

A EXPLOTACIÓN DUN CORPUS PARA HUMANIDADES E CIENCIAS SOCIAIS: PARLAMINT

OBRADOIRO DE ADINA VLADU

Universidade de Santiago de Compostela

Os debates parlamentarios veñen de constituírse nunha fonte importante de datos para unha ampla gama de disciplinas do ámbito das Humanidades e as Ciencias Sociais. Non obstante, a falta de homoxeneidade dos corpus de debates parlamentarios existentes dificultaba a investigación comparativa entre os diferentes parlamentos europeos. Co fin de eliminar este obstáculo púxose en marcha [ParlaMint](#), un dos proxectos clave de [CLARIN](#), cuxo propósito foi a creación de corpus comparables, enriquecidos con anotación homoxénea e útiles para a investigación multidisciplinar.

A entrega máis recente deste proxecto, **ParlaMint 4.0**, contén 29 corpus nacionais e suma máis de mil millóns de palabras. Estes corpus están dispoñibles para descarga directa en tres versións: [en formato TEI](#); [con anotación lingüística](#); e [traducidos automaticamente ao inglés e enriquecidos con anotación semántica](#). Ademais, están integrados nas plataformas [noSketch](#), [kontext](#) e [teitok](#), que facilitan a súa consulta.

Este obradoiro ten como obxectivo animar o persoal investigador de dentro e fóra das Humanidades Dixitais, que poida ter pouca ou ningunha experiencia coa análise de corpus, a explorar o potencial dos corpus ParlaMint como un recurso lingüístico rico e coherente para a investigación do discurso parlamentario a nivel nacional e transnacional, así como para fomentar a cooperación interdisciplinaria.

BIONOTA

Adina Vladu (Instituto da Lingua Galega, USC) é doutora en Filoloxía pola Universidade de Bucarest (UB), na que tamén exerceu a docencia.

Os seus intereses de investigación céntranse na morfoloxía derivativa, tradutoloxía, lingüística de corpus e máis recentemente na lingüística computacional.

Actualmente é investigadora asociada no Proxecto Nós, onde traballa no desenvolvemento de recursos e ferramentas de voz para a lingua galega, centrándose nos aspectos prácticos e éticos da colección e procesamento de recursos e na creación de corpus voz-texto para síntese de voz e recoñecemento da fala.