

# O galego no proxecto ESMAS-ES<sup>+</sup>: retos e oportunidades para a *anotación semántica*

Un laboratorio para validar estratexias de etiquetaxe semántico-ontolóxica multilingüe

Iván Arias Arias · Vitor Míguez Rego · Lola Mosquera Sánchez · María Teresa Sanmarco Bande  
Universidade de Santiago de Compostela — Instituto da Lingua Galega

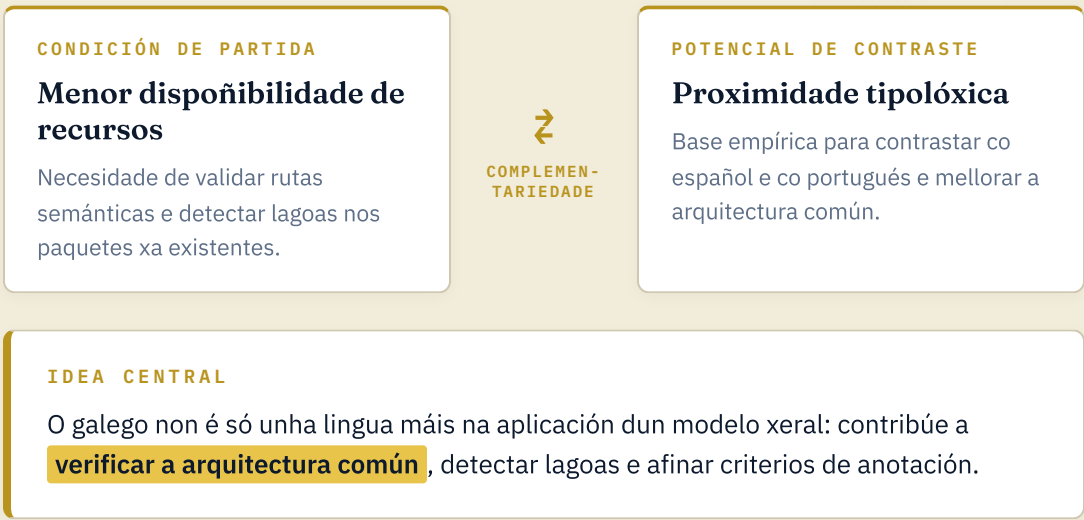


## 1 · PUNTO DE PARTIDA

### O galego nun etiquetador semántico multilingüe

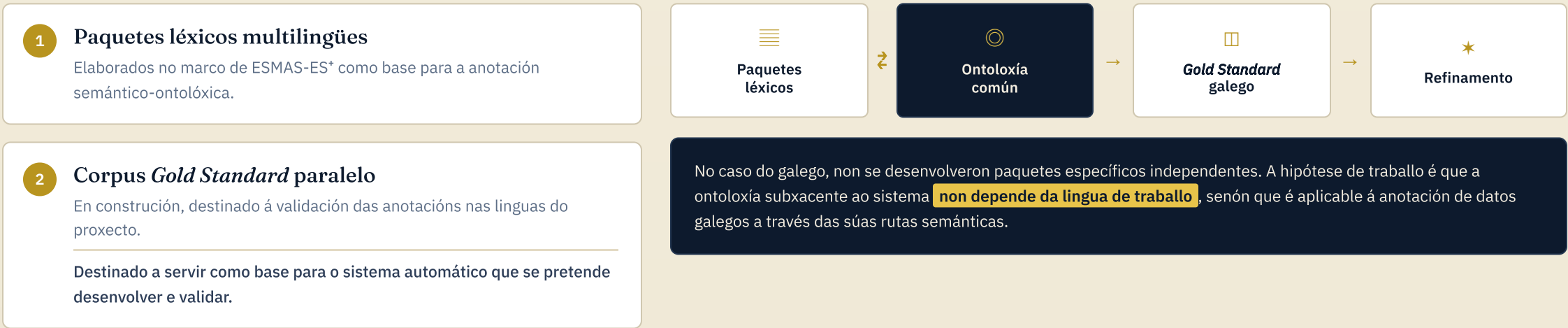
O proxecto ESMAS-ES<sup>+</sup> ten como obxectivo desenvolver e validar un prototipo de etiquetador semántico multilingüe, automático e sostible para alemán, español, francés e galego.

No marco do proxecto, o galego ocupa unha posición metodoloxicamente relevante: conta con menor dispoñibilidade de recursos ca outras linguas do proxecto, mais presenta tamén oportunidades derivadas da súa proximidade tipolóxica co español e co portugués.



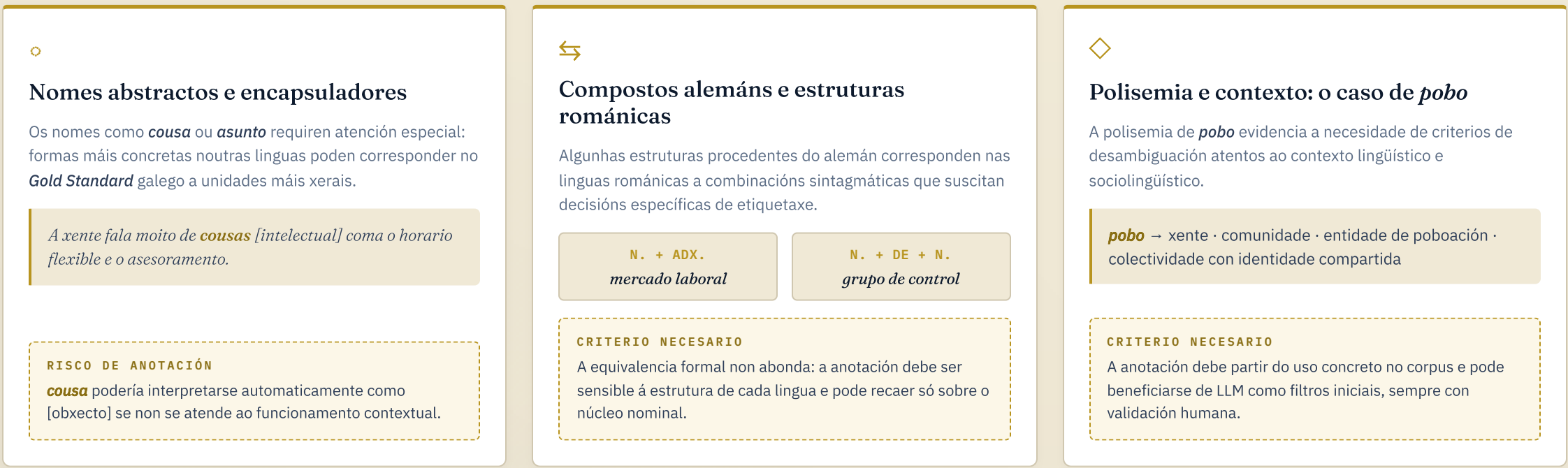
## 3 · MATERIAIS E MÉTODO

### Paquetes léxicos, *Gold Standard* e ontoloxía compartida



## 4 · FENÓMENOS OBSERVADOS

### Tres zonas sensibles da anotación multilingüe



## 5 · ACHEGAS

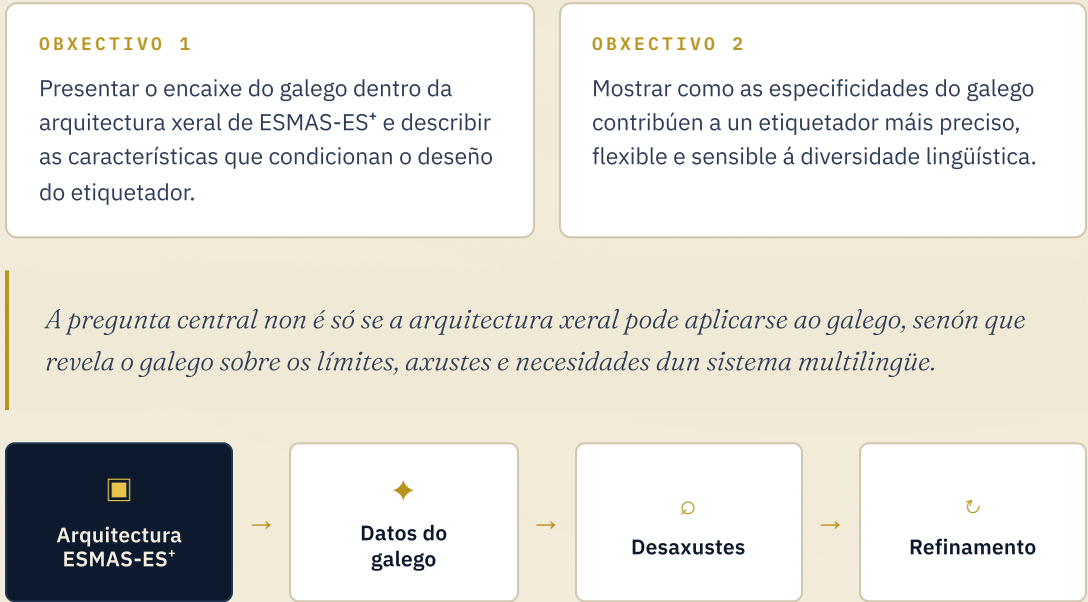
### O galego: banco de proba multilingüe

- Comprobar a robustez dunha ontoloxía concibida como independente da lingua.
- Detectar lagoas e desaxustes nos paquetes xa existentes.
- Revisar rutas semánticas a partir de datos empíricos.
- Formular criterios de anotación máis sensibles á diversidade lingüística.
- Contribuír a unha ferramenta máis robusta para linguas con diferente grao de dispoñibilidade de recursos.

O galego funciona como **proba metodolóxica** e como **laboratorio privilexiado** para validar estratexias de anotación semántica multilingüe.

## 2 · OBXECTIVOS

### Do encaixe arquitectónico á sensibilidade lingüística



## 6 · CONCLUSIÓN

### Da adaptación á mellora do modelo

**Máis ca unha aplicación periférica**  
A incorporación do galego a ESMAS-ES<sup>+</sup> mostra que a anotación semántica multilingüe non pode resolverse só mediante equivalencias formais ou transferencia automática entre linguas.

A análise dos datos galegos permite avanzar cara a un etiquetador máis preciso, sostible e atento á diversidade lingüística. A súa menor dispoñibilidade de recursos non debe entenderse só como unha limitación: permite someter a arquitectura xeral do sistema a unha validación máis esixente.

#### REFERENCIAS ESENCIAIS

- Dominguez Vázquez, M. J. (2025). *Revista de Investigación Lingüística*, 28, 175–192. DOI: 10.6018/rii.662171.
- Dominguez Vázquez, M. J., Simões, A., Bardanca Outeirño, D., Caiña Hurtado, M., & Iglesias Allones, J. L. (2024). *Natural Language Processing*, 1–25. DOI: 10.1017/nlp.2024.32.
- Míguez-Rego, V. (2026). *Open Linguistics*, 12(1), 20250078. DOI: 10.1515/opli-2025-0078.

