

KIParla corpus

Das Korpus umfasst ca. 130 Stunden Sprachaufnahmen und besteht aus zwei Subkorpora (*moduli*). Im 'KIP' Subkorpus enthalten sind Interviews, Gespräche, Prüfungen, Sprechstunden und Unterrichtsstunden von Muttersprachlern mit mindestens einem Bachelor-Abschluss, die bis zu 65 Jahre alt sind. Das 'ParlaTO' Subkorpus besteht aus gesprochenen Interviews und umfasst eine weitere Altersgruppe sowie Sprecher mit unterschiedlichen Schul- und Universitätsabschlüssen. ParlaTO wird erst ab 2020 online verfügbar, KIP kann jetzt online mithilfe des NoSketchEngines durchsucht werden.

| | |
|---------------------------------|---|
| Sprache | Italienisch |
| Varietät | (Regional-)Standard |
| Sprachliche Realisierung | mündlich |
| Umfang | ca. 130 Stunden, momentan sind ca. 660.000 Tokens online |
| Medium | transkribierte Sprachaufnahmen, die in Bologna und Turin aufgenommen wurden |
| Geographischer Ursprung | Bologna, Turin |
| Zeitliche Einordnung | 2016–2019 |
| Form der Daten | Transkripte von Sprachaufnahmen, online durchsuchbar und in verschiedenen Formaten herunterladbar, die dazugehörigen Audiodateien, umfassende Metadaten |
| Format | HTML |
| Annotation | noch nicht Verfügbar: PoS-Annotation und Lemmatisierung sind noch geplant |
| Mögliche Suchabfragen | Das Korpus ist mit NoSketchEngine durchsuchbar, unterschiedliche Suchabfragen sind möglich: Kollokationen, Key Word in Context, usw. Eine ausführliche Anleitung finden Sie hier . Die Volltexte sind auch abrufbar, und die Suche kann durch unterschiedliche Metadaten (Dokumenttyp, Alter, usw.) eingeschränkt werden. |
| Quelle /Herausgeber | Zusammenarbeit zwischen der Università di Bologna & Torino |
| Nutzungsvoraussetzungen | Zugang frei |
| Link | http://kiparla.it/search/ |
| Zum Zitieren: | Mauri, Caterina, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti & Francesco Suriano. 2019. <i>KIParla corpus: a new resource for spoken Italian</i> . In: Bernardi, Raffaella, Roberto Navigli & Giovanni Semeraro (eds.), <i>Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it</i> . |