
I Jornadas de Innovación en Lengua, Empleabilidad e Inteligencia Artificial

Traducción automática, inteligencia artificial y el futuro de traductores e intérpretes

Antoni Oliver (aoliverg@uoc.edu)

Objetivos del seminario

- Presentar la evolución de la TA especialmente desde las metodologías basadas en corpus hasta la irrupción de la IA
- Ver como cada nueva metodología produce grandes expectativas y decepciones
- Principios básicos de IA aplicada a la traducción
- Calidad TAN - LLM
- Otras tareas relacionadas con la traducción
- Análisis de la evolución del mercado de la traducción, de los estudios de traducción y del uso de la tecnología en la traducción
- Conclusiones

Historia de la traducción automática

1930	1940	1950	1960	1970	1980	1990	2000	2010	2020	
Precursors i pioners			Grans expectatives	ALPAC	Dècada silenciosa	Sistemes operatius i comercials	TA estadística		TA neuronal	LLMs

Amenazas, desencantos y vaticinios



Traductores e intérpretes, ¿las próximas víctimas de la inteligencia artificial?

- El desarrollo de las redes neuronales artificiales amenaza a un entero sector

La inteligencia artificial sacude el negocio de la traducción: “Las máquinas son más baratas que las personas”

En 3 años la mayoría de traductores podrían ser sustituidos por la IA, según dicen algunas previsiones

SOFTWARE

El presidente chino es 'Mr. Mierdas': el vergonzoso error por el que Facebook pide perdón

CIENCIA Y TECNOLOGÍA

¿Muerte por traducción automática?

El exceso de confianza en las herramientas de traducción automática ha provocado un mal uso por parte de autoridades en situaciones de alto riesgo. La solución pasa por mejorar los sistemas, pero también por educar en su uso.



VISTO EN REDES

Error grotesco en la carta de postres en catalán de un restaurante: “Nada que hacer”

- El traductor automático hizo que un establecimiento ofrezca un postre surrealista en el menú

Elon Musk predice las profesiones más perjudicadas por la IA

Las 10 profesiones que la inteligencia artificial podría reemplazar en la próxima década, según Microsoft

De acuerdo con un informe reciente de la firma tecnológica, la IA está redefiniendo las tareas laborales y expone a ciertos empleos a un alto riesgo de automatización en un futuro

Las 10 profesiones que más peligro corren con la IA, según Microsoft

TECNOLOGÍA | INTELIGENCIA ARTIFICIAL

Las 40 profesiones que van a desaparecer por culpa de ChatGPT

Tanto si quieres dar un giro de 180° a tu carrera laboral o vas a empezar a formarte en una profesión, estos son los trabajos que pueden extinguirse por culpa de la inteligencia artificial.

https://www.nationalgeographic.com.es/tecnologia/40-profesiones-que-van-a-desaparecer-por-culpa-chatgpt_25974

Intérpretes y traductores

Historiadores

Auxiliares de pasajeros

Representantes de ventas de
servicios

Escritores y autores

Representantes de servicio al cliente

Programadores de herramientas
CNC

Operadores telefónicos

Agentes de billetes y empleados de
viajes

Locutores de radio y DJs

Empleados de corretaje

Educadores en gestión agrícola y
doméstica

Teleoperadores

Conserjes

Politólogos

Analistas de noticias, reporteros,
periodistas

Matemáticos

Redactores técnicos

Correctores y correctores de textos

Anfitriones y anfitrionas

Editores

Profesores de negocios, educación
postsecundaria

Especialistas en relaciones públicas

Demostradores y promotores de
productos

Agentes de ventas de publicidad

Nuevos empleados de cuentas

Asistentes estadísticos

Empleados de mostrador y de alquiler

Científicos de datos

Asesores financieros personales

Archivistas

Archivistas

Profesores de Economía, Educación
Postsecundaria

Desarrolladores web

Analistas de gestión

Geógrafos

Modelos

Analistas de investigación de mercado

Telecomunicadores de seguridad
pública

Operadores de centralita

Profesores de Bibliotecología,
Educación Postsecundaria

TECNOLOGÍA | GLOBAL

¿Va a estallar la burbuja de las IA?

Nik Martin

07/11/2025

Miles de millones de dólares han sido invertidos en la IA, lo que ha impulsado al alza las cotizaciones en la bolsa. Sin embargo, ya empiezan a aparecer grietas, como el aumento de los costos y los escasos retornos.

IA, la burbuja y el peligro del exceso de expectativas

MERCADOS E INVERSIONES · BOLSA

La barrera de los cinco billones de Nvidia y la burbuja de la IA: ¿lista para explotar?

Algunos inversores se están alejando de los valores más cotizados o tratando de diversificar sus carteras ante la posibilidad de que las empresas del sector estén cotizando por encima de su rendimiento real

Conceptos básicos de IA aplicada a la traducción

Inteligencia artificial

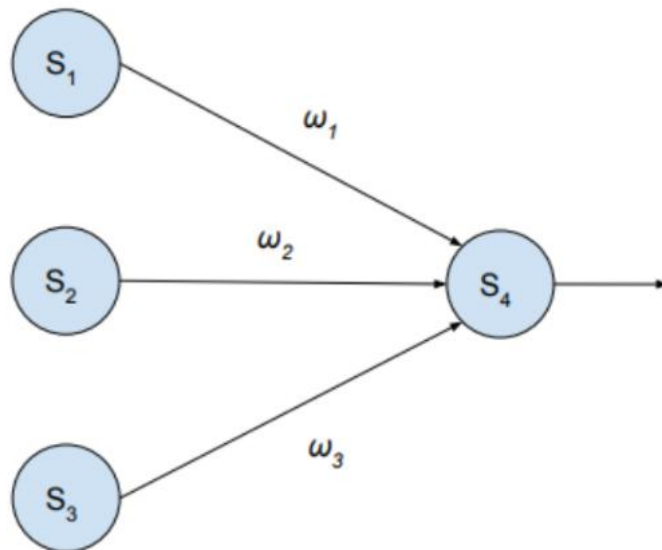
Una definición simplista de la Inteligencia Artificial (IA):

La capacidad de las máquinas o programas informáticos para imitar la inteligencia humana y realizar tareas que normalmente requieren pensamiento y aprendizaje.

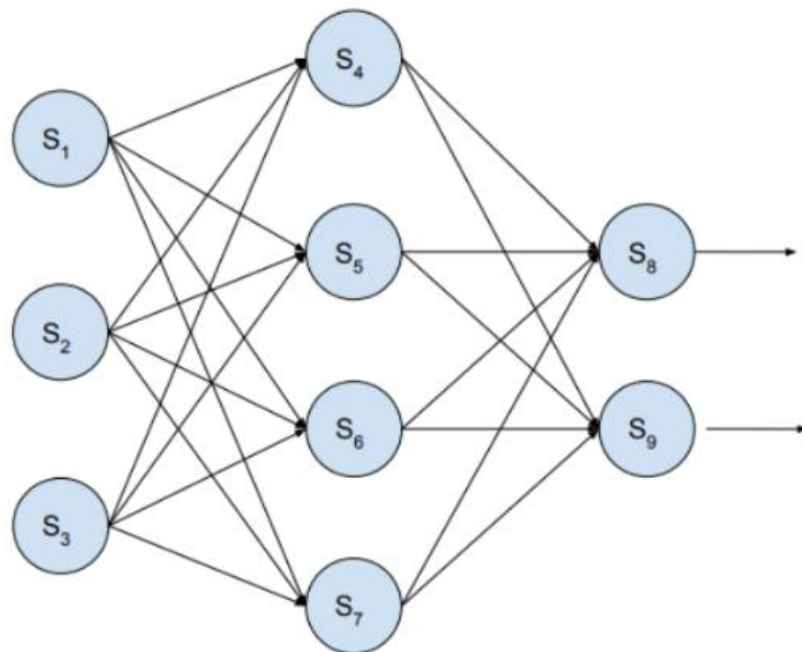
En traducción:

- Sistemas estadísticos?
- Sistemas neuronales
- Sistemas basados en LLM

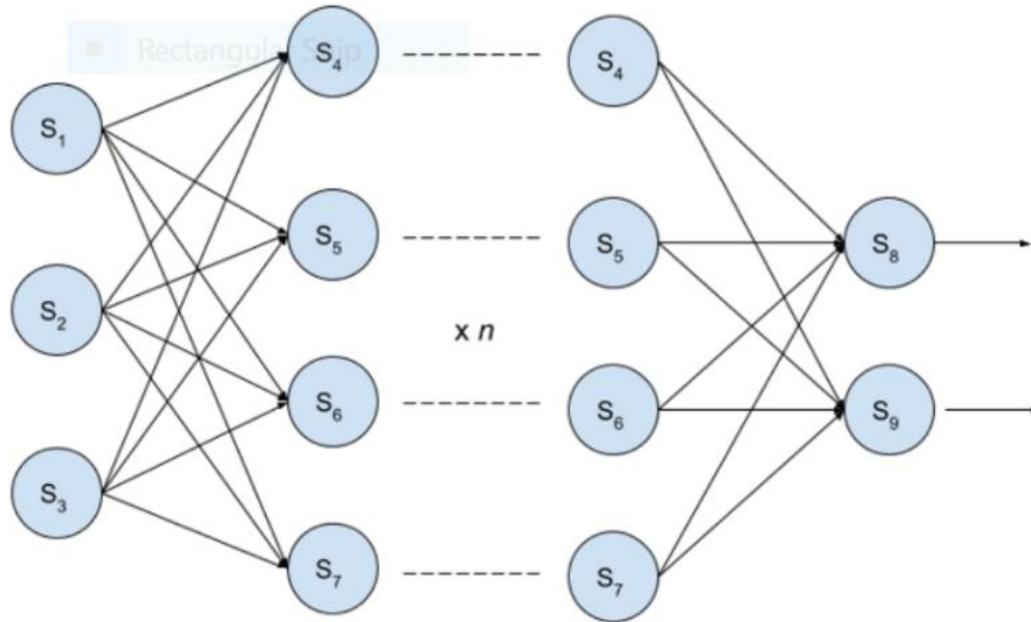
Neurona artificial



Red neuronal



Red neuronal profunda - *deep learning*



Word embeddings

coche 0.26469058 0.11690102 0.064389735 0.0999779 -0.4663968 -
0.039075375 -0.24742055 -0.3318912 0.13531172 0.25524357... 0.4009233
0.4096302 0.015230008 0.23592739 -0.256997 0.8789512 -0.5697038 -
0.044996012 -0.3101102 -0.55564547

Word embeddings y significado de las palabras

similitud entre perro y gato: 0.81765985

similitud entre perro y grifo: 0.14417386

rey - hombre + mujer:

reina 0.5088128447532654

zar 0.5039949417114258

conde 0.5027994513511658

príncipe 0.5018768906593323

monarca 0.49871590733528137

trono 0.495973140001297

corte 0.4910587966442108

Napoleón 0.4826952815055847

duque 0.4769935607910156

papa 0.4703159034252167

Madrid - España + Francia

París 0.6241433024406433

Nápoles 0.5885446071624756

Suecia 0.5847496390342712

Bélgica 0.5675634145736694

Polonia 0.549278736114502

Alemania 0.5441193580627441

Dinamarca 0.5427573919296265

Italia 0.5388016104698181

Londres 0.5377974510192871

Suiza 0.5317127108573914

Subpalabras (subwords)

- Problema: cada palabra (forma) necesita un word embedding
- Solución: palabras frecuentes y fragmentos de palabras
- Cada subpalabra tiene asociado un word embedding

```
_El _p í v ot _del _Barça _Lassa _comparte _el _p ich ic hi _de _la _liga  
_con _Ricard in ho , _con _5 _goles _en _3 _jornadas
```

```
_El _p iv ot _del _Barça _Lassa _comparteix _el _P itx itx i _de _la _Lliga  
_amb _Ricard in ho , _amb _cinc _gols _en _tres _jornades
```

- Lenguas con el mismo alfabeto: subpalabras comunes
- Lenguas con alfabetos diferentes: subpalabras diferentes
- En IA a cada subpalabra se suele denominar *token*
- Capacidades multilingües de los LLM

few-shot y zero-shot

- Few-shot: Capacidad de realizar una tarea con muy pocos ejemplos
- Zero-shot: Capacidad de realizar una tarea sin ningún ejemplo

Traducción automática neuronal

- Utiliza técnicas de inteligencia artificial
- Configuración de codificador-decodificador (encoder-decoder)
 - Codificador procesa la frase de entrada para crear una representación vectorial
 - Decodificador utiliza esta representación para generar la frase traducida.
- Se entrenan con grandes corpus paralelos
- Existen técnicas para utilizar también corpus monolingües
 - Corpus sintéticos
 - *Backtranslation*

Proceso TAN

Oración de entrada: El pívot del Barça Lassa comparte el pichichi de la liga con Ricardinho, con 5 goles en 3 jornadas.

Encoder: _El _p í v ot _del _Barça _Lassa _comparte _el _p ich ic hi
_de _la _liga _con _Ricard in ho , _con _5 _goles _en _3 _jornadas
</s>

Decoder: _El _p iv ot _del _Barça _Lassa _comparteix _el _P itx itx i
_de _la _Lliga _amb _Ricard in ho , _amb _cinc _gols _en _tres
_jornades

Oración traducida: El pivot del Barça del Barça Lassa comparteix el Pitxixi de la Lliga amb Ricardinho, amb cinc gols en tres jornades

Modelos extensos de lenguaje (LLM)

- Generalmente utilizan una configuración de solo decodificador (decoder-only)
 - Optimizada para la generación de texto secuencial prediciendo el siguiente token en función de todo el contexto anterior.

Entrenamiento de LLMs

- preentrenamiento (pretraining)
- ajuste fino por instrucciones (instruct fine tuning)

Conjuntos de instrucciones con respuestas (por ejemplo Alpaca)

```
"instruction": "What are the three primary colors?"
```

```
"input": "",
```

```
"output": "The three primary colors are red, blue, and yellow."
```

```
"instruction": "Summarize the following text within a couple of sentences.",
```

```
"input": "Many people believe that it is too complicated and expensive to eat healthily.  
After all, organic food is often much more expensive than its non-organic counterpart and  
requires more preparation time.",
```

```
"output": "Many people think that eating healthily is difficult and costly, as organic  
food is often more expensive than non-organic food and requires more effort in  
preparation."
```

```
"instruction": "Translate the following sentence into Spanish.",
```

```
"input": "The blue sky is so beautiful.",
```

```
"output": "El cielo azul es tan hermoso."
```

Ajustar todavía más para traducción

Tower Instruct

```
"instruction": "Translate the following English source text to Spanish:\nEnglish: Let's  
get started\nSpanish:",
```

```
"output": "Vamos a empezar."
```

SalamandraTA

```
"instruction": "Translate the following text from Spanish into Catalan.\nSpanish: Esto es  
un ejemplo de traducción automática.\nCatalan:",
```

```
"output": "Això és un exemple de traducció automàtica."
```

Responder/Traducir es una tarea de generar las palabras siguientes

Who invented the telephone?

Translate the following text from Spanish into English.

Spanish: Esto es un ejemplo de traducción automática.

English:

Corpus y parámetros

	Corpus (en tokens)	Parámetros modelos
TAN pequeño (OpusMT)	50M-500M	60M-100M
TAN grande	50M-500M	200M
LLM pequeño (Llama3 8B)	1000M-1500M	7000M-8000M
LLM grande (GPT-3, Llama3 70B)	3000M-7000M	70000M-175000M

Mejora de calidad

TAN eng - spa vs basada en reglas

Sistema	BLEU	chrF	TER
Apertium	12.1	44.1	73.1
OpusMT	26.6	55.0	58.2
NLLB	27.2	55.7	57.4
GoogleT	30.5	58.1	54.7
DeepL	28.6	56.2	56.0

TAN spa - eng vs basada en reglas

Sistema	BLEU	chrF	TER
Apertium	11.9	46.5	86.7
OpusMT	27.5	58.2	61.1
NLLB	29.5	59.5	58.1
GoogleT	28.3	54.0	63.3
DeepL	33.4	62.1	54.3

TAN eng - spa LLMs

Sistema	BLEU	chrF	TER
SalamandraTA-2b	30.0	57.4	55.5
SalamandraTA-7b	31.7	58.3	54.7
TowerInstruct-7B	28.5	53.6	56.9

TAN spa - eng LLMs

Sistema	BLEU	chrF	TER
SalamandraTA-2b	33.3	61.6	54.8
SalamandraTA-7b	36.0	63.1	51.7
TowerInstruct-7B	31.9	61.3	57.0

Otras tareas relacionadas con la traducción

Tutorial: exploración de LLMs con Ollama y MTUOC-Ollama-tester

<https://github.com/mtuoc/tutorials/wiki/Tutorial:-exploraci%C3%B3n-de-LLMs-con-Ollama-y-MTUOC%E2%80%90Ollama%E2%80%90tester>

Prompt:

"You're a professional English to Spanish translator specialized in economy. Translate the following sentence providing the translation and nothing else. "An interest rate is a percentage that represents the cost or return generated by a loan or certain saving instruments."

Resultado:

Una tasa de interés es un porcentaje que representa el costo o el rendimiento generado por un préstamo o ciertos instrumentos de ahorro.

Prompt:

You're a professional English to Spanish translator specialized in economy. Translate the following sentence providing the translation and nothing else. "An interest rate is a percentage that represents the cost or return generated by a loan or certain saving instruments." Use the following glossary: "interest rate": "tipo de interés"

Resultado:

Un tipo de interés es un porcentaje que representa el coste o el rendimiento generado por un préstamo o ciertos instrumentos de ahorro.

Prompt:

You're a professional English to Spanish translator specialized in economy. Translate the following sentence providing the translation and nothing else. "An interest rate is a percentage that represents the cost or return generated by a loan or certain saving instruments." Use the following glossary: "interest rate": "tipo de interés". Use the following example "A mortgage loan is a loan used either by purchasers of real property to raise funds to buy real estate, or by existing property owners to raise funds for any purpose while putting a lien on the property being mortgaged.": "Un préstamo hipotecario se define como un préstamo utilizado por compradores de bienes inmuebles para recaudar fondos para comprar bienes inmuebles, o por propietarios de propiedades existentes para recaudar fondos para cualquier propósito mientras se coloca un gravamen sobre la propiedad que se hipoteca."

Resultado:

Un tipo de interés se define como un porcentaje que representa el coste o el rendimiento generado por un préstamo o ciertos instrumentos de ahorro.

Prompt:

You're a professional English to Spanish translator specialized in economy. Translate the following sentence providing the translation and nothing else. "An interest rate is a percentage that represents the cost or return generated by a loan or certain saving instruments." Use the following glossary: "interest rate": "tipo de interés". Use the following example "A mortgage loan is a loan used either by purchasers of real property to raise funds to buy real estate, or by existing property owners to raise funds for any purpose while putting a lien on the property being mortgaged.": "Un préstamo hipotecario se define como un préstamo utilizado por compradores de bienes inmuebles para recaudar fondos para comprar bienes inmuebles, o por propietarios de propiedades existentes para recaudar fondos para cualquier propósito mientras se coloca un gravamen sobre la propiedad que se hipoteca.". Avoid the use of "préstamo" in the translation.

Resultado:

Un tipo de interés se define como un porcentaje que representa el coste o el rendimiento generado por una operación de crédito o ciertos instrumentos de ahorro.

SalamandraTA

Posedición automática:

Please fix any mistakes in the following {source}-{target} machine translation or keep it unedited if it's correct.

Source: {source_sentence}

MT: {machine_translation}

Corrected:"

Traducción a nivel de párrafo:

Please translate this text from {source} into {target}.

{source}: {paragraph}

{target}:

█

Reconocimiento de entidades nombradas:*

Analyse the following tokenized text and mark the tokens containing named entities.

Use the following annotation guidelines with these tags for named entities:

- ORG (Refers to named groups or organizations)
- PER (Refers to individual people or named groups of people)
- LOC (Refers to physical places or natural landmarks)
- MISC (Refers to entities that don't fit into standard categories).

Prepend B- to the first token of a given entity and I- to the remaining ones if they exist.

If a token is not a named entity, label it as O.

Input: {list of words in a sentence}

Marked:

Corrección gramatical:

Please fix any mistakes in the following {source} sentence or keep it unedited if it's correct.

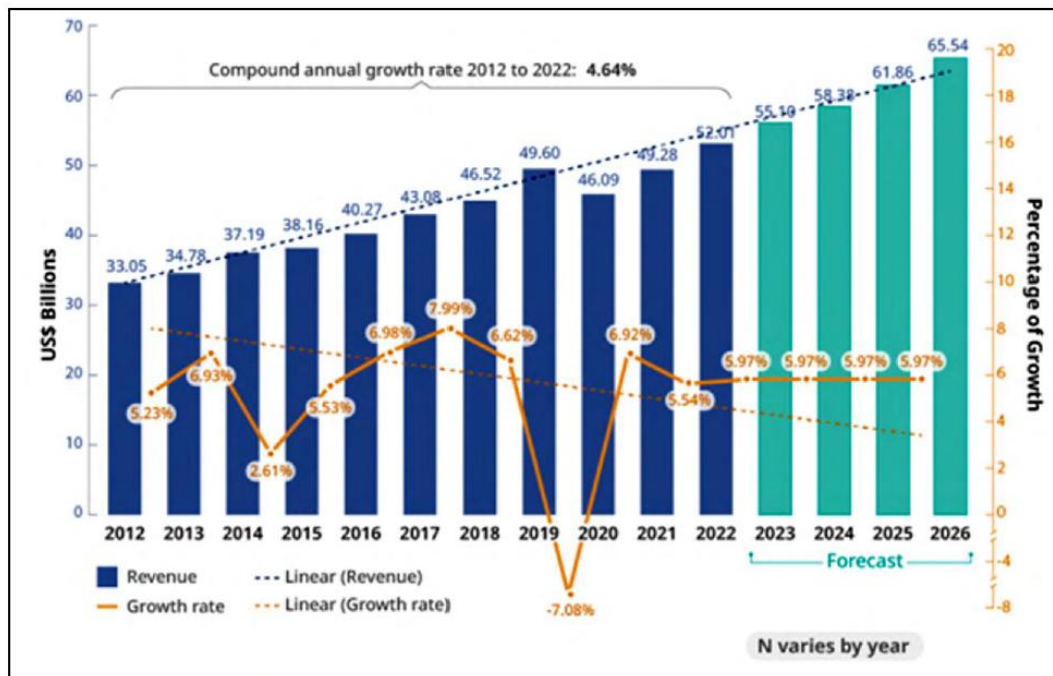
Sentence: {sentence}

Corrected:

Estado de la profesión, los estudios y el uso de la tecnología

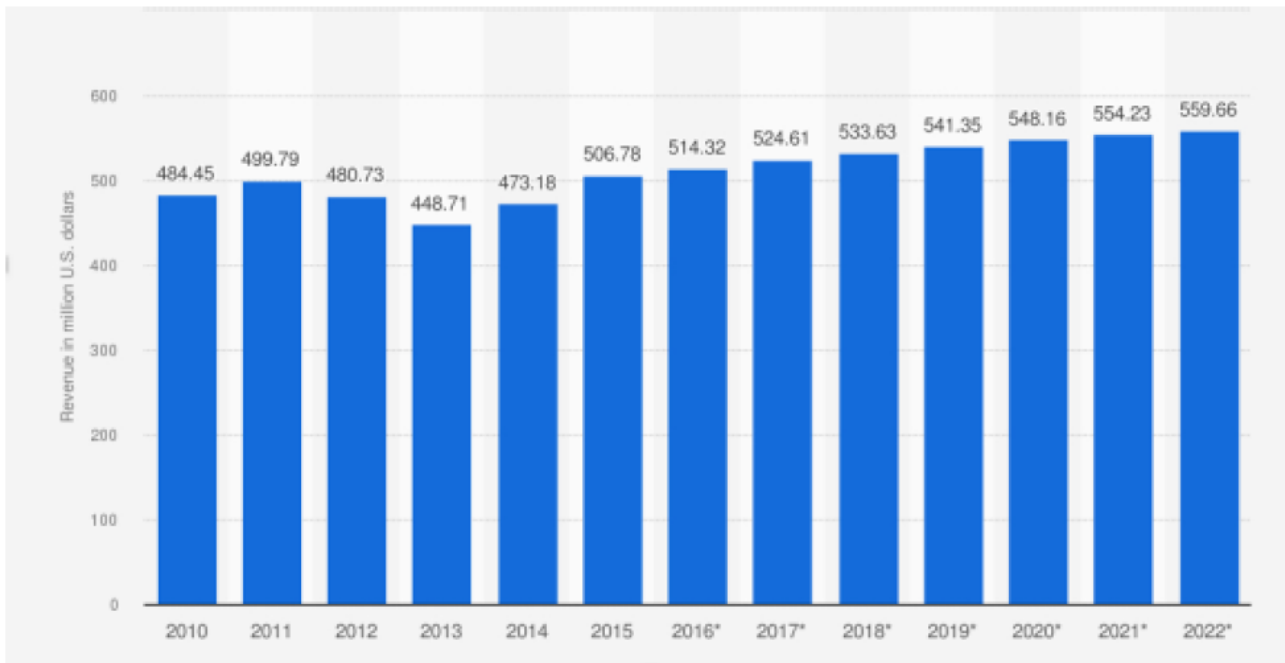
Situación y perspectivas del mercado de los servicios lingüísticos y la comunicación multilingüe en España. Análisis estratégico. - ANETI y ASPROSET

Gráfico 1. Ingresos de la industria de la lengua a escala global (2012-2022) y proyección hasta 2026



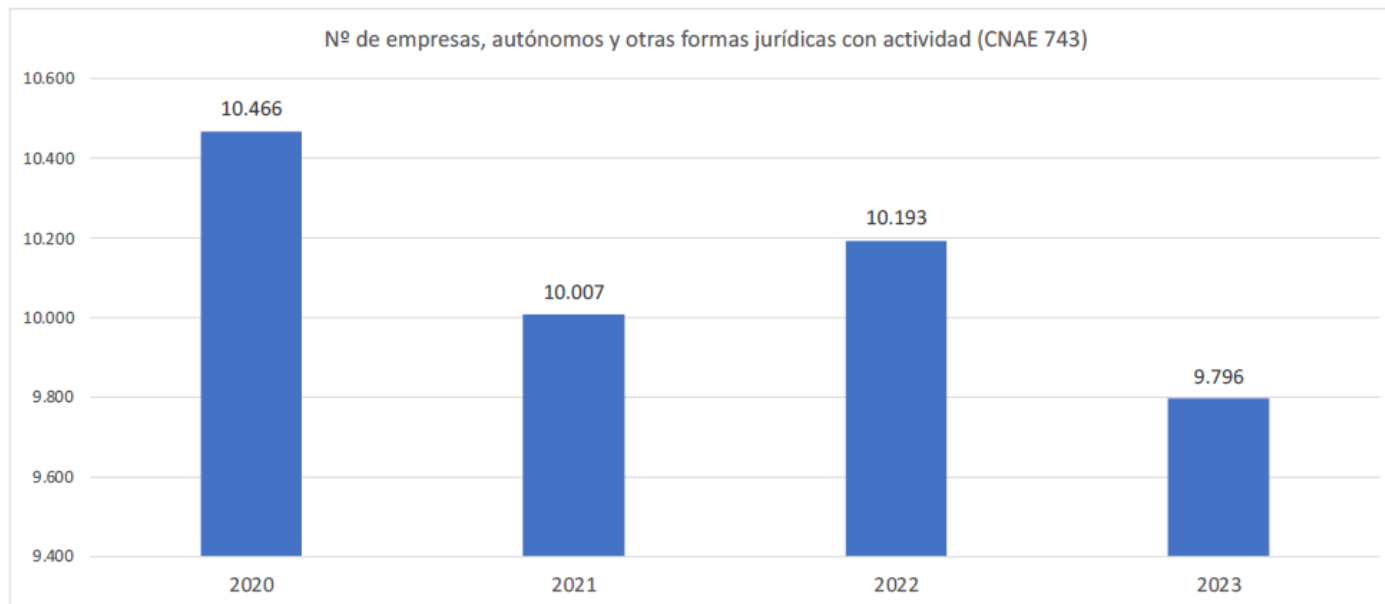
Fuente: CSA Research, 2023

Gráfico 3. Ingresos del mercado de los servicios lingüísticos en España (2010-2022)

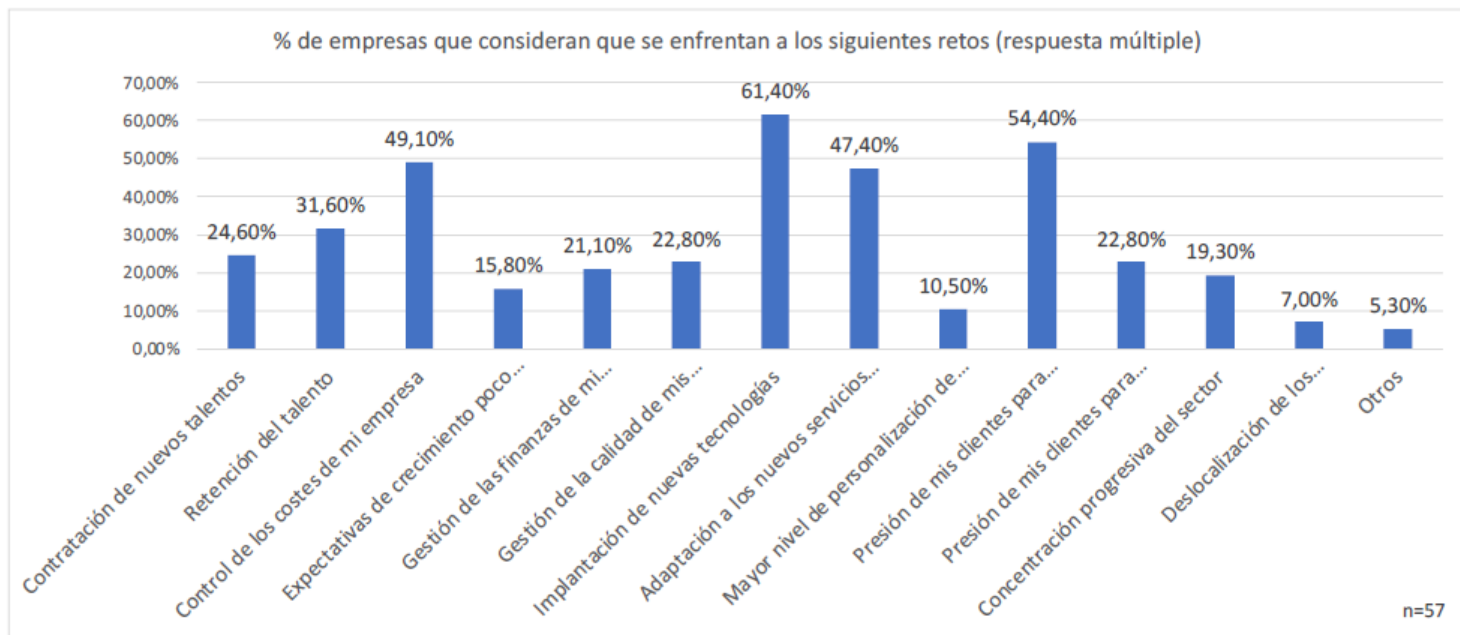


Fuente: Statista, 2023

Gráfico 4. Evolución del número de proveedores de servicios lingüísticos en España



Fuente: Elaboración a partir de INE, 2023



La tecnología como oportunidad

En general, se observa un consenso respecto a esta cuestión: la tecnología y en concreto la inteligencia artificial se perciben como una **oportunidad** entre todos los entrevistados, aunque también aparecen en el discurso términos como «**miedo**», «**incertidumbre**» o «**adaptación**», especialmente entre los autónomos.

...hay que tener en cuenta que **este sector siempre ha estado sometido a los cambios** provocados por la innovación tecnológica.

...todos los profesionales entrevistados consideran que es una **oportunidad** para el sector, pero también que requerirá **cambios en los perfiles de los profesionales** y un período de adaptación:

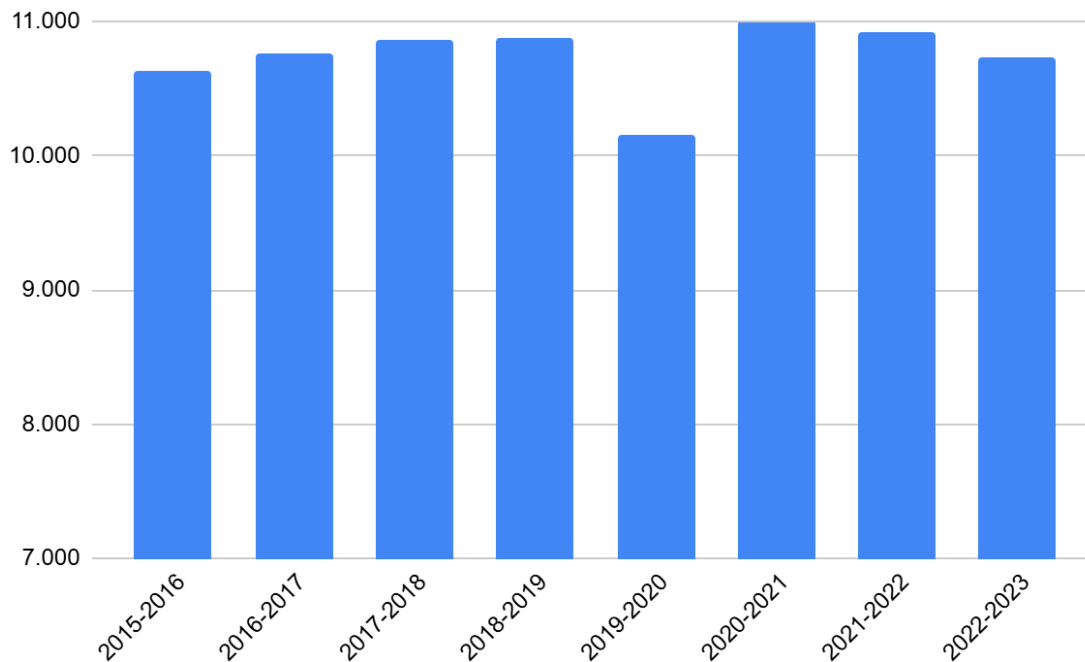
Los nuevos perfiles profesionales

el consenso es casi mayoritario —al menos entre clientes finales y grandes empresas del sector de los servicios lingüísticos— es en el **cambio** que la inteligencia artificial ha provocado en los **perfiles profesionales**

Estos perfiles profesionales son **revisores o poseedores**: personas que, tras una traducción realizada con inteligencia artificial tienen que aplicarse a la labor de poseer —es decir, revisar y adaptar el resultado de la traducción automática— y corregir el texto o de darle el sentido que una máquina no puede afinar de modo que el **resultado sea lo más parecido a una traducción humana**.

Estudiantes traducción e interpretación en España

fuelle: <https://estadisticas.universidades.gob.es/>



Uso de la tecnología en empresas de traducción

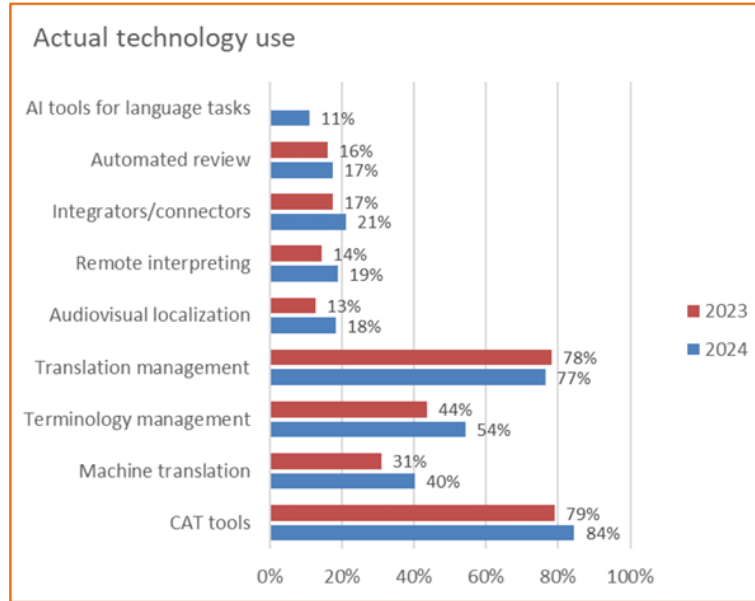


Figure 89 – Actual technology use - language companies

Impacto de la IA

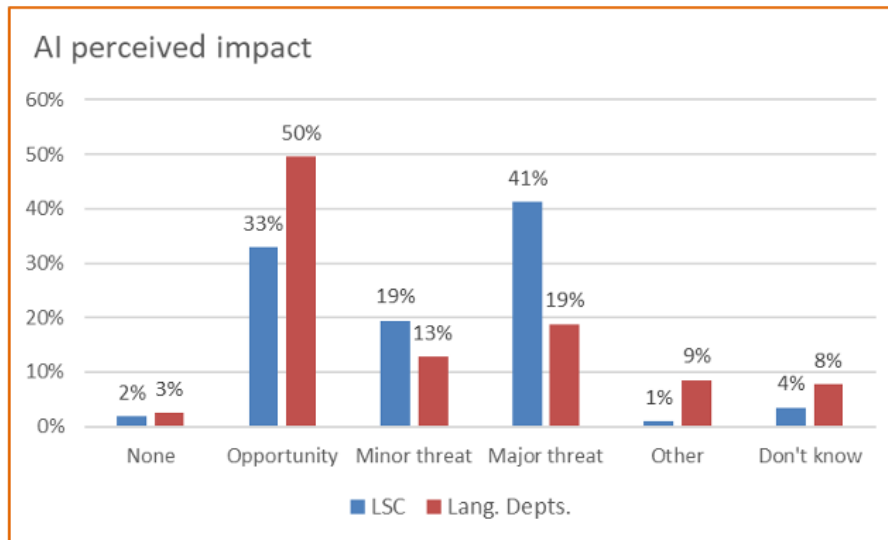


Figure 98 - AI perceived impact

ELIS 2024 European Language Industry Survey

Conclusiones

- No hay "una inteligencia artificial" si no una serie de modelos
- Modelos comerciales / modelos libres
- Aspectos éticos y medioambientales
 - ¿De quién son los datos que se utilizan?
 - Consumo energético y recursos (agua)
 - ¿Sustituirán a algunos trabajadores?
- El sector de la traducción está acostumbrado a nuevas herramientas y nuevas amenazas
- Dos situaciones:
 - La persona que sirve para corregir a la máquina
 - La persona que utiliza a la máquina
 - Que conoce la tecnología
 - Que escoge qué tecnología utilizar, cuándo y cómo

Gracias por vuestra atención

Antoni Oliver (aoliverg@uoc.edu)