que el encuestado conteste.
  - Ejemplo Abanico: ¿A qué institución acude usted para tramitar las ayudas? 1)A nadie, 2)Asesor, 3)Sindicato, 4)Organismo certificador, 5)No sabe

Experimentación en lingüística (11)

Finalmente, en el caso de la investigación en corpus, como ustedes lo han visto, el diseño de experimentos consiste en el desarrollo de estrategias y herramientas que faciliten la búsqueda de patrones determinados que sean representativos de algún fenómeno lingüístico determinado.

Estos patrones pueden ser identificados haciendo uso de las llamadas **concordancias**, que son secuencias de palabras asociadas de forma recurrente, las cuales pueden reflejar fenómenos lingüísticos específicos.



Experimentación en lingüística (12)

Una de las ventajas que aportan los corpus lingüísticos es que permiten combinar criterios cualitativos y cuantitativos, e igualmente tener en cuenta un marco teórico que guie nuestra investigación.

```
>>> cfd = nltk.ConditionalFreqDist(
...             (genre, word)
...             for genre in brown.categories()
...             for word in brown.words(categories=genre))
>>> genres = ['news', 'religion', 'hobbies', 'science_fiction', 'romance', 'humor']
>>> modals = ['can', 'could', 'may', 'might', 'must', 'will']
>>> cfd.tabulate(conditions=genres, samples=modals)
                 can could  may might must will
           news   93   86   66    38   50  389
       religion   82   59   78    12   54   71
        hobbies  268   58  131    22   83  264
science_fiction   16   49    4    12    8   16
        romance   74  193   11    51   45   43
          humor   16   30    8     8    9   13
```

Gracias al uso de corpus lingüísticos, en los últimos años se ha consolidado el uso de herramientas computacionales para analizar estas colecciones de datos, dada las cantidades de palabras que contienen.

Experimentación en lingüística (13)

Asimismo, el uso de métodos estadísticos es otro aporte relevante de la lingüística de corpus, los cuales permiten validar la pertinencia o no de una hipótesis a partir de su evidencia empírica, contabilizando todos aquellos datos que se correspondan con dicha hipótesis:

$$P^*(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n)+1}{C(w_{n-1})+V}$$

	i	want	to	eat	chinese	food	lunch	spend
i	0.0015	0.21	0.00025	0.0025	0.00025	0.00025	0.00025	0.00075
want	0.0013	0.00042	0.26	0.00084	0.0029	0.0029	0.0025	0.00084
to	0.00078	0.00026	0.0013	0.18	0.00078	0.00026	0.0018	0.055
eat	0.00046	0.00046	0.0014	0.00046	0.0078	0.0014	0.02	0.00046
chinese	0.0012	0.00062	0.00062	0.00062	0.00062	0.052	0.0012	0.00062
food	0.0063	0.00039	0.0063	0.00039	0.00079	0.002	0.00039	0.00039
lunch	0.0017	0.00056	0.00056	0.00056	0.00056	0.0011	0.00056	0.00056
spend	0.0012	0.00058	0.0012	0.00058	0.00058	0.00058	0.00058	0.00058

Análisis como los que ven en la imagen, la cual representa un modelo de **n-gramas**, nos ayudan no solamente a hacer cálculos de frecuencias, sino también a predecir qué tan probable o no es que ocurra un fenómeno determinado en un corpus nuevo o diferente al que analizamos.

Gracias por su atención

Blog del curso:
http://cesaraguilar.weebly.com/presentacioacuten.html