



I Xeira CLARIAH-GAL
2 de mayo de 2024

Recursos sobre escritura académica

Margarita Alonso-Ramos



Introducción

■ Motivación

- calidad de los textos académicos mejorable (Y NO SOLO DE LOS ESTUDIANTES)
- no hay hablantes nativos del discurso académico
- estudiantes universitarios españoles tienen **más recursos para inglés académico** a su disposición que para **otras lenguas del estado**
- el concepto de **escritura** (académica) está en proceso de cambio
 - Herramientas de ayuda a la escritura cada vez mejores pero SOBRE TODO PARA INGLÉS
 - Aparición de las IAs generativas como ChatGPT

Introducción

- **Proyectos de investigación**

FFI2016-78299-P

PID2019-109683GB-C21

PID2022-142843NB-I00

orientados a la creación de recursos

orientado a la evaluación de textos

- **Objetivo de esta presentación**

- Presentar algunos recursos fruto de estos proyectos
- Reflexionar sobre el papel de las IA en el cambio de concepción de la escritura

Plan de la presentación

1. Muestra de calidad de textos académicos
2. Presentación de recursos
 - 2.1. Corpus
 - 2.2. Herramientas de uso
3. Evaluando textos
4. Exploración de IAs generativas
5. Conclusiones y trabajos futuros

1. Muestra de calidad de textos académicos

Inadecuaciones fraseológicas (entre otras cosas)

- **Como ya se sabe**, Marco Común Europeo de Referencia para las Lenguas es un documento que sirve de referencia para el aprendizaje de una lengua
- **Siendo así como el objetivo principal de este trabajo, es** realizar un diagnóstico de las políticas de inclusión
- La investigación que se presenta como Trabajo Fin de Máster **asume un objetivo** que ha **sido demandado** a lo largo de los años
- **Los autores relataron** que el 75% de los pacientes...
- ...proponemos, abajo, una **recopilación de objetivos**
- **como conclusión final se extrajo** que las ortesis plantares
- Se han convertido en una **fuentes seria de preocupación**
- En conclusión, se deduce que la **hipótesis está equivocada**, ya que el número de redes sociales utilizadas no tiene relación
- De esta forma, **la presente publicación, relata** los aspectos centrales recogidos en la edición de dicho seminario.
- **Estudios recientes** han demostrado que aún con un adecuado control de calidad, la citología convencional ...

2. Presentación de recursos: corpus

- **Corpus Experto** (~ 21 millones de palabras)
 - SERAC (Spanish-English Research Article Corpus, facilitado por Interlae)
 - 360 artículos de revista
 - Cuatro secciones temáticas (AyHum, BioyCSalud, FiseIng, CSoc y Ed)
 - Iberia
- **Corpus de noveles** (~ 2 millones de palabras)
 - TFG, TFM (UDC y Universidad de León)
- **Evaluación general** (y superficial)
 - Calidad del corpus experto: muy mejorable (especialmente los de C. de la Salud)
 - Calidad del corpus de noveles: mejor de lo esperado (¿mano de los directores?)

2. Presentación de recursos: corpus

Corpus anotado con funciones discursivas

The screenshot displays the INCEpTION interface for a corpus document. The top navigation bar includes the INCEpTION logo, 'Projects', 'Dashboard', 'Help', 'Estela Mosqueira', 'Log out', and '29 min'. The document title is 'Estela Mosqueira: harta/fold_0000_harta_exp_bio.csv' and the current position is '508-517 / 5000 sentences [doc 1 / 34]'. The main text area shows a document with several paragraphs and annotations. A tooltip for a 'Named entity' with ID '196055' is visible over the text 'Por otra parte'. The sidebar panel on the right shows the 'Annotation' details for the selected text 'Por otra parte', including the layer 'Named entity' and the value 'EST_ADDINFO'.

INCEpTION Projects Dashboard Help Estela Mosqueira Log out 29 min

Estela Mosqueira: harta/fold_0000_harta_exp_bio.csv 508-517 / 5000 sentences [doc 1 / 34]

508 FIGURA 2 .
509 Distribución de áreas ANEP en Recercat

510 Se observa una distribución desigual, donde el área de Economía aparece en el primer puesto de forma muy destacada .
511 La Ingeniería y las Ciencias sociales le siguen en número de registros .
512 También resalta un pequeño agrupamiento de áreas en la zona alta de la curva de distribución , concretamente del 4º puesto (243 registros) al 8º (200 registros) .

513 Por otra parte , en la Figura 3 se analiza el estilo de firma para cada una de las áreas ANEP .

514 Las áreas con una mayor cantidad de firmas con el estilo recomendado corresponden a Economía y Ciencias Sociales (1ª y 3ª en número total de registros) , mientras que áreas más tecnológicas presentan grados muy bajos en el uso de este estilo de firma .

515 FIGURA 3 .
516 Distribución del estilo de firma por área ANEP en Recercat

517 No obstante , estos datos están influenciados por la gran cantidad de registros en estas áreas pues , pese a todo , el estilo de firma dominante en éstas es mediante dos apellidos .

Annotation: Delete Clear

Layer: Named entity

Text: Por otra parte

No links or relations connect to this annotation.

value: EST_ADDINFO

2. Presentación de recursos: herramientas

- **Combinaciones léxicas académicas (CLA)**
 - **Dominio de la fraseología** es clave para el dominio de la escritura académica
 - **Colocaciones:** *extraer conclusiones, confirmar una hipótesis, este trabajo se centra en, etc.*
 - **Fórmulas:** *en conclusión, parece razonable concluir que, como ya se ha señalado, etc.*
- **HARTA** (Herramienta de Ayuda a la Redacción de Textos Académicos)
- **HARTA/TAILA** (coordina español y euskera)

2. Presentación de recursos: HARTA

- Acceso por expresión académica



HARTA Herramienta de Ayuda a la Redacción de Textos Académicos

Expresión académica | Búsqueda en Corpus | Funciones discursivas

estudio 🔍

- estudio
- el presente estudio
- en el presente estudio
- en este estudio
- en los estudios de
- en nuestro estudio
- estudios previos
- los resultados de este estudio
- los resultados del estudio
- nuestro estudio
- para el estudio de

¿Qué puedo buscar aquí?

- Una colocación, a partir de un nombre como p. ej. *análisis*
análisis + adjetivo *análisis detallado, profundo*.
análisis + verbo *el análisis indica, muestra, revela*.
verbo + análisis *efectuar, hacer, llevar a cabo un análisis*.
- Una fórmula de varias palabras recurrentes en el discurso académico
p.ej. *en el presente estudio, podemos concluir que, en lo que respecta a*

2. Presentación de recursos: HARTA

- Colocaciones V+ *estudio*

estudio

¿Qué puedo buscar aquí?

- Una colocación, a partir de un nombre como p. ej. **análisis**
análisis + adjetivo **análisis detallado, profundo**.
análisis + verbo **el análisis indica, muestra, revela**.
verbo + análisis **efectuar, hacer, llevar a cabo un análisis**.
- Una fórmula de varias palabras recurrentes en el discurso académico
p.ej. *en el presente estudio, podemos concluir que, en lo que respecta a*

Colocaciones

Ver colocaciones para: estudio [Ver todas](#) | estudio + adjetivo | [verbo + estudio](#) | estudio + verbo | nombre de estudio

Glosa	Colocativos
tratar el ~	abordar, plantear
dirigir la temática del ~ hacia algo	centrar
tratar algo en el ~	considerar, describir, detectar, determinar, evidenciar, incluir, observar, obtener, tratar
causar que el ~ exista	desarrollar

2. Presentación de recursos: HARTA

- Descripción de una colocación específica

Colocación: *abordar estudio*

Detalles de la Colocación

Glosa
tratar el ~

Dominio

Biblioteconomía Ciencias de la Tierra Economía y empresa Educación
Física y química Ingeniería Lingüística Aplicada Literatura
Ciencias de la Salud Sociología

Ejemplos

Además, cabe distinguir la existencia de diferentes enfoques para abordar el **estudio** de la creación de empresas, dado que no se ha alcanzado todavía un consenso entre los autores sobre este fenómeno. (acad.)

Desde que en la década de los noventa el campo de la lingüística aplicada abordara el **estudio** de los géneros, muchos han sido los trabajos que han girado en torno a su análisis, tanto discursivo como retórico y léxico-sintáctico. (acad.)

Datos cuantitativos

Arte y Humanidades: 33 apariciones (32.04%) | Ciencias Físicas: 34 apariciones (33.01%) | Ciencias Sociales: 34 apariciones (33.01%) | Biología y Ciencias de la Salud: 2 apariciones (1.94%)

Categoría	Apariciones	Porcentaje
Arte y Humanidades	33	32.04%
Ciencias Físicas	34	33.01%
Ciencias Sociales	34	33.01%
Biología y Ciencias de la Salud	2	1.94%

Total: 103 apariciones en el corpus y 0.026 por artículo



2. Presentación de recursos: HARTA

- Descripción de una fórmula

Fórmula: *en el presente estudio*

Detalles de la fórmula

Funciones discursivas

Hacer referencia al propio trabajo

Dominio

Información y Documentación | Lingüística | Biología | Medicina
Ciencias de la Tierra | Física y Química | Ingeniería mecánica
Economía y Empresa | Educación | Sociología

Ejemplos

En el presente estudio, se analizarán en primer lugar los dos enfoques por separado, pero también se considerarán aquellas aportaciones en las que ambos enfoques son considerados conjuntamente.
SPBM11.xml | Economía y empresa

De acuerdo a los resultados de selectividad del método QuECh- ERS sin modificar, se evidencia que existe un número relativamente elevado de compuestos que presentan interferencias significativas; estos equivalen a cerca del 21% del total de compuestos analizados **en el presente estudio**.
SUPERQU12.xml | Física y química

Datos cuantitativos

Biología y Ciencias de la Salud: 74 apariciones (68.52%) | Ciencias Físicas: 17 apariciones (15.74%) | Ciencias Sociales: 12 apariciones (11.11%) | Arte y Humanidades: 5 apariciones (4.63%)

Número de apariciones por millón de palabras

Sección	Número de apariciones por millón de palabras
Resumen	80
Introducción	5
Cuerpo	15
Conclusión	220
Anexos	105

Total: 108 apariciones en el corpus y 0.183 por artículo

2. Presentación de recursos: HARTA

- Acceso por corpus

Harta Herramienta de Ayuda a la Redacción de Textos Académicos

Expresión académica **Búsqueda en Corpus** Funciones discursivas

Forma
 Lema
 Combinación de palabras

analizar

2. Presentación de recursos: HARTA

- Acceso por funciones discursivas

Harta Herramienta de Ayuda a la Redacción de Textos Académicos

Expresión académica Búsqueda en Corpus **Funciones discursivas**

Búsqueda por función discursiva

Estructurar el texto ▶	Función discursiva Hacer hincapié
Referirse al contenido de la investigación ▶	
Posicionarse y dirigirse al lector ▶	

Fórmulas

- a tener en cuenta
- cabe destacar
- cabe señalar
- en todo momento
- es importante
- es importante destacar
- es relevante
- hay que destacar
- hay que tener en cuenta
- hay que tener en cuenta que
- llama la atención
- se debe tener en cuenta
- se destacan
- sobre todo



2. Presentación de recursos: HARTA/TAILA

- Combinaciones léxicas en español y en euskera

HARTA/TAILA

Herramienta de Ayuda a la Redacción de Textos Académicos
Testu Akademikoen Idazketarako Laguntza

Búsqueda por funciones discursivas

Estructurar el texto ▾

Referirse al contenido de la investigación ▾

Posicionarse y dirigirse al lector ▾

EU ES

Búsqueda de expresiones académicas

EU ES

Búsqueda en corpus

LEMA **EU** ES

2. Presentación de recursos: HARTA/TAILA

- Comparación de fórmulas en español y en euskera

Función discursiva

Hacer hincapié



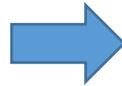
3. Hacia la evaluación de los textos académicos

- **Objetivo general (ambicioso)**
 - Definir de un modo objetivable y operativo en qué consiste “escribir bien” en el ámbito académico
- **Parámetros de medición**
 - Complejidad léxica
 - Diferentes medidas sobre repetición de palabras en un texto, uso de palabras poco frecuentes y uso de palabras con contenido léxico
 - Complejidad sintáctica
 - Sofisticación y variedad de las construcciones sintácticas

3. Hacia la evaluación de los textos académicos

- Herramienta de medición (modo beta)

Resultado del análisis de un texto de alumno de 1Bac



Estadísticas

- Número de oraciones: 58
- Número de cláusulas (= verbos finitos): 115
- Número de palabras: 847
- Número de palabras distintas: 321
- Número medio de palabras por oración: 14.60
- Número medio de verbos por oración: 3.40
- Número medio de verbos en forma finita por oración: 1.98
- Número de lemas distintos: 247
- Riqueza léxica: 37.90%
- Riqueza léxica (lemas): 29.16%
- Sin stopwords:
 - Número de palabras: 427
 - Número de palabras distintas: 261
 - Número de lemas distintos: 205
 - Riqueza léxica: 61.12%
 - Riqueza léxica (lemas): 48.01%

Acciones disponibles

- [Oraciones ordenadas por número de palabras](#)

• Oraciones que contengan la relación:



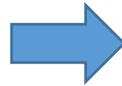
3. Hacia la evaluación de los textos académicos

- **Parámetros de medición**
 - Complejidad colocacional
 - *hacer un análisis / llevar a cabo un análisis*
 - Complejidad discursiva
 - Conectores y marcadores discursivos, entre otras cosas

3. Hacia la evaluación de los textos académicos

- Herramienta de medición: CoColAca (modo beta)

Resultado del
análisis de un TFG



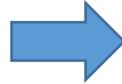
Puntuación	Nivel medio (2/3)
Nº palabras texto	5459

Relación	N+Adj	V+Obj	TOTALES
Tokens colocacionales	25	23	48
Lemas colocacionales	16	22	38
			MEDIA
Diversidad 1	0.64	0.96	0.80
Diversidad 2	3.20	4.59	3.89
Sofisticación muy alta	0.04	0.00	0.02
Sofisticación alta	0.84	0.43	0.64
Sofisticación media	0.08	0.09	0.08
Sofisticación baja	0.04	0.48	0.26

3. Hacia la evaluación de los textos académicos

- CoColAca

Resultado del análisis de un TFG:
Listado de colocaciones por franjas



F4: Sofisticación muy alta

Tipo	Colocación	Frecuencia
N+Adj	presencia,mayoritario	1

F3: Sofisticación alta

Tipo	Colocación	Frecuencia
N+Adj	influencia,mayor	1
N+Adj	usuario,potencial	1
N+Adj	publicación,científico	4
N+Adj	literatura,científico	1
N+Adj	disponibilidad,mayor	1
N+Adj	análisis,estadístico	2
N+Adj	número,mayor	1
N+Adj	parte,mayor	1
N+Adj	relevancia,especial	1
N+Adj	número,mínimo	1
N+Adj	pérdida,mayor	1
N+Adj	trabajo,presente	1
N+Adj	tamaño,medio	5
V+Obj	indicador,obtener	1

4. Exploración de las IAs generativas

- **Objetivo general**
 - Evaluar tanto la calidad de sus textos como su **capacidad de ser evaluadora de textos humanos**
- **Objetivo más específico (y ambicioso)**
 - Desarrollar un sistema informático capaz de evaluar automáticamente los ensayos
 - PROPOR'24 Competition on Automatic Essay Scoring of Portuguese Narrative Essays
- **Empecemos probando ChatGPT y similares**

4. Exploración de las IAs generativas

Copilot

¿Puedes detectar cuál es el problema de escritura en la siguiente frase en español y explicármelo: "**Los alumnos los cuales vienen a clase aprueban mas fácilmente**"

Copilot

Claro, puedo ayudarte con eso. La frase “Los alumnos **los cuales** vienen a clase aprueban más fácilmente” contiene un error de redundancia. La palabra “los cuales” es innecesaria en este contexto. La corrección sería simplemente decir: “Los alumnos que vienen a clase aprueban más fácilmente”. De esta manera, la frase es más concisa y gramaticalmente correcta¹. ¡Espero que esta explicación te haya sido útil! 😊

Más información

1magisnet.com2tekmaneducation.com3webdelmaestrocmf.com



4. Exploración de las IAs generativas

Copilot

y podrías explicar el error de estas frases: "Sino vienes ahora, me gustaría que no vinieras. Además, él no es listo si no listillo".

Copilot

Claro, con gusto te explico los errores en esas frases:

“Sino vienes ahora, me gustaría que no vinieras.”

- El error aquí está en el uso de la palabra “sino”. La forma correcta es separarla en dos palabras: “si no”.
- La corrección sería: “Si no vienes ahora, me gustaría que no vinieras.”

2. “Además, él no es listo si no listillo.”

- En esta frase, el error está en la redundancia. “Listo” y “listillo” son sinónimos, por lo que no es necesario usar ambos.
- La corrección sería: “Además, él no es listo.”

4. Exploración de las IAs generativas

ChatGPT

Problemas de puntuación con explicación



Tú

Corrige el error en la frase "Me gusta como vistes"



ChatGPT

El error en la frase "Me gusta como vistes" radica en la incorrecta utilización del pronombre relativo "como". La forma correcta de escribir la frase sería: "Me gusta cómo vistes". En este caso, "cómo" se utiliza para introducir una pregunta indirecta o una oración subordinada que modifica el verbo "gustar", indicando de qué manera vistes y no "como", que sería una comparación.

4. Exploración de las IAs generativas

ChatGPT

Prompt: Mejora el texto

Texto original: La fibrosis pulmonar idiopática es una enfermedad respiratoria que causa inflamación y cicatrización en los pulmones. **Presenta** síntomas similares a otras muchas enfermedades respiratorias, por lo que para un correcto diagnóstico de la enfermedad es necesaria una biopsia y una tomografía computarizada de alta resolución. **Tiene** una muy elevada mortalidad y los pacientes suelen fallecer 3 años después del diagnóstico

Propuesta: La fibrosis pulmonar idiopática es una enfermedad respiratoria caracterizada por la inflamación y cicatrización de los pulmones. **Sus síntomas** pueden asemejarse a los de otras enfermedades respiratorias, por lo que un diagnóstico preciso requiere una biopsia y una tomografía computarizada de alta resolución. **Esta enfermedad** tiene una elevada tasa de mortalidad, con los pacientes generalmente **falleciendo** aproximadamente tres años después del diagnóstico

5. Conclusiones

1. Saber escribir es una competencia transversal e interlingüística que no se puede dar por sentada
2. Recursos para lenguas que no sean inglés son escasos
3. La aportación de la IA puede hacer que la situación cambie radicalmente
 1. IA como recurso de ayuda
 2. IA como generadora de textos
 3. Línea divisoria entre “escribir” y “editar” un texto empieza a ser borrosa

5. En progreso / Trabajo futuro

- Corpus de textos procedentes de la ABAU
 - Analizarlos con técnicas de PLN
 - Buscar correlaciones con la calificación y los rasgos lingüísticos
 - Experimentar con modelos de lenguaje con técnicas de “deep learning” para ver si se pueden evaluar automáticamente
- Test diagnóstico de la calidad de escritura académica (test CEA)
- Experimentos de usabilidad con diferentes herramientas de ayuda a la escritura

Gracias por vuestra atención

margarita.alonso@udc.es

www.dicesp.com

<http://www.dicesp.com:8083/>

www.grupolys.com

