

X Congreso Asiático de Hispanistas
Universidad Hankuk de Estudios Extranjeros, Corea del Sur
(25-26 de junio de 2021)

Aspectos estadísticos del español oral
Análisis del corpus sociolingüístico de
Santander (España)

Inmaculada Martínez (CIESE- Comillas)

Hiroto Ueda (Universidad de Tokio)

Objetivos

GENERAL

- Acercamiento a la variación lingüística del español :

A. Corpus de lengua oral

B. Estilometría

ESPECÍFICOS

- Caracterización lingüística cuantitativa del corpus PRESEEA-Santander.

- Análisis estadísticos (Léxico, Gramática y Morfología)

Necesidad

- **Objetividad** en los análisis lingüísticos y estilísticos.
- **Auge en el mundo hispánico** (/ámbito anglosajón).

Hipótesis

Rasgos gramaticales

- Tendencia caracterizada por el alto grado de regularidad.

Rasgos léxicos

- Presentan frecuencias peculiares cuya regularidad es limitada.

GUIÓN

1. Objetivos
- 2. Marco Teórico**
3. Metodología: muestra y materiales
4. Discusión de los datos
5. Conclusiones

Marco teórico

- Sociolingüística
- Estilometría

GUION

1. Objetivos
2. Marco Teórico
- 3. Metodología: muestra**
4. Discusión de los datos
5. Conclusiones

El proyecto PRESEEA

- 1993: nace PRESEEA
- Corpus sociolingüísticos sincrónicos
- Lengua española hablada
- Variedad geográfica y social
- Núcleos urbanos
- 50 equipos de investigación
- Metodología común



PRESEEA
PROYECTO PARA EL ESTUDIO SOCIOLINGÜÍSTICO DEL ESPAÑOL DE ESPAÑA Y DE AMÉRICA

Inicio Noticias Equipos Metodología Corpus Contacto

Presentación

PRESEEA es un proyecto para la creación de un corpus de lengua española hablada representativo del mundo hispánico en su variedad geográfica y social. Esos materiales se reúnen atendiendo a la diversidad sociolingüística de las comunidades de habla hispanohablantes.

PRESEEA agrupa a cerca de 40 equipos de investigación sociolingüística. Es el fruto del trabajo coordinado de investigadores comprometidos con una metodología común para reunir un banco de materiales coherente que posibilite su aplicación con fines educativos y tecnológicos.

WHAT IS PRESEEA?
PRESEEA is the "Project for the Sociolinguistic Study of Spanish from Spain and America". The goal is to coordinate sociolinguistic researchers from Spain and the Hispanic America in order to make possible comparisons between different studies and materials, as well as a basic information exchange. The main aim is to create a spoken language Corpus with sufficient guarantees and rich in terms of linguistic information.

Equipos

Universidad de Alcalá

HC IAS

REAL ACADEMIA ESPAÑOLA

CIESE

<https://preseea.linguas.net/>

El corpus PRESEEA-Santander

- 2013: nace PRESEEA-Santander
- Requisitos mínimos

	Generación 1		Generación 2		Generación 3	
	H	M	H	M	H	M
Grado de instrucción 1	3	3	3	3	3	3
Grado de instrucción 2	3	3	3	3	3	3
Grado de instrucción 3	3	3	3	3	3	3

Muestra: 1 en cada celda

Santander, ciudad

SOCIAL:

- 173 375 habitantes
- meseta y mar
- sector terciario
- ciudad del siglo XX: joven
- población envejecida



Santander, ciudad

LINGÜÍSTICA:

- conciencia de habla
- homogeneidad



Módulo	Variables morfosintácticas
Saludos	Formas de tratamiento, atención al uso de <i>tú / usted</i> .
El tiempo	Usos de los verbos <i>haber</i> y <i>hacer</i> .
Lugar donde vive	Usos de los verbos <i>haber</i> y <i>estar</i> .
Familia y amistad	Usos de los verbos <i>ser</i> y <i>estar</i> .
Costumbres	Impersonalidad.
Peligro de muerte	Valores y usos del sistema verbal: indicativo, subjuntivo / condicionales
Anécdotas importantes en la vida	Valores y usos del sistema verbal: presente / pretérito con valor de pasado.
Deseo de mejora económica	Valores y usos del sistema verbal: indicativo, subjuntivo / condicionales

GUION

1. Objetivos
2. Marco Teórico
3. Metodología: muestra
- 4. Discusión de los datos**
5. Conclusiones

Análisis estadístico

Sexo:	H: hombre			H.total	M: Mujer			M.total	Total
Edad: Nivel	E1	E2	E3		E1	E2	E3		
Nivel: N1	6 380	6 395	8 160	20 935	6 347	6 971	5 141	18 459	39 394
Nivel: N2	4 250	8 319	7 633	20 202	3 337	10 278	5 459	19 074	39 276
Nivel: N3	7 699	5 437	5 888	19 024	4 343	3 710	9 352	17 405	36 429
Total	18 329	20 151	21 681	60 161	14 027	20 959	19 952	54 938	115 099

- $FN = FA \text{ (frecuencia absoluta)} * 100\,000 / \text{Suma}$

Análisis estadístico

3.1. Verbo

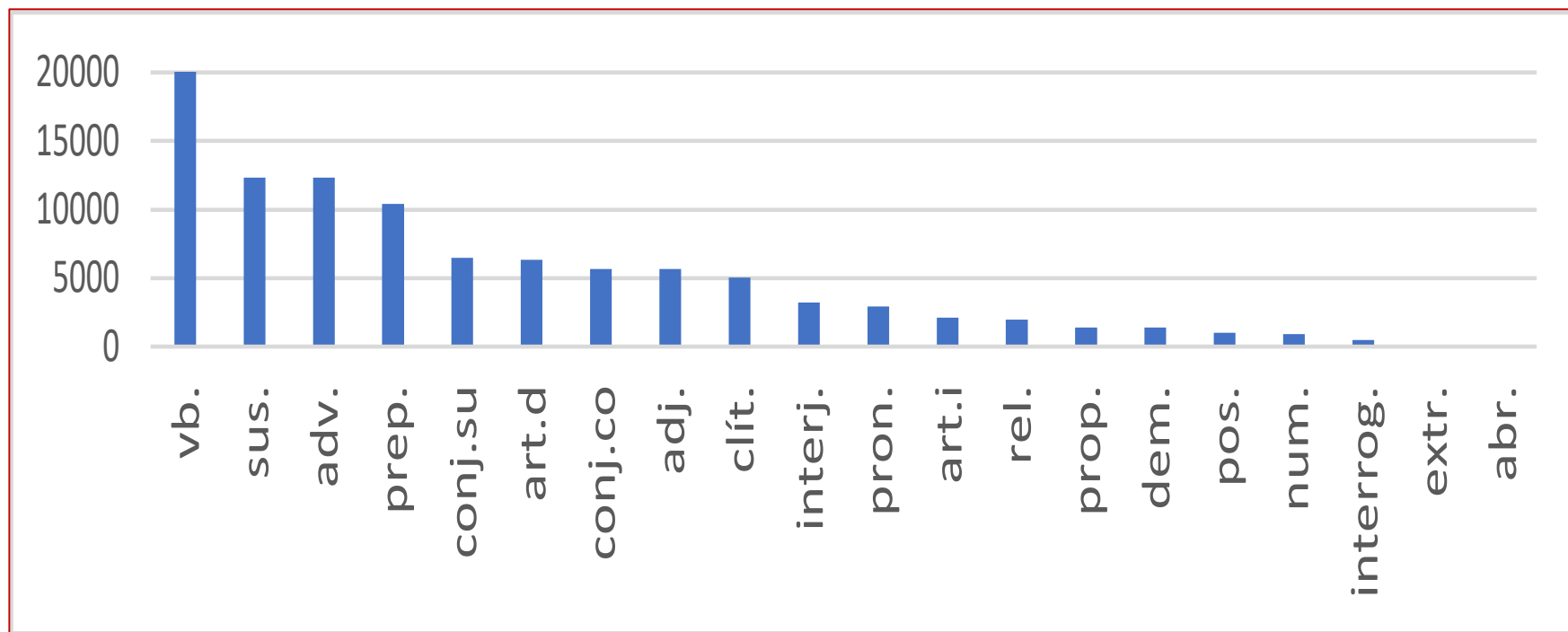


Fig. 1. Categorías gramaticales (total).

Análisis estadístico

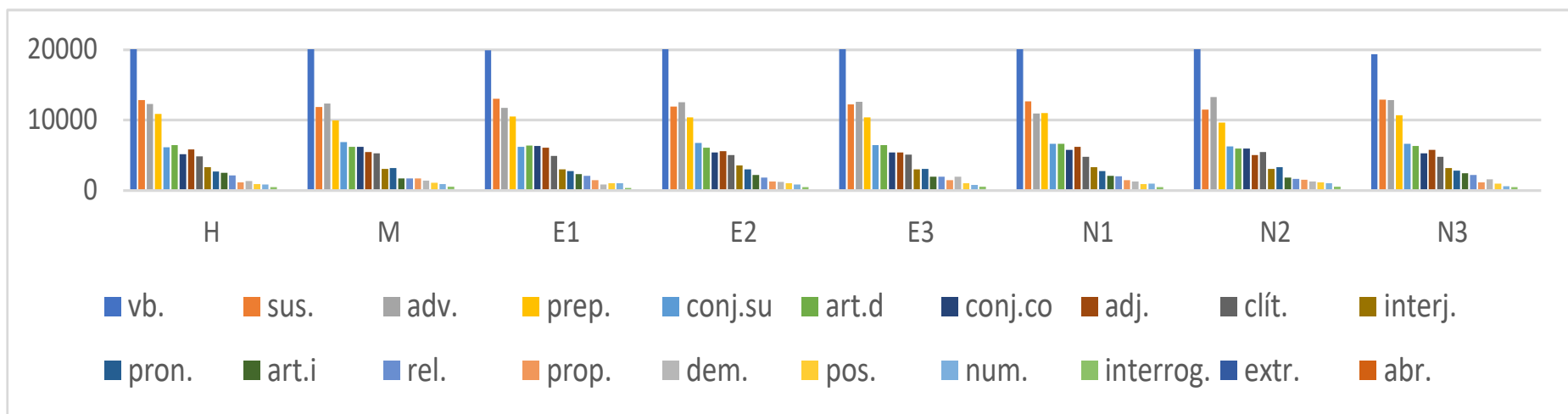


Fig. 2. Categorías gramaticales (grupos).

Casi total coincidencia de las curvas descendentes en todos los grupos de sexo (H, M), edad (E1, E2, E3) y nivel educativo (N1, N2, N3).

Análisis estadístico

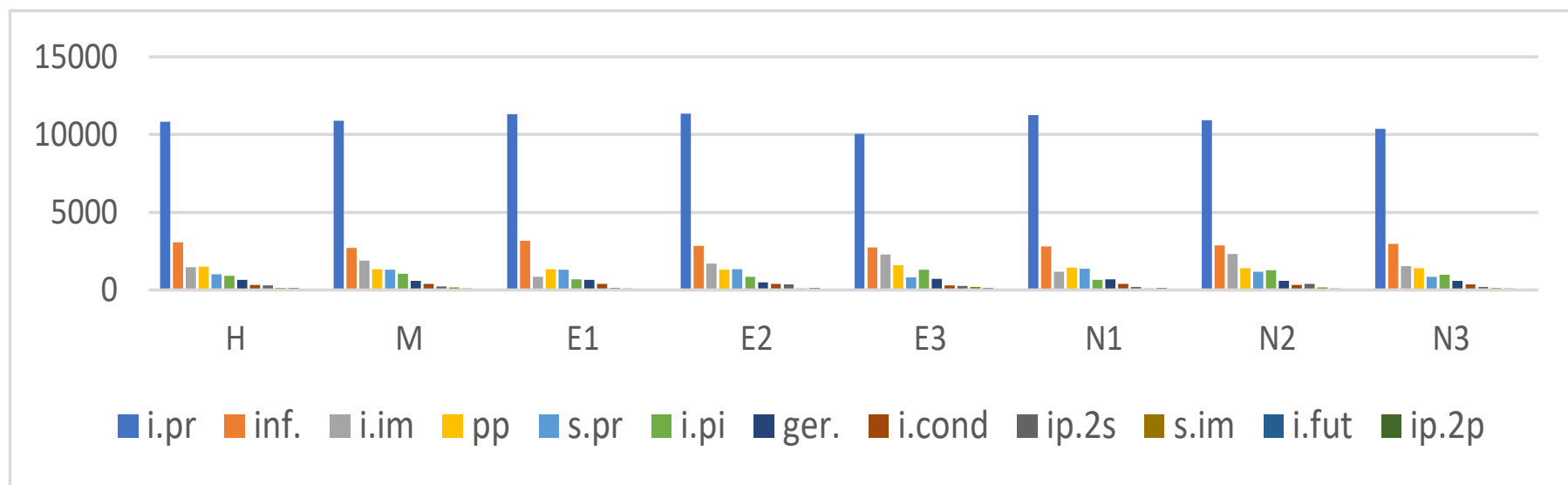


Fig. 3. Verbos. Modo y tiempo.

- Orden de frecuencia de [i.pr] (indicativo, presente) → [i.im] (indicativo, imperfecto) → [s.pr] (subjuntivo, presente) → [s.im] (subjuntivo, imperfecto)
- Carácter bímembre de {modo} (indicativo, subjuntivo) y {tiempo} (presente, imperfecto)

Análisis estadístico

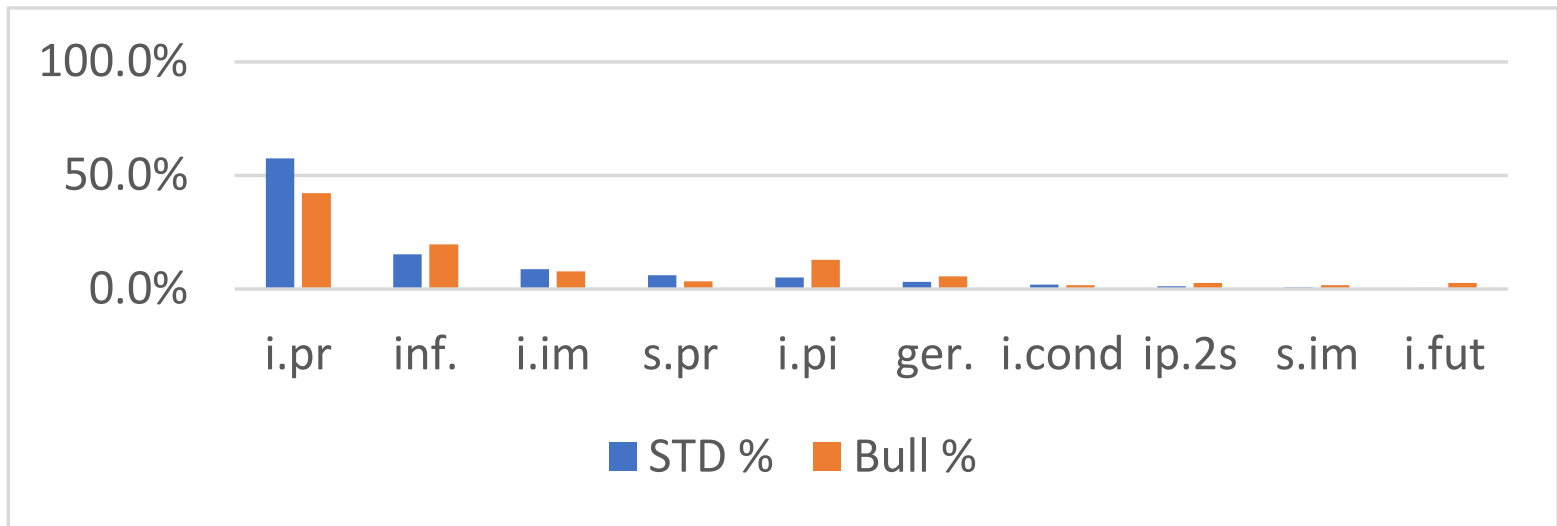


Fig. 4. Verbos. Persona y número. PRESEEA-Santander (STD) y Bull (1947).

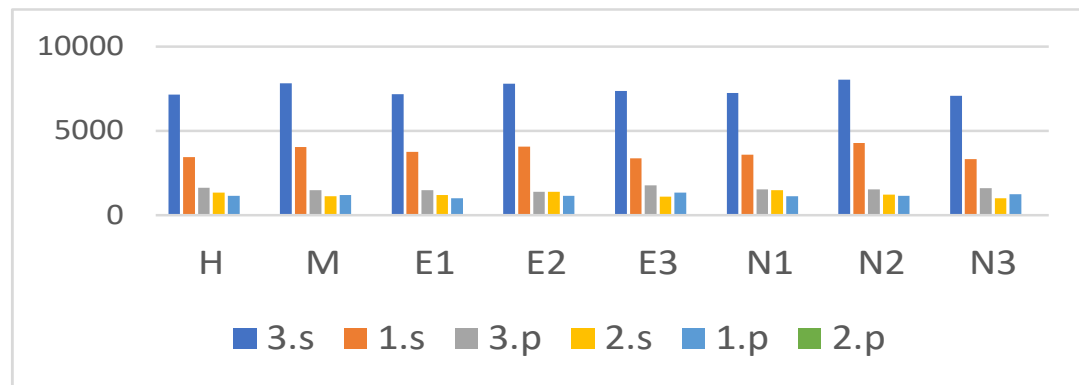


Fig. 5. Verbos. Persona y número.

Análisis estadístico

- La tercera persona (término no marcado) a costa de la primera y la segunda (términos marcados).
- La forma singular por encima de la plural.
- La composición binaria de dos morfemas {persona} y {número}, antes que la unitaria de ambas categorías {persona-número} (RAE y AALE 2009: 182).
- Tanto el orden de persona (3, 1, 2) como el de número (singular y plural) son casi constantes entre los textos.
- El carácter bimembre de {persona} (3, 1, 2) y {número} (singular, plural), en lugar del conjunto {persona-número} (3s, 1s, ...).

Análisis estadístico

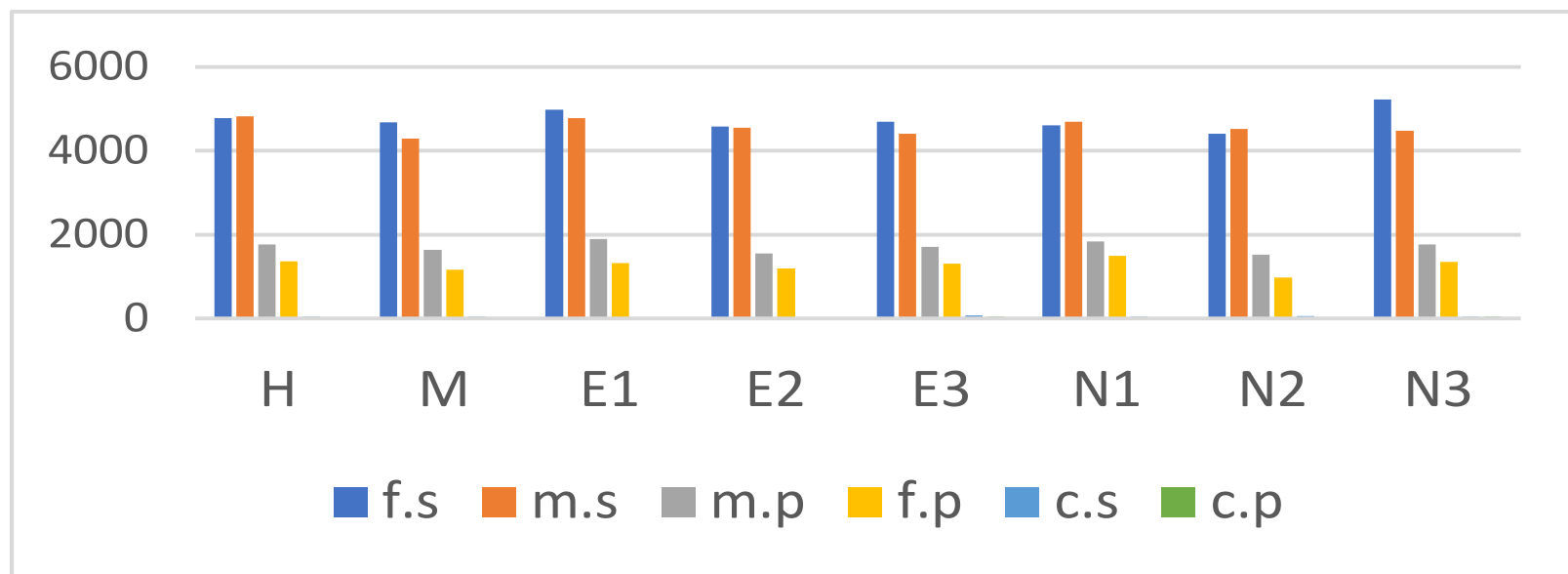


Fig. 6. Sustantivo. Género y número.

- Los dos géneros, masculino y femenino, compiten en sus frecuencias, casi igualadas.

Análisis estadístico

- Esta igualdad numérica de uso entre los sustantivos femeninos y los masculinos se debe preferentemente a la característica léxica de género, más que a su génesis gramatical.
- El cambio de género, *hijo* → *hija*, no se realiza en la flexión gramatical, sino en la derivación léxica, lo mismo que *rey* → *reina*, *actor* → *actriz*, *libro* → *librería*.
- Las bases de derivación, *hijo*, *rey*, *actor*, *libro*, que son términos no marcados, suelen ser más frecuentes que las derivadas, *hija*, *reina*, *actriz*, *librería*, términos marcados por la sufijación.
- A pesar de la diferencia de frecuencia, el léxico total se divide en dos géneros equitativamente.
- En cambio, el número del sustantivo, singular y plural, sí que es una cuestión de carácter gramatical y siempre la forma singular, el término no marcado, es más frecuente que la plural, el término marcado. A nue

Análisis estadístico

3.3. Adjetivo

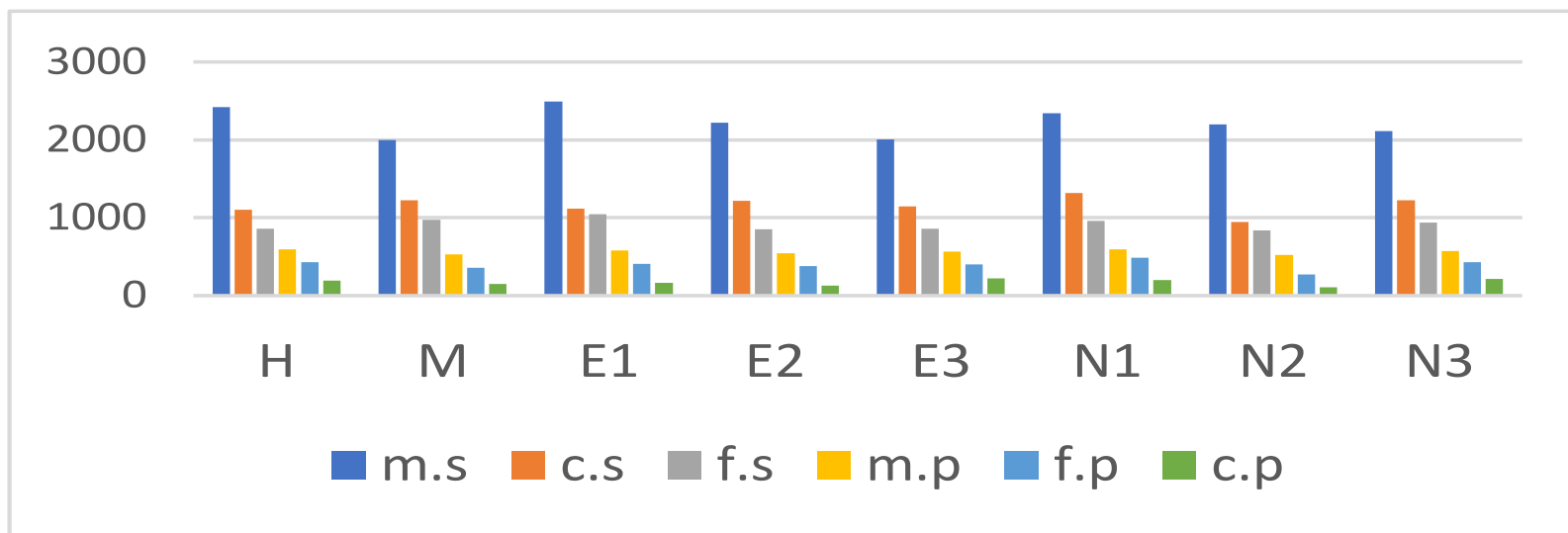


Fig. 7. Adjetivo. Género y número.

Análisis estadístico

- El género del adjetivo no es inherente al léxico, sino que es concordante con el sustantivo al que modifica, de modo que es un elemento gramatical.
- El mayor uso del género masculino con respecto al femenino se debe al hecho de que el masculino es un término no marcado, normal de género, frente al femenino, término marcado, especial de género (RAE y AALE 2009: 85-94).
- El género femenino marcado aparece solo cuando modifica al sustantivo femenino, y en todos los demás casos, aparece el masculino no marcado, por ejemplo, *bueno* en el contexto de *lo ___ que*, o *es ___ que*, etc., además del concordante con el sustantivo masculino: *libro ___*.
- Se observa la homogeneidad perfecta de frecuencias de distintas formas de género entre los textos, lo que es natural por tratarse de variantes gramaticales.

Análisis estadístico

• 3.4. Léxico

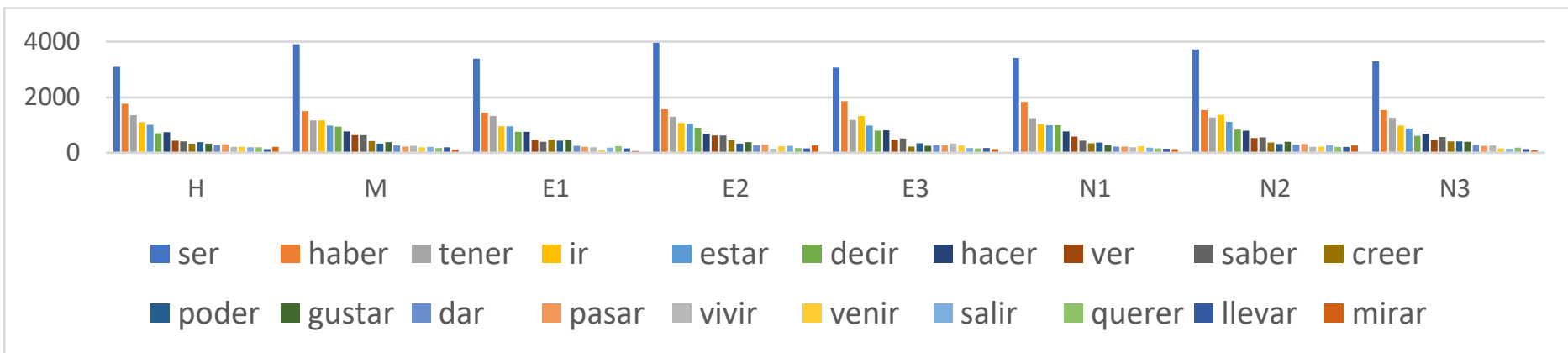


Fig. 8. Verbos frecuentes.

- Es de destacar la coincidencia casi perfecta de las curvas descendentes que se observa entre los distintos tipos de textos.
- La razón de la coincidencia suponemos encontrarla en las propias características del léxico.

Análisis estadístico

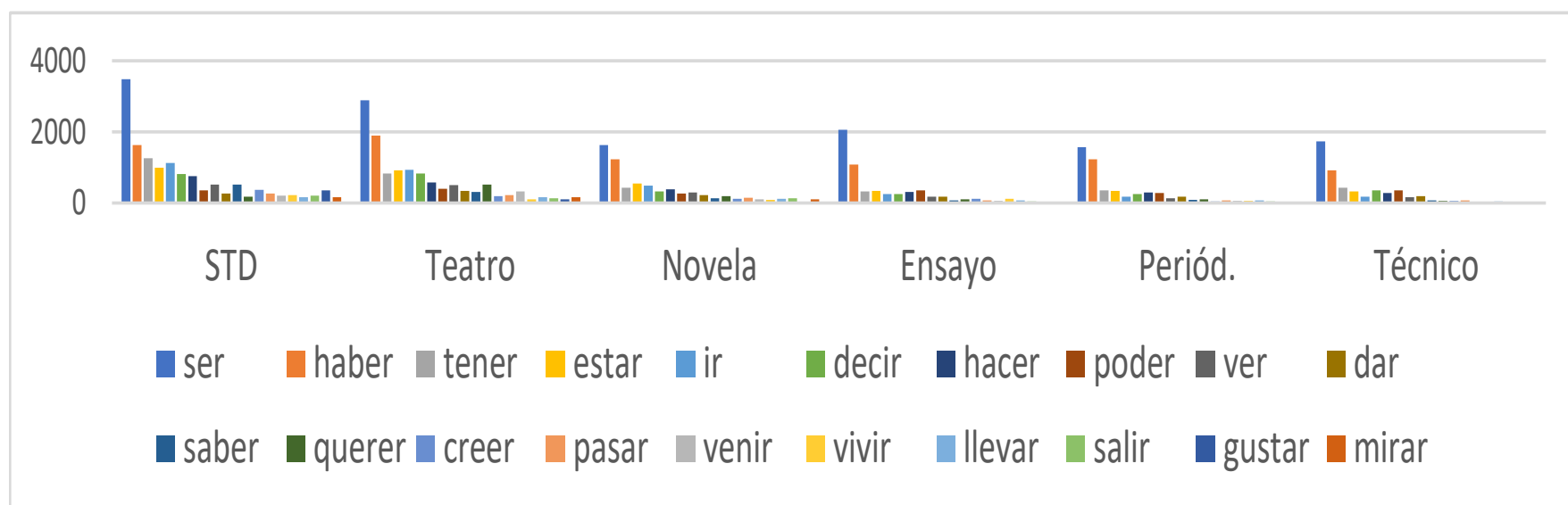


Fig. 9. Verbos frecuentes. PRESEEA-Santander (STD) / Juilland y Chang-Rodríguez (1964).

Los tres verbos más frecuentes, *ser*, *haber* y *tener*, son siempre frecuentes sin distinción de tipos de textos, por ser útiles y necesarios para todo propósito de expresión.

Análisis estadístico

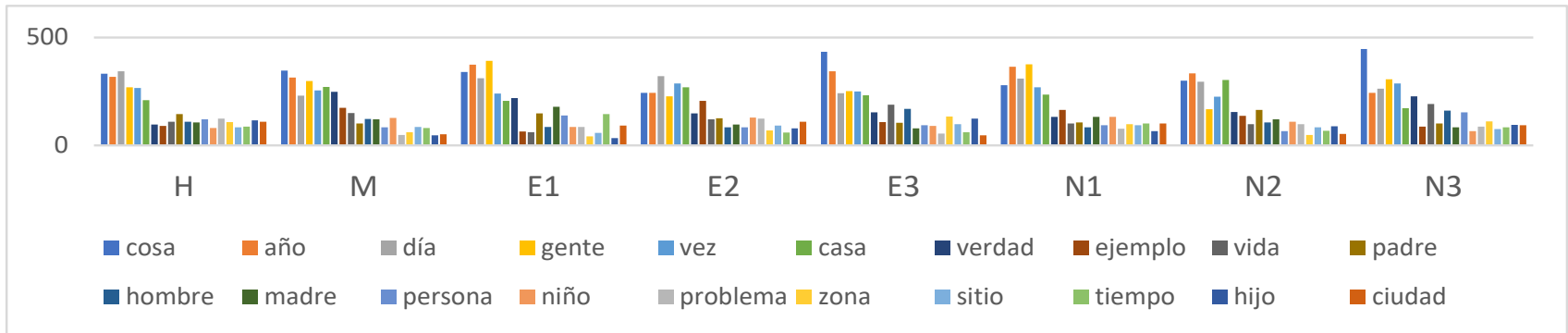


Fig. 10. Sustantivos frecuentes.

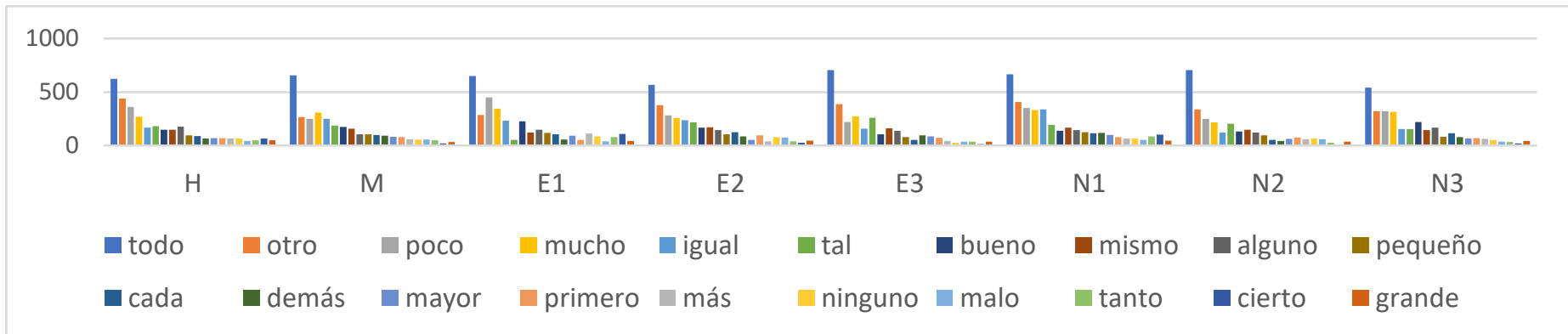


Fig. 11. Adjetivos frecuentes.

Análisis estadístico

- Los verbos y adjetivos más frecuentes son los vocablos instrumentales, necesarios, en general, para todo tipo de expresión, mientras que los sustantivos, frecuentes o poco frecuentes, son vocablos temáticos específicos cuyo uso depende mayoritariamente del tema tratado en los diálogos o textos de distintos tipos.
- Creemos que la distinción entre el léxico instrumental y el léxico temático es importante en el análisis práctico de la lengua.

Conclusiones

- El corpus oral estratificado de PRESEEA-Santander.
- Una aproximación desde el punto de vista de la lingüística cuantitativa práctica en lugar de la cualitativa teórica.
- Basándonos en la práctica de la lengua, hemos apoyado la hipótesis sobre el género léxico derivacional del sustantivo y el género flexivo del adjetivo y la interpretación bimembre de {modo} y {tiempo}, {persona} y {número} de las desinencias verbales.
- La instrumentalidad léxica manifestada en los verbos y adjetivos frecuentes.
- Vocablos instrumentales residen en su alto grado de utilidad a través de diferentes tipologías de texto.
- Los vocablos temáticos, sustantivos y palabras de contenido de reducida frecuencia y de limitada dispersión, son empleados en cada texto en particular.

Bibliografía

- Alarcos Llorach, Emilio (1994): *Gramática de la lengua española*, Madrid, Espasa Calpe.
- Ambadiang, Théophile (1993): *La morfología flexiva*, Madrid, Santillana.
- _____. (1999): “La flexión nominal. Género y número”, en Ignacio Bosque y Violeta Demonte (dir.) *Gramática descriptiva de la lengua española*, vol 3. 4843-4913.
- Ayuntamiento de Santander (2020): *Plan estratégico Santander 2010-2020*, recuperado en <http://www.planestrategicosantander.com/documentos.php>
- Bull, William E. (1947): “Modern Spanish Verb-Form Frequencies”, *Hispania* (American Association of Teachers of Spanish and Portuguese), vol. 30, pp. 451-466.
- Davies, Mark (2006): *A Frequency Dictionary of Spanish*. New York, Routledge.
- García Hoz, Víctor (1951): *Vocabulario usual, vocabulario común y vocabulario fundamental*. Madrid, Consejo Superior de Investigaciones científicas.
- Juilland, Alphonse and E. Chang-Rodríguez (1964): *Frequency Dictionary of Spanish Words*, The Hague, Mouton.
- López Morales, Humberto (1993): *Sociolingüística*. 2ª ed., Madrid, Gredos.
- Moreno Fernández, Francisco (2021): *Metodología del “Proyecto para el estudio sociolingüístico del español de España y América” (PRESEEA)*. Universidad de Alcalá de Henares.
- Moreno Fernández, Francisco y Hiroto Ueda (1986): “El género en los sustantivos del español: Sobre su naturaleza gramatical”, *Boletín de la Academia Puertorriqueña de la Lengua Española*, vol. 14, 79-109.
- RAE y AALE. Real Academia Española y Asociación de Academias de la Lengua Española (2009): *Nueva gramática de la lengua española. Morfología, Sintaxis I*, Madrid, Espasa Libros.

¡Muchas gracias!

martinezi@fundacioncomillas.es

uedahiroto@jcom.home.ne.jp

X Congreso Asiático de Hispanistas
Universidad Hankuk de Estudios Extranjeros, Corea del Sur
(25-26 de junio de 2021)

Aspectos estadísticos del español oral
Análisis del corpus sociolingüístico de
Santander (España)

Inmaculada Martínez (CIESE- Comillas)

Hiroto Ueda (Universidad de Tokio)