

Corpus de Referência do Português Contemporâneo (CRPC)

Portugiesisch in Brasilien, Afrika und Asien. 311 Mio. Wörter, unterschiedliche Textsorten auch der gesprochenen Sprache.

Sprache	Portugiesisch
Varietät	(Regional-)Standard
Sprachliche Realisierung	schriftlich, mündlich
Umfang	ca. 311 Mio. Wörter
Medium	diverse Textsorten zum aktuellen Sprachgebrauch der geschriebenen Sprache (Literatur, Journalismus, Technik, Wissenschaft, Didaktik, Politik) und der gesprochenen Sprache im formellen und informellen Bereich
Geographischer Ursprung	Portugal, Brasil, Angola, Cabo Verde, Guiné-Bissau, Moçambique, São Tomé e Príncipe, Goa, Macau, Timor-Leste
Zeitliche Einordnung	ca. 1850-2006, die meisten Texte wurden nach 1970 verfasst
Form der Daten	digitalisierte Texte / Transkripte der Sprachaufnahmen; das Teilkorpus zur geschriebenen Sprache ist online durchsuchbar
Format	HTML
Annotation	tokenisiert, lemmatisiert, part-of-speech-annotiert
Mögliche Suchabfragen	Der QCPweb (Corpus Query Processor) ermöglicht die Suche nach Wörtern, Wortfolgen, Lemmata und Wortarten unter Verwendung regulärer Ausdrücke. Zudem können Frequenzlisten abgerufen werden.
Quelle /Herausgeber	Centro de Linguística da Universidade de Lisboa
Nutzungsvoraussetzungen	freier Zugang für die einfache Suche; für die Gliederung in Subkorpora ist eine kostenlose Registrierung erforderlich
Link	http://www.clul.ul.pt/en/resources/183-crpc#cqp
Literatur	Généreux, Michel, Iris Hendrickx, Amália Mendes (2012): "Introducing the Reference Corpus of Contemporary Portuguese On-Line". In: Proceedings of the Eighth International Conference on Language Resources and Evaluation - LREC 2012, Istanbul, 21-27 de Maio de 2012, 2237-2244. Bacelar do Nascimento, M. F. (2000): " <i>Corpus de Referência do Português Contemporâneo</i> ". In: BILGER, M. (ed.) <i>Corpus, Metodologie et Applications Linguistiques</i> , Paris, H. Champion et Presses Universitaires de Perpignan (2000), 25-30. weitere Literaturtips auf der Korpus-Homepage