

The GenDJus corpus

SUBCORPUS	LANGUAGE	TOKENS	WORDS		TEXTS	
GenDJus_ACTHPR_EN	English	20,874	16,771	2		judgments
GenDJus_CC_IT	Italian	715,753	563,691	104	88	judgments
					16	orders
GenDJus_CJEU_EN	English	436,898	355,143	73	70	judgments
					3	orders
GenDJus_CJEU_ES	Spanish	479,358	397,802	73	70	judgments
					3	orders
GenDJus_CJEU_IT	Italian	451,246	366,727	73	70	judgments
					3	orders
GenDJus_CSC_IT	Italian	790,757	646,233	167	137	judgments
					30	orders
GenDJus_ECtHR_EN	English	3,557,550	2,981,120	454	216	judgments
					17	decisions (admissibility)
					221	separate opinions
GenDJus_IACtHR_EN	English	3,972,538	3,111,991	110	47	judgments
					3	advisory opinions
					60	separate opinions
GenDJus_IACtHR_ES	Spanish	4,859,905	3,927,537	128	53	judgments
					1	order
					3	advisory opinions
Tot.		15,284,879	12,367,015	1184	71	separate opinions

Pontrandolfo, G., Danisi, C., Gilleri, G., Miyamoto, E., & Sarni, C. (2024). The GenDJus corpus [Data set]. The GenDJus Project - PRIN 2022 PNRR. <https://doi.org/10.5281/zenodo.14761912>