Europarl: European Parliament Proceedings Parallel Corpus

Sammlung von Berichten des Europäischen Parlaments in 21 Sprachen, jeweils paarweise gegenübergestellt mit dem Englischen, ursprünglich zum Zweck der maschinellen Übersetzung.

Sprache	21 europäische Sprachen, darunter Französisch, Italienisch, Spanisch, Portugiesisch, Rumänisch, Englisch
Sprachliche Realisierung	schriftlich
Umfang	bis zu 60 Mio. Wörter pro Sprache
Medium	Berichte des Europäischen Parlaments. Zum Zweck der maschinellen Übersetzung wurden die Versionen der verschiedenen Sprachen mit der englischen Version satzweise aligniert.
Geographischer Ursprung	Europäische Union
Zeitliche Einordnung	1996-2011
Form der Daten	Vergleichskorpora zwischen dem Englischen und einer weiteren EU-Sprache, zum Download verfügbar und weiter formatierbar
Format	XML
Annotation	Annotation von Satzgrenzen, Metadaten in XML
Quelle /Herausgeber	Philipp Koehn, School of Informatics, University of Edinburgh
Nutzungsvoraussetzungen	Zugang frei
Link	http://www.statmt.org/europarl/
Literatur	Philipp Koehn (2005): Europarl: A Parallel Corpus for Statistical Machine Translation, MT Summit 2005. Download