

¿Una IA en *fabla aragonesa*?: Inyección de conocimiento lingüístico en modelos basados en Transformer para transferencia croslingüe

Miguel López-Otal, Jorge Gracia

Afiliación: Sistemas de Información Distribuidos (SID)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: mlopezotal@unizar.es

Resumen

Los modelos del lenguaje basados en Transformer, base de la actual revolución en IA, solo están disponibles para lenguas mayoritarias, dado el gran número de textos necesarios para entrenarlos. En este trabajo proponemos métodos para adaptar modelos preexistentes a lenguas minoritarias, especialmente el aragonés, por medio de inyección de conocimiento lingüístico estructurado.

Motivación

Los modelos del lenguaje preentrenados (PLMs), basados en la arquitectura Transformer [6], han alcanzado una gran aceptación en la industria y la investigación, gracias a sus demostradas capacidades en numerosos *benchmarks* y aplicaciones prácticas (por ejemplo, ChatGPT). Sin embargo, estos modelos están diseñados para modelar su conocimiento del mundo y de la lengua a partir de aprender la co-ocurrencia estadística de *tokens* en una gran cantidad de textos. Esto implica la necesidad de corpus textuales masivos para entrenar uno de estos modelos con éxito. Si bien esto no es un problema para lenguas mayoritarias, como el inglés o el español, supone una barrera de acceso para las lenguas minoritarias sin esos grandes corpus. Para remediar esta situación, ha habido propuestas para intentar adaptar modelos ya entrenados, en una lengua o varias, a otras lenguas, por medio de una metodología denominada “transferencia croslingüe” (o *cross-lingual transfer* en inglés). De estos tenemos el ejemplo de la *framework* MAD-X [5] o la tesis doctoral de Marinela Parović [4], que propone varias soluciones en este sentido. Muchas propuestas en este sentido, sin embargo, se basan en el reentrenado de modelos con textos de una lengua minoritaria, que no aportan una interpretabilidad de los resultados. En este contexto proponemos la creación de métodos interpretables y más controlables de transferencia lingüística para modelos Transformer, que utilicen para ello reglas gramaticales explícitamente

definidas de una lengua minoritaria. Dichas reglas estarán formateadas preferentemente como Datos Enlazados Lingüísticos (*Linguistic Linked Data* o LLD) [1] y codificarán, en un formato estructurado, normas gramaticales (morfología, sintaxis, vocabulario, etc.) de una lengua minoritaria. Estas servirán como fuente de conocimiento externo para modificar y adaptar modelos preexistentes a una lengua minoritaria, usando técnicas de inyección de conocimiento (*knowledge injection*). Como base práctica para el desarrollo de esta propuesta, se realizarán experimentos de transferencia croslingüe desde modelos entrenados en español a la lengua aragonesa, en su variante estándar. El aragonés carece de suficientes textos para entrenar un modelo desde cero, si bien sus reglas gramaticales son bien conocidas y estudiadas [2]. Esta lengua presenta una serie de marcadas diferencias frente al español (tales como diferente morfología o léxico), pero también similitudes respecto a dicha lengua (tales como una sintaxis similar), que facilitaría un proceso de transferencia de modelos de una lengua a otra. Los resultados obtenidos a lo largo de esta investigación servirán para ser extrapolados a otras lenguas, si bien los recursos desarrollados expresamente para el caso del aragonés –datos enlazados lingüísticos sobre su gramática, datasets de evaluación en dicha lengua, etc.– solo serán utilizables en ese contexto concreto.

Representación de reglas gramaticales

Para poder realizar el proceso de modelado de una ontología gramatical en un tiempo razonable, se propone modelar exclusivamente una serie de reglas gramaticales que diferencien al castellano del aragonés, para así ayudar en su transferencia lingüística. Si bien esto limita la propuesta a lenguas tipológicamente similares, esto sería contrarrestado por la potencial controlabilidad e interpretabilidad del proceso de transferencia. Se propone utilizar como base la Gramática del Aragonés [2] y el diccionario morfológico en aragonés de la plataforma

Apertium. Se abarcarán principalmente el modelado, en formato de datos enlazados, del léxico y reglas de la morfología de dicha lengua, ya que son sus principales rasgos diferenciadores frente al español. Respecto a la información léxica, esta se representará por medio de diccionarios bilingües español-aragonés, en formato Ontolex-Lemon. La representación del contenido morfológico de la lengua aragonesa se representará, en cambio, por medio del estándar Ontolex-Morph, actualmente en desarrollo, pero prácticamente ya finalizado.

Resultados preliminares

Por el momento se ha concluido un esfuerzo de revisión del estado de la cuestión sobre interpretabilidad en PLMs respecto a sus capacidades lingüísticas [3], como base teórica para facilitar la inyección de conocimiento lingüístico. Este trabajo, enviado a la revista *ACM Computing Surveys*, está actualmente bajo revisión. También podemos presentar los resultados de la aplicación de una serie de modelos entrenados en una serie de *baselines* preexistentes para transferencia croslingüe (MAD-X [5] y BAD-X [4], al igual que un modelo RoBERTa experimental entrenado con corpus disponibles en aragonés) en una serie de *benchmarks* de carácter lingüístico en lengua aragonesa (Tabla 1). Estos últimos han sido creados de manera preliminar, como *silver standards*, a partir de la traducción automática de datasets comúnmente usados para evaluar PLMs. También se han realizado experimentos en un modelo RoBERTa en español con los datasets aragoneses. Los resultados obtenidos, comparados a su vez con el último *baseline*, nos permiten concluir la necesidad de recurrir a soluciones de *cross-lingual transfer* dedicadas, dadas las bajas puntuaciones de los modelos base en los *benchmarks* en aragonés.

Tabla 1. Resultados preliminares de algunas *baselines* de *cross-lingual transfer* para el aragonés

<i>Modelo</i>	<i>Setting</i>	COPA	ESXN LI (github.com/artexem/esxnli)	PAWS
MAD-X (XLM-R+ Language Adapter [LA] <i>an</i> +Task Adapter [TA])	<i>TA</i> train + test <i>an</i>	0.49, 0.45, 0.48	0.332, 0.362	0.805
	<i>TA</i> train <i>es</i> + test <i>an</i>	0.51, 0.44, 0.52	0.488, 0.5	0.810

BAD-X (XLM-R+ Language Adapter [LA] <i>es-an</i> + Task Adapter [TA])	<i>TA</i> train + test <i>an</i>	0.51, 0.43, 0.56	0.328, 0.346	0.821
	<i>TA</i> train <i>es</i> + test <i>an</i>	0.48, 0.37, 0.52	0.334, 0.346	0.824
RoBERTa-BNE (modelo <i>es</i> , <i>baseline</i> ; fine-tuneado, train+test <i>an</i>)		0.49	0.334	0.577
RoBERTa-AN (modelo <i>an</i> ; fine- tuneado; train+test <i>an</i>)		0.48	0.334	0.577
RoBERTa-BNE (modelo <i>es</i> , <i>baseline</i> ; <u>fine-tuneado</u> , <u>train+test <i>es</i></u>)		0.6	0.334	0.577

REFERENCIAS

- [1] CIMIANO, P., CHIARCOS, C., MCCRAE, JP, AND GRACIA, J. *Linguistic linked data*. Springer International Publishing, 2020.
- [2] ESTUDIO DE FILOLOGÍA ARAGONESA. *Gramatica basica de l'aragonés*. Prensas de la Universidad de Zaragoza, 2021.
- [3] LÓPEZ-OTAL, M., GRACIA, J., BERNAD, J., BOBED, C., PITARCH-BALLESTEROS, L., AND ANGLÉS-HERRERO, E. Linguistic Interpretability of Transformer-based Language Models: a systematic review. 2025. *[Under review]*
- [4] PAROVIĆ, M. Improving Parameter-Efficient Cross-Lingual Transfer for Low-Resource Languages. Apollo – University of Cambridge Repository, 2024.
- [5] PFEIFFER, J., VULIĆ, I., GUREVYCH, I., AND RUDER, S. MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: ACL, 2020, pp. 7654-7673.
- [6] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., AND POLOSUKHIN, I. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA: Curran Associates, Inc, 2017, pp. 6000-6010.

