

2 Diseño de informe y tipos de datos

Métodos empíricos 2

19/04/2022

Hoy

- Componentes de un análisis
- Datos
- Vocabulario de diseño experimental
- Formatos y manipulación de datos

Componentes de análisis

Componentes

- Pregunta(s) del análisis
- Plan de diseño
- Plan de muestreo (sampling plan)
- Especificación de variables
- Plan de análisis

Componentes

- **Pregunta(s) del análisis**
 - Pueden variar en grado de granularidad
 - Pueden variar en naturaleza: Análisis confirmatorio vs. exploratorio
- Plan de diseño
- Plan de muestreo (sampling plan)
- Especificación de variables
- Plan de análisis

Componentes

- Pregunta(s) del análisis
- **Plan de diseño**
 - Descripción general del análisis.
 - Entre otros, especifica qué medirás, y bajo que circunstancias
- Plan de muestreo (sampling plan)
- Especificación de variables
- Plan de análisis

Componentes

- Pregunta(s) del análisis
- Plan de diseño
- **Plan de muestreo (sampling plan)**
 - Cuántas observaciones recolectarás? Cómo? Por qué?
- Especificación de variables
- Plan de análisis

Componentes

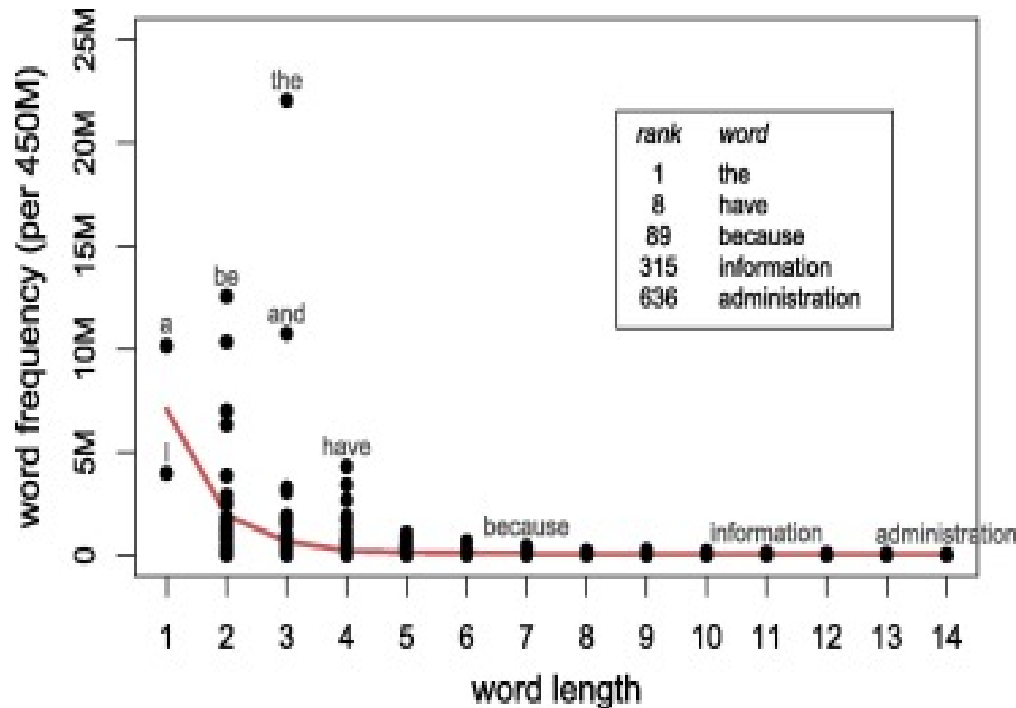
- Pregunta(s) del análisis
- Plan de diseño
- Plan de muestreo (sampling plan)
- **Especificación de variables**
 - Descripción todas las variables que juegan un rol en tu análisis
- Plan de análisis

Componentes

- Pregunta(s) del análisis
- Plan de diseño
- Plan de muestreo (sampling plan)
- Especificación de variables
- **Plan de análisis**
 - Descripción de cómo analizarás los datos

Zipf's Law of Abbreviation

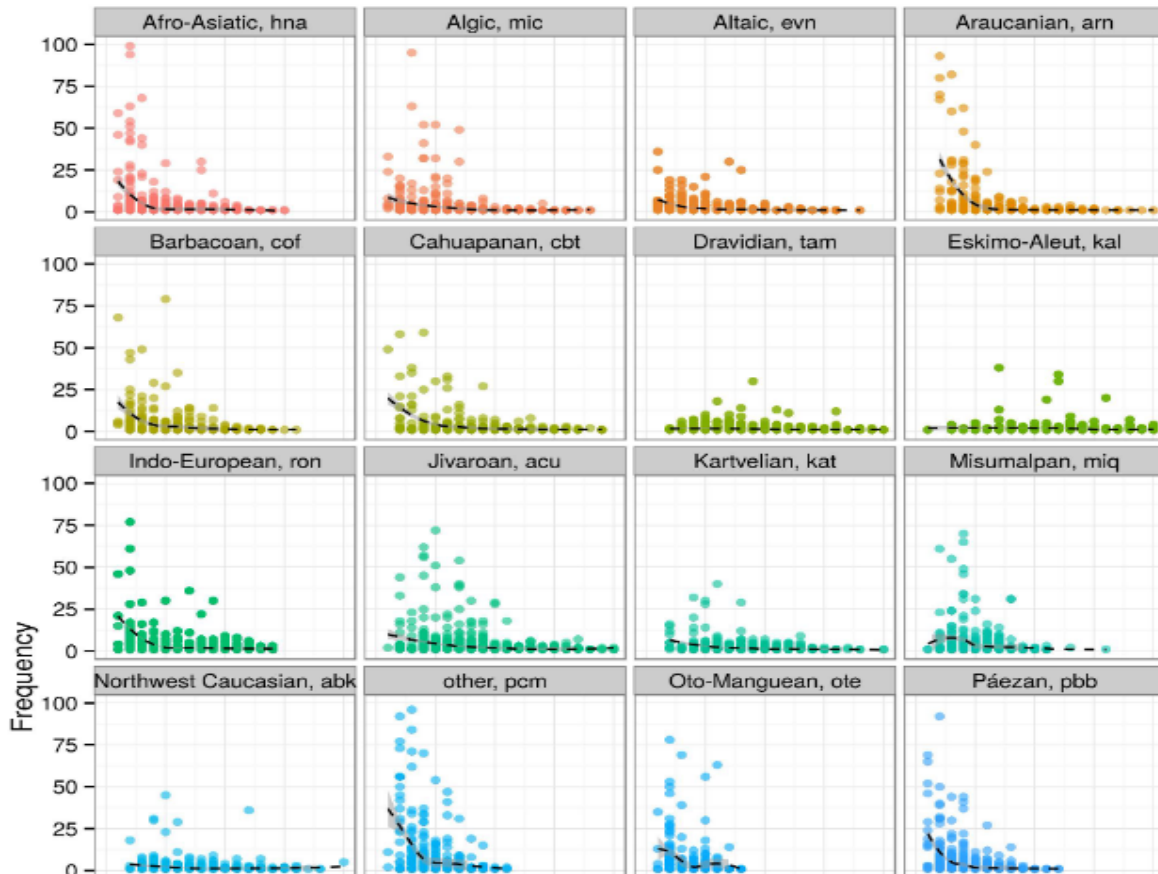
Zipf's Law of Abbreviation: Formas más frecuentes tienden a ser más cortas



G.K. Zipf (1935) *The psycho-biology of language*

Kanwal et al. (2017) *Zipf's Law of abbreviation as a language universal*

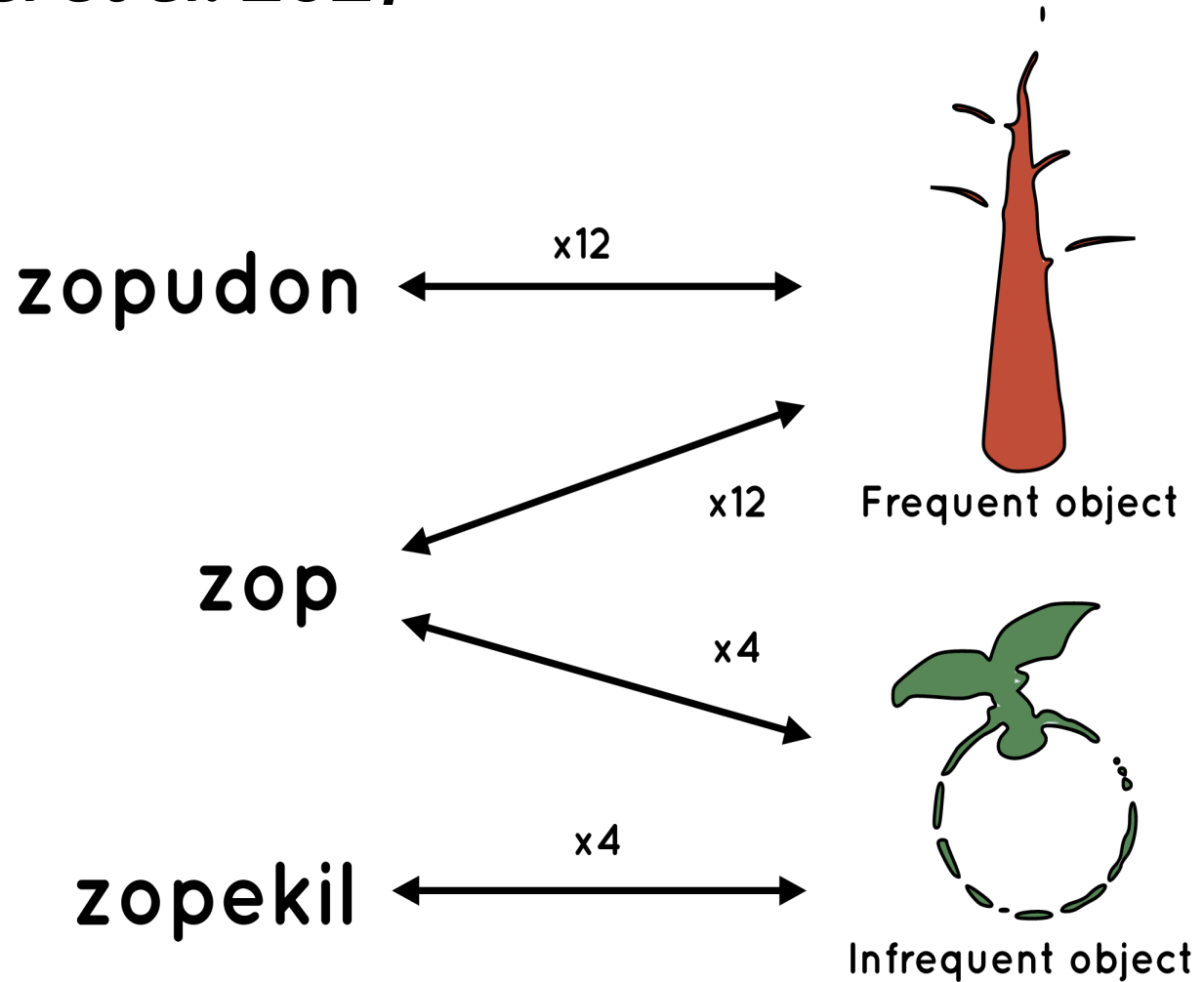
Zipf's Law of Abbreviation: Formas más frecuentes tienden a ser más cortas



G.K. Zipf (1935) *The psycho-biology of language*

Benz & Ferrer-i-Cancho (2016) *Zipf's Law of abbreviation as a language universal*

Kanwal et al. 2017



Kanwal et al. 2017: Condiciones

1. Comunicación con otro jugador, con diferencial de tiempo entre mensajes
2. Comunicación con otro jugador, sin diferencial de tiempo
3. Sin otro jugador, con diferencial de tiempo entre mensajes
4. Sin otro jugador, sin diferencial de tiempo

Kanwal et al. 2017: Resultado

Sólo con diferencial de tiempo entre mensajes:

mensaje más corto \Rightarrow significado más frecuente

Componentes (Kanwal et al. 2017)

- Pregunta(s) del análisis
- Plan de diseño
- Plan de muestreo (sampling plan)
- Especificación de variables
- Plan de análisis

Componentes (Kanwal et al. 2017)

- **Pregunta(s) del análisis**
 1. Emerge la ley de abreviación de una interacción entre dos presiones lingüísticas: brevedad y comunicación?
 2. Emerge (1) en un dialogo artificial de laboratorio en un par de rondas?
 3. Qué aprendemos de (1), si (2) es correcta?
- Plan de diseño
- Plan de muestreo (sampling plan)
- Especificación de variables
- Plan de análisis

Componentes (Kanwal et al. 2017)

- Pregunta(s) del análisis
- **Plan de diseño**
 - Analizaremos si un uso de expresiones referenciales que refleja la ley de abreviación de Zipf (más frecuente → más corta) emerge en función a una interacción y presencia de dos presiones lingüísticas: presión por comunicación y por brevedad.
 - Controlaremos la presencia/ausencia de las presiones en un experimento de laboratorio, usando una lengua artificial para también controlar el inventario léxico a disposición de cada hablante
 - Habrán 4 condiciones: $[\pm \textit{comunicación}] \times [\pm \textit{brevedad}]$
 - Esperamos que uso de expresiones referenciales en línea con la ley de abreviación de Zipf sólo emerja si ambas presiones están presentes: $[+\textit{comunicación}] \times [+\textit{brevedad}]$
- Plan de muestreo (sampling plan)
- Especificación de variables

Componentes (Kanwal et al. 2017)

- Pregunta(s) del análisis
- Plan de diseño
- **Plan de muestreo (sampling plan)**
 - 124 participantes: 31 por condición
 - Reclutados a través de MTurk
 - Criterios de exclusión
 - ...
- Especificación de variables
- Plan de análisis

Componentes (Kanwal et al. 2017)

- Pregunta(s) del análisis
- Plan de diseño
- Plan de muestreo (sampling plan)
- **Especificación de variables**
 - Uso de expresión corta
 - Referente a comunicar
 - ID de sujeto y par
 - Turno
- Plan de análisis

Componentes (Kanwal et al. 2017)

- Pregunta(s) del análisis
- Plan de diseño
- Plan de muestreo (sampling plan)
- Especificación de variables
- **Plan de análisis**
 - Regresión logística con expresión como variable dependiente, y frecuencia del referente como variable independiente

Datos y terminología

Tipos de estudios

- Observacionales
- Experimentales
- Virtuales (simulación, predicción)

Tipos de variables

- **Nominales** (género; idioma materno; lugar de origen)
- **Ordinales** (años cursados; edad)
- **Binarias**
- **Métricas** (número de usos de pronombres; tono)

Representación de datos

- Cada **variable** forma una columna
- Cada **observación** forma una línea
- (Cada **típo de unidad observacional** forma una tabla)

"Tidy representation"

Wickham, Hadley. 2014. "Tidy Data." Journal of Statistical Software 59 (10). Foundation for Open Access Statistic. <https://doi.org/10.18637/jss.v059.i10>.

Próxima sesión

- Entrega de "Assignment 2" (08:00 AM 26/04)
-

- **Recolección de datos y muestras**