1.	Record Nr.	UNINA9910300610603321
	Autore	Dash Niladri Sekhar
	Titolo	History, Features, and Typology of Language Corpora / / by Niladri Sekhar Dash, S. Arulmozi
	Pubbl/distr/stampa	Singapore : , : Springer Singapore : , : Imprint : Springer, , 2018
	ISBN	981-10-7458-5
	Edizione	[1st ed. 2018.]
	Descrizione fisica	1 online resource (XXIX, 293 p. 75 illus., 17 illus. in color.)
	Disciplina	410.188
	Soggetti	Corpora (Linguistics)
		Natural language processing (Computer science)
		Language and languages—Study and teaching
		Corpus Linguistics
		Natural Language Processing (NLP)
		Language Teaching
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Nota di contenuto	1. Definition of Corpus -- 2. Features of Corpus -- 3. Genre of Text -- 4. Nature of Data -- 5. Type and Purpose of Text -- 6. Nature of Text Application -- 7. Parallel Translation Corpus -- 8. Web Text Corpus -- 9. Pre-Digital Corpora (Part-I) -- 10. Pre-Digital Language Corpora (Part-2) -- 11. Digital Text Corpora (Part-I) -- 12. Digital Text Corpora (Part-II) -- 13. Digital Speech Corpora -- 14. Utilization of Language Corpora -- 15. Limitations of Language Corpora.
	Sommario/riassunto	This book discusses key issues of corpus linguistics like the definition of the corpus, primary features of a corpus, and utilization and limitations of corpora. It presents a unique classification scheme of language corpora to show how they can be studied from the perspective of genre, nature, text type, purpose, and application. A reference to parallel translation corpus is mandatory in the discussion of corpus generation, which the authors thoroughly address here, with a focus on Indian language corpora and English. Web-text corpus, a new development in corpus linguistics, is also discussed with elaborate reference to Indian web text corpora. The book also presents a short history of corpus generation and provides scenarios before and after

the advent of computer-generated digital corpora. This book has several important features: it discusses many technical issues of the field in a lucid manner; contains extensive new diagrams and charts for easy comprehension; and presents discussions in simplified English to cater to the needs of non-native English readers. This is an important resource authored by academics who have many years of experience teaching and researching corpus linguistics. Its focus on Indian languages and on English corpora makes it applicable to students of graduate and postgraduate courses in applied linguistics, computational linguistics and language processing in South Asia and across countries where English is spoken as a first or second language.
.