

Spanish Trends

Corpus con metadatos temporales en Sketch Engine

Corpus preparados para los estudios diacrónicos

¿Qué es la función Trends?

Sketch Engine permite rastrear la evolución temporal del vocabulario mediante Trends (Tendencias).

Neologismos

Detecta cuándo una palabra empieza a ganar frecuencia y en qué momento se consolida en el uso.

Historia de la lengua

Permite rastrear el auge o la caída de términos a lo largo de siglos o décadas con datos empíricos.

Datación de términos

Ayuda a situar cronológicamente el pico/valle de uso de una palabra

Requisito previo: el corpus debe estar estructurado con un metadato temporal (year, periodo, etc.) para que Trends funcione.

Las palabras del año

Vamos a ver si *Spanish Trends* (prensa hispánica, 2022-2026) refleja el aumento de frecuencia de algunas de estas palabras

2013 <i>escrache</i>	2014 <i>selfi</i>	2015 <i>refugiado</i>
2016 <i>populismo</i>	2017 <i>aporofobia</i>	2018 <i>microplástico</i>
2019 <i>emojis</i>	2020 <i>confinamiento</i>	
2021 <i>vacuna</i>	2022 <i>inteligencia artificial</i>	
2023 <i>polarización</i>	2024 <i>dana</i>	2025 <i>arancel</i>

Evolución de *arancel* entre 2022 y 2026 en Spanish Trends

arancel: palabra del año 2025

lema *arancel* • 196.713

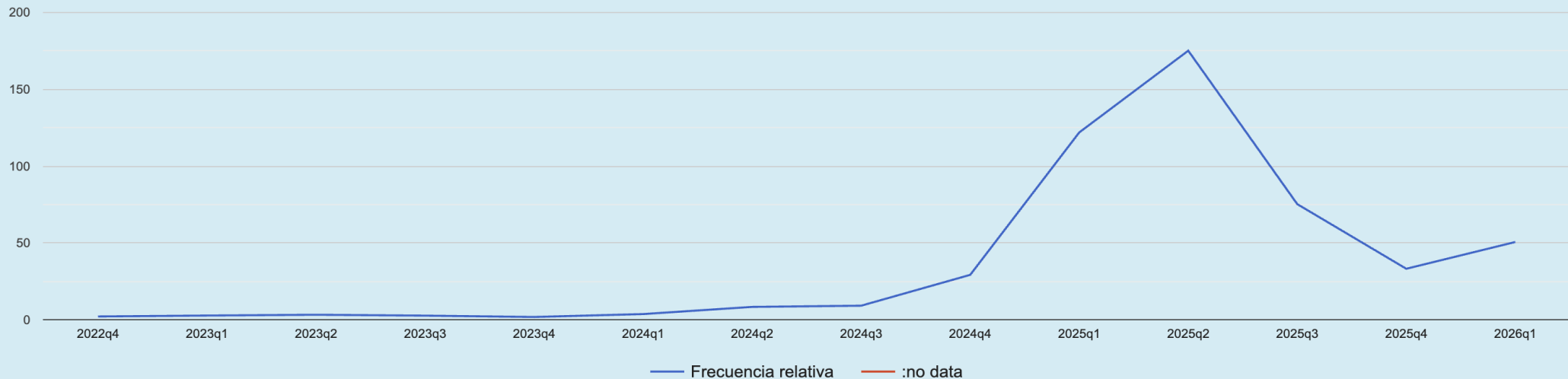
40,8 por millón tokens • 0.0041%



Timeline

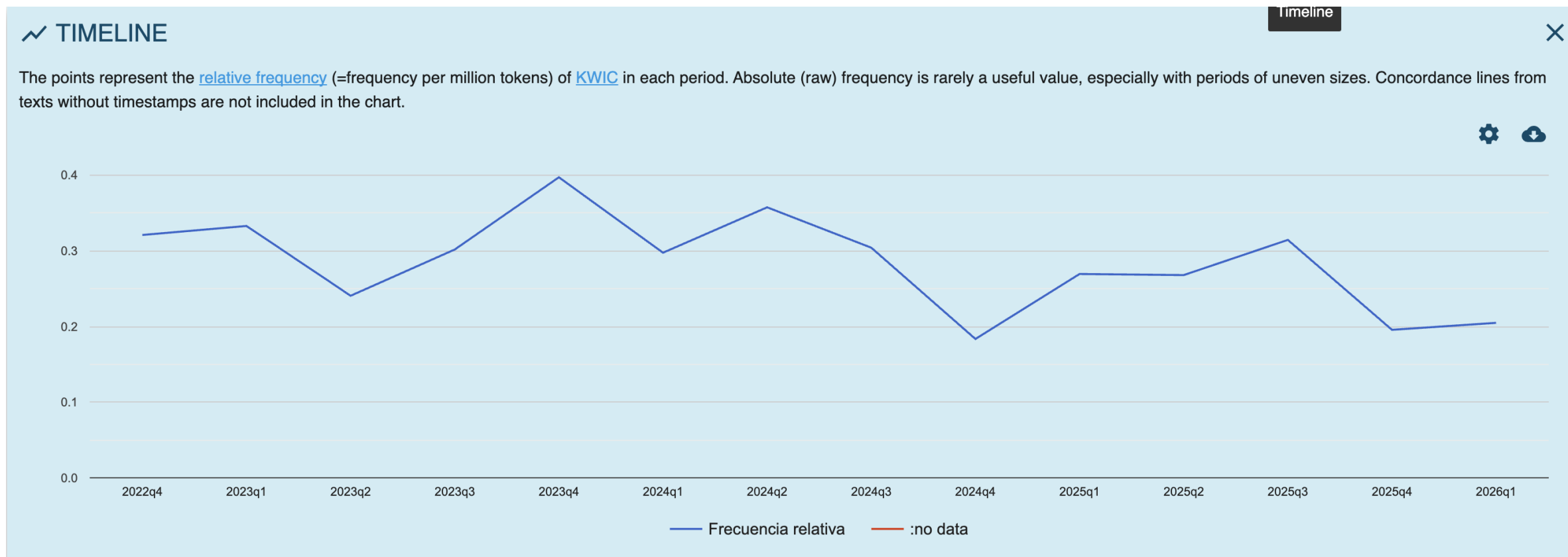
TIMELINE

The points represent the [relative frequency](#) (=frequency per million tokens) of [KWIC](#) in each period. Absolute (raw) frequency is rarely a useful value, especially with periods of uneven sizes. Concordance lines from texts without timestamps are not included in the chart.



Evolución de *selfi* entre 2022 y 2026 en Spanish Trends

selfi: palabra del año en 2014



Evolución de *inteligencia artificial* entre 2022 y 2026 en Sp. Trends

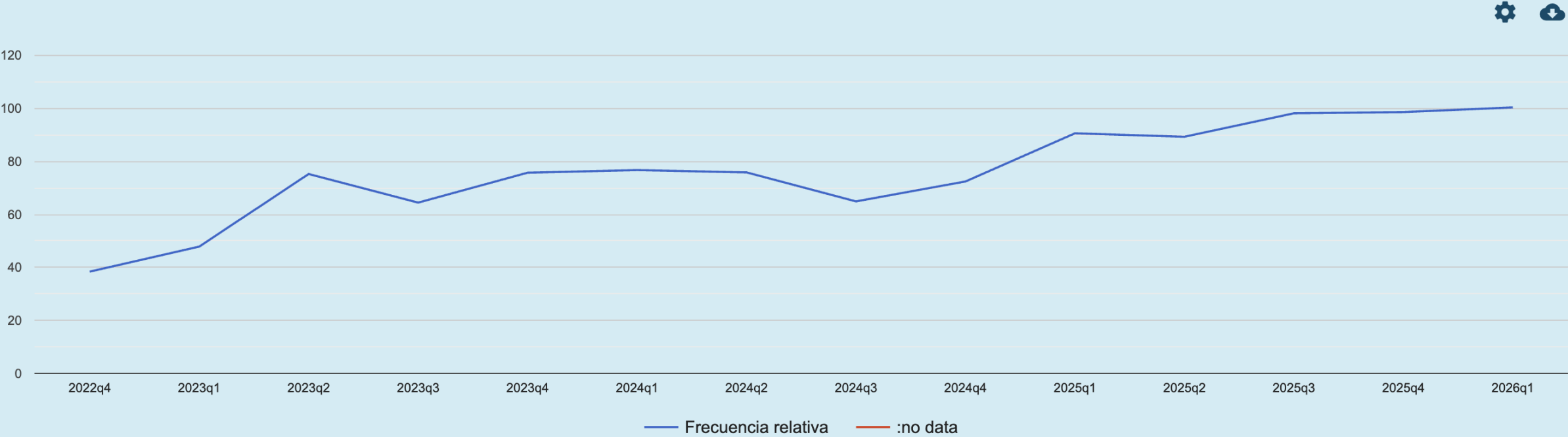
inteligencia artificial: palabra del año en 2022

frase **inteligencia artificial** • 421.616
87,44 por millón tokens • 0.0087%



TIMELINE

The points represent the [relative frequency](#) (=frequency per million tokens) of [KWIC](#) in each period. Absolute (raw) frequency is rarely a useful value, especially with periods of uneven sizes. Concordance lines from texts without timestamps are not included in the chart.



Evolución de *machirulo* entre 2022 y 2026 en Sp. Trends

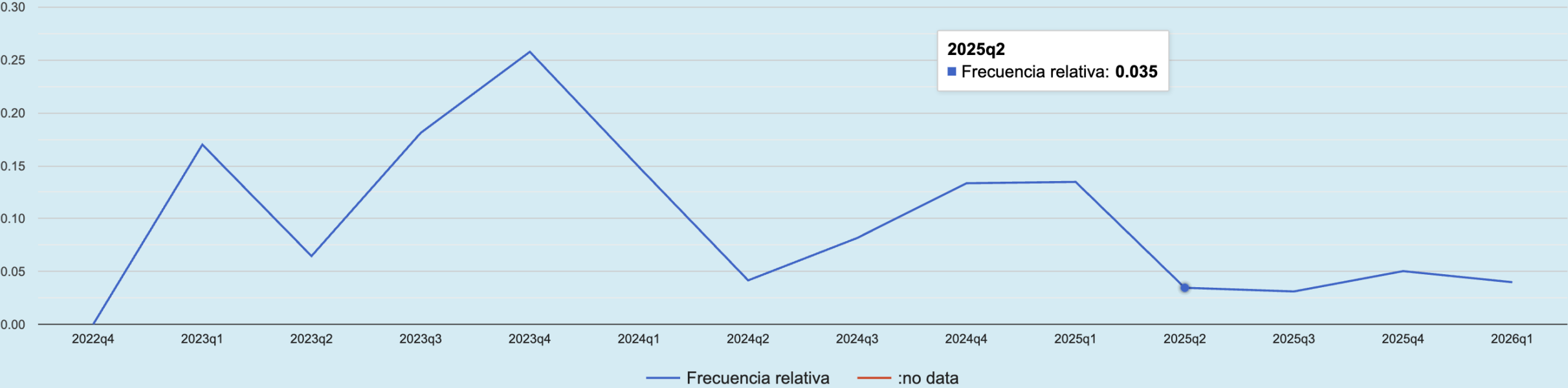
simple **machirulo** • 388

0,08 por millón tokens • 0.000008%



TIMELINE


The points represent the [relative frequency](#) (=frequency per million tokens) of [KWIC](#) in each period. Absolute (raw) frequency is rarely a useful value, especially with periods of uneven sizes. Concordance lines from texts without timestamps are not included in the chart.



referido al videoarbitraje y que son siglas en inglés que se refieren al 'árbitro asistente por vídeo'. En cuanto a '**machirulo**', habla de "una persona, especialmente un hombre, que exhibe una actitud machista", además de añadir que es un

Consultar la evolución de varias palabras

Datos de significación estadística (valor de p).

TENDENCIAS

Spanish Trends

BÁSICO

AVANZADO

Atributos de estructura

Quarter

Subcorpus

ningún (todo el corpus)

Ordenar ítems por

Tendencias

Filtro

todo

que empieza con

que termina con

que contiene

que no contiene

coincide con este registro

emojis

confinamiento

vacuna

polarización

dana

Pegar la lista aquí, una palabra por línea

de esta lista:

	Lema	Tendencia ↓	Frecuencia	Valor-p	Muestra	
1	confinamiento	↘ -1,00	20.858	0.00068		...
2	dana	↗ 0,84	52.619	0.0044		...
3	vacuna	↘ -0,51	106.961	0.021		...
4	polarización	↗ 0,49	41.165	0.079		...
5	emojis	↘ -0,45	9301	0.011		...

Descenso significativo de *confinamiento* entre 2022 y 2026 valor de $p = 0.00068$ (umbral de signif. $p < 0.05$)

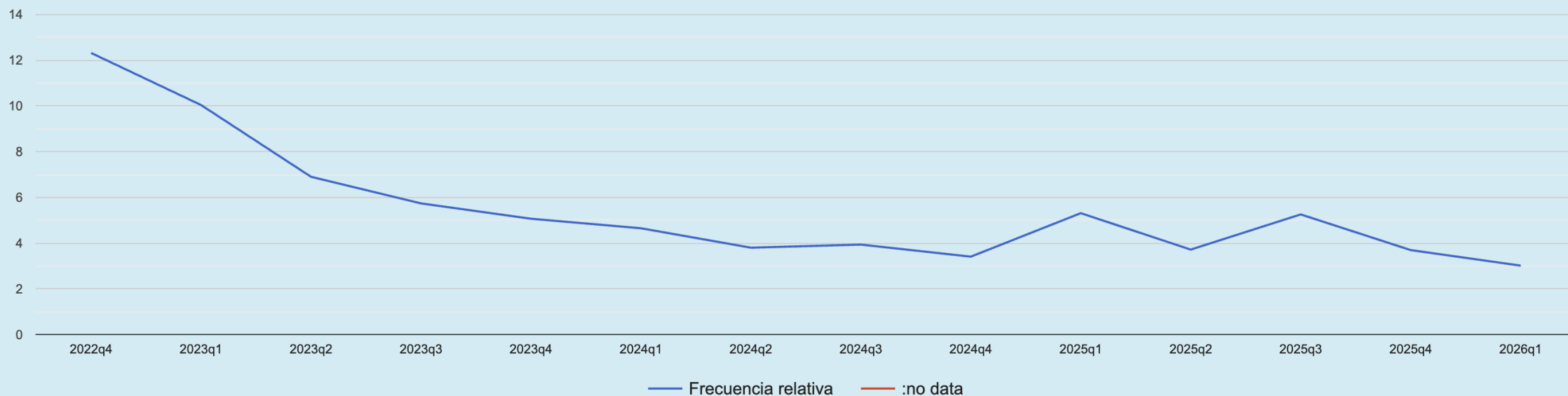
CQL [lemma="confinamiento"] • 20.858
4,33 por millón tokens • 0.00043%



TIMELINE



The points represent the [relative frequency](#) (=frequency per million tokens) of [KWIC](#) in each period. Absolute (raw) frequency is rarely a useful value, especially with periods of uneven sizes. Concordance lines from texts without timestamps are not included in the chart.



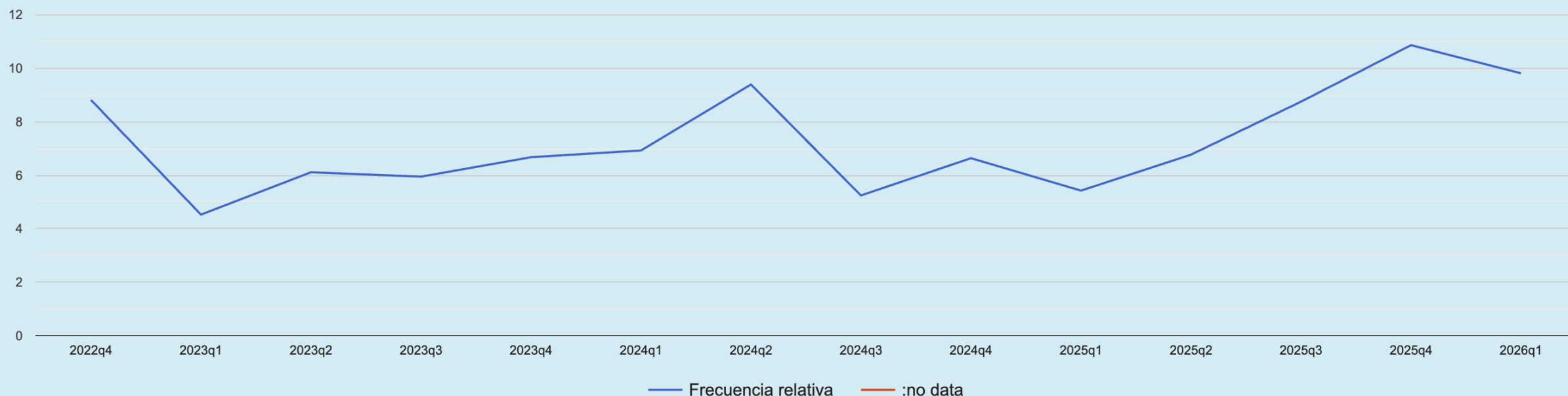
Aumento no significativo de *polarización* entre 2022 y 2026 valor de $p = 0.079$

CQL [lemma="polarización"] • 41.165
8,54 por millón tokens • 0.00085%



TIMELINE

The points represent the [relative frequency](#) (=frequency per million tokens) of [KWIC](#) in each period. Absolute (raw) frequency is rarely a useful value, especially with periods of uneven sizes. Concordance lines from texts without timestamps are not included in the chart.



Creación de corpus diacrónicos:

Generar corpus en Sketch Engine con la función Trends

Material complementario

Material necesario para este bloque:
archivos en las carpetas discursos_inv_txt y discursos_inv_trends

Preparar los documentos: estructura de los archivos .txt

Cada texto se guarda en un archivo .txt independiente. Todos deben seguir una estructura XML mínima obligatoria.

ESTRUCTURA OBLIGATORIA

```
<text year="1979" periodo="1982" partido="UCD">
```

```
    el texto  
    publicado en ese año
```

```
</text>
```

tres metadatos: year (año de publicación)

partido: partido del presidente

periodo: fecha de inicio del mandato del siguiente partido

⚠ El atributo year o periodo (uno de los dos) es imprescindible. Sin él, Trends no puede calcular la distribución temporal.

Subir los documentos y configurar el metadato *year*

La subida es la habitual, pero hay un paso crítico que debe hacerse ANTES de compilar.

1 Subir los documentos

Ir a "Tengo textos" en Sketch Engine y arrastrar los archivos .txt a la ventana de carga. No compilar todavía.

2 Abrir Opciones expertas → marcar el metadato elegido (en este caso, periodo)

ANTES de compilar: ir a Opciones expertas > marcar "Text" > pulsar el botón + > seleccionar el metadato periodo. Verificar que aparece correctamente (ver diapositiva siguiente). Pulsar Guardar.

3 Compilar el corpus

Solo después de guardar la configuración del metadato, pulsar Compilar. Si se compila antes, habrá que reconfigurar.

Opciones expertas antes de compilar: activar metadatos (text)

AÑADIR MÁS TEXTOS

COMPILAR

Opciones expertas ▴

OPCIONES EXPERTAS

Contenido duplicado ?

☐ Remover contenido duplicado

Language filtering ?

☐

Estructuras y atributos a mantener ?

todo | ninguno

✓ g (15.191)

✓ s (4560)

✓ text (14) —

✓ year

✓ partido

✓ período

Gramática de sketch ?

☒ Spanish (Freeling) 2.0 (recomendado) ⓘ

☐ Universal-generic-1.0 ⓘ

☐ Ninguna (sin word sketch)

+

Gramática de términos ?

☒ Spanish terms (Freeling) 2.0 ⓘ

☐ Spanish terms (Freeling) 1.4 ⓘ

☐ Ninguna (sin extracción terminológica)

+

Estructura para documentos ?

doc

Estructura para contar documentos ?

Lo mismo como nombre de documentos ▾

CANCELAR

GUARDAR Y COMPILAR

Después de mactivar los metadatos textuales, pulsar guardar y compilar

Activar los atributos diacrónicos en la gestión del corpus

Una vez compilado el corpus, hay que indicarle a Sketch Engine qué atributo representa el tiempo.

Gestión del corpus → Configurar → Opciones expertas → Diachronic attributes = text.periodo → Guardar

CONFIGURAR

investidura



CORPUS: investidura (Spanish)

Nombre

discurso_inv

Descripción

Opciones expertas ▲

Diachronic attributes

text.periodo ▼

Configuración de subcorpus ?

ATRÁS

GUARDAR

Resultado: la función *Trends* disponible en el menú lateral

Si la configuración es correcta, Trends (Tendencias) aparecerá como opción activa en el menú del corpus.



Hacer gráficos de tendencias cronológicas

esfuerzo

democracia

nación

futuro

libertades

patria

hijos

primera persona del plural

solidaridad, etc.

ver proyectos de Elena Álvarez Mellado (lirondos) sobre discursos de Navidad

Crear subcorpus: UCD, PSOE, PP

Analizar palabras (únicas y multipalabras) clave del PSOE frente al resto
del PP frente al resto
de UCD frente al resto

Por ejemplo: UCD frente al resto: reconversión industrial, transición política, etc.
PSOE frente al resto: justicia social, salario mínimo, pensión mínima, etc.

Búsquedas complejas en CQL

[lc = "señores"] within <text year>="1980" & year<="2000"/>

[lc="señores"]within <text year>="2000" & year<="2020" & partido="PP"/>

[lc="señores"] within <text partido="PSOE"/>

Forma fácil de encontrar sustantivos, adjetivos, verbos o adverbios

LISTA DE PALABRAS

[Obtén más almace](#)[BÁSICO](#)[AVANZADO](#)[LEARN](#)

encuentra ?

palabras

lemas

adjective

adverb

conjunction

noun

numeral

preposition

todo

que empieza con

que termina con

que contiene

coincide con este regex

de esta lista:

☐ excluir siguientes palabras:☒ Incluir no palabras ?☒ A = a ?

Frecuencia mín. ?

1

Frecuencia máx. ?

0

visualizar como

☒ Lista simple ?☐ Mostrar como ?


Subcorpus ?

PP








Comparar los
adjetivos, sustantivos,
adverbios más
frecuentes en cada
subcorpus


Buscar adjetivos seguidos de sustantivos

BÁSICO	AVANZADO	CÓMO FUNCIONA
<p>Tipo de consulta </p> <div><p>simple</p><p>lema</p><p>frase</p><p>forma</p><p>caracter</p><p>CQL</p></div>		<p>CQL</p> <p><u>[tag = "AQ.+"] [tag = "NC.+"]</u></p> <p>Atajos <div><div>[]</div><div>{ }</div><div>< ></div><div>" "</div><div>&</div><div>\</div></div></p> <p>Atributo predeterminado [?]</p> <p><u>lema</u> ▼</p>

Sustantivos más frecuentes en el primer periodo

LISTA DE PALABRAS  

[Obtén más almacenamiento](#)   

BÁSICO **AVANZADO** LEARN 

encuentra ?

- palabras
- lemas
- adjective
- adverb
- conjunction
- noun**
- numeral
- preposition

todo

- que empieza con
- que termina con
- que contiene
- coincide con este regex
- de esta lista:

☐ excluir siguientes palabras:

☒ Incluir no palabras ?

☒ A = a ?



Frecuencia mín. ? Frecuencia máx. ?

visualizar como


☒ Lista simple ?


☐ Mostrar como ?

Subcorpus ?


ningún (todo el corpu...  



Tipos de texto (1) ? ^

text.year v 

text.partido v 

text.periodo ^

1982 x 

1996

2004

2016

2020

Expandir

Voyant Tools como alternativa a Sketch Engine

Generar corpus en Voyant Tools. Opciones de visualización

Material complementario

Material necesario para este bloque:
archivos en la carpeta: discursos_inv_trends
archivo: stopwords-es.txt


Voyant Tools

Una alternativa a Sketch Engine



Añadir textos

Escriba una o más URL en líneas separadas o péguelas en un texto completo.

 Abrir

☒ Revelar

*Voyant Tools es un entorno de lectura y análisis basado en web para textos digitales .
Traducción al español por Manuel Vargas Alegría*

Announcing the [Voyant Consortium](#). Register to be a [member here](#).

Guardar el enlace para seguir consultando el corpus
sin necesidad de cargarlo de nuevo

<https://voyant-tools.org/?corpus=3db5afd17860e60b7e86bace2854d6f2>

Excluir palabras gramaticales (stop words). **LISTA ADAPTABLE A LAS NECESIDADES**
Revisar la lista (puede interesar no eliminar algunas: no, solo, yo, ahora, etc.)

Opciones

Palabras
excluidas:

Nueva lista -definida por el

Auto-detección

Lista blanca :

Ninguno

Categories:

Nueva lista -definida por el
usuario-

Editar la lista de palabras excluidas



Esta es la lista de palabras excluidas, cada término en una línea.

n
nada
nadie
ni
ninguna

Salvar

Cancelar

DEMASIADAS PANTALLAS A LA VEZ. Seleccionar una herramienta para que ocupe toda la pantalla (icono con la flecha)



 Lector

☐ TérminosBerry

Exportar hacia URL, datos, referencia bibliográfica o para incrustar esta herramienta.



Escala: Corpus y documentos concretos
analizar todo el corpus o un subcorpus concreto



25 palabras más frecuentes del primer discurso de Zapatero

Evolución (tendencia) de *libertad* en el corpus

