

**DISCURSOS TRADICIONALES
Y PARTICULARES**

Festschrift für Johannes Kabatek zum 60. Geburtstag

CRISTINA BLEORȚU

CARLOTA DE BENITO MORENO

DAVID PAUL GERARDS

PHILIPP OBRIST

ÁLVARO S. OCTAVIO DE TOLEDO Y HUERTA

ALBERT WALL

(eds.)

DISCURSOS TRADICIONALES Y PARTICULARES

Festschrift für Johannes Kabatek
zum 60. Geburtstag

CRISTINA BLEORȚU
CARLOTA DE BENITO MORENO
DAVID PAUL GERARDS
PHILIPP OBRIST
ÁLVARO S. OCTAVIO DE TOLEDO Y HUERTA
ALBERT WALL
(eds.)



IBEROAMERICANA - VERVUERT - 2025

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra.

(www.conlicencia.com; 91 702 19 70 / 93 272 04 47)

Reservados todos los derechos

© Iberoamericana, 2025

Amor de Dios, 1 – E-28014 Madrid

Tel.: +34 91 429 35 22

© Vervuert, 2025

Elisabethenstr. 3-9 – D-60594 Frankfurt am Main

Tel.: +49 69 597 46 17

info@iberoamericanalibros.com

www.iberoamericana-vervuert.es

ISBN 978-84-9192-493-7 (Iberoamericana)

ISBN 978-3-96869-698-0 (Vervuert)

ISBN 978-3-96869-700-0 (PDF)

ISBN 978-3-96869-699-7 (EPUB)

Depósito Legal: M-1471-2025

Cubierta: Diego Calvo Cubero

Impreso en España

Índice

PROEMIUM

CRISTINA BLEORȚU, CARLOTA DE BENITO MORENO, DAVID PAUL GERARDS, PHILIPP OBRIST, ÁLVARO S. OCTAVIO DE TOLEDO Y HUERTA, ALBERT WALL	13
---	----

LINGVISTICA COȘERIANĂ ȘI TEORIA LINGVISTICĂ

PRECISIONES SOBRE LA CONCEPTUALIZACIÓN DE LA “LENGUA FUNCIONAL” EN LA LINGÜÍSTICA COSERIANA MIGUEL CASAS GÓMEZ	29
--	----

LA TIPOLOGÍA SEGÚN E. COSERIU (Y LAS OTRAS) ADOLFO ELIZAINCÍN	37
--	----

LA CORRESPONDENCIA ENTRE RAMÓN MENÉNDEZ PIDAL Y EUGENIO COSERIU INÉS FERNÁNDEZ-ORDÓÑEZ	43
--	----

LUCIAN BLAGA ȘI LIMBA „TĂCUTĂ” CRISTINEL MUNTEANU	53
--	----

„PURISMUS“ — FRAME-THEORETISCHE MODELLIERUNG SPRACHGESCHICHTLICHER KONZEPTUALISIERUNGEN SARAH DESSÌ SCHMID	59
--	----

LANGUAGE AS A SIGN: A NOTE ON SAUSSURE JÜRGEN TRABANT	67
--	----

PE LIMBA LUI COȘERIU, SAU DIN NOU DESPRE A <i>PRIORI</i>-UL ISTORIC CORNEL-DUMITRU VÎLCU	75
--	----

HELMUT GIPPER AND EUGENIO COSERIU ON BASIC COLOUR WORDS KLAAS WILLEMS.....	83
--	----

TRADICIONES DISCURSIVAS

QUEBRA DE CONCORDÂNCIA DE GÊNERO E/OU NÚMERO NO HEBRAICO BÍBLICO E FIGURAS DE LINGUAGEM: PARANOMÁSIA NUMÉRICA, FIGURA DO TODO E SUAS PARTES, ANALOGIA ALESSANDRA CASTILHO DA COSTA.....	91
---	----

¿QUÉ SON 60 AÑOS EN LA VIDA... DE UN MANUSCRITO? O EL ANÓMALO CASO DEL FUERO DE ALCARAZ MÓNICA CASTILLO LLUCH	99
---	----

EL CAMBIO LINGÜÍSTICO ENTRE LA INTERTEXTUALIDAD Y LA TRADICIONALIDAD DISCURSIVA SANTIAGO DEL REY QUESADA	105
--	-----

ALGUMAS CONTRIBUIÇÕES DA ROMANÍSTICA ALEMÁ NO BRASIL VERENA KEWITZ	111
--	-----

SOBRE LA FÓRMULA DISCURSIVA DE DESPEDIDA <i>CUÍDATE</i> ANA LLOPIS CARDONA	117
--	-----

EL NIVEL INDIVIDUAL DEL LENGUAJE Y LAS TRADICIONES DISCURSIVAS. UN BREVÍSIMO EJERCICIO DE ANASTILOSIS ARACELI LÓPEZ SERENA.....	123
---	-----

ROMANISTIK IM SESSEL? NOCH EINMAL ZUR HISTORIZITÄT ALS WESENSMERKMAL VON DISKURSTRADITIONEN UND EINZELSPRACHEN ESME WINTER-FROEMEL	129
--	-----

METHODE UND KORPUS

ALGÚNS PROBLEMAS NA EXPLOTACIÓN DOS CORPUS TEXTUAIS DO GALEGO PARA A ELABORACIÓN DO <i>DICIONARIO HISTÓRICO ETIMOLÓXICO DA LINGUA GALEGA</i> MANUEL GONZÁLEZ GONZÁLEZ	139
---	-----

LA ENCAPSULACIÓN ANAFÓRICA EN CLAVE EXPERIMENTAL CELIA HERNÁNDEZ PÉREZ, MATHIS TEUCHER, ÓSCAR LOUREDA	145
SOBRE LA COMPLEJIDAD SINTÁCTICA DE LOS TEXTOS LITERARIOS DEL ESPAÑOL A TRAVÉS DEL TIEMPO ROBERT HESSELBACH.....	155
FINDING WHAT'S NOT THERE: HOW TO RETRIEVE 'NOVEL' BARE NOUNS FROM CORPORA MARIANNE HUNDT.....	167
THE FALLACY OF STRUCTURAL SIMILARITY. EXAMPLES FROM SPANISH AND PORTUGUESE VOLKER NOLL	175
ASTÉRIX CHEZ LES ROMANISTES. COMICFORSCHUNG UND DIE FREIBURGER GRAL-SAMMLUNG CLAUS D. PUSCH.....	183
UN BREVE APUNTE SOBRE LAS FRECUENCIAS LÉXICAS GUILLERMO ROJO	189
EL VALOR DE LA RIQUEZA LÈXICA EN ELS CORPUS TEXTUALS JOAN TORRUELLA	195
INTUICIONES METALINGÜÍSTICAS. ¿QUÉ NOS PUEDEN APORTAR? SENTA ZEUGIN	201
LINGUISTIQUE ROMANE	
GRICE MEETS FLAMENCA: ON THE ART OF INTERPRETING MINIMAL UTTERANCES ANDREAS DUFTER.....	209
« DECURSIS OMNIBUS »/ « TOUTES CHOSES DISCOURSES » : UN ASPECT INCONNU DE LA SÉMANTIQUE DE M. FR. DISCOURIR ET L'HISTOIRE DU CONCEPT PHILOSOPHIQUE DE DISCOURS FRANZ LEBSANFT	217

À PROPOS DES CONCEPTIONS DIVERGENTES DU NÉOLOGISME DANS LES ÉTUDES DE NÉOLOGIE ROMANE CHRISTOPHE GÉRARD.....	223
LA LINGÜÍSTICA ROMÀNICA ENTRE LA COMPARACIÓ I LA HISTÒRIA ÁNGEL LÓPEZ GARCÍA-MOLINS.....	233
WILHELM MEYER-LÜBKE E O ENSINO DA FILOLOGIA PORTUGUESA E ROMÂNICA NA UNIVERSIDADE DE COIMBRA CLARINDA MAIA.....	241
LINGUAS MINORITARIAS E MINORIZADAS	
ASPECTOS POLÍTICOS DOS ESTUDOS DE GALEGO NA REPÚBLICA DEMOCRÁTICA ALEMÁ: O CASO DE LEIPZIG LUÍS CALVO SALGADO	251
EL HABLA DE LA FRONTERA URUGUAYO-BRASILEÑA: ¿ARTIGAS, ALDEA LINGÜÍSTICA GLOBAL? BÁRBARA GARRIDO SÁNCHEZ-ANDRADE	257
WO IN RAUM UND SPRACHE? ELVIRA GLASER.....	263
ORIGEN DE LA VOZ GALLEGA ‘MIXIRIQUEIRO’. O PORTUGUÉS FRONTE AO GALEGO EN MARTIN SARMIENTO, CONTRA DUARTE NUNEZ DE LEÁO HENRIQUE MONTEAGUDO	269
DOM IN ENGADINESE AND THE ROLE OF REFERENTIALITY GEORG A. KAISER, KLAUS VON HEUSINGER	275
CAN A MIXED LANGUAGE BE CREATED ON THE FLY? PUTTING ECUADORAN MEDIA LENGUA TO THE TEST JOHN M. LIPSKI.....	281
OS LINGÜISTAS ANTE O QUE CHAMA(BA)MOS UNHA LINGUA XOSÉ LUÍS REGUEIRA.....	289
VLACH AROMANIAN IN KALOMOIRA (GREECE): A FIRST APPROACH IOANNA SITARIDOU, DAVID GARDNER	295

LINGÜÍSTICA PORTUGUESA

A DIALETOLOGIA PODE SER ACÚSTICA E A FONÉTICA PODE SER DIALETOLÓGICA FERNANDO BRISSOS	305
AS MÚLTIPLAS CARAS DA PRIMEIRA PESSOA DO PLURAL: ENTRE INDEXICALIDADE E GENERECIDADE YOSELIN HENRIQUES PESTANA.....	313
DAR COR À NEGAÇÃO: PATAVINA E PINGO ANA MARIA MARTINS.....	319
A ESCRITA ANTES DA NORMA: A EMERGÊNCIA DA LÍNGUA VERNACULAR NO FORAL DE JUNQUEIRA DE VILARIÇA JOANA SERAFIM	327
O PORTUGUÊS DOS BRASILEIROS SÃO MUITOS, VIU? JOSÉ DA SILVA SIMÕES.....	333
PARA A ABORDAGEM SOCIOCOGNITIVA DO PLURICENTRISMO DO PORTUGUÊS: AGENDA E DESAFIOS AUGUSTO SOARES DA SILVA.....	339
HISPANIC LINGUISTICS	
NOTAS SOBRE CONTACTO LINGÜÍSTICO E HIBRIDACIÓN: A PROPÓSITO DEL ALEMAÑOL HÉCTOR ÁLVAREZ MELLA.....	349
UN PLAGIO DE TORRES VILLARROEL PEDRO ÁLVAREZ DE MIRANDA	355
LA CONFIGURACIÓN HISTÓRICA DE LOS APÉNDICES GENERALIZADORES: EL CASO DE Y TODO ESO MARGARITA BORREGUERO ZULOAGA	361
CONVERSACIÓN, DIÁLOGO, COLOQUIO: SU ENTRADA EN CASTELLANO RAFAEL CANO AGUILAR.....	369

HACIA UNA SEMÁNTICA INTEGRAL MANUEL CASADO VELARDE.....	375
SOBRE LA PERVIVENCIA DE UN MITO LINGÜÍSTICO: LA DISIMILACIÓN DE [L] – [L] EN COMBINACIONES DE CLÍTICOS DE OBJETO ANDRÉS ENRIQUE-ARIAS.....	385
REFLEXIONES ACERCA DEL ORIGEN JURÍDICO DEL PRO-NOMBRE PROPIO <i>FULANO</i> WILTRUD MIHATSCH.....	391
CUANDO <i>ESO</i> SIGNIFICA <i>AQUELLO</i> MANUEL PÉREZ SALDANYA.....	399
TRES ETAPAS EN EL ESTUDIO DEL ESPAÑOL COLOQUIAL SALVADOR PONS BORDERÍA.....	405
NOTA MÍNIMA ESDRÚJULA SOBRE LOS CAMBIOS PROSÓDICOS POR ELABORACIÓN LOLA PONS RODRÍGUEZ.....	411
CALIDAD DE LA ESCRITURA EN INFORMES DE PRÁCTICA PROFESIONAL: UNA APROXIMACIÓN DESDE LA LINGÜÍSTICA EMPÍRICA COMPUTACIONAL RENÉ VENEGAS, FERNANDA JARAMILLO BECKER.....	417
TABULA GRATULARORIA.....	427

Un breve apunte sobre las frecuencias léxicas

Guillermo Rojo

Universidad de Santiago de Compostela / Real Academia Española

La lingüística de corpus (LC) ha supuesto una importante revolución metodológica en los estudios lingüísticos, que han modificado fuertemente sus modos de trabajo incluso en las visiones más formalistas, tradicionalmente alejadas del análisis directo de los actos lingüísticos realizados. Una de las características más destacadas de la LC es, sin duda, su carácter marcadamente cuantitativo, centrado en los últimos años en estudiar las diferentes frecuencias con las que un elemento o fenómeno se presenta en distintos corpus o subcorpus. Resulta por ello realmente curioso que los diccionarios de frecuencias (DF) no hayan experimentado un cambio que refleje y explote las nuevas opciones que surgen de la posibilidad de obtener datos procedentes de conjuntos formados por cientos o miles de millones de palabras y que habitualmente están aceptablemente codificados y anotados. En Rojo (2023) se muestra que los DF posteriores a la publicación del *Frequency Dictionary of Spanish Words* (Juilland y Chang-Rodríguez 1964) superan las deficiencias que presenta esta obra en el volumen de corpus manejado (solo 500 000 palabras), pero no lo mejoran o incluso resultan inferiores a él en algunos otros aspectos. A mi modo de ver, este defecto procede de la insuficiente explotación de las posibilidades que brindan los corpus y, asociado a ella, el peso que, con independencia de los formatos de salida, todavía tiene la versión impresa de los DF. En los párrafos siguientes trataré de mostrar los cambios que produce la explotación de los corpus textuales.

Si nos restringimos al módulo léxico, los DF y las listas de frecuencias han sido utilizados fundamentalmente para establecer el volumen y la secuencia de introducción del léxico en los cursos de lengua extranjera. Desde esta perspectiva, la “utilidad” de un lema se mide básicamente en la cobertura que proporciona en los textos, esto es, en el porcentaje de un texto que se puede entender si se conoce el significado de un cierto conjunto de lemas. En este sentido, se alude habitualmente a las leyes de Zipf y Pareto, con las que se explica que en cualquier texto o conjunto de textos hay siempre unas pocas palabras que tienen frecuencia muy alta, muchas palabras con frecuencias bajas o muy bajas y también un alto porcentaje que, independientemente del tamaño del corpus manejado, aparecen solo una vez (son hápax). Tomando los datos del DF del español incluido en la versión 1.1 del CORPES, se comprueba que con solo 50 lemas (caracterizados también por la clase de palabras) se da cuenta del 50% de todo el corpus manejado (unos 180 millones de palabras procedentes de textos de prensa). Naturalmente, esa observación va seguida inmediatamente de la restricción esperable: los lemas más frecuentes son artículos, preposiciones, conjunciones, etc., de modo que la comprensión de los textos que se puede lograr con el conocimiento de su significado es más bien escasa. En la bibliografía referida al inglés, hay acuerdo general en considerar que para alcanzar el reconocimiento del 98% de los lemas presentes en un texto (se entiende que sin tener en cuenta nombres propios, cifras y algunos otros tipos de elementos) se requiere el dominio de unas 8000 o 9000 familias de lemas.¹ Los datos de los que dispongo para español y gallego, referidos a lemas y con inclusión de la clase de palabras son los que figuran en la tabla 1.

1 Las familias de lemas agrupan los relacionados mediante procedimientos gramaticales, como, por ejemplo, la derivación. Así *limpio*, *limpiar*, *limpieza*, *limpiador*, *limpiamente* y algunos otros son lemas distintos, pertenecientes a distintas clases de palabras, pero forman parte de la misma familia. Pues, bien, según indica Nation (2006: 79): If we take 98% as the ideal coverage, a 8,000-9,000 word-family vocabulary is needed for dealing with written text, and 6,000-7,000 families for dealing with spoken text.

TABLA 1: Porcentajes de lemas en diferentes corpus de español y gallego
Fuente: CORPES, CORGA y ESLORA. Elaboración propia.

Los primeros x lemas	Todo el lemario			Adjetivos, adverbios, sustantivos y verbos		
	CORPES 1.0 (prensa)	CORGA 4.1 (2001-2022)	ESLORA 2.2	CORPES 1.0 (prensa)	CORGA 4.1 (2001-2022)	ESLORA 2.2
10	38,82	38,65	30,78	10,56	12,13	30,06
25	47,81	48,27	46,93	14,6	17,68	42,64
50	52,58	54,04	60,85	18,81	22,86	54,09
100	57,24	59,45	72,9	24,9	29,12	63,79
500	71,36	72,75	88,68	47,57	49,84	82,11
1000	78,85	79,34	92,83	60,81	61,75	88,16
2000	86,24	86,19	96	74,19	74,29	93,24
3000	90,01	89,71	97,43	81,15	80,83	95,63
4000	92,37	91,92	98,26	85,43	84,96	97,02
5000	93,86	93,45	98,78	88,31	87,83	97,94
6000	94,95	94,58	99,14	90,36	89,96	98,55
7000	95,76	95,45	99,41	91,9	91,58	99,05
8000	96,39	96,13	99,58	93,1	92,86	99,33
9000	96,9	96,68	99,72	94,07	93,89	99,61
10 000	97,31	97,13	99,86	94,85	94,74	99,89
15 000	98,54	98,54		97,21	97,39	
20 000	99,12	99,23		98,33	98,67	
25 000	99,44	99,6		98,92	99,34	
50 000	99,89	100		99,8	100	
Total lemas	116 707	49 603	10 990	112 971	46 148	10 386
Hápax	29 172 (25%)	7700 (15,52%)	3719 (33,83%)		6510 (14,10%)	3606 (14,60%)
Total subcorpus	184 108 153	25 831 840	728 038	92 013 231	12 643 355	358 032

La parte izquierda de la tabla 1 muestra la coincidencia en este punto entre español y gallego en el número de lemas necesarios para alcanzar una cobertura del 95% y el 98%. En cambio, en ESLORA, formado por textos orales, las cifras respectivas bajan a 2000 y 4000 lemas, respectivamente. Las cifras son válidas en sí mismas, pero es necesario tener mucha precaución en la comparación, puesto que son corpus que tienen características de composición y tamaño bastante diferentes.

En la tabla se muestra también un aspecto al que, precisamente por las insuficiencias del modelo clásico de DF, no se ha prestado atención suficiente. Si nos limitamos a excluir de nuestra consideración elementos como nombres propios, cifras, fechas y similares, los porcentajes de cobertura arrojarán irremediablemente resultados en los que los elementos de significado gramatical seguirán teniendo un gran peso como consecuencia de la alta frecuencia que tienen muchos de ellos. No es seguro, pues, que la imagen que proyectan estos resultados sobre la comprensión del contenido de un texto (en realidad, del reconocimiento de los lemas) refleje realmente el funcionamiento de ese módulo. Podemos intentar evitar esos efectos distorsionadores reduciendo nuestra consideración a los lemas que tienen significado léxico en el sentido más estricto, es decir, sustantivos, verbos, adjetivos y adverbios.

Eso es lo que muestran las columnas de la parte derecha de la tabla 1. En los tres corpus considerados la frecuencia de uso de los lemas pertenecientes a estas cuatro clases se sitúa en torno al 50% del total, pero su frecuencia de inventario (cf. Rojo 2011) ronda el 95%. Esas cifras son sumamente significativas y adelantan lo que sucede con la cobertura: en los textos de prensa del CORPES, el reconocimiento del 98% de los lemas de estas cuatro clases requiere unos 18 500, que es un 150% de los necesarios para todo el lemario (12 500). Algo parecido sucede en el CORGA y la cifra asciende a unos 5000 en el caso de ESLORA.

Esta diferencia, que considero relevante, se basa en el carácter diferente de las unidades léxicas consideradas, pero también a un hecho relacionado con la estructura estadística de los textos al que no hemos prestado la atención necesaria hasta ahora. Cualquier comparación de los lemas contenidos en una serie de textos muestra que los porcentajes de los que aparecen en todos ellos es bastante baja. En un experimento que me parece realmente demostrativo, Miller y Biber compararon los lemas contenidos en diez textos universitarios de introducción a la psicología utilizados en universidades estadounidenses. El corpus produce un total de 32 598 le-

mas distintos (sin contar nombres propios, como es lógico). A pesar de la identidad de tema, objetivo y nivel, los presentes en todos los textos son el 13% y el 41% se encuentra en solo uno de los textos. Es lógico pensar que esa distribución tan escasamente homogénea procede, sobre todo, de elementos pertenecientes a estas cuatro clases.

El que hemos tocado en esta breve nota de homenaje a nuestro querido colega es uno de los varios aspectos que debemos reformar para que las frecuencias léxicas den resultados realmente útiles en la enseñanza de lenguas extranjeras y en nuestra comprensión general del funcionamiento de las lenguas.

BIBLIOGRAFÍA

- JUILLAND, Alphonse y CHANG-RODRÍGUEZ, Eugenio (1964): *Frequency dictionary of Spanish words*. Den Haag: Mouton.
- MILLER, Don y BIBER, Douglas (2015): “Evaluating reliability in quantitative vocabulary studies. The influence of corpus design and composition”, en *IJCL*, 20/1: pp. 30-53.
- NATION, Paul (2006): “How large a vocabulary is needed for reading and listening?”, en *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63/1: pp. 59-82.
- ROJO, Guillermo (2011): “Frecuencia de inventario y frecuencia de uso”, en *Revista Española de Lingüística*, 41/1: pp. 5-43.
- (2023): “Hacia un nuevo concepto de diccionario de frecuencias”, en Dolores Corbella, Josefa Dorta y Rafel Padrón (eds.), *Perspectives de recherche en linguistique et philologie romanes*. Strasbourg: ELIPHI, pp. 45-63.

CORPUS MENCIONADOS EN EL TEXTO

- CORGA = Centro Ramón Piñeiro para a Investigación en Humanidades. Corpus do Galego Actual. Dirs. Guillermo Rojo y María Sol López Martínez, <<http://www.cirp.gal/corga/>>. Versión 4.1.
- CORPES XXI = Real Academia Española y Asociación de Academias de la lengua española. Corpus del Español del Siglo XXI, <<http://rae.es/recursos/banco-de-datos/corpes-xxi>>. Versión 1.1.
- ESLORA = Corpus para el Estudio del Español Oral. Dir. Victoria Vázquez Rozas, <<http://eslora.usc.es/>>. Versión 2.2.