1. Record Nr. UNINA9910453312303321

Titolo	Automatic treatment and analysis of learner corpus data / / edited by Ana Diaz-Negrillo, Nicolas Ballier, Paul Thompson
Pubbl/distr/stampa	Amsterdam ; ; Philadelphia : , : John Benjamins Publishing Company, , [2013] ©2013
ISBN	90-272-7095-3
Descrizione fisica	1 online resource (320 p.)
Collana	Studies in corpus linguistics ; ; volume 59
Altri autori (Persone)	BallierNicolas Diaz NegrilloAna ThompsonPaul
Disciplina	410.1/88
Soggetti	Corpora (Linguistics) Second language acquisition Electronic books.
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Automatic Treatment and Analysis of Learner Corpus Data; Editorial page; Title page; LCC data; Table of contents; Section 1. Introduction; Introduction; References; Learner corpora; 1. Introduction; 2. Corpora types, processing and annotation; 2.1 Types of learner corpora; 2.2 Annotation; 3. Uses and users of learner corpus data; 3.1 Overview; 3.2 Foreign language teaching; 3.3 Second language acquisition research; 3.4 Corpus and computational linguistics; 4. Looking forwards; References; Section 2. Compilation, annotation and exchangeability of learner corpus data Developing corpus interoperability for phonetic investigation of learner corpora1. Introduction; 2. Processing and annotating spoken data; 2.1 A tentative typology of spoken learner corpora; 2.2 Existing annotation layers in phonetic corpora, corpus comparability and interoperability; 2.3 Comparing with native corpora; 3. Some of the limits of automatisation; 3.1 The limits of phonetic annotation (forced alignments); 3.2 Some syllabification issues; 3.3 Prosodic annotation; 3.4 Speaker-dependent models?; 3.5 The uses of automation (caveats);

4. Challenges and recommendations

4.1 Tokenisation and categorisation of realisations and learner phonetic errors4.2 Modelling; 4.3 Comparing with native data (corpus interoperability); 5. From spoken learner corpora to spoken learner databases; 5.1 Textual datasets; 5.2 XML and XML tools; 5.3 Working with customized interface of Praat; 5.4 An alternative stance: WinPitch; 5.5 An incoming mixed model?; 6. The advent of spoken databases vs. speech databases; References; Learner corpora and second language acquisition; 1. Introduction; 2. Learner corpora in SLA research; 2.1 A bias in second language research

2.2 Corpora in language acquisition research2.3 An overview of learner corpora and learner corpus research; 2.4 L2 Spanish learner corpora: Introducing CEDEL2; 3. Design principles in learner corpora for SLA purposes: CEDEL2, a case study; 3.1 Principle 1. Content selection; 3.2 Principle 2. Representativeness; 3.3 Principle 3. Contrast; 3.4 Principle 4. Structural criteria; 3.5 Principle 5. Annotation; 3.6 Principle 6. Sample size; 3.7 Principle 7. Documentation; 3.8 Principle 8. Balance; 3.9 Principle 9. Topic; 3.10 Principle 10. Homogeneity; 3.11 Conclusion; 4. Current status of CEDEL2

4.1 Data collection4.2 Data distribution; 4.3 Source of data; 4.4 Preliminary segmentation and annotation; 4.5 CEDEL2: Next steps; 5. Learner corpora: The way forward; 6. Conclusion; References; Appendices; Competing target hypotheses in the Falko corpus; 1. Introduction: Why corpus architecture matters; 2. What kind of information should a learner corpus provide and what kind of data is needed?; 2.1 POS & lemmas; 2.2 Target hypotheses; Error exponent; Conflicting spans; 2.3 Stand-off models; 3. Case study: Falko; 3.1 Target hypotheses in the Falko essay corpus; 3.2 Automatic error tagging

3.3 Manual error tagging

Sommario/riassunto

This paper is an overview of several basic statistical tools in corpus-based SLA research. I first discuss a few issues relevant to the analysis of learner corpus data. Then, I illustrate a few widespread quantitative techniques and statistical visualizations and exemplify them on the basis of corpus data on the genitive alternation - the of-genitive vs. the s-genitive from German learners and native speakers of English. The statistical methods discussed include a test for differences between frequencies (the chi-squared test), tests for differences between means/medians (the