

NUNC: Newsgroups UseNet Corpora

"Newsgroups UseNet Corpora" mit Texten aus 5 europäischen Sprachen, unterteilt nach Genre (z.B. Kochen, Business oder Fotografie). Das Korpus befindet sich noch in der Betaphase, aber kann schon durchsucht werden.

Sprache	Italienisch, Spanisch, Englisch (Britisch, Australisch), Französisch, Deutsch
Sprachstufe	Standard
Sprachliche Realisierung	schriftlich
Umfang	ca. 600 Mio. Tokens pro Sprache
Medium	Sprachdaten von <i>newsgroups</i> (Online-Foren)
Geographischer Ursprung	Italien, Spanien, Großbritannien, Australien, Frankreich, Deutschland
Form der Daten	Die Ergebnisse einer Suche werden als Keyword-in-context angezeigt
Format	Nur Online
Annotation	lemmatiziert, POS-Tags
Mögliche Suchabfragen	Die Korpora können mit der CQP-Query-Language durchsucht werden (nach Tokens, Lemma, POS-Tags)
Quelle/Herausgeber	Università di Torino
Nutzungsvoraussetzungen	Zugang frei
Link	http://www.corpora.unito.it/index_nunc.php
Zum Zitieren:	Barbera M., Marelli C. 2008. Tra scritto-parlato, Umgangssprache e comunicazione in rete: i corpora NUNC. In Accademia della Crusca (Eds.), <i>Studi di grammatica italiana</i> – vol. 27, 157-185.