



SPAIN AI
AI WORKSHOPS

WORKSHOP GRATUITO EN ESPAÑOL

NLP de 0 a 100 con Hugging Face

Apúntate en **eventbrite**



 7 SESIONES MARTES (quincenal)
DEL 13/07 AL 5/10 (18-18:40 CET)

Imparten:

- María Grandury
- Manuel Romero
- Omar Sanseviero
- Lewis Tunstall

¡SÍGUENOS!

@NLP_en_ES

@Spain_AI_

Enlaces útiles



[@nlp-en-es/nlp-de-cero-a-cien](#)



[#nlp-de-cero-a-cien](#)



[playlist: NLP de 0 a 100 con 🙌](#)



[@spain_ai](#)

[@nlp en es](#)



[@company/spainai](#)

[@company/nlp-en-es](#)



[spain-ai.com](#)

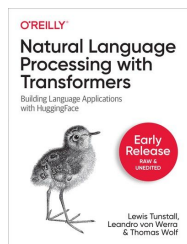
[nlp-en-es.org](#)




Sobre Lewis Tunstall @_lewtun



huggingface.co/course/



[NLP with Transformers](#)




Hugging Face

The AI community building the future.

📍 NYC + Paris 🔗 <https://huggingface.co/> Verified


[Overview](#) [Repositories 189](#) [Packages](#) [People 100](#) [Teams 6](#) [Projects 4](#) [Sponsoring 4](#)

Pinned

 **transformers**


📖 Transformers: State-of-the-art Natural Language Processing for Pytorch, TensorFlow, and JAX.

🐍 Python ⭐ 50.3k 📄 11.9k

 **datasets**


📖 The largest hub of ready-to-use datasets for ML models with fast, easy-to-use and efficient data manipulation tools

🐍 Python ⭐ 8.8k 📄 1.1k

 **tokenizers**


📖 Fast State-of-the-Art Tokenizers optimized for Research and Production

🦀 Rust ⭐ 4.8k 📄 375

 **knockknock**


📖 Knock Knock: Get notified when your training ends with only two additional lines of code

🐍 Python ⭐ 2.2k 📄 189

 **accelerate**

📖 A simple way to train and use PyTorch models with multi-GPU, TPU, mixed-precision

🐍 Python ⭐ 1.7k 📄 86

 **huggingface_hub**

all the open source things related to the Hugging Face Hub.

🐍 Python ⭐ 191 📄 34



Plan de ataque

Conceptos básicos

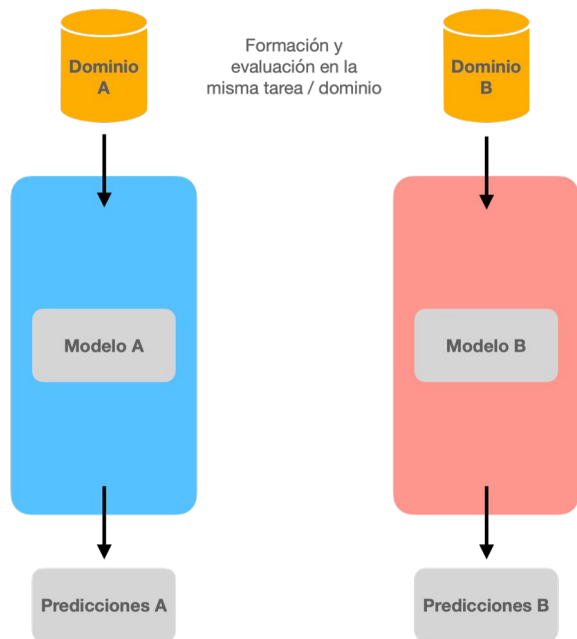
- aprendizaje de transferencia (“transfer learning”)
- modelado del lenguaje y modelos preentrenados (“pretrained models”)
- ajuste fino (“fine-tuning”)

Aplicación

- clasificación de texto de las reseñas de Amazon
- [Google Colab link](#)



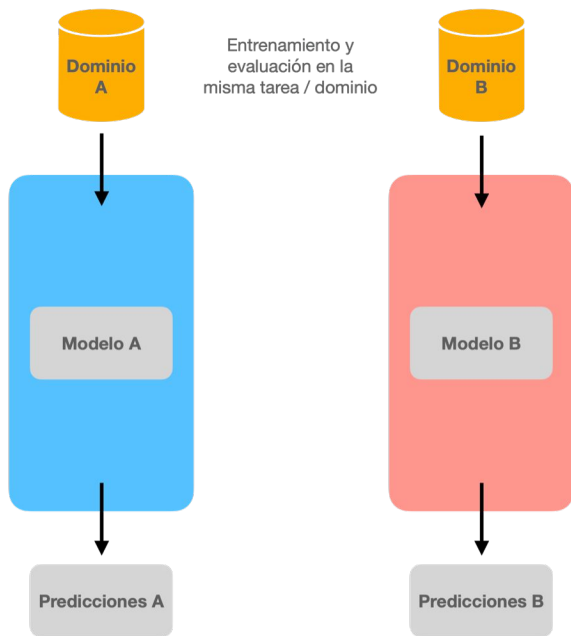
Aprendizaje de transferencia



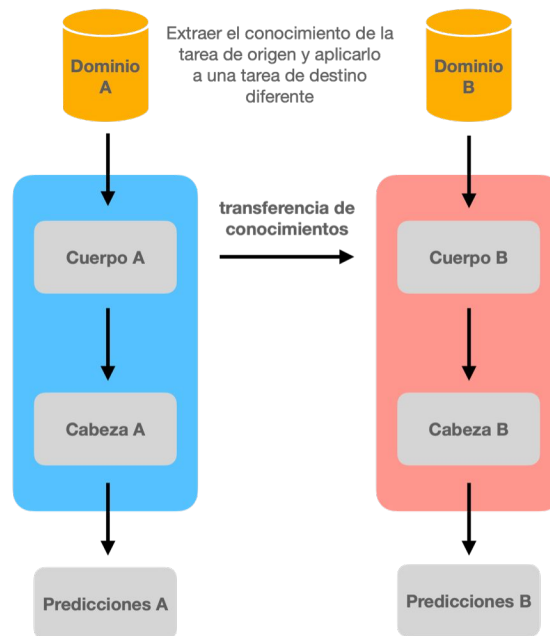
Aprendizaje supervisado



Aprendizaje de transferencia



Aprendizaje supervisado



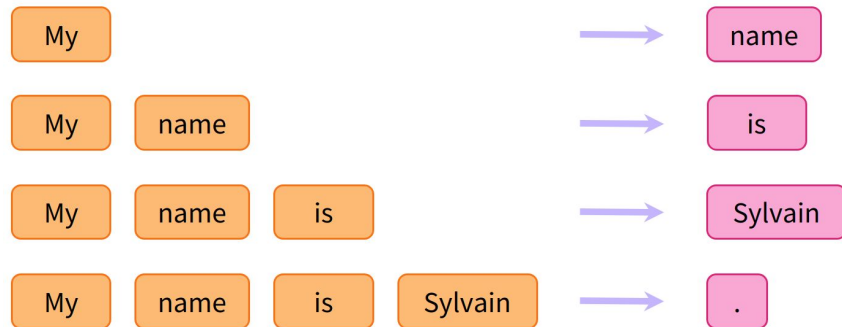
Aprendizaje de transferencia



¿Cuál es la tarea de origen?

Modelado del lenguaje

- “aprendizaje autosupervisado”, ¡no necesita etiquetas!
- dos estilos principales: *autorregresivo* y *enmascarado*



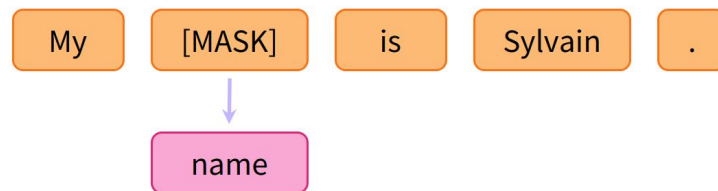
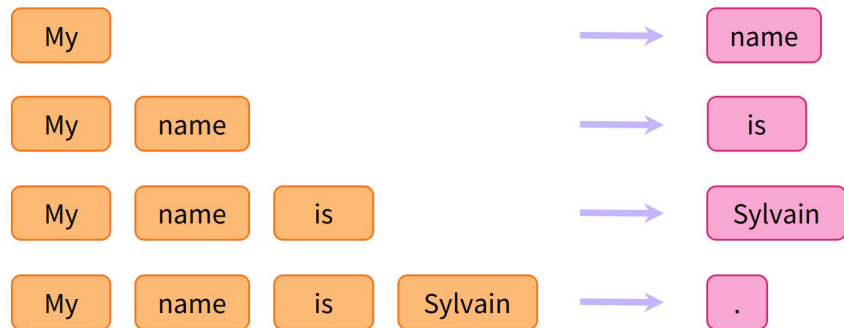
[Fuente de imágenes](#)



¿Cuál es la tarea de origen?

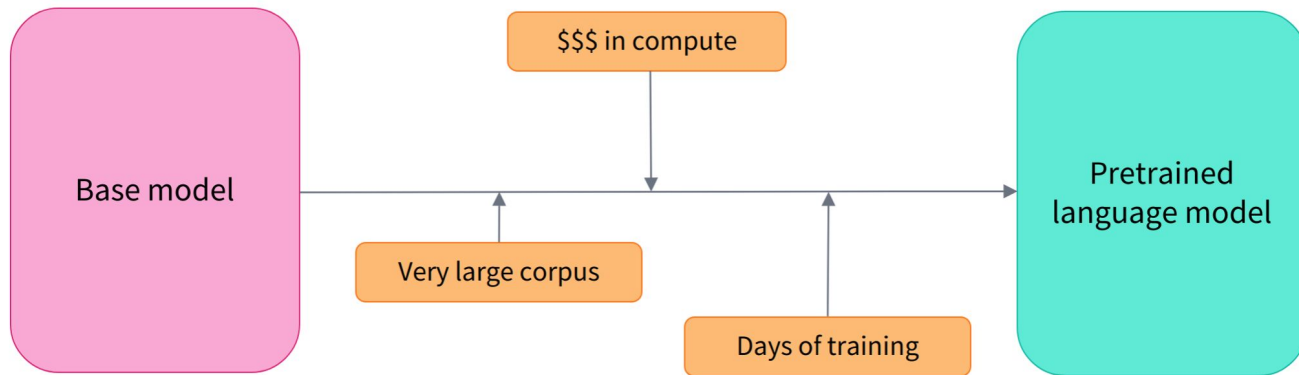
Modelado del lenguaje

- “aprendizaje autosupervisado”, ¡no necesita etiquetas!
- dos estilos principales: *autorregresivo* y *enmascarado*



Pre-entrenamiento

- en este paso entrenamos un modelo *desde cero*
- requiere *muchos datos y mucha computación*
- tras el entrenamiento, los “weights” se *reutilizan* en diversas tareas (clasificación de textos etc.)

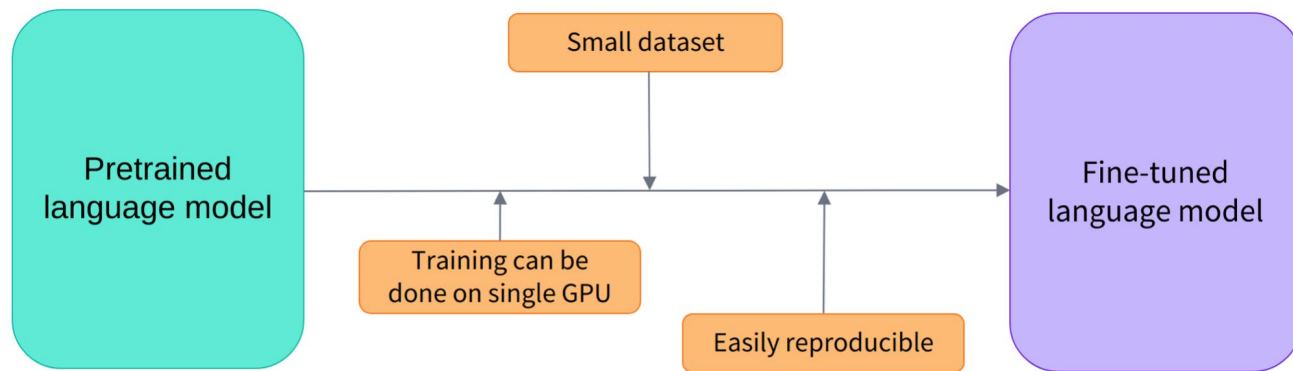


[Fuente de imagen](#)



Ajuste fino

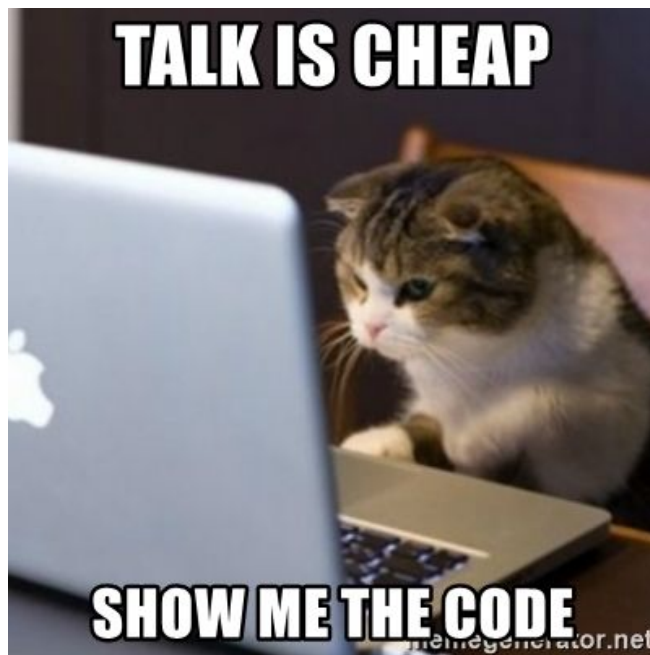
- utilizar los pesos del modelo preentrenado como “cuerpo”
- añadir una “cabeza” específica para la tarea (por ej. fully connected network + softmax)
- afinar los weights de la cabeza con un entrenamiento supervisado



[Fuente de imagen](#)



¡Hora de programar!



[Google Colab link](#)

@NLP_en_ES

