

Frasesología en el discurso académico: presentación del recurso HARTA

M. Alonso Ramos



XXI Jornada AETER
Universidad politécnica de Madrid
11 de noviembre 2022

Introducción

● Motivación

- no hay “hablantes nativos del discurso académico”
- escasez de **recursos de aprendizaje** de español académico

● Léxico académico

- para referir a las actividades que caracterizan el trabajo académico, organizar el discurso científico y construir la retórica de los textos académicos (Paquot 2010: 28)
- no se trata de terminología
- interdisciplinar (o transdisciplinar)
- *análisis, examinar, hipótesis, confirmar una hipótesis, este trabajo se centra en, por otra parte, como se ha señalado previamente*

● Fraseología académica

- dominio de la fraseología **es clave para el dominio de la escritura académica (EA)**

Introducción

● Objetivo general de la investigación

(Proyecto MINECO FFI2016-78299-P)

(Proyecto MCIN PID2019-109683GB-C21)

Herramienta de **Ayuda a la Redacción de Textos Académicos** apuntando a las ***combinaciones léxicas académicas (CLA)***

● Objetivo de esta presentación

1. describir la tipología de CLA con sus funciones discursivas
2. presentar
 - arquitectura de HARTA
 - tests sobre las necesidades de sus potenciales usuarios

Plan de la presentación

1. **Objeto de HARTA: combinaciones léxicas académicas (CLA)**
2. **Funciones discursivas en HARTA**
3. **Arquitectura de HARTA**
4. **Test de uso de la herramienta HARTA**
5. **Conclusiones y trabajo futuro (y en curso!)**

**1. Objeto de HARTA:
Combinaciones léxicas académicas:**

Noción de CLA

● CLA

- segmentos de palabras **recurrentes**
- específicas de **género académico escrito**
- composicionales o no
- algunas con una **función discursiva** (contrastar, concluir, reformular, expresar certeza o posibilidad, etc.)

● **Lexicología explicativa y combinatoria (Mel'čuk 2015):**

- composicionalidad (\neq transparencia)
- combinatoria restringida
 - la restricción puede ser en diferentes niveles

Tipología de CLA

● Locuciones (una UL)

- no composicionales y no libres (un sentido → una UL)
- necesitan una entrada léxica
- ***en conclusión, al contrario***

● Colocaciones (dos UL, base y colocativo)

- composicionales y no libres (elección léxica del colocativo restringida)
- descritas en la entrada de la base de la colocación
- ***sacar una conclusión, el problema estriba, conclusión lógica***

● Fórmulas (un subtipo de *clichés*, Mel' čuk 2015)

- composicionales y no libres (elección del sentido y de la forma está restringida): ***en otras palabras / ? diciendo de un modo diferente***
- casi nunca entrada en los diccionarios
- ***en otras palabras, hay que destacar, es importante subrayar***

Tipología de CLA

● Fórmulas

- ¿qué tipo de entidad léxica?
 - *in other words*
 - *en lo que **respecta a**, por lo que **respecta a***
 - *el objetivo de este trabajo, este estudio ha mostrado que*
- semejantes a los llamados *lexical bundles* (Biber et al. 1999)
- todas las fórmulas son *lexical bundles*, pero lo inverso no es cierto:
 - *la probabilidad de que* no es interesante fraseológicamente
- Lo que es relevante es su **función discursiva**, no su significado composicional
- función discursiva: ‘**para hacer hincapié**’
 - *hay que destacar, es importante subrayar, merece mención especial*

2. Funciones discursivas en HARTA

Funciones discursivas

● Clasificaciones previas de funciones discursivas

- Hyland, 2008; Simpson-Vlach & Ellis, 2010; Salazar, 2014
- Algunos problemas:
 - largas listas de subfunciones difícilmente interpretables para el usuario de una herramienta

● Propuesta de tipología de funciones discursivas

- García-Salido et al. 2019
- enfoque **bottom-up**
- asignación **manual** necesaria, **PERO** **NUEVOS EXPERIMENTOS**

● Estructura de la tipología

- Estructurar el texto
- Referirse al contenido de la investigación
- Posicionarse y dirigirse al lector

Funciones discursivas

● Objetivo

- una **base de datos de CLA** con funciones discursivas asociadas
- a partir de la lista, un **corpus anotado** discursivamente

● Muestra de funciones discursivas

- Para **hacer hincapié**: *hay que destacar, debemos señalar*
- Para **reformular**: *esto es, es decir, dicho de otro modo,*
- Para **reenviar** a trabajos de otros: *de acuerdo con, siguiendo a*
- Para **reenviar** a otras partes del texto: *como ya se ha visto anteriormente, como se ha señalado previamente,*
- Para **reenviar** a elementos no textuales: *como se aprecia en la tabla x, según la tabla X*
- Para **introducir** ideas relacionadas: *en lo que respecta a, en cuanto, en lo que concierne a*

Funciones discursivas

● Corpus anotado con funciones discursivas

The screenshot displays the INCEption web interface. The top navigation bar includes 'INCEption', 'Projects', and 'Dashboard'. The user is logged in as 'Estela Mosqueira'. The document being viewed is 'Estela Mosqueira: harta/fold_0000_harta_exp_bio.csv', with 508 sentences out of 5000. The main text area shows a list of sentences from 508 to 517. Sentence 512 is highlighted, and a tooltip for the annotation 'Por otra parte' (ID: 196055) is visible. The text of sentence 512 is: 'También resulta un pequeño agrupamiento de áreas en la zona alta de la curva de distribución, concretamente del 4º puesto (243 registros) al 8º (200 registros)'. Other annotations include 'REF_PRESDATA', 'EST_ADDINFO', 'EST_RESEND', 'REF_PRESMETH', 'REF_EXPAMOUNT', 'EST_EXPOPOS', 'REF_EXPAMOUNT', and 'EST_EXPCONCE'. The right sidebar shows the 'Annotation' panel with a 'Delete' button, 'Layer' set to 'Named entity', and 'Text' set to 'Por otra parte'. The 'value' field contains 'EST_ADDINFO'.

INCEption Projects Dashboard Help Estela Mosqueira Log out

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 resulta 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

Layer
Named entity

Text
Por otra parte

No links or relations connect to this annotation.

value
EST_ADDINFO

3. Arquitectura de HARTA

Descripción general de HARTA

● Características principales

- híbrido = diccionario + corpus
 - límites borrosos entre diccionario y corpus
 - corpus-driven lexicography → “lexically-driven corpora interfaces”
- en línea
- onomasiológico y semasiológico

● Modos de búsqueda

- por expresión académica
- por función discursiva
- en la base de datos léxica o en el corpus

Macroestructura de 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*

Microestructura de HARTA

● Colocaciones del nombre *estudio*

HARTA

Herramienta de Ayuda a la Redacción de Textos Académicos

[Bienvenida](#) **[Consulta](#)** [¿Cómo usar HARTA?](#) [Proyecto HARTA](#)

Expresión académica [Funciones discursivas](#) [Búsqueda en Corpus](#)

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](#)

Microestructura de 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

tratar el ~

dirigir la temática del ~ hacia algo

tratar algo en el ~

causar que el ~ exista

Colocativos

[abordar](#), [plantear](#)

[centrar](#)

[considerar](#), [describir](#), [detectar](#), [determinar](#), [evidenciar](#), [incluir](#), [observar](#), [obtener](#)
[, tratar](#)

[desarrollar](#)

Microstructure de HARTA

Descripción de la colocación *abordar + estudio*

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

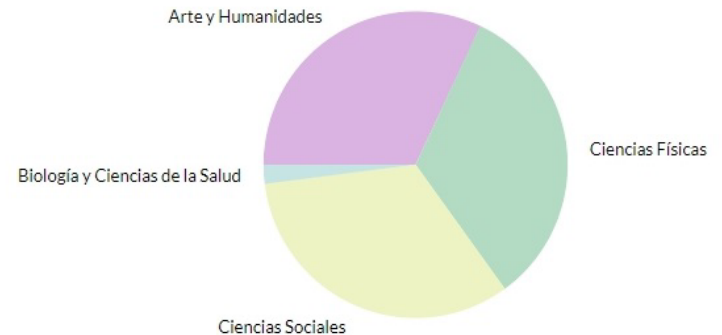
Ejemplos

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%)



Total: 103 apariciones en el corpus y 0.026 por artículo

Microestructura de HARTA

● Entrada para *en el presente estudio*

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

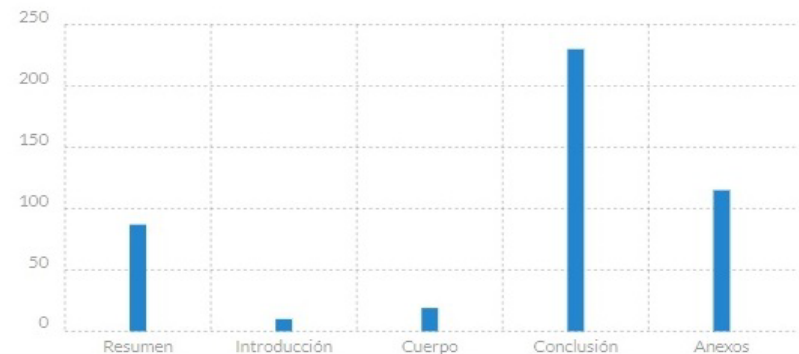
De acuerdo a los resultados de selectividad del método QuECh- ERS sin modificar, se evidencia que existe un número relativamente elevado de compuestos que presentan interferencias significativas; estos equivalen a cerca del 21% del total de compuestos analizados en el presente estudio.

SUBFISQUI23.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



Total: 108 apariciones en el corpus y 0.183 por artículo

Macrostructure de HARTA

● Acceso par 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 ▶

Referirse al contenido de la investigación ▶

Posicionarse y dirigirse al lector ▶

Función discursiva

Hacer hincapié

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

Macroestructura de HARTA

● Acceso par corpus

Harta

Herramienta de Ayuda a la Redacción de Textos Académicos

Expresión académica

Búsqueda en Corpus

Funciones discursivas

Forma

Lema

Combinación de palabras

analizar|

Buscar

4. Test de usabilidad de HARTA

Testando HARTA

- **Objetivo:**

- Identificar las necesidades de los usuario
- Evaluar si HARTA puede satisfacer esas necesidades

- **Preguntas:**

1) ¿Cuáles son las necesidades de los usuarios en cuanto a las CLA?

a) ¿*conclusión* va con qué verbo?; ¿otra expresión similar a *hay que subrayar*?

b) ¿saben cómo encontrar esas combinaciones en HARTA?

2) ¿El **corpus** es un medio de resolver las necesidades vinculadas a las CLA?

3) ¿Pueden **mejorar** su actuación en una tarea de producción **con HARTA** sin formación previa?

Testando HARTA

- **Diseño del estudio**

1) Tarea

10 oraciones: 5 con colocaciones y 5 con fórmulas

Buscar alternativa a **la expresión subrayada** (sin y con HARTA)

Item 1: **colocación:**

*Actualmente se sigue trabajando en la metodología para **extraer conclusiones** más precisas*

Item 2. **fórmula:**

Con el objetivo de promover la dimensión social de la integración regional...

2) Participantes

12 (8 estudiantes Grado, Master, Doctorado y 4 profesionales)

3) Protocolo “think aloud”

debían grabar su pantalla y su voz en cuanto empezaran la tarea

Testando HARTA

1) Respuesta a la pregunta 1

a) Varios usuarios desconocen el modo de seleccionar colocativos

- colocativo \neq sinónimo
 - *extraer* = *arrancar* en algunos contextos
- información colocacional está codificada en la entrada de la base
- Dificultad en HARTA debido al reenvío al DiCE

b) Dudas sobre el estatus léxico de las fórmulas

- « Ah, es una cadena entera que va todo junto »
- Algunas dificultades en reconocer la función discursiva

2) Respuesta a la pregunta 2

- El usuario que busca en el corpus en HARTA es menos eficaz
- La interfaz del corpus en HARTA se concibió para encontrar palabras o cadenas de palabras (fórmulas, no colocaciones)

Testando HARTA

3) Respuesta a la pregunta 3

- a) un **buen nivel medio de puntuación**, incluso sin formación previa
 - 11/12 una puntuación más alta de 15 sobre 30, aunque solo han visto unos breves vídeos tutoriales
 - Las tres mejores puntuaciones; 27, 28 y 30
- b) **mejor resultado para las fórmulas** que para las colocaciones
- c) Algunos usuarios cuya producción era floja **encontraron resultados** en HARTA

Testando HARTA

Recapitulando,

Identificación más precisa de las **necesidades de los usuarios** en el momento de la producción del texto

- ***Function-related needs*** (Tarp 2009)
 - Las necesidades que surgen en una situación **extra/pre-lexicográfica** →
 - Varios usuarios tienen **dificultades en ofrecer alternativas**
- ***Usage-related needs*** (Tarp 2009)
 - Las necesidades que surgen durante el **proceso de consulta** →
 - Varios usuarios **no saben qué estrategias** son necesarias para encontrar una combinación léxica fraseológica pero composicional

5. Conclusiones y trabajo futuro

Conclusiones

HARTA es un primer intento para poder comparar el vocabulario académico español con el inglés (LEAD) o el francés (LST)

- Metodología mixta: lingüística de corpus, lingüística computacional y fraseología
- Apuesta por una entrada léxica rica en lugar de por listas de secuencia ordenadas por frecuencias
- Después del test
 - HARTA ha recibido mejoras
 - Mejor integración de la información colocacional
 - Hipervínculo con la función discursiva
 - Una toma de conciencia de las necesidades potenciales de los usuarios potenciales de recursos como HARTA

En progreso

PERO COMENZAMOS A EXPLORAR...

✓ Exploración de técnicas computacionales basadas en la **semántica distribucional** para la identificación de fórmulas

✓ Contexto : HARTA para el **vasco, lengua aglutinante**

Si una expresión vasca presenta un grado de semejanza distribucional como una fórmula española, la probabilidad de que se trate de una fórmula es elevada

A partir de la lista de fórmulas españolas, podemos generar **embeddings** de cada fórmula y emplearlas para los equivalentes de una sola palabra vasca

Más exploratorio todavía: examinar si los modelos del lenguaje identifican un sentido discursivo como el de las fórmulas (**ROBERTA**)

En progreso

- ✓ Red de recursos interlingües
- ✓ [HARTA / TAILA](#) para el vasco

The screenshot shows the HARTA/TAILA website interface. At the top, the title "HARTA/TAILA" is displayed in large blue letters, with the subtitle "Testu Akademikoen Idazketarako Laguntza" below it in green. A search bar is present, with "EU" and "ES" language selection buttons on the right. Below the search bar are three main navigation buttons: "Corpus bilaketa" with a list icon, "Formula bilaketa" with a double-headed arrow icon, and "Kolokazio bilaketa" with a puzzle piece icon.

Trabajo futuro

HARTA

Herramienta de Ayuda a la Redacción de Textos Académicos

Bienvenida

Consulta


¿Cómo usar HARTA?


Proyecto HARTA

Expresión académica

Funciones discursivas

Búsqueda en Corpus

Analiza tu texto 



En este capítulo con el que se inicia el desarrollo del trabajo se llevará a cabo el estudio más profundo y formal sobre los elementos estéticos incluidos en las series de televisión, analizando no solo la corrección o el error con el que estos se han introducido como parte de la ambientación, sino que también se tendrán en cuenta las circunstancias que han podido determinar estos factores, principalmente la facción de público a la que esté destinado el programa. Esto se considerará teniendo en cuenta que el objetivo primordial de la totalidad de las series de televisión es el de resultar atractivas para la audiencia a la que se dirigen. Asimismo, con el fin de llevar a cabo este análisis de forma sencilla y estructurada se ha optado por realizar previamente un proceso de selección de aquellas series cuya estética presentara los ejemplos más representativos de error y fantasía en la inclusión de elementos artísticos y arquitectónicos.

« < 1 2 3 4 5 ... 8 > »

Gracias por vuestra atención

margarita.alonso@udc.es

www.dicesp.com

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

www.grupolys.com

