



Aracne: el observatorio del lenguaje periodístico del siglo XX en España

Leticia Martín-Fuertes Moreno
Lingüista en Molino de ideas
@nimbusaeta
leticia@molinodeideas.es

14 de noviembre de 2015
#EOL2



Proyecto



Objetivo del proyecto

Observar los cambios en los textos periodísticos desde 1914 hasta 2014 para analizar **cómo ha evolucionado el lenguaje** de la prensa española a lo largo del último siglo.

Responder a preguntas como: ¿qué palabras son las **más frecuentes**? ¿Qué términos han **aparecido** y cuáles han **desaparecido**? ¿Han proliferado los **eufemismos**? ¿Se está **empobreciendo** el vocabulario? ¿Cuáles han sido los **conceptos** más relevantes y cómo han cambiado?

Cómo funciona

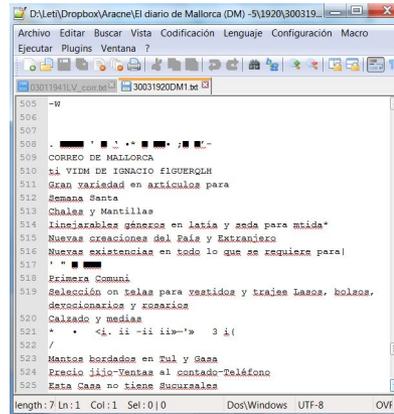
1

Limpieza

A

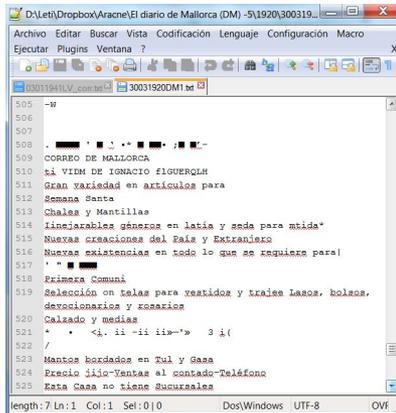


PDF



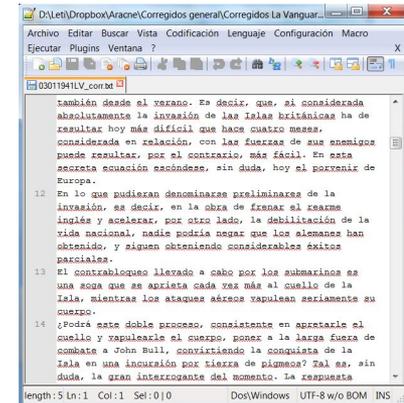
Texto generado por OCR

B



Texto generado por OCR

Aracne

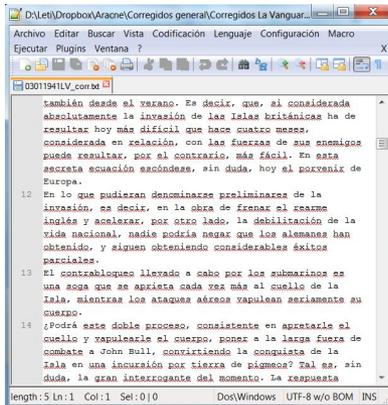


Texto plano corregido

2

Cómo funciona

Fontanería de recopilación de datos



Un programa trocea cada archivo en artículos. En todo momento queda guardado el año, el mes, el día, el periódico y el artículo al que pertenecen las oraciones o las palabras.

Guarda los artículos en una tabla y les asigna un número (ID) a cada uno

Guarda las palabras, los lemas, las categorías y el orden en que la palabra aparece en la oración en otra tabla. Cada palabra tiene asignado otro ID.

Envía cada oración a lematizar (API de Freeling)

Separa cada artículo en oraciones, las guarda en otra tabla y les asigna otro ID

3

Análisis

Fase sin
empezar:
primero
determinaremos
qué preguntas
queremos
contestar y
luego haremos
las búsquedas
pertinentes en la
base de datos.

Cómo funciona

palabras que desaparecen?: copioso, laborar, estelar, tener efecto, cumplimentar, faccioso, ulterior, parisién

después de + participio ("después de pronunciado su discurso")

aparte + nombre / aparte de + nombre (aparte Finlandia)

Buscar cuándo aparecen más palabras que denominan corrientes políticas:

socialdemo-cracia/crático, marx-ismo/ista

anteanoche / anoche

Creaciones modernas: desregulación, revalorizar, prejubilación, inconstitucional, pyme, estrellato, actualizar, rediseñar

En 2000-9 / en el 2000-9

Ver cuándo aparece: Beijing (2008 es lo esperado), gurú, progre, facha, chat, naif, peperero, collage, pop

Distribución de Yugoeslavia/Yugoslavia y Checoeslovaquia/Checoslovaquia (1960?)

en + gerundio (en viendo...)

es fuerza|

Palabras que empiezan por subs- / sus-

Distribución de encarnado y rojo (encarnado se usaba para rojo antes)

la totalidad de (más antiguo tengo la sensación) / todos/as los/as (más moderno)

grande (sin apocopar): antes salía más precediendo al nombre?

Eufemismos: buscar palabras como inútil, discapacitado, lisiado, cojo, discapacidad, movilidad

Orden de palabras: mejor comprender / comprender mejor, en caso tal / en tal caso

Medir la longitud de las oraciones a lo largo del tiempo (¿lo esperable es que sean cada vez más cortas?), evolución de la media de palabras por oración por año

Medir cuándo se usa más el orden oracional SVO / VSO o al menos la posición en que aparece el verbo dentro de la oración ("Ingresó el paciente en el hospital Clínico", "Ocurrió el suceso en la calle del Clot" 1917)