



# **Beyond Frequency: On the Dictionarization of New Words in Spanish**

Judit Freixa, Sergi Torner  
Pompeu Fabra University

# Neómetro, neologicity and dictionarizability

- **NEÓMETRO:** “The measurement of neologicity and dictionarizability of Spanish neologisms”

Dir. Judit Freixa & Sergi Torner - 2017-2020 - [<https://www.upf.edu/web/neometro>]

Universitat Pompeu Fabra in Barcelona

Funded by: Spanish Research Agency (AEI) and European Regional Development Fund (ERDF)

- **Neologicity:** the gradable quality of a word of being perceived as more or less neological
- **Dictionarizability:** the quality of a word of being more or less included in dictionaries

[Sablayrolles 2003, Cabré et al. 2004, Ben Hariz 2009, Auger 2010, Bouzidi 2010, Freixa 2010, Sánchez Manzanares 2013, Estopà 2015, Bernal et al. in press]

# Neómetro at Observatori de Neologia

The Observatori de Neologia (**OBNEO**) is a public-funded consolidated group of research at Universitat Pompeu Fabra (Barcelona) that analyzes the phenomenon of the appearance of neologisms, both for Catalan and Spanish, since **1989**.

[<https://www.upf.edu/web/obneo>]

## Main resources

- **BOBNEO** (public database for Catalan and Spanish): <http://obneo.iula.upf.edu/bobneo/index.php>
- **Neology networks coordination** in Spain, Europe and America: <https://www.upf.edu/web/obneo/xarxes>
- **Neolosfera** (Catalan): <https://neolosfera.wordpress.com/>
- **Martes Neológico** (Spanish): <https://blogscvc.cervantes.es/martes-neologico/>
- **Antenarío** (Spanish): <https://antenario.wordpress.com/>
- **NADIC** (Catalan): <https://www.upf.edu/web/nadic>

# Dictionarization and institutionalization

- Dictionarization is the official institutionalization of a word (Hohenhaus 2005)
- Dictionarization and institutionalization are very related concepts, but not synonyms.
- Examples of highly institutionalized but non dictionarized words:
  - loans: *selfie* (*autofoto*), *celebrity* (*celebridad*), *online* (*en línea*), and *bullying* (*acoso escolar*)
  - derivatives: *chavismo*, *uribista*, *kirchnerismo*;
  - compounds: *verdiblanco* (*green-and-white*), *rojinegro* (*red-and-black*), *blanquiazul* (*white-and-blue*)
  - etc.

# Neologism: somewhere between creation and dictionarization

Three perspectives and three stages of the establishment of new words.

Perspectives / Stages	Structural perspective	Socio-pragmatic perspective	Cognitive perspective
<b>Creation</b>	(Product of) nonce-formation	(Process of) nonce-formation	Pseudo-concept
<b>Consolidation</b>	Stabilization	Spreading	(Process of) hypostatization
<b>Establishing</b>	Lexicalized lexeme	Institutionalized lexeme	Hypostatized concept

[Schmid 2008]

# Dictionarization criteria

- Literature about the incorporation of neologisms into dictionaries:
  - Barnhart 1985, Algeo 1993, Ishikawa 2006, O'Donovan and O'Neil 2008 (English dictionaries).
  - Freixa et al. 2003, Adelstein and Freixa 2013, Bernal et al. 2018, Freixa in press (Romance lexicography).
- Criteria: word-formation, semantic, documentary and pragmatic. Frequency always having a prominent place.

# Frequency and its dimensions

- General agreement: frequency of use is a good indicator: Barnhart 1985, Ishikawa 2006, O'Donovan and O'Neill 2008, Cook 2010, Adelstein and Freixa 2013, Freixa 2016.
- Decisive factor / prior filter
- Limitations of total frequency of use: some words are used more frequently than others, in a unit's total frequency of use different parameters converge
- Frequency dimensions:
  - Time axis: stability in use
  - Social axis: diversity of communicative situations
  - Geolectal axis: geographical extension

# The data

- **Objective:**

- To analyse the relationship between frequency in corpus and data of perception by speakers.

- **Sample:**

- 132 neologisms

- **Selection criteria:**

- Frequency
- Date
- Formation type



# The data

- **Data base:**

- BOBNEO

- Observatori de neologia, Pompeu Fabra University
    - Lexicographical neologisms from 1989 onwards

- Lexicographical **definition of neology:**

- A *neologism* is a word which is found in the mainstream press but is not included in our reference dictionaries:
    - *Diccionario de la Lengua Española* [DRAE]
    - *Diccionario de uso del español de América y España* [VOXUSO].

# The analysis

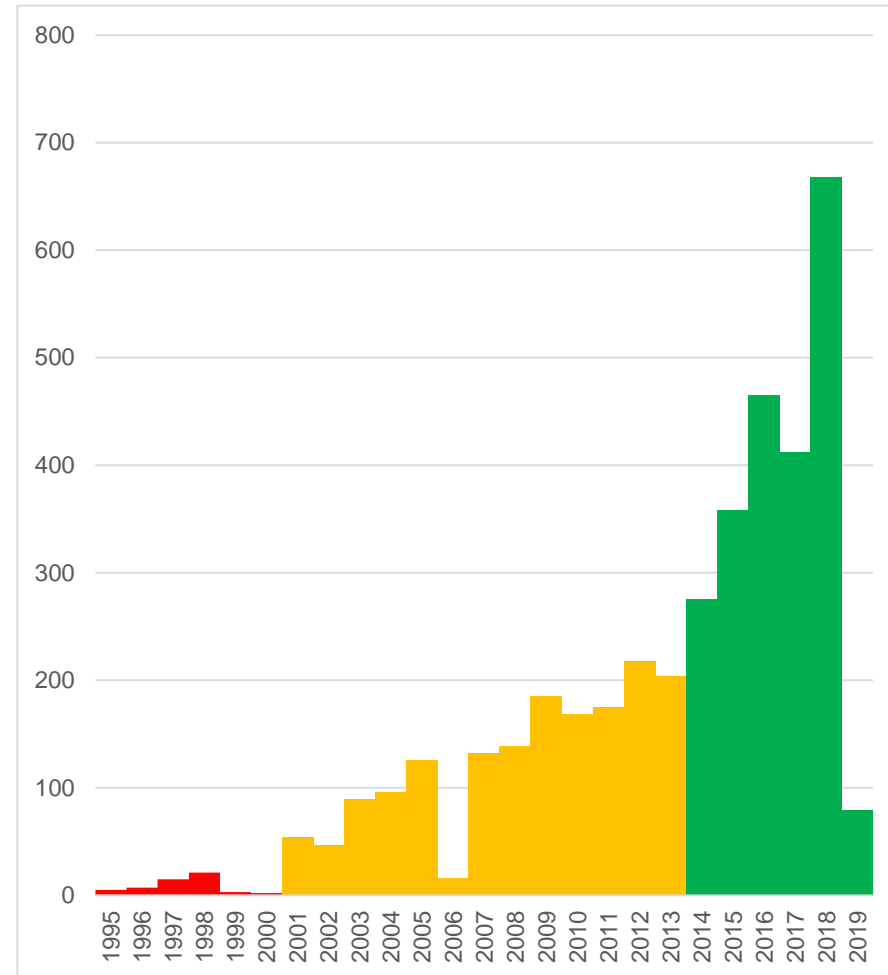
- Analysis: contrast between objective frequency and subjective perception
- Frequency analysis:
  - Dow Jones Factiva (Global News Database)
- Perception analysis:
  - On-line questionnaire
  - Each neologism presented in a short context
  - Question: “Do you think this word should appear in a general dictionary?”
  - 100 participants

# Perception analysis

- It does not allow direct interpretation:
  - Speakers are not lexicographers.
- But:
  - Speakers have intuitive knowledge about the frequency with which a word is used.
  - Speakers have certain expectations about the type of information found in lexicographic works.

# Frequency analysis

- Prototypical pattern of institutionalization:
  - In red: creation
  - In yellow: consolidation
  - In green: institutionalization



Histogram of *rebautizar* ('rename')

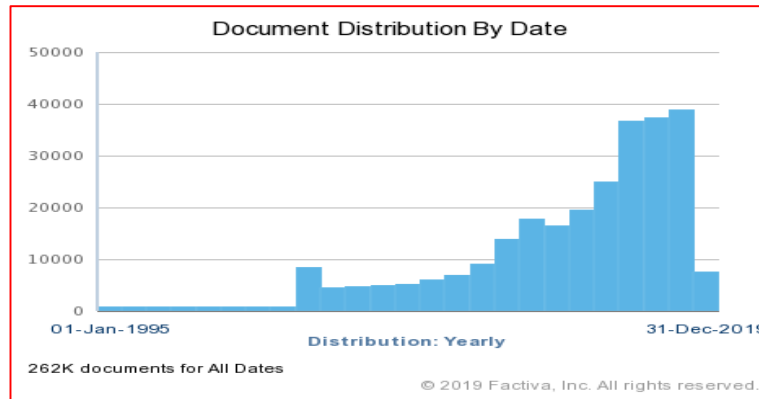
# Frequency vs. perception

Neologism	GC	1st doc Factiva	Frequency in Factiva	% of affirmative responses
hoja de ruta	f	1995	261768	80.6
low cost	loc	1999	96259	38.7
carril bici	m	1995	67841	88.17
línea roja	f	1995	54138	72
big data	m	2007	52154	49.4
selfie	f	2013	41126	68.8
internet de las cosas	m	2004	36318	31.1
gap	m	1994	34817	15
intifada	f	1994	30863	74.19
fetiché	adj	1995	27065	81.7
mensaje de texto	m	1997	21310	75.27
biopic	m	1995	19364	43
feeling	m	1995	19228	44.1
economía colaborativa	f	2003	16584	67.7
hackeo	m	1997	16093	71
precuela	f	1998	14626	76.3
relectura	f	1995	14535	63.4
provisionar	v	1994	13916	59.4
top manta	m	2001	13570	68.8
cláusula suelo	f	2009	13480	60.2
techo de cristal	m	1995	13344	76.3
agencia calificadora	f	1995	13086	35.5
café para todos	m	1994	11916	34.4
puntocom	m	1996	11295	30.1

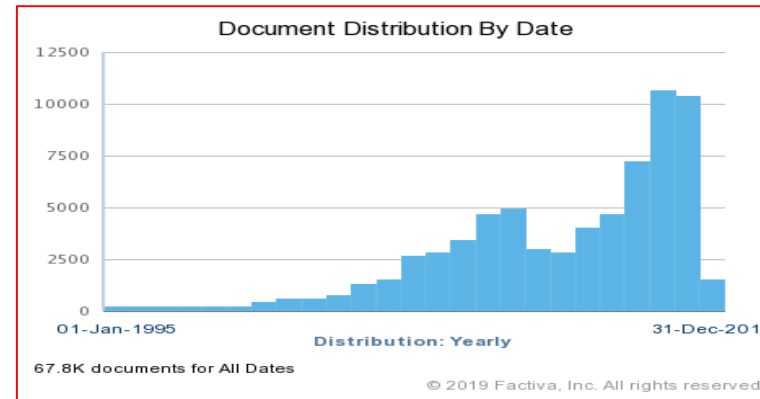
24 more frequent neologisms according to Factiva

# Frequency vs. perception

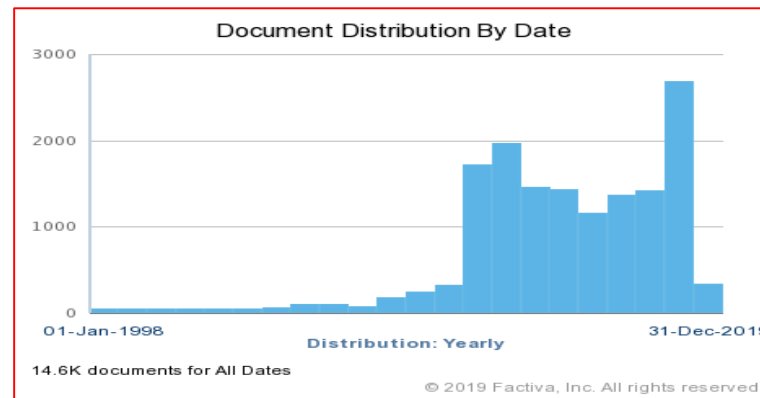
- In all cases, these words' histograms follow prototypical patterns of institutionalization



histogram of *hoja de ruta* ('road map')



histogram of *carril bici* ('bike lane')



histogram of *precuela* ('prequel')

# Frequency vs. perception

Neologism	GC	1st doc Factiva	Frequency in Factiva	% of affirmative responses
hoja de ruta	f	1995	261768	80.6
low cost	loc	1999	96259	38.7
carril bici	m	1995	67841	88.17
línea roja	f	1995	54138	72
big data	m	2007	52154	49.4
selfie	f	2013	41126	68.8
internet de las cosas	m	2004	36318	31.1
gap	m	1994	34817	15
intifada	f	1994	30863	74.19
fetiché	adj	1995	27065	81.7
mensaje de texto	m	1997	21310	75.27
biopic	m	1995	19364	43
feeling	m	1995	19228	44.1
economía colaborativa	f	2003	16584	67.7
hackeo	m	1997	16093	71
precuela	f	1998	14626	76.3
relectura	f	1995	14535	63.4
provisionar	v	1994	13916	59.4
top manta	m	2001	13570	68.8
cláusula suelo	f	2009	13480	60.2
techo de cristal	m	1995	13344	76.3
agencia calificadora	f	1995	13086	35.5
café para todos	m	1994	11916	34.4
puntocom	m	1996	11295	30.1

In 9 cases less than 50% of speakers considered that they should be included in the dictionary.

# Frequency vs. perception

- Similar mismatches can be observed if we go down the list following the order of frequency according to Factiva.

Neologism	GC	1st doc Factiva	Frequency in Factiva	Order in Factiva	% of affirmative responses
superestrella	f	1994	10881	26	49.4
frame	m	1995	7012	31	12.9
enfant terrible	m	1994	6475	32	41
macrobotellón	m	2001	5852	34	30.1
pricing	m	1995	2473	45	15



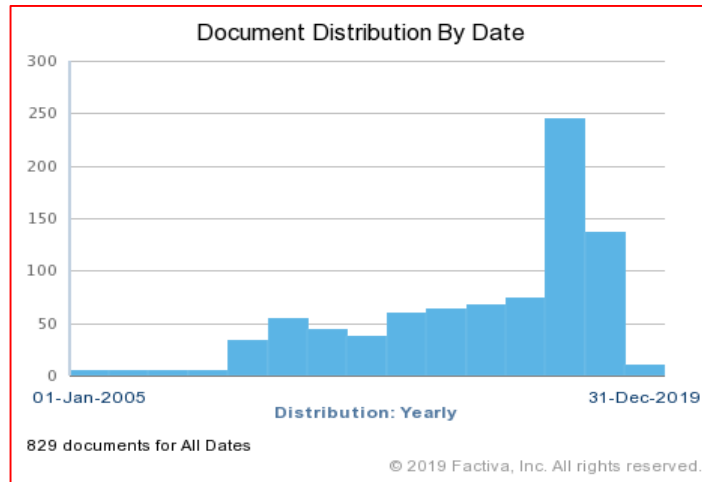
# Frequency vs. perception

- Other inconsistencies between objective and subjective data can be observed in cases in which speakers perceive words with low frequency as dictionary-izable:

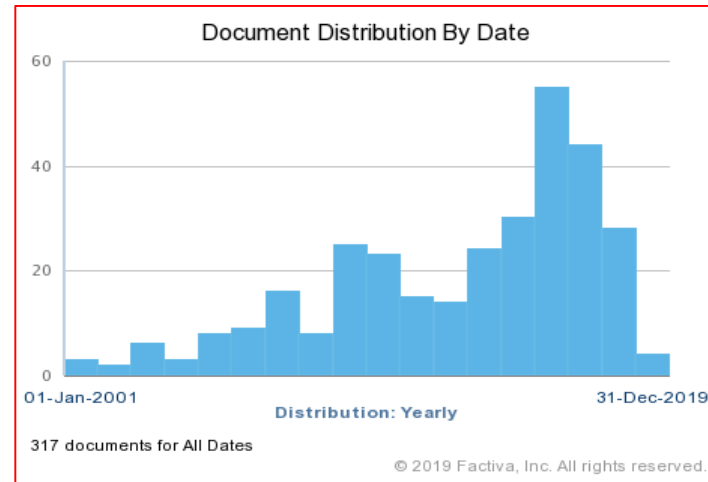
Neologism	GC	1st doc Factiva	Frequency in Factiva	Order in Factiva	% of affirmative resp.
vermutería	f	2006	906	65	74.19
pixelar	v	2005	829	66	89.2
hipersexualización	f	2004	733	67	55.9
psicoestimulación	f	1999	725	68	54.8
infoxicación	f	2002	461	75	51.6
isquiotibial	m	2001	317	78	80.6
apoyapiés	m	1996	54	96	72

# Frequency vs. perception

- The histogram of these neologisms does not allow us to determine their institutionalization because of their low frequency range and the short period of use in which to observe tendencies:



histogram of *pixelar* ('to piexel')



histogram of *isquiotibial* ('hamstring muscle')

# Explaining perception

- Loanwords ↓
  - *big data, biopic, enfant terrible, feeling, frame, gap, low cost, pricing*
- But ↑
  - *selfie, emoji, nikab*
- Derivatives formed with a borrowed base ↑
  - *hackeo* ('hacking'), *pixelar* ('to pixelate')

# Explaining perception

- Compounds ↑
  - *hipersexualización* ('hypersexuality'), *psicoestimulación* ('psicostimulation'), *infoxicación* ('infoxication'), *isquiotibial* ('hamstring muscle')
- Derivatives ↓
  - *superestrella* ('superstar'), *macrobotellón* ('booze-up party', lit. 'big bottle')
- Sintagmation ↑
  - *hoja de ruta* ('road map'), *carril bici* ('bike lane'), *línea roja* ('red line'), *mensaje de texto* ('text message'), *economía colaborativa* ('collaborative economy'), *cláusula suelo* ('base clause in mortgages')
- But ↓
  - *internet de las cosas* ('Internet of Things'), *café para todos* ('one size fits all policy' lit. 'coffe for all')

# Conclusions

- Frequency of use plays an important role in determining the degree of institutionalization of neologisms.
- But frequency is a necessary but not a determining condition.
- Denominative need seems to be a determining factor.
  - Terminological units.
- Some units are not considered highly dictionarizable:
  - Colloquial words
  - Loans
  - Units with expressive value

Thank you for your attention

neómetro

Judit Freixa and Sergi Torner  
Universitat Pompeu Fabra

<<https://www.upf.edu/web/neometro>>