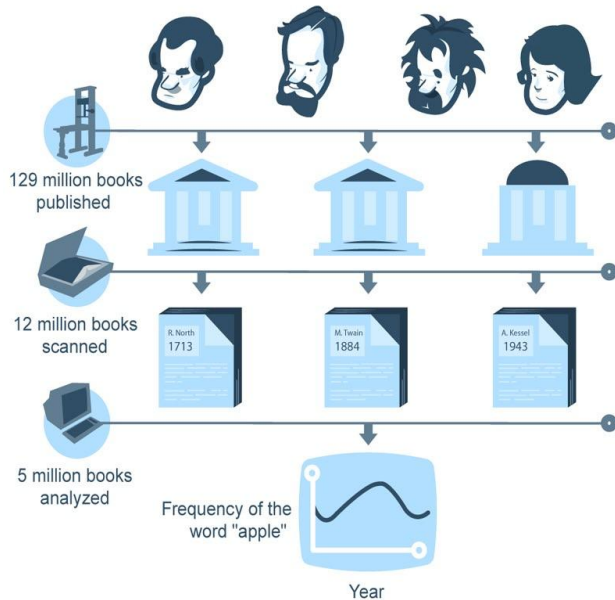


Procesamiento del Lenguaje Natural como ventana a la mente y a la sociedad

Edgar Altszyler
Programa de Ciencia de Datos
<http://www.fundacionsadosky.org.ar/programas/pcd/>

Culturomics: Análisis cuantitativo de la cultura a gran escala

Google n-grams, 15 millones de libros escaneados



Michel, J.B. et. Al, Quantitative analysis of culture using millions of digitized books, Science, (2011)

La matemática de las palabras

¿25?

0



100

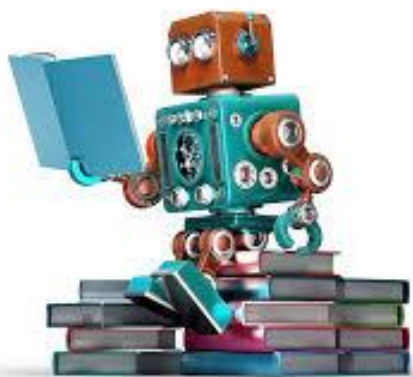
¿felicidad?

amor

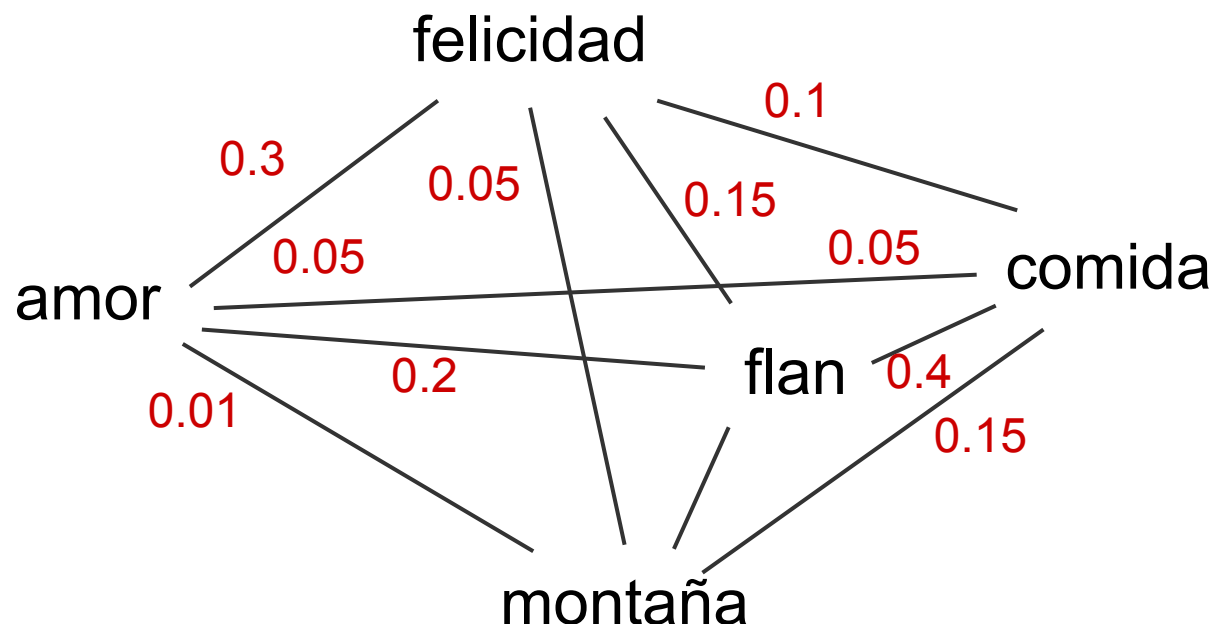


comida

Similaridad Semántica



Corpus de textos



Point-Wise Mutual Information (PMI)



WIKIPEDIA
The Free Encyclopedia

sliding window (size=2)

Choripan. The Argentine *choripán* consists of a sausage made out of beef and pork, hot off the grill, split down the middle, and served on a roll. The *chorizo* may be used whole or cut in half lengthwise, in which case it is called a *mariposa* (butterfly). It is customary to add sauces on the bread, most likely *chimichurri*.

Church and Hanks (1990) Word Association Norms, Mutual Information, and Lexicography

Point-Wise Mutual Information (PMI)

Choripán Chimichurri

$$\text{PMI}(w, c) = \log \left(\frac{P(w, c)}{P(w) \cdot P(c)} \right)$$



WIKIPEDIA
The Free Encyclopedia

sliding window (size=2)

Choripan. The Argentine *choripán* consists of a sausage made out of beef and pork, hot off the grill, split down the middle, and served on a roll. The *chorizo* may be used whole or cut in half lengthwise, in which case it is called a *mariposa* (butterfly). It is customary to add sauces on the bread, most likely *chimichurri*.

Church and Hanks (1990) Word Association Norms, Mutual Information, and Lexicography



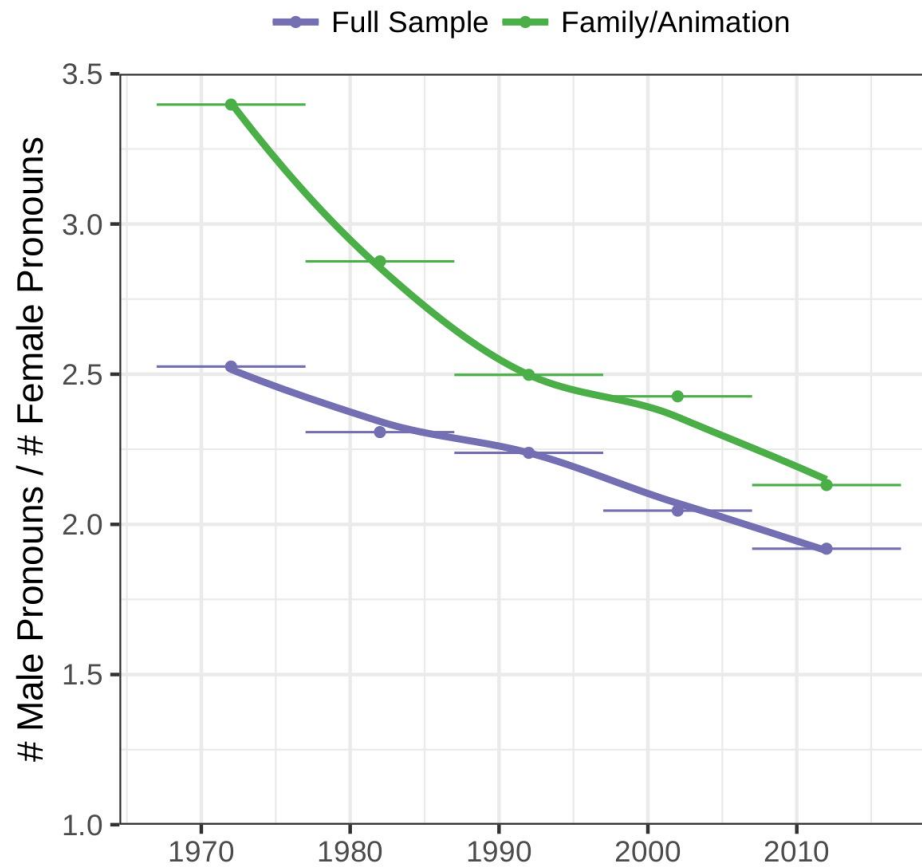
Observatorio del Cine



Más de 11.000
subtítulos



Estereotipos de Género en el Cine



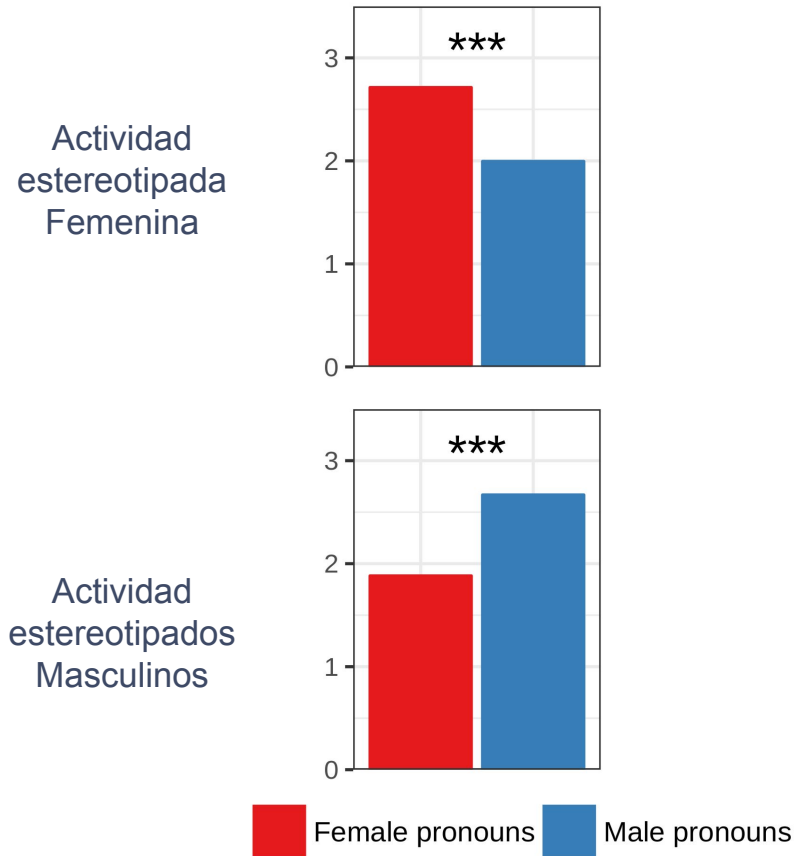
Gálvez R., Tiffenberg V. and Altszyler E. (2019) Half a century of stereotyping associations between gender and intellectual ability in films

Estereotipos de Género en el Cine

- **Feminine** stereotyping roles
dancer, decorator, designer, dietician, florist, homemaker, housekeeper, model, nanny, typist...
- **Masculine** stereotyping roles
engineer, programmer, physicist, architect, detective, pilot, firefighter, inventor, mechanic, officer...

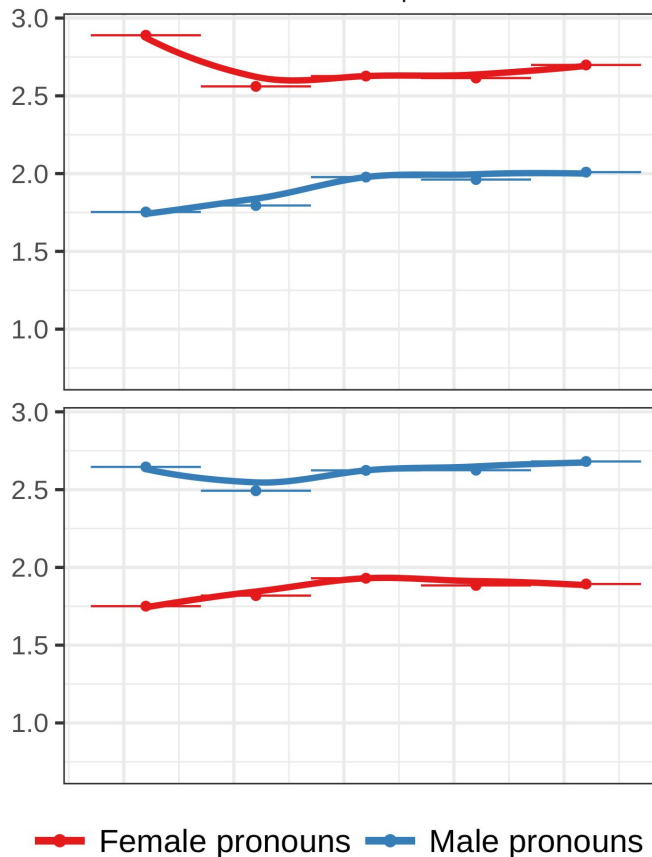
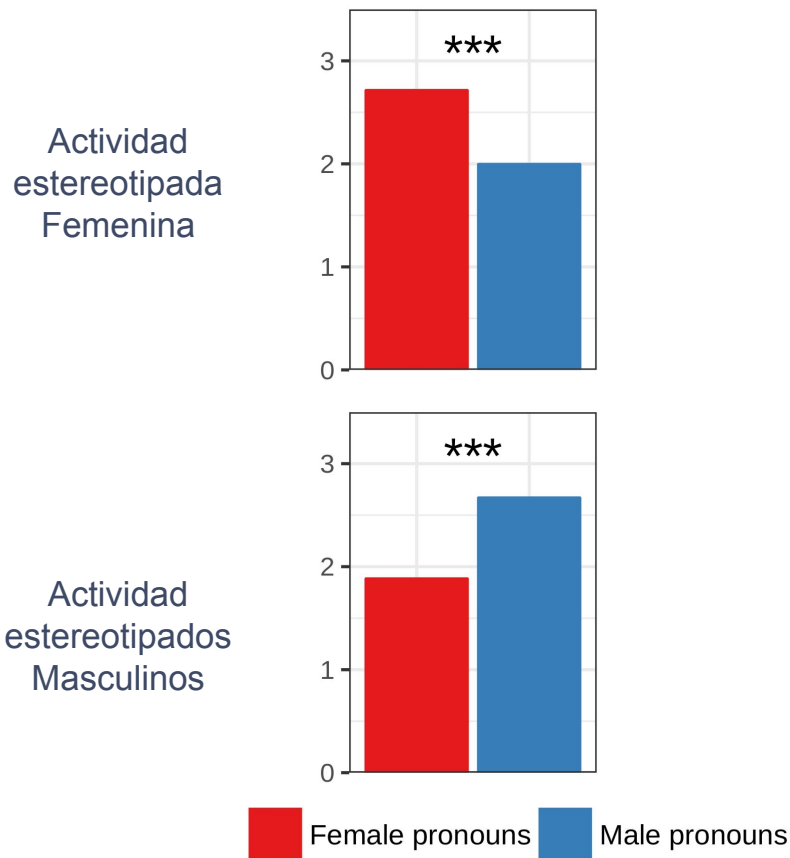
A. P. Lenton, C. Sedikides, M. Bruder, Sex Roles 60, 269 (2009).

Estereotipos de Género en el Cine



Gálvez R., Tiffenberg V. and Altszyler E. (2019) Half a century of stereotyping associations between gender and intellectual ability in films

Estereotipos de Género en el Cine



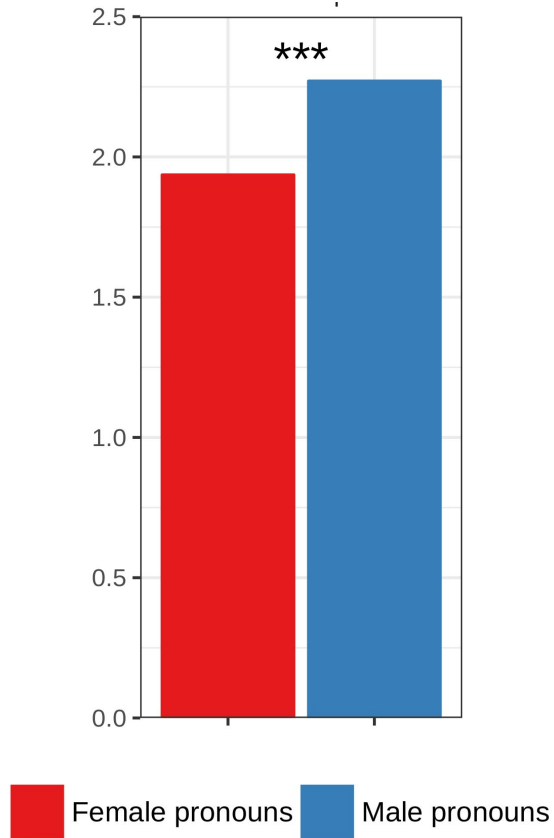
Gálvez R., Tiffenberg V. and Altszyler E. (2019) Half a century of stereotyping associations between gender and intellectual ability in films

Estereotipos de Género en el Cine

- **Intelligence** associated words:
Intelligent, intelligence, intelligently, Ingenious, genius, ingeniousness, ingeniously, bright, brightness, brightly, brilliant, brilliance, brilliantly, clever, cleverness, cleverly.

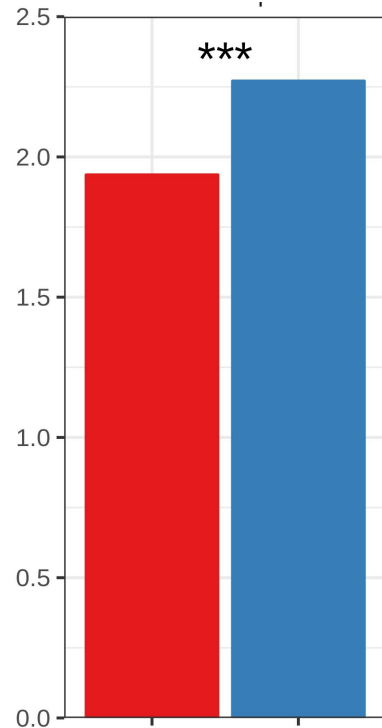
Gálvez R., Tiffenberg V. and Altszyler E. (2019) Half a century of stereotyping associations between gender and intellectual ability in films

Estereotipos de Género en el Cine

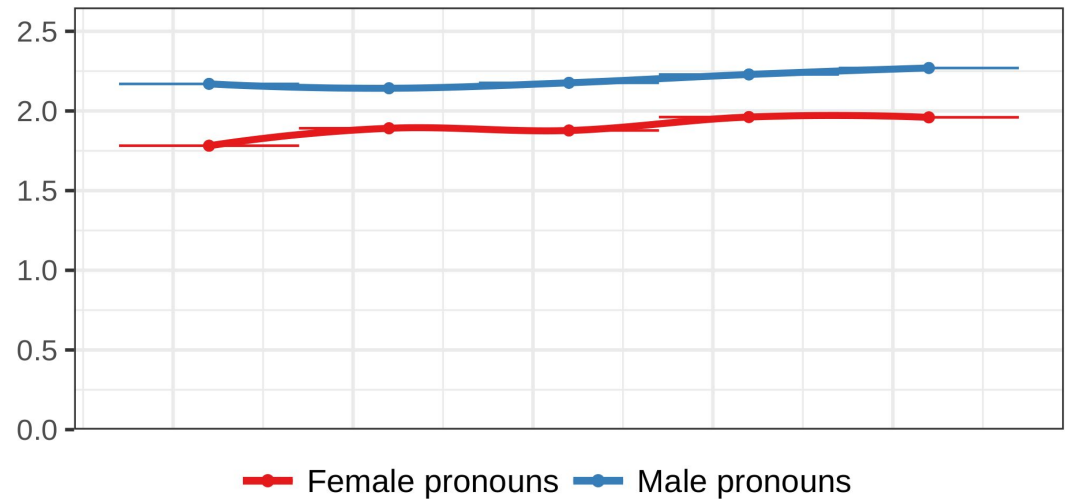


Gálvez R., Tiffenberg V. and Altszyler E. (2019) Half a century of stereotyping associations between gender and intellectual ability in films

Estereotipos de Género en el Cine



Female pronouns Male pronouns



Gálvez R., Tiffenberg V. and Altszyler E. (2019) Half a century of stereotyping associations between gender and intellectual ability in films

Word Embeddings más allá de contar co-ocurrencias



Ministerio de Educación,
Cultura, Ciencia y Tecnología
Presidencia de la Nación



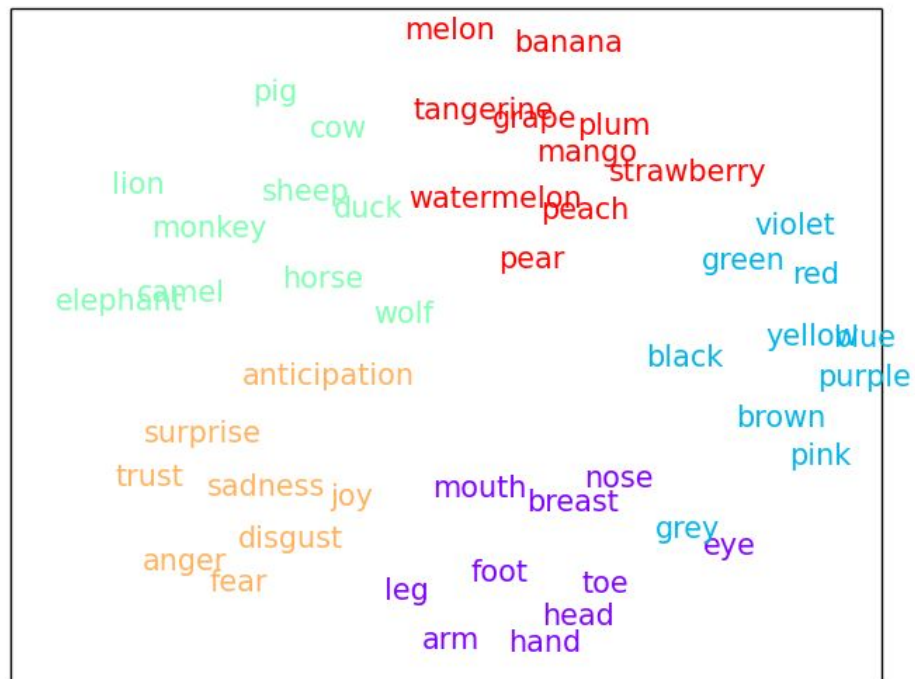
La matemática de las palabras



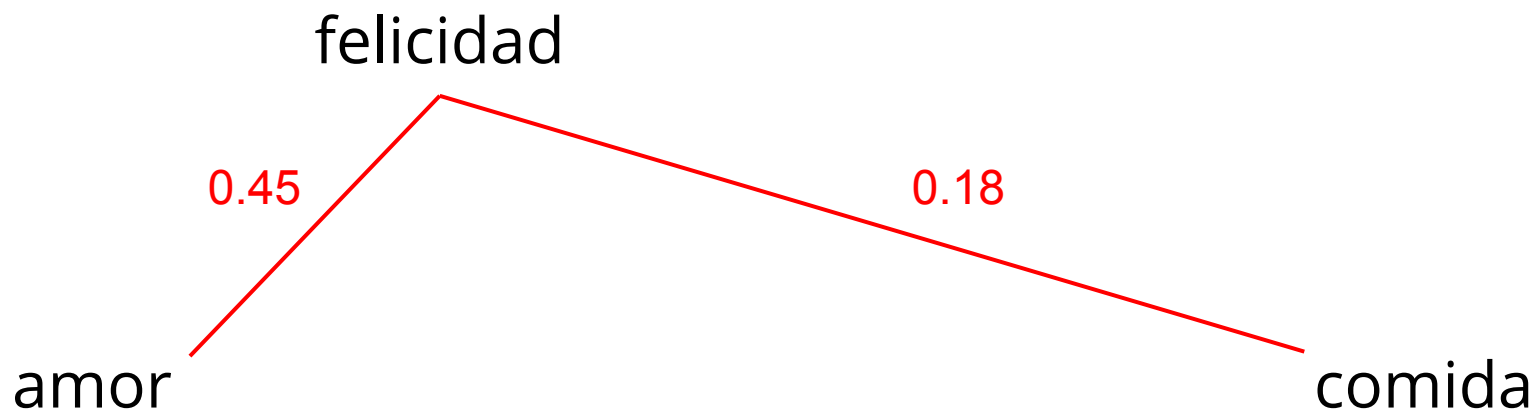
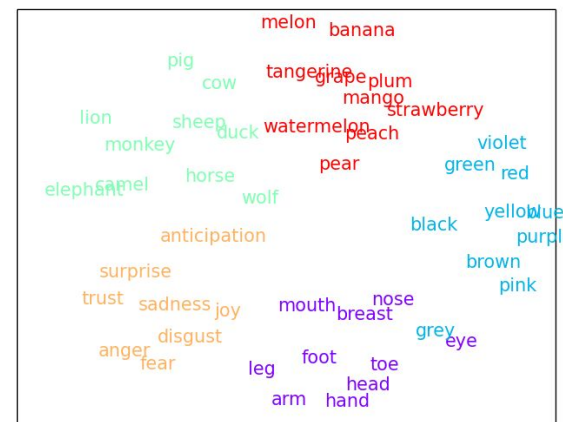
Corpus de textos



Word-embeddings



Cálculo de **cercanía** semántica entre palabras



Latent Semantic Analysis (LSA)

Term-Documents Matrix

	Doc 1	Doc2	Doc 3	Doc 4	Doc 5	Doc 6	...	Doc N
chimichurri	1	0	0	2	0	0	...	0
choripan	0	2	0	0	0	0	...	0
costanera	0	1	2	0	0	0	...	0
asado	1	0	0	0	0	2	...	1
frutilla	0	0	0	0	3	0	...	1
...
kiwi	0	0	0	0	1	2	...	0

A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Landauer and Dumais (1997)

Latent Semantic Analysis (LSA)

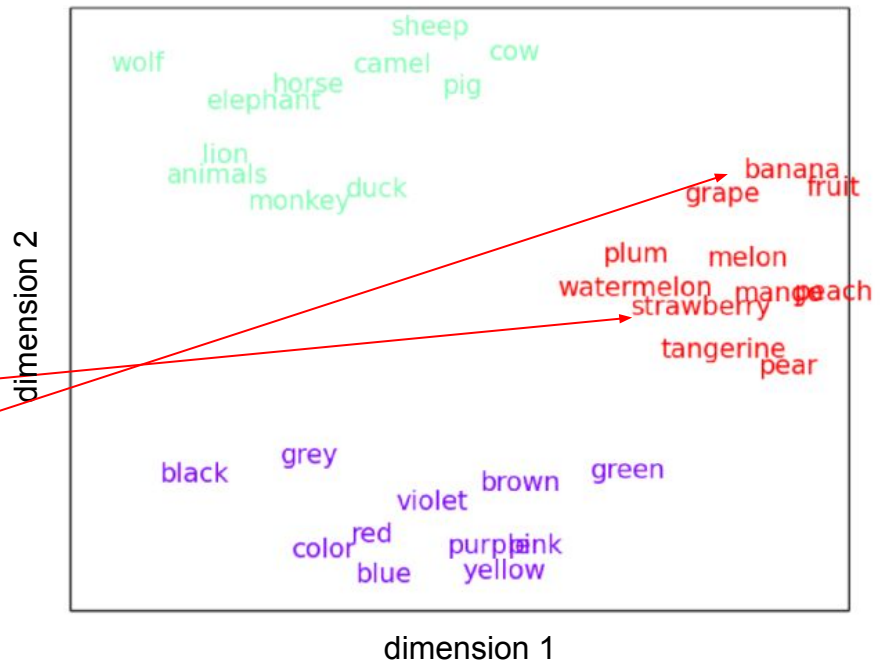
Singular Value Decomposition

chimichurri	0.22	-0.11	...	0.12	-0.11
choripan	0.20	-0.07	...	0.30	-0.07
costanera	0.24	0.04	...	0.21	0.04
barcos	0.40	0.26	...	0.40	0.06
asado	0.64	-0.07	...	0.04	-0.17
perro	0.07	0.01	...	0.27	0.11
casa	0.07	0.11	...	0.27	0.11
frutilla	0.00	0.29	...	0.30	-0.14
kiwi	0.01	0.47	...	0.21	0.27
banana	0.01	0.39	...	0.01	0.49
licuadora	0.04	0.22	...	0.04	0.62
arbol	0.03	-0.35	...	0.03	0.45

Vector de "k" elementos

frutilla = (0.12, 0.97, 0.35, -0.51, ..., 0.11)

Word-embeddings



El texto como ventana a los Sueños

DreamBank
www.dreambank.net

Adam Schneider & G. William Domhoff
Psychology Dept., UC Santa Cruz

Más de 20.000 reportes de sueños



Schneider, A., & Domhoff, G. W. (2016). The Quantitative Study of Dreams.

El texto como ventana a los Sueños

DreamBank
www.dreambank.net

Adam Schneider & G. William Domhoff
Psychology Dept., UC Santa Cruz

Más de 20.000 reportes de sueños



	run
Bay Area girls: Grades 7-9 (n=154)	12 (7.8%)

Category	Women's Reports (N=246)	Men's Reports (N=95)
Parents/Siblings	26.8%	14.7%
Spouses/Partners	19.9%	13.7%
Other Family	10.2%	4.2%
Friends	46.7%	44.2%
Any Familiar Character	75.2%	62.1%
Travel/Vacation	16.3%	6.3%
Sports	7.3%	9.5%
Entertainment/shows	7.7%	9.5%
Parties/Cafes/Bars	11.4%	6.3%
Shopping	9.8%	2.1%
Any Leisure Activity	42.3%	27.4%
School/Work/Politics	20.3%	29.5%
Dreams With No Familiar Elements	12.6%	20.0%

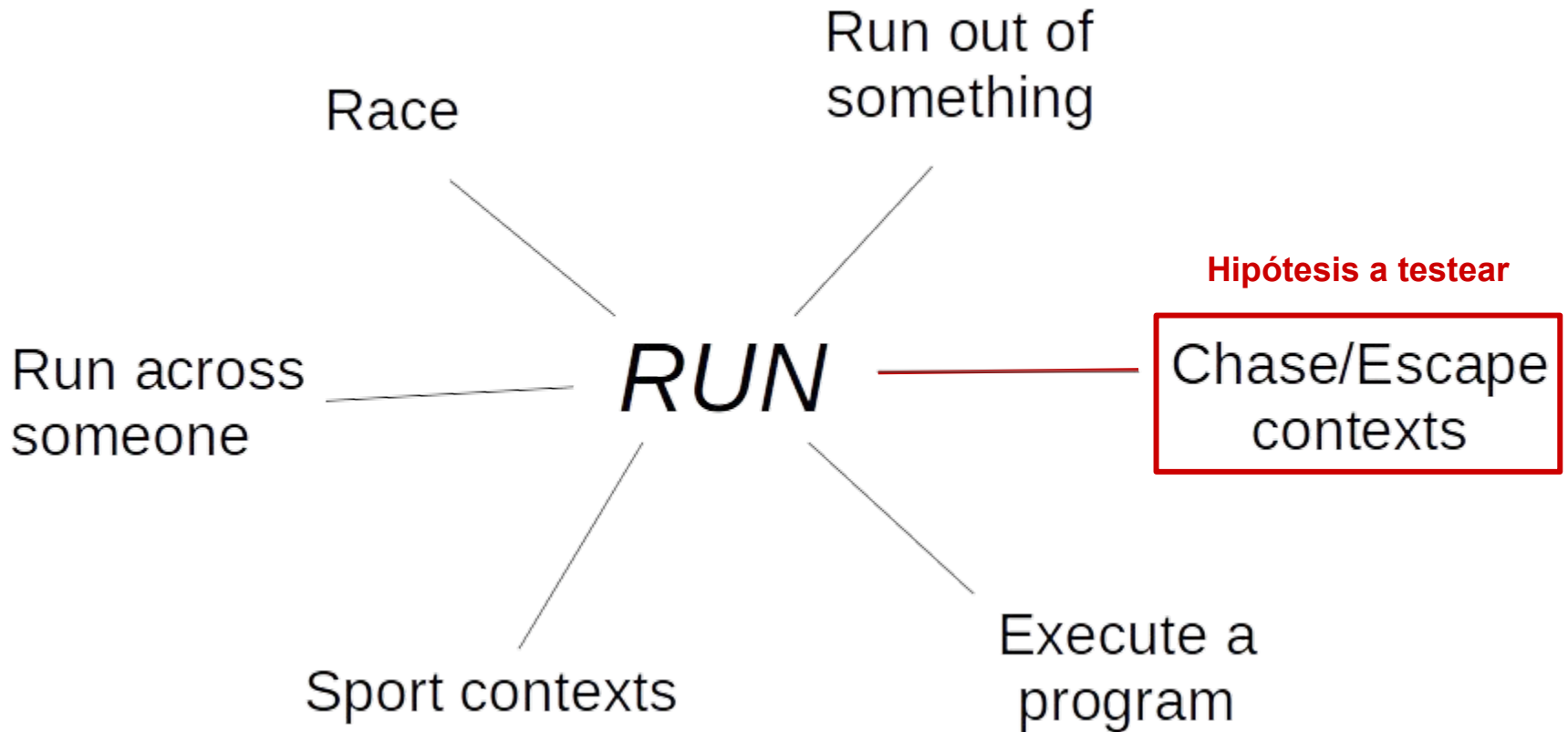
Schneider, A., & Domhoff, G. W. (2016). The Quantitative Study of Dreams.

Teoría de la simulación de amenazas



Antti Revonsuo (2000) The reinterpretation of dreams: An evolutionary hypothesis of function of dreaming

De que corremos en los sueños?



Altszyler et al. (2017) The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text

De que corremos en los sueños?

Palabras más cercanas a “run”

Asociaciones a peligro



Sueños	Web
chase	running
running	dash
scream	jumping
chasing	jump
escape	yell
runs	workouts
chases	kick
grab	jogging
screaming	workout
nazi	stretch
hide	tiring
chased	fun
yells	fast
safety	repetitions
wolf	throw

Altszyler et al. (2017) The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text

Procesamiento de Lenguaje Natural



PLN



Ministerio de Educación,
Cultura, Ciencia y Tecnología
Presidencia de la Nación

fundación
SADOSKY
investigación y desarrollo en TIC

cessi
Argentina

