

Korpusy językowe

#PrzydatneNarzędzia

✳ **Corpus of Contemporary American English COCA** ➡ <https://www.english-corpora.org/coca/>

Jedyny zrównoważony (zawierający szeroki zakres tekstów z różnych źródeł, który dobrze odzwierciedla rzeczywiste codzienne użycie języka) i jeden z największych korpusów amerykańskiego angielskiego.

✳ **British National Corpus (BNC)**

➡ <https://www.english-corpora.org/bnc/> lub <http://www.natcorp.ox.ac.uk/>
Największy korpus brytyjskiego angielskiego. BNC został pierwotnie stworzony przez Oxford University Press w latach 80-tych i wczesnych 90-tych i zawiera 100 milionów słów tekstu z szerokiego zakresu gatunków.

✳ **Narodowy Korpus Języka Polskiego (NKJP)** ➡ <http://www.nkjp.uni.lodz.pl/>

Podstawowy korpus referencyjny języka polskiego. Stworzony przez Instytut Podstaw Informatyki PAN (koordynator), Instytutu Języka Polskiego PAN, Wydawnictwo Naukowe PWN oraz Zakład Językoznawstwa Komputerowego i Korpusowego Uniwersytetu Łódzkiego.

✳ **Korpus Języka Polskiego PWN**

➡ <https://sjp.pwn.pl/korpus/>
Popularny korpus języka polskiego z przyjaznym interfejsem. Zawiera ponad 70 milionów słów. Oparty na tekstach książek, czasopism, druków ulotnych i akcydensowych, stron internetowych oraz tekstów mówionych, zrównoważony tematycznie i gatunkowo.

✳ **Korpusy Digitales Wörterbuch der deutschen Sprache** ➡

<http://www.dwds.de/ressourcen/korpora/>
Zbiór korpusów języka niemieckiego połączony ze słownikami, zawierający łącznie 2,5 mld wyrazów. Dodatkowo zawiera także wyjaśnienia wyrazów, dane o etymologii i artykuły tezaurusowe z kolokacjami.

✳ **Polsko-niemiecki / niemiecko-polski korpus równoległy PolGerCorp** ➡

<http://diaspol.uw.edu.pl/polniem/>
Korpus stworzony pod kierownictwem profesorów z Uniwersytetu Warszawskiego oraz Uniwersytetu Gutenberga w Moguncji. Zawiera ok. 10 mln słów z tekstów pochodzących z literatury pięknej (46 proc.) oraz z literatury faktu, z prasy i prawa (54 proc.).

✳ **Russian National Corpus / Национальный корпус русского языка** ➡ <https://ruscorpora.ru/>

Narodowy korpus języka rosyjskiego. Zawiera ponad 2 miliardy segmentów.

✳ **Translation Equivalents Database TREQ** ➡ <https://treq.korpus.cz/>

Na koniec prawdziwa perełka, choć nie jest to stricte korpus językowy, a zbiór słowników dwujęzycznych, tworzonych automatycznie na podstawie równoległego korpusu InterCorp. Słowniki są dwukierunkowe, ze wszystkimi językami reprezentowanymi w korpusie (w tym PL) po jednej stronie i czeskim, angielskim lub hiszpańskim po drugiej stronie.