

Implementación de sistemas conexionistas: estudio de la polisemia en lengua española a través de grandes modelos de lenguaje

Pablo Pérez Rodríguez

`pab@usal.es`

Máster Universitario en Sistemas Inteligentes
Universidad de Salamanca

Junio de 2026

Tabla de contenidos

1 Introducción

2 Metodología

3 Resultados

- MONO
- XL-WSD-es

4 Conclusiones

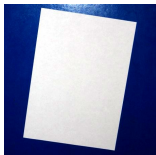
Polisemia y desambiguación

hoja *f.*



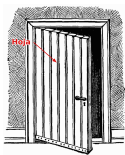
vegetal

Wikimedia Commons



papel

Wikimedia Commons



puerta

Wikimedia Commons



cuchilla

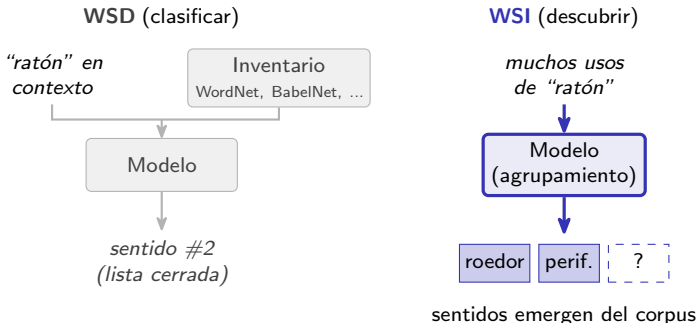
Wikimedia Commons

	A	B	C	D
1	Nombre	Edad	Deuda total	
2	Juan	66	114,00 €	
3	Paula	65	23,00 €	
4	Enrique	64	58,00 €	
5	Carla	60	568,00 €	
6	Miguel	35	1.245,00 €	
7	Total		2.008,00 €	
8				
9				
10				
11				
12				
13				
14				

de cálculo

Elaboración propia

Tareas de WSD y WSI



- WSD: ¿cuál de los sentidos conocidos tiene una palabra en un contexto?
- WSI: gran lista de ejemplos -¿ separarlos por significado (sin inventario)

Objetivos y contribuciones

Objetivo: diseñar y evaluar un sistema de **WSI no supervisado para el castellano** basado en *embeddings* contextuales de BERT.

Objetivos específicos:

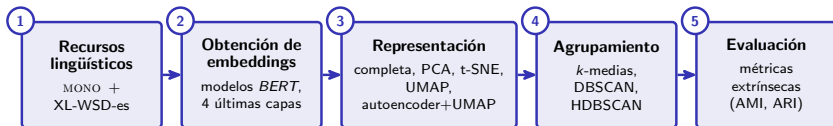
- Comparar modelos del castellano: *mBERT*, *BETO*, *RoBERTa BNE*.
- Evaluar métodos de agrupamiento y reducción de dimensionalidad.
- Visualizar los clústeres en 2D y validar los sentidos cualitativamente.
- Evaluar sobre un *benchmark* y comparar con trabajos anteriores.

Estado de la cuestión



- Wiedemann et al. [4] muestran que *BERT* **separa los sentidos** en el espacio \Rightarrow WSI = clústering de *embeddings* contextuales.
- **Sustitución léxica** domina en inglés (SemEval-2010: AVG 53.6 vs. 24.5 del clústering directo).
- *Benchmarks* de WSI **sólo en inglés** (SemEval). En castellano sólo hay tareas de WSD, y los modelos (*BETO*, *RoBERTa BNE*, *mBERT*) se han analizado en WSD, no en WSI.

Flujo de trabajo



Los hiperparámetros de agrupamiento se ajustan con **métricas intrínsecas**:

- *k*-medias(*k*): maximizar **silueta** en $[2, n_{\text{sentidos}} \cdot 3]$.
- DBSCAN (ϵ): candidatos de la curva de *k*-distancias, maximizar **DBCV**, o codo (Kneedle), o mediana.
- HDBSCAN (`min_cluster_size`, `min_samples`): rejilla, maximizar **DBCV**, o más clústeres y menos ruido.

Conjuntos de datos

Conjunto	Inst.	Palabras	Inst./ palabra	Sentidos/ palabra	Long. (palabras)
MONO (construido)	693	1	693.00	4.00	14.61
XL-WSD-es (filtrado)	213	36	5.92	2.19	31.35

MONO — aportación propia

*Mostrario de Oraciones con un mismo
Nexo Ortográfico*

- Gemini, 4 sentidos de *mono*.
- Publicado en Kaggle.
- Gran volumen de ejemplos para facilitar análisis.

XL-WSD-es — conjunto del mundo real

- Corpus natural (XL-WSD, [2]), filtrado instancias > 1 sentido correcto, palabras con < 2 sentidos o < 3 ejemplos.
- Pequeño, difícil: pocos ejemplos.
- Comparación directa con Kokosinskii et al. [1].

Las acepciones utilizadas en MONO

Significado

Ejemplo



'Simio'

«El pequeño *mono* saltó entre los árboles buscando algo de fruta fresca».



'Bonito, lindo'

«Llevas un peinado muy *mono*, te favorece muchísimo con esa cara».



'Prenda de una pieza'

«El astronauta se preparó para la misión ajustándose bien su *mono* especial».



'Síndrome de abstinencia'

«Llevo toda la mañana de mal humor porque tengo *mono* de café».

Espacio de pruebas

Modelos (7)

mBERT (con/sin may.)
BETO (con/sin may.)
RoBERTa BNE
(base/large)
MrBERT-es

×

Representación (5)

Original (768/1024 dim)
Reducida a 10 dim:
PCA, t-SNE, UMAP,
autoencoder+UMAP
(Rother et al. [3])

×

Agrupamiento (3)

k-medias
DBSCAN
HDBSCAN

$7 \times 5 \times 3 = 105$ combinaciones

Evaluado mediante **AMI**¹ (principal) y **ARI**².

¹**AMI**: cuánta información comparten los clústeres con los sentidos reales, ajustada por azar (1 = perfecto, 0 = aleatorio).

²**ARI**: de todas las parejas de ejemplos, en cuántas coinciden clúster y sentido (juntas en ambos o separadas en ambos), ajustada por azar (1 = perfecto, 0 = aleatorio). Penaliza más los clústeres de tamaños muy distintos.

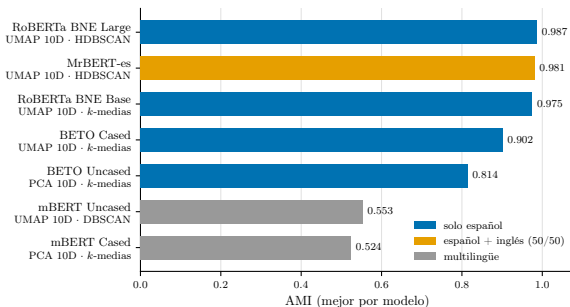
Comparativa global

Modelo	Proyección	Método	k	AMI	ARI	Ruido
<i>RoBERTa BNE Large</i>	UMAP 10D	HDBSCAN	4	0.987	0.992	0.00 %
<i>MrBERT-es</i>	UMAP 10D	HDBSCAN	4	0.981	0.987	0.00 %
<i>RoBERTa BNE Large</i>	PCA 10D	k -medias	4	0.980	0.988	—
<i>RoBERTa BNE Base</i>	UMAP 10D	k -medias	4	0.975	0.979	—
<i>RoBERTa BNE Base</i>	t-SNE 10D	k -medias	4	0.952	0.964	—
<i>RoBERTa BNE Base</i>	Completo	k -medias	4	0.941	0.955	—
<i>RoBERTa BNE Base</i>	PCA 10D	k -medias	4	0.937	0.951	—
<i>MrBERT-es</i>	t-SNE 10D	k -medias	4	0.929	0.942	—
<i>RoBERTa BNE Large</i>	Completo	k -medias	5	0.911	0.885	—
<i>MrBERT-es</i>	Rother et al.	k -medias	4	0.907	0.926	—

Diez mejores combinaciones de modelo, proyección y método, ordenadas por AMI.

Mejores modelos

MONO

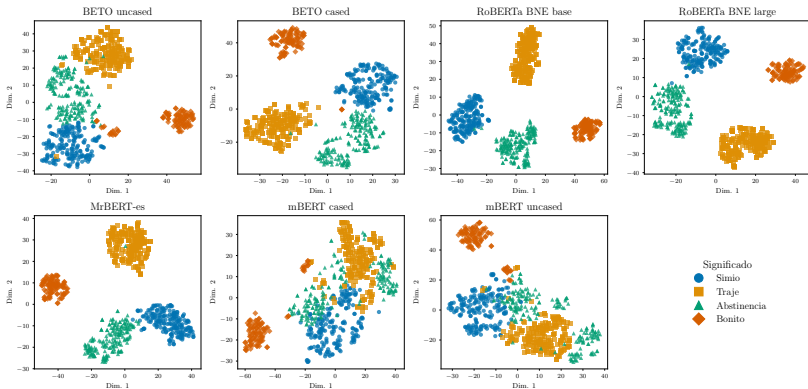


Mejor AMI por modelo, coloreado según los datos de entrenamiento.

Modelos entrenados **sólo en español** tienden a separar los sentidos mejor que los **multilingües**. *RoBERTa BNE Large* **AMI 0,987**.



Análisis pre-aglomerativo

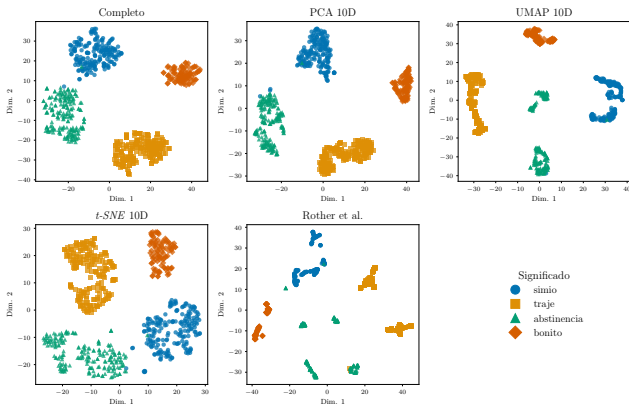


Proyección t-SNE 2D del espacio completo, coloreada por el sentido real.

■ Fuertes: nubes **separadas**.

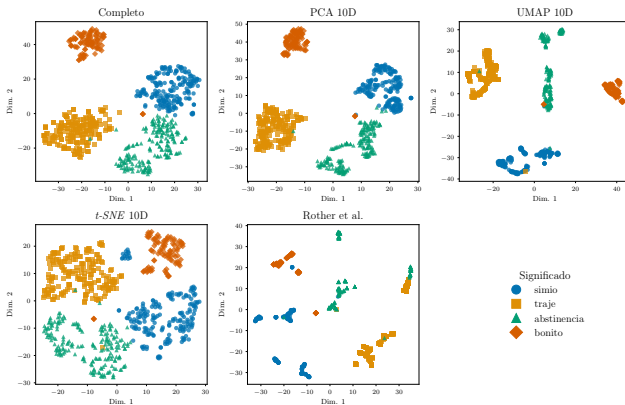
■ Débiles: sentidos **mezclados**.

Efecto de la reducción (1)



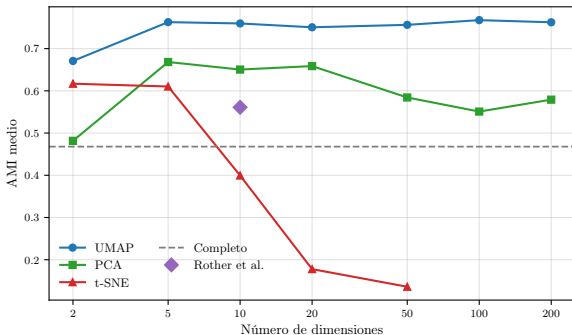
RoBERTa BNE Large por técnica de reducción (t-SNE 2D, sentido real).

Efecto de la reducción (2)



BETO cased por técnica de reducción (t-SNE 2D, sentido real).

Reducción: barrido de dimensiones



AMI medio frente al número de dimensiones, por técnica de reducción.

Las **10 dimensiones** del experimento principal son una elección razonable.

Efecto del método de agrupamiento

Método	AMI medio	ARI medio	k medio	Ruido medio
k -medias	0.672	0.605	3.8	—
HDBSCAN	0.535	0.409	6.3	18 %
DBSCAN	0.487	0.321	13.6	21 %

Medias por método de agrupamiento (todas las combinaciones).

- k -medias gana **de media** y es el **más fiable**.
- Pero el **mejor caso** individual es de HDBSCAN (*RoBERTa BNE Large*, UMAP 10D, AMI 0,987).
- Densidad: mucho **ruido** (18–21 %) y **sobre-segmenta** (DBSCAN hasta 56 grupos).
- k -medias (3,8 grupos) se acerca a los **4 sentidos**.

Sentidos compactos y convexos: escenario **ideal para k -medias**. La densidad iguala en su mejor caso, pero k -medias es más **fiable**.

Interacción modelo–reducción

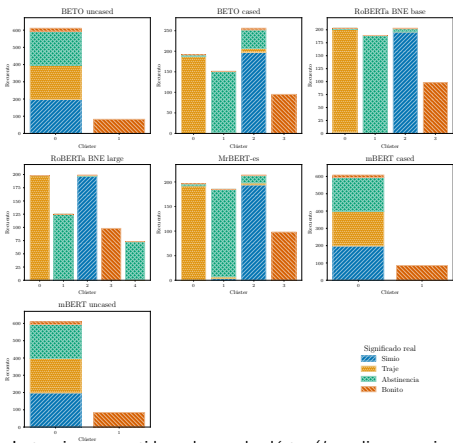
Modelo	Completo	PCA 10D	UMAP 10D	t-SNE 10D	Rother et al.
<i>RoBERTa BNE Base</i>	0.941	0.937	0.975	0.952	0.863
<i>MrBERT-es</i>	0.857	0.857	0.821	0.929	0.907
<i>RoBERTa BNE Large</i>	0.911	0.980	0.828	0.839	0.788
<i>BETO cased</i>	0.776	0.796	0.902	0.636	0.727
<i>BETO uncased</i>	0.346	0.814	0.340	0.387	0.625
<i>mBERT cased</i>	0.363	0.524	0.357	0.261	0.396
<i>mBERT uncased</i>	0.357	0.357	0.357	0.314	0.502

AMI por modelo y técnica de reducción (k -medias, 10D). Negrita: mejor reducción de cada modelo.

- **Fuertes:** ya altos en completo; reducción **apenas** ayuda.
- **Débiles:** saltan con la reducción **adecuada**.

La **mejor reducción depende del modelo**: el mejor resultado depende de la combinación de ambos factores.

Composición de los clústeres

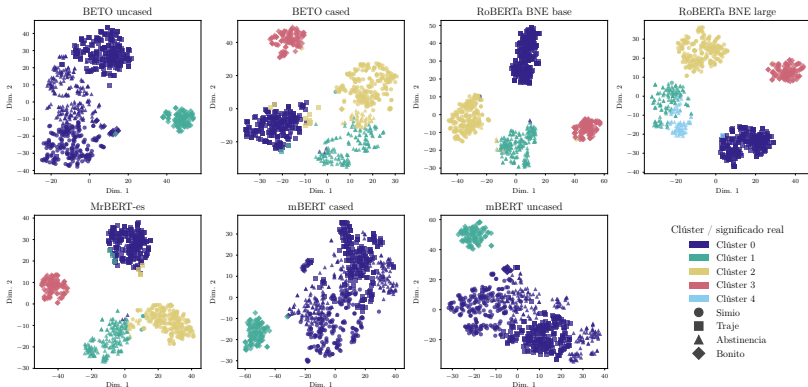


- k -medias separa los sentidos **casi a la perfección** en los modelos fuertes.

Algunos modelos no tienen una relación 1:1 con los sentidos reales:

- *RoBERTa BNE Large* **subdivide** 'abstinerencia'.
- Modelos débiles juntan varios sentidos, separación depende de la **categoría gramatical** (sustantivo / adjetivo).

Visualización de los clústeres agrupados



Proyección t-SNE 2D del espacio completo, coloreada por clúster (k -medias).

Subdivisión de 'abstinencia'

RoBERTa BNE Large parte el sentido 'abstinencia' en dos grupos separados por diferencias **semánticas** y de **construcción sintáctica**.

Sin complemento (sintagma nominal)

- «está pasando el mono»
- «los síntomas del mono»
- «aguantar el mono»

Padecimiento físico de la abstinencia en general.

Con complemento (*mono de algo*)

- «mono de café»
- «mono de nicotina»
- «mono de videojuegos»

Necesidad o antojo del objeto que nombra.

Admite reemplazo por *necesidad* («**necesidad de café**»).

No son **errores**: el algoritmo encuentra distinciones **reales**, aunque la métrica la penalice porque 'abstinencia' es un solo sentido.

XL-WSD-es: un conjunto del mundo real

XL-WSD-es

- Conjunto real y difícil: sólo 5,92 ejemplos por palabra \Rightarrow sólo espacio completo, sin reducción.
- De nuevo, k -medias es el mejor método de agrupamiento.

Modelo	Método	ARI	AMI
<i>BETO cased</i>	k -medias	0.303	0.324
<i>MrBERT-es</i>	k -medias	0.300	0.332
<i>BETO uncased</i>	k -medias	0.296	0.332
<i>mBERT uncased</i>	k -medias	0.291	0.322
<i>RoBERTa BNE Base</i>	k -medias	0.289	0.302
<i>mBERT cased</i>	k -medias	0.253	0.267
<i>RoBERTa BNE Large</i>	k -medias	0.198	0.232

Mejor resultado por modelo (espacio completo), ordenado por ARI.

XL-WSD-es: comparación con el trabajo previo

XL-WSD-es

Sistema	ARI
Base (<i>BERT</i> LSDP)	0.205
Este trabajo (<i>BETO</i> <i>cased</i> , <i>k-medias</i>)	0.303
WCM-en SDP-en [1]	0.380
WCM-ru SDP [1]	0.393

Comparación con Kokosinskii et al. [1].

- **Supera** la línea base (+0,098 ARI).
- Por debajo de sus mejores sistemas de sustitución léxica (−0,077 / −0,090).

Conclusiones

- 1 Modelos **sólo español** > **multilingües** (AMI 0,987).
- 2 k -medias **fiable**; densidad (DBSCAN, HDBSCAN) **inestable**.
- 3 **Reducción** (UMAP) ayuda si hay **suficientes puntos**.
- 4 **Competitivo** con el trabajo previo.
- 5 Clústering de *embeddings* de *BERT* **funciona para la WSI en castellano**: techo MONO (AMI \approx 0,99), suelo XL-WSD-es (ARI \approx 0,33).

Contribuciones

- **Estudio comparativo de WSI en castellano**, campo dominado por el inglés: varios modelos, reducciones de dimensionalidad y métodos de agrupamiento.
- **Línea base reproducible** para la WSI no supervisada en castellano, resultados sobre un *benchmark* real (XL-WSD-es).
- **Nuevo conjunto de datos**: MONO publicado en Kaggle.

Limitaciones y trabajo futuro

Limitaciones

- MONO es **construido**: los usos son arquetípicos, ejemplos muy claramente distintos. En corpus reales, puede que haya más dispersión, no sabemos si anisótropa.
- XL-WSD-es es **pequeño** (213 instancias, 36 palabras, 5,92 ejemplos/palabra): impide la reducción y eleva la varianza.

Trabajo futuro

- Construir un **conjunto anotado más grande** y real para el castellano.
- **Análisis de varianza** de la contribución de cada factor (modelo, reducción, método) a la calidad del clústering.
- Explorar la **sustitución léxica**, dominante en inglés, frente al clústering directo.

Gracias por su atención.

Referencias I

- [1] Denis Kokosinskii y Nikolay Arefyev. “Multilingual Substitution-based Word Sense Induction”. En: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Ed. por Nicoletta Calzolari et al. Torino, Italia: ELRA and ICCL, mayo de 2024, págs. 11859-11872. (Visitado 26-05-2026).

Referencias II

- [2] Tommaso Pasini, Alessandro Raganato y Roberto Navigli. “XL-WSD: An Extra-Large and Cross-Lingual Evaluation Framework for Word Sense Disambiguation”. En: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.15 (mayo de 2021), págs. 13648-13656. ISSN: 2374-3468, 2159-5399. DOI: [10.1609/aaai.v35i15.17609](https://doi.org/10.1609/aaai.v35i15.17609). (Visitado 07-06-2026).

Referencias III

- [3] David Rother, Thomas Haider y Steffen Eger. “CMCE at SemEval-2020 Task 1: Clustering on Manifolds of Contextualized Embeddings to Detect Historical Meaning Shifts”. En: *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. Ed. por Aurelie Herbelot et al. Barcelona (online): International Committee for Computational Linguistics, dic. de 2020, págs. 187-193. DOI: [10.18653/v1/2020.semeval-1.22](https://doi.org/10.18653/v1/2020.semeval-1.22). (Visitado 25-04-2026).
- [4] Gregor Wiedemann et al. “Does BERT Make Any Sense? Interpretable Word Sense Disambiguation with Contextualized Embeddings”. En: (2019).

Resultados completos en MONO

Modelo	Proyección	Método	k	AMI	ARI	Ruido
<i>RoBERTa BNE Large</i>	UMAP 10D	HDBSCAN	4	0.987	0.992	0.00 %
<i>MrBERT-es</i>	UMAP 10D	HDBSCAN	4	0.981	0.987	0.00 %
<i>RoBERTa BNE Large</i>	PCA 10D	k -medias	4	0.980	0.988	—
<i>RoBERTa BNE Base</i>	UMAP 10D	k -medias	4	0.975	0.979	—
<i>RoBERTa BNE Base</i>	t-SNE 10D	k -medias	4	0.952	0.964	—
<i>RoBERTa BNE Base</i>	Completo	k -medias	4	0.941	0.955	—
<i>RoBERTa BNE Base</i>	PCA 10D	k -medias	4	0.937	0.951	—
<i>MrBERT-es</i>	t-SNE 10D	k -medias	4	0.929	0.942	—
<i>RoBERTa BNE Large</i>	Completo	k -medias	5	0.911	0.885	—
<i>MrBERT-es</i>	Rother et al.	k -medias	4	0.907	0.926	—

Diez mejores combinaciones de modelo, proyección y método, ordenadas por AMI.

Mejor configuración por modelo (MONO)

Modelo	Proyección	Método	k	AMI	ARI	Ruido
<i>RoBERTa BNE Large</i>	UMAP 10D	HDBSCAN	4	0.987	0.992	0.00 %
<i>MrBERT-es</i>	UMAP 10D	HDBSCAN	4	0.981	0.987	0.00 %
<i>RoBERTa BNE Base</i>	UMAP 10D	k -medias	4	0.975	0.979	—
<i>BETO cased</i>	UMAP 10D	k -medias	5	0.902	0.877	—
<i>BETO uncased</i>	PCA 10D	k -medias	4	0.814	0.840	—
<i>mBERT uncased</i>	UMAP 10D	DBSCAN	3	0.553	0.448	0.00 %
<i>mBERT cased</i>	PCA 10D	k -medias	3	0.524	0.438	—

- Los modelos entrenados **sólo en español** (*BETO*, *RoBERTa BNE*) superan a los **multilingües** (*mBERT*).
- k -medias aparece como el método más fiable en casi todos los modelos.

AMI por modelo y reducción (k -medias)

Modelo	Completo	PCA 10D	UMAP 10D	t-SNE 10D	Rother et al.
<i>RoBERTa BNE Base</i>	0.941	0.937	0.975	0.952	0.863
<i>MrBERT-es</i>	0.857	0.857	0.821	0.929	0.907
<i>RoBERTa BNE Large</i>	0.911	0.980	0.828	0.839	0.788
<i>BETO cased</i>	0.776	0.796	0.902	0.636	0.727
<i>BETO uncased</i>	0.346	0.814	0.340	0.387	0.625
<i>mBERT cased</i>	0.363	0.524	0.357	0.261	0.396
<i>mBERT uncased</i>	0.357	0.357	0.357	0.314	0.502

Los modelos débiles (*BETO uncased*, *mBERT uncased*) mejoran mucho con ciertas reducciones; los fuertes ya parten de un AMI alto en el espacio completo.

Métricas por método de agrupamiento (MONO)

Método	AMI medio	ARI medio	k media	Ruido medio (%)
k -medias	$0,672 \pm 0,248$	$0,605 \pm 0,314$	$3,8 \pm 1,7$	—
HDBSCAN	$0,535 \pm 0,265$	$0,409 \pm 0,323$	$6,3 \pm 6,7$	$18,03 \pm 26,27$
DBSCAN	$0,487 \pm 0,229$	$0,321 \pm 0,275$	$13,6 \pm 15,5$	$21,08 \pm 29,18$

- k -medias gana de media; los métodos de densidad generan **mucho ruido** (18–21 %) y demasiados grupos.
- Su mejor caso individual (HDBSCAN, AMI 0,987) iguala a k -medias, pero es menos fiable.

Resultados completos en XL-WSD-es

Modelo	Método	ARI	AMI
<i>BETO cased</i>	<i>k-medias</i>	0.303	0.324
<i>MrBERT-es</i>	<i>k-medias</i>	0.300	0.332
<i>BETO uncased</i>	<i>k-medias</i>	0.296	0.332
<i>mBERT uncased</i>	<i>k-medias</i>	0.291	0.322
<i>RoBERTa BNE Base</i>	<i>k-medias</i>	0.289	0.302
<i>mBERT cased</i>	<i>k-medias</i>	0.253	0.267
<i>RoBERTa BNE Large</i>	<i>k-medias</i>	0.198	0.232

- Mejor resultado por modelo (espacio completo, todos los métodos), ordenado por ARI.
- El orden **cambia** respecto a MONO: *BETO cased* pasa a ser el mejor; *RoBERTa BNE* cae. Todos los mejores usan *k-medias*.

Generación de MONO

Oraciones generadas con **Gemini**, una definición (en inglés) por acepción:

Acepción	Definición utilizada
'Simio'	«The word "mono" means a primate or monkey».
'Bonito, lindo'	«The word "mono" is an adjective meaning cute, pretty or nice-looking».
'Prenda de una pieza'	«The word "mono" means a one-piece garment. A jumpsuit, overalls, onesie or work suit».
'Síndrome de abstinencia'	«The word "mono" means drug or substance withdrawal syndrome (craving / abstinence)».

Se descartaron las oraciones sin *mono* o con la palabra repetida, para evitar ambigüedades en la evaluación.

Hiperparámetros y reproducibilidad

Barrido de dimensiones (2–200) con PCA, UMAP y t-SNE; $k \in [2, 3 n_{\text{sent}}]$.
Versiones principales (sobre Python 3.14):

Biblioteca	Versión	Biblioteca	Versión
numpy	2.3.5	hdbscan	0.8.42
pandas	2.3.3	kneed	0.8.6
scipy	1.16.3	dbcv	0.1.1
transformers	4.57.3	matplotlib	3.10.8
scikit-learn	1.8.0	seaborn	0.13.2
umap-learn	0.5.12	altair	6.0.0

Detalle del agrupamiento

Tres algoritmos; sus hiperparámetros se eligen con **métricas intrínsecas** (sin etiquetas), de forma coherente con la pipeline no supervisada.

Algoritmo	Hiperparámetro	Selección	Respaldo
k -medias	$k \in [2, 3 n_{\text{sent}}]$	máx. silueta (compacta y convexa, afín a k -medias)	—
DBSCAN	ε (<code>min_samples= 5</code> , dist. coseno)	barrido de candidatos de la curva k -NN, máx. DBC CV	codo (Kneedle) → mediana k -NN
HDBSCAN	<code>min_cluster_size</code> × <code>min_samples</code> (rejilla)	máx. DBC CV	máx. n° clústeres, mín. ruido

Detalle del agrupamiento (cont.)

- **DBC**V no está definida con < 2 clústeres \Rightarrow los respaldos garantizan asignación a todo modelo (ninguno se descarta).
- **HDBSCAN**: coseno aproximado por euclídea sobre normalización L2 (\equiv distancia angular), ya que la biblioteca no admite coseno.
- **Ruido** (DBSCAN/HDBSCAN): tratado como clúster adicional en AMI/ARI/silueta; DBCV lo penaliza de forma natural.
- El codo (Kneedle) sobre la inercia para k se probó y descartó: peor que silueta.

La maldición de la dimensionalidad I

Al crecer el número de dimensiones, el volumen del espacio crece de forma **exponencial** y los datos quedan **dispersos**: para mantener la misma densidad harían falta exponencialmente más muestras (Bellman [1]; efecto Hughes [2]).

- **Las distancias colapsan.** En alta dimensión, la distancia al vecino más cercano se acerca a la del más lejano: «cercanía» pierde poder discriminativo.
- **Afecta al agrupamiento.** k -medias (distancias) y los métodos de densidad (DBSCAN, HDBSCAN) y la silueta dependen de distancias bien definidas.
- **Nuestro caso.** Los *embeddings* tienen cientos de dimensiones (768–1024) frente a pocas muestras por palabra.

Por eso se reduce la dimensionalidad antes de agrupar: PCA, UMAP y t-SNE recuperan estructura y mejoran las métricas (ver el barrido de dimensiones).

La maldición de la dimensionalidad II

Por qué colapsan las distancias: en alta dimensión el volumen **huye a los bordes**.

Dim. d	$V_{\text{esfera}}/V_{\text{cubo}}$	Cáscara exterior 10 %
2	0.785	0.19
3	0.524	0.27
10	0.0025	0.65
100	$\sim 10^{-70}$	0.9997

- **Esquinas.** La esfera inscrita en el hipercubo ocupa una fracción del volumen que $\rightarrow 0$: casi todo queda en las esquinas, a distancia \sqrt{d} del centro.
- **Cáscara.** En la esfera, casi todo el volumen se concentra en una fina corteza exterior: casi todos los puntos viven cerca de la superficie.

La maldición de la dimensionalidad III

Los puntos acaban casi todos a una distancia parecida \Rightarrow vecino cercano \approx vecino lejano, y la noción de «cercanía» se diluye.

Reducción: barrido de dimensiones (números)

Proyección	Dim.	AMI medio	Desv.
UMAP	10	0.704	0.256
PCA	10	0.636	0.230
Rother et al.	10	0.600	0.170
Completo	-	0.468	0.209
t-SNE	10	0.414	0.307

AMI a 10 dimensiones (todas las técnicas igualadas).

Proyección	Dim.	AMI medio	Desv.
UMAP	100	0.767	0.213
PCA	5	0.668	0.229
t-SNE	2	0.617	0.332
Rother et al.	10	0.561	0.191
Completo	-	0.468	0.209

AMI con la dimensión óptima de cada técnica.

Con su **dimensión óptima**, t-SNE pasa de la última posición a superar al espacio completo: toda reducción ayuda.

Referencias (apéndice) I

- [1] R.E. Bellman. *Dynamic Programming*. Dover Books on Computer Science Series. Dover Publications, 2003. ISBN: 978-0-486-42809-3.
- [2] G. Hughes. “On the Mean Accuracy of Statistical Pattern Recognizers”. En: *IEEE Transactions on Information Theory* 14.1 (ene. de 1968), págs. 55-63. ISSN: 1557-9654. DOI: [10.1109/TIT.1968.1054102](https://doi.org/10.1109/TIT.1968.1054102). (Visitado 10-06-2026).