

La ratio de categoría gramatical en español. Rasgos estilísticos y sociolingüísticos

Inmaculada Martínez-Martínez. Universidad de Cantabria

Hiroto Ueda. Universidad de Tokio

Plan de trabajo

1. Introducción
2. Marco teórico
3. Métodos estadísticos
4. Análisis
5. Conclusión

1. Introducción

- Escasez de estudios de RCG en Sociolingüística
 - Estilística literaria, análisis de autor o géneros discursivos.
 - RCG no muestra, *a priori*, variación sociolingüística.
- Corpus de 54 entrevistas (sexo, edad, nivel de instrucción) de PRESEEA-Santander.

2. Marco teórico

A. Lingüística de corpus:

- Biber (1988): primer acercamiento al análisis estadístico de las categorías gramaticales
- Koskela (1998): uso de frases nominales, más frecuente en textos de especialidad.
- Rojo (2021): en CORPES, verbos (+) / sustantivos, adjetivos y adverbios (-) con los años

B. Estructura cuantitativa de CG:

- Zipf (1936): clasificación de la CG en dos grupos: palabras funcionales / contenido.

C. Estilometría:

- Hatano (1965): Yokomitsu y Tanizaki – menos nombres el primero y verbos, igual.

D. Sociolingüística:

- Rangel y Rosso (2013): uso de las CG por género en comentarios de Facebook. Hombres usan más preposiciones / mujeres usan más pronombres e interjecciones.

3. Métodos estadísticos

1. Frecuencia ajustada:

- Frecuencia absoluta – Frecuencia relativa – Porcentaje – **Frec. relat. ajustada** ($FA = R / \text{sum}(R) * \text{sum}(D)$)

2. Probabilidad estandarizada:

- Estandarización: se aproxima a la distribución normal estándar (M: 0, SD: 1)
- Mérito: probabilidad constante, no afectada por la magnitud de los datos.

3. Concentración diagonal:

- los valores altos se concentran en la zona diagonal de la matriz de datos

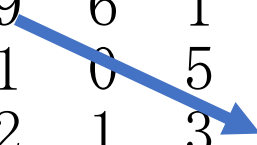
Concentración diagonal

$D = \text{Matrix}(c(1, 3, 2, 5, 0, 5, 1, 4, 6, 1, 9, 3), 3, b=T)$; D # Data matrix ... (9)

	c1	c2	c3	c4
r1	1	3	2	5
r2	0	5	1	4
r3	6	1	9	3

➤ $D_c = DC(D)$; D_c

	c3	c1	c2	c4
r3	9	6	1	3
r2	1	0	5	4
r1	2	1	3	5



- Reordenación de filas y columnas
- Movimiento de las frecuencias en horizontal y vertical al mismo tiempo

4. Análisis

4.1. Datos

4.2. Parámetros (sexo, edad, nivel educativo)

4.3. Totalidad

4.1. Datos

n: *año* (1,169), *gente* (1,030), *cosa* (991), *día* (982), *casa* (898), *vez* (786), *Santander* (677), *hombre* (534), *verdad* (505), *padre* (491)... [N: 4,494; Total: 48,233].

nm: *dos* (640), *tres* (314), *cuatro* (267), *primero* (215), *cinco* (188), *diez* (129), *ocho* (125), *siete* (122), *seis* (118), *veinte* (110)... [N: 57; Total: 3,541].

p: *de* (10,417), *a* (8,034), *en* (6,104), *por* (2,768), *con* (2,561), *para* (1,946), *al* (1,435), *del* (870), *hasta* (387), *como* (366)... [N: 26; Total: 36,143].

4.2. Análisis por parámetros (Sexo)

- Resultado de la concentración diagonal de frecuencia relativa:

Rc	n	p	ar	a	ij	r	nm	ir	av	pn	c	v
H	0.145	0.110	0.083	0.075	0.022	0.020	0.011	0.004	0.127	0.086	0.116	0.201
M	0.140	0.104	0.075	0.071	0.017	0.019	0.010	0.005	0.128	0.095	0.123	0.213

Fig. 11. Frecuencia relativa concentrada (P). Sexo.

- Movimiento de las categorías gramaticales.
- Gradación continua: de mayor a menor

4.2. Análisis por parámetros (Edad)

El orden de categorías gramaticales es:

A	a	ar	av	c	ij	ir	n	nm	p	pn	r	v
E1	0.872	0.854	0.835	0.714	0.636	0.433	0.133	0.187	0.205	0.295	0.281	0.275
E2	0.345	0.452	0.521	0.722	0.782	0.859	0.797	0.424	0.387	0.269	0.283	0.289
E3	0.230	0.175	0.153	0.124	0.130	0.183	0.610	0.860	0.867	0.876	0.876	0.876

Fig. 16. Probabilidad concentrada (C). Edad.

4.2. Análisis por parámetros (Nivel educativo)

*	c	nm	ir	v	pn	av	p	r	ij	ar	a	n
N1	0.862	0.816	0.802	0.761	0.721	0.157	0.669	0.415	0.160	0.352	0.291	0.279
N2	0.415	0.570	0.600	0.668	0.715	0.840	0.126	0.191	0.496	0.225	0.273	0.285
N3	0.191	0.141	0.135	0.126	0.124	0.506	0.760	0.862	0.842	0.872	0.876	0.876

Fig. 21. Probabilidad concentrada (C). Nivel de educación.

4.3. Análisis de la totalidad

H: 6 / M: 3

N3: 5 / N1y2: 2

M: 6 / H: 3

N1y2: 4 / N3: 1

C	a	ij	r	av	ar	n	p	nm	c	ir	pn	v
H-E2-N3	0.792	0.999	0.132	0.961	0.330	0.285	0.258	0.597	0.157	0.122	0.137	0.039
H-E1-N3	0.948	0.470	0.856	0.427	0.718	0.513	0.373	0.248	0.831	0.084	0.102	0.308
M-E2-N3	0.949	0.446	0.841	0.152	0.808	0.895	0.900	0.530	0.355	0.098	0.121	0.247
M-E3-N2	0.807	0.901	0.566	0.764	0.195	0.410	0.131	0.434	0.272	0.357	0.711	0.198
H-E2-N1	0.790	0.580	0.974	0.119	0.852	0.830	0.874	0.543	0.376	0.755	0.197	0.217
H-E3-N3	0.463	0.873	0.887	0.099	0.973	0.980	0.927	0.091	0.125	0.621	0.102	0.150
H-E1-N2	0.775	0.287	0.356	0.701	0.379	0.739	0.905	0.123	0.500	0.192	0.206	0.227
H-E1-N1	0.757	0.234	0.624	0.079	0.812	0.497	0.865	0.905	0.859	0.078	0.440	0.406
M-E1-N3	0.661	0.196	0.582	0.360	0.770	0.843	0.533	0.330	0.780	0.297	0.162	0.446
H-E3-N1	0.377	0.277	0.413	0.510	0.686	0.567	0.816	0.873	0.246	0.489	0.478	0.531
M-E1-N2	0.282	0.302	0.380	0.930	0.050	0.677	0.342	0.516	0.427	0.621	0.482	0.547
M-E3-N3	0.366	0.254	0.777	0.862	0.684	0.511	0.294	0.012	0.109	0.755	0.782	0.418
H-E3-N2	0.180	0.600	0.582	0.761	0.562	0.081	0.142	0.816	0.359	0.804	0.876	0.569
H-E2-N2	0.320	0.298	0.252	0.648	0.658	0.574	0.621	0.434	0.048	0.455	0.697	0.747
M-E2-N1	0.292	0.323	0.396	0.386	0.147	0.146	0.568	0.237	0.780	0.522	0.753	0.892
M-E3-N1	0.114	0.264	0.380	0.821	0.131	0.034	0.062	0.833	0.809	0.986	0.922	0.810
M-E2-N2	0.045	0.488	0.061	0.252	0.127	0.101	0.102	0.734	0.972	0.837	0.807	0.983
M-E1-N1	0.072	0.377	0.019	0.115	0.131	0.335	0.204	0.935	0.888	0.857	0.959	0.954

Fig. 23. Probabilidad concentrada. Totalidad.

5. Conclusión

- Aporte de métodos estadísticos fiables y permanentes
- Cierta variación según los parámetros tratados
- Escala de elevación:

<más elevado> a (adjetivo) ij (interjección), r (relativo), av (adverbio),
ar (artículo), n (nombre), p (preposición), nm (numeral), c (conjunción),
ir (interrogativo), pn (pronombre), v (verbo): <menos elevado>.

- Aunque puede variar, se mantendría esta escala *estándar*
- Gran cantidad de datos
- Realidad lingüística fiable

Referencias

- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge University Press.
- Hatano, Kanji. 1966. "Psicología de estilo" (en japonés). en *Introducción a la estilística*. Tokio. Sanseido. pp. 22-47.
- Justicia, Fernando. 1995. *El desarrollo del vocabulario. Diccionario de frecuencias*. Universidad de Granada.
- Koskela, Merja. 1998. "Verbs and Noun Phrases - Two Tendencies in Philosophical Essays", Lukka, M.-R., S. Salla and H. Dufva (eds.). *Puolin ja toisin*. The Finnish Association for Applied Linguistics Yearbook 1998. Publications of the Finnish Association for Applied Linguistics, No. 56. Jyväskylä. pp. 159-170.
- Martínez Martínez, Inmaculada y Ueda, Hiroto. 2025. *Inventario léxico de PRESEEA-Santander*. Proyecto LYNEAL. <https://h-ueda.sakura.ne.jp/lyneal/il/std/>
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Rangel, Francisco and Paolo Rosso. 2013. "Use of Language and Author Profiling: Identification of Gender and Age", *Natural Language Processing and Cognitive Science*, 177, 56-66
- Rojo, Guillermo. 2021. *Introducción a la lingüística de corpus en español*. London. Routledge.
- Rangel, Francisco and Paolo Rosso. 2013. "Use of Language and Author Profiling: Identification of Gender and Age", *Natural Language Processing and Cognitive Science*, 177, 56-66
- Rojo, Guillermo. 2021. *Introducción a la lingüística de corpus en español*. London. Routledge.
- Zipf, George Kingsley. 1936. *The Psycho-biography of Language. An Introduction to Dynamic Philology*. London. Routledge.

Muchas gracias

La ratio de categoría gramatical en español.

Rasgos estilísticos y sociolingüísticos

Inmaculada Martínez-Martínez. Universidad de Cantabria

Hiroto Ueda. Universidad de Tokio