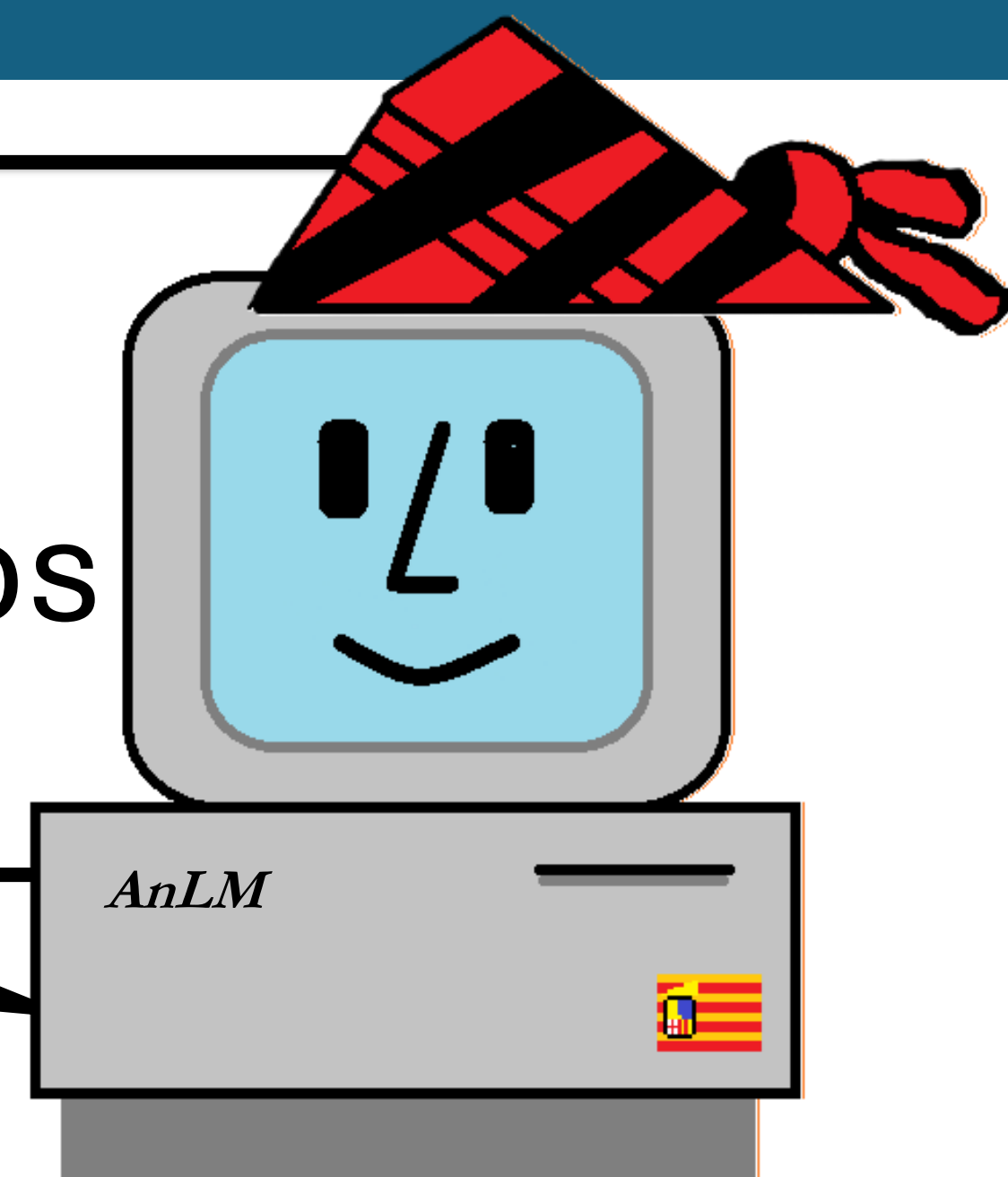


¿Una IA en *fabla aragonesa*?:

Inyección de conocimiento lingüístico en modelos basados en Transformer para transferencia croslingüe

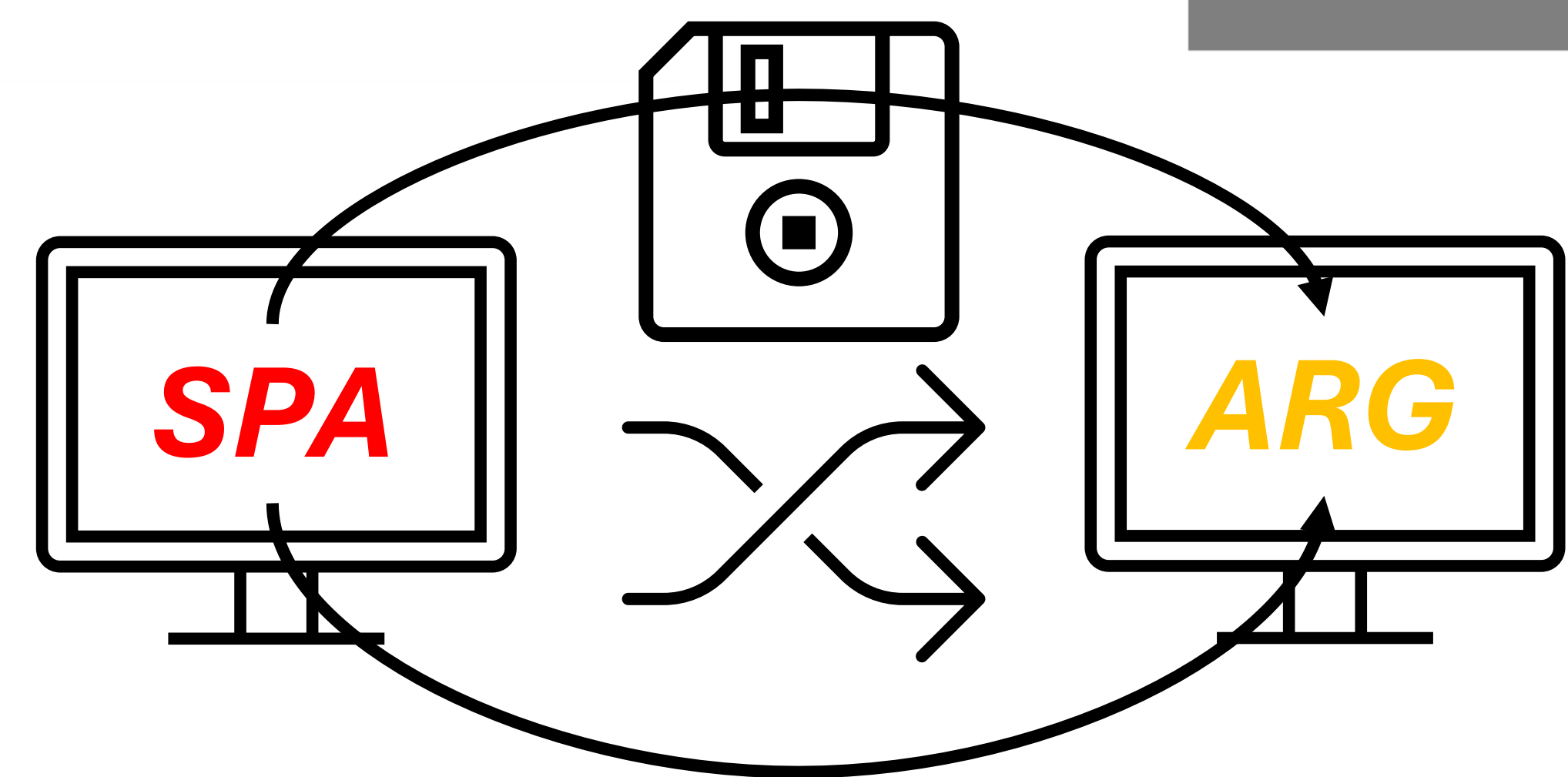
Miguel López-Otal, Jorge Gracia



Problema

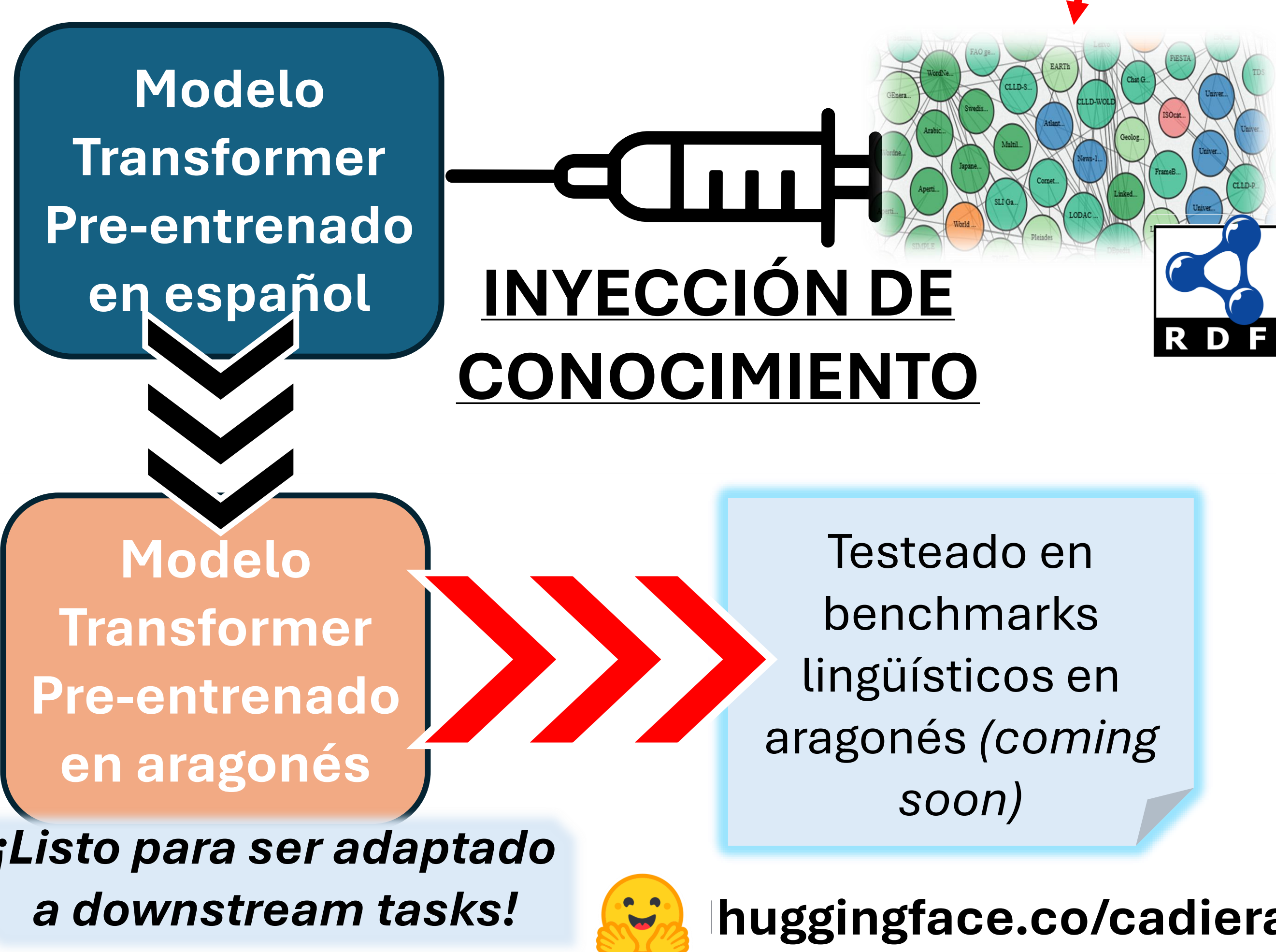
Modelos del lenguaje basados en Transformer **no existen para lenguas minoritarias**

➤ **RAZÓN:** gran cantidad de textos requeridos para entrenamiento + corpus inexistentes en esas lenguas



PROPUESTA PARA CROSS-LINGUAL TRANSFER

- Existen recursos lingüísticos de lenguas minoritarias en la Web Semántica, creados por expertos.
 - Ej.: diccionarios bilingües, corpus de textos...
 - Formatos estandarizados (e.g. Ontolex-Lemon) y **enlazados** → **Linguistic Linked Data**
 - Datos formateados como grafos (*knowledge graphs* lingüísticos!)



SOLUCIÓN: Cross-lingual transfer

Transferir las capacidades lingüísticas aprendidas por un modelo desde una lengua a otra

Modelo		COPA-silver	ESXNLI-silver	PAWS-silver
MAD-X (XLM-R+ Language Adapter [LA] an+Task Adapter [TA])	Setting TA train + test an	0.49, 0.45, 0.48	0.332, 0.362	0.805
	TA train es + test an	0.51, 0.44, 0.52	0.488, 0.5	0.810
BAD-X (XLM-R+ Language Adapter [LA] es-an + Task Adapter [TA])	TA train + test an	0.51, 0.43, 0.56	0.328, 0.346	0.821
	TA train es + test an	0.48, 0.37, 0.52	0.334, 0.346	0.824
RoBERTa-BNE(modelo es, baseline; fine-tuneado, train+test an)		0.49	0.334	0.577
RoBERTa-BNE (modelo es, baseline; adapter, train+test an)*		0.56, 0.54, 0.54	0.558, 0.53, 0.53	0.773
RoBERTa-AN (modelo an; fine-tuneado; train+test an)		0.48	0.334	0.577
RoBERTa-AN (modelo an; adapter; train+test an)*		0.48, 0.48, 0.48	0.342, 0.324, 0.324	
RoBERTa-BNE (modelo es, baseline; fine-tuneado, train+test es)		0.6	0.334	0.577
Lauscher et al. (2020) + Apertium RDF ES-AN*	Lauscher Adapter + Task Adapter	0.56, 0.54, 0.54	0.59, 0.584, 0.584	0.804
Lauscher et al. (2020) + Apertium RDF ES-AN + Texto en arg. (usado en MAD-X)*	Lauscher Adapter + Task Adapter	0.62, 0.53, 0.53	0.604, 0.578, 0.578	0.801

* Resultados obtenidos con posterioridad a los publicados en el artículo

¿Por qué cross-lingual transfer a aragonés? ¿Por qué así?

- Lenguas similares, pero con diferencias marcadas → **Baseline** para crear técnicas de cross-lingual transfer **más interpretables y controlables**
- Evitar el reentrenamiento con corpus textuales masivos → uso de reglas lingüísticas manejables y modificables, **con efectos contrastables** en los modelos tratados

NEW