

Peer review (Part 1)

1. What is your general research question?

Sabemos que la Ley de Abreviación de Zipf establece una relación entre la frecuencia de aparición de diferentes palabras. Nuestra investigación pretende determinar si la Ley de Zipf se cumple, independientemente del público al que se dirija un texto o discurso. Esto se realizaría mediante un análisis estadístico de textos literarios, concretamente, de obras escritas en inglés dirigidas a dos públicos diferentes (juvenil y adulto). A simple vista, no parece lógico pensar que esta ley se cumpla de la misma forma en los dos tipos de texto. De hecho, aparentemente, podríamos suponer que esta ley no se cumple en determinados textos.

2. Why do you think this question is interesting? What does an answer to it tell us?

La obtención de resultados de la investigación, principalmente, nos serviría para comprobar si la Ley de Zipf se cumple en textos, independientemente de su tipología textual. Si estos resultados son afirmativos, podrían derivar a diversas conclusiones interesantes para la lingüística y el aprendizaje de lenguas. Por un lado, si conseguimos extraer el léxico más frecuente, nos aproximaría a la creación de un corpus textual, el cual que nos puede servir como vocabulario esencial para el aprendizaje de la lengua inglesa. Incluso se podría aplicar a otras lenguas extranjeras. Por otro lado, los resultados del análisis podrían servir como reflexión sobre la naturaleza del lenguaje humano, sobre el funcionamiento de la producción de textos, e incluso servirían como datos para generar otros estudios relacionados con la lingüística aplicada.

3. What is your specific research question?

¿Se cumple la Ley de Abreviación de Zipf de la misma manera en textos de literatura inglesa tanto para jóvenes como para adultos? Una aproximación de las palabras más frecuentes en obras literarias y comparación de patrones de frecuencia.

4. What kind of data would you use to address Question 2 if you had unlimited resources?

Se emplearían transcripciones de obras originales y sus adaptaciones al público juvenil, y se generaría un corpus textual con las palabras más frecuentes. También se usaría el lenguaje R para el análisis de distribución de frecuencia.

5. What kind of data are you planning to use to address Question 2 within the scope of this class? How will you obtain it? How much will you collect? Do you think that is enough data to address Question 2? Why (not)?

Emplearía transcripciones en diferentes formatos textuales (PDF, txt, rtf, odt, doc-docx, wps, html, etc.) y el lenguaje R mediante Google Collab para el análisis estadístico. Analizaría un mínimo de 20 obras y un máximo de 100. Un mínimo de 20 obras para asegurar una heterogeneidad de obras que permita una estimación pertinente, y un máximo de 100 obras para que la investigación no se exceda con tanta información y esfuerzos innecesarios.