

Corpus Histórico do Português Tycho Brahe

Historische Texte aus dem 14.-20. Jahrhundert, zum Teil POS- und syntaktisch annotiert.

Sprache	Portugiesisch
Sprachliche Realisierung	schriftlich
Umfang	ca. 2,7 Mio. Tokens
Medium	historische Texte aus 7 Jahrhunderten
Geographischer Ursprung	Portugal, Brasilien
Zeitliche Einordnung	14.-20. Jh.
Form der Daten	HTML, txt (zip)
Format	digitalisierte Texte, zum Download verfügbar, zum Teil annotiert und online durchsuchbar
Annotation	part-of-speech (für 1,5 Mio. Tokens), syntaktisches Parsing (für ca. 672.000 Tokens)
Mögliche Suchabfragen	Suche nach Wörtern, Wortarten und syntaktischen Kategorien
Quelle/Herausgeber	Helena Britto, Charlotte Galves, Universidade Estadual de Campinas
Nutzungsvoraussetzungen	kostenlose Registrierung erforderlich
Link	http://www.tycho.iel.unicamp.br/~tycho/corpus/index.html