WebLesp: Corpus de comunicación digital especializada en español

WebLesp compilation was funded by the **Università "G. D'Annunzio" di Chieti-Pescara (Fondi FARS)**. The project is the result of ongoing collaboration between the **Dipartimento di Lingue, Letterature e Culture Moderne dell'Università "G. D'Annunzio" di Chieti-Pescara** (**Prof. Sara Piccioni**) and the **Dipartimento di Scienze Giuridiche, del Linguaggio, dell'Interpretazione e della Traduzione (IUSLIT) dell'Università di Trieste** (**Prof. Gianluca Pontrandolfo**). Corpus compilation and annotation have been carried out with IT support from Dr. Eros Zanchetta (Dipartimento di Interpretazione e Traduzione, DIT, Università di Bologna).

CORPUS ACCESS: https://corpora.dipintra.it/unich/

WebLesp is an electronic corpus of texts representative of **specialized digital communication in Spanish**. The corpus is made up of four sections corresponding to four macro-disciplinary fields (**natural sciences, law, economics, medicine**). Each section includes a subcorpus of **popularised texts** aimed at non-expert audiences and a subcorpus of **specialist genres** of expert-to-expert communication. All the texts included in the corpus are **digital** in nature, in that they are published and read on the **web**.

The architecture of the corpus is as follows:

Discipline	Topis	Popularised subcorpus	Specialised subcorpus
Natural sciences	Climate change	[Ciencia DIV] - Popular blogs - Reports - Websites Size: **786,420 tokens**	[Ciencia ESP] - Research articles - Reports - PhD dissertations Size: **657,760 tokens**
Law	Divorce in Spain	[Derecho DIV] - Popular blogs - General press articles - Websites Size: **759,609 tokens**	[Derecho ESP] - Normative texts (articles of the Spanish Civil Code + legislation) - Interpretative texts (research articles, PhD dissertations) - Applicative texts (judgments) Size: **761,538 tokens**
Economics	Spanish vs. EU economy	[Economía DIV] - Popular blogs - General press articles Size: **752,793 tokens**	[Economía ESP] - Reports - Research articles Size: **852,419 tokens**
Medicine	Breast cancer	[Medicina DIV] - General press articles - Specialised press articles - Websites Size: **886,772 tokens**	[Medicina ESP] - Abstracts - Research articles - Specialised blogs - Clinical practice guidelines - PhD dissertations Size: **955,546 tokens**

By accessing WebLesp through the platform provided, users agree to query the corpus for research purposes only.