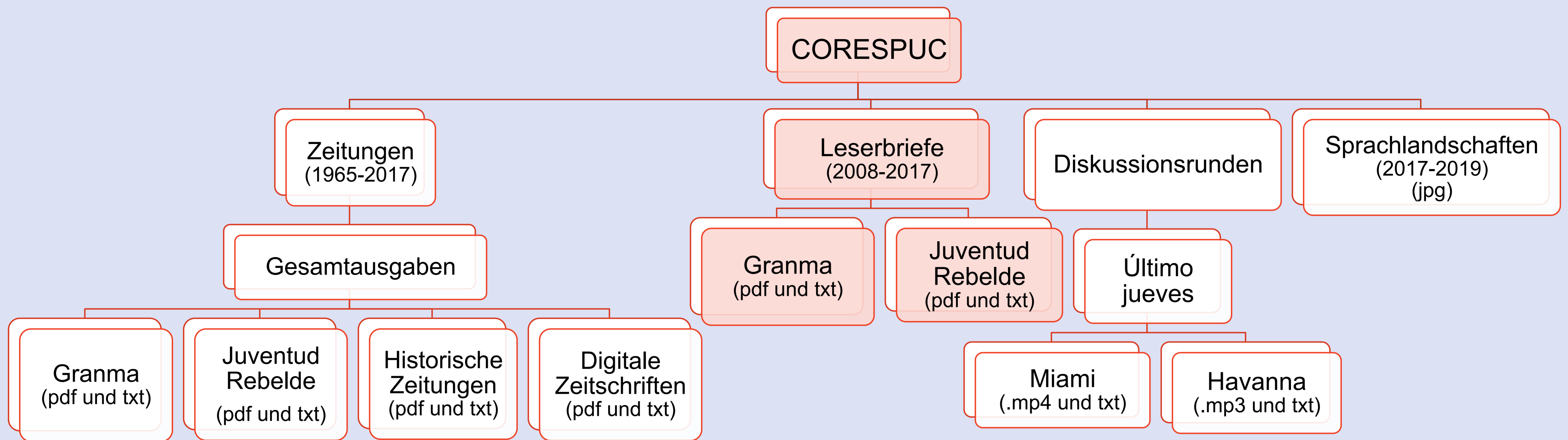


Struktur der Datenbank **Corpus del Español Público en Cuba**



Gesamtausgaben

Granma (2011-2017): Tokens: 15.869.117
 Juventud Rebelde (2009-2017): Tokens: 23.137.970

Leserbriefe

Granma (2008-2017): Tokens: 1.009.100
 Juventud Rebelde (2009-2017): Tokens: 1.297.274

Forschungsfragen

- **Welcher Zusammenhang besteht zwischen gesellschaftlichem Wandel und Sprachvariation bzw. Sprachwandel in Kuba?**
- **Welche sprachlichen Merkmale weist der öffentliche Sprachgebrauch in Kuba auf?**
 - Herausbildung und Entwicklung des politischen Wortschatzes
 - Sprachliche Musterbildung
 - Spanisch als Nullsubjektsprache
 - Informationsstruktur
 - Diskursmarker
 - Pragmatische Kategorien der Subjektivität und Objektivität
 - Textsortenvariation und -wandel
- **Wie werden kommunikative Interaktionen politischer und kommerzieller Art in öffentlichen Räumen diskursiv konstruiert?**
 - Sprachlandschaften und kommunikative Räume
- **Wie werden soziale Kategorien im öffentlichen Diskurs repräsentiert?**
 - Mediendiskurse, Sprachlandschaften

Nachhaltige Speicherung: Medienrepositorium der HU



Zugänglichkeit

- Die Daten werden zunächst nur den Mitgliedern der Forschergruppe zugänglich gemacht, die Freigabe der Daten ist aber geplant.
- Die Verantwortlichen für die Teildatenbanken und die InformatikerInnen haben Zugangsrechte.
- An der Universität Havanna kann es wegen technischer Mängel zu zeitweisen Nutzungseinschränkungen kommen.
- Die juristischen Voraussetzungen werden derzeit geprüft. Die Genehmigungen der Zeitungen liegen teilweise bereits vor.
- Eine Nutzungsvereinbarung für das Medienrepositorium zwischen der Humboldt-Universität und der Universität Havanna ist in Vorbereitung.

Analysewerkzeuge und-programme

- Bisher werden die Daten mit dem Hilfsprogramm Antconc quantifiziert. Mit der korpuslinguistischen Aufbereitung der Datenbank wurde im Bereich Informatik der Universität Havanna begonnen. Geplant sind:
 - automatische Analysen: Wortarten- und Lemmazuweisung
 - weitere manuelle Annotationen (Benennungen für soziale Kategorien, Pronomina, Diskursmarker, Fragesatz, finale Konstruktionen usw.)
 - die Annotationswerkzeuge sind
 - TreeTagger (Spanisch; Schmid 1994)
 - EXMARaLDA
 - Importe der Daten ins Suchsystem ANNIS (Zeldes et al. 2009).

Bibliographie:
 Schmid, Helmut (1994): *Probabilistic Part-of-Speech Tagging Using Decision Trees*. In: Proceedings of the International Conference on New Methods in Language Processing. Manchester, 44-49. [<http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf>].
 Zeldes, Amir, Ritz, Julia, Lüdeling, Anke and Chiaros, Christian (2009): "ANNIS: A Search Tool for Multi-Layer Annotated Corpora". In: *Proceedings of Corpus Linguistics 2009*. Liverpool, UK. [<http://edoc.hu-berlin.de/oa/conferences/reS4Xo05sncZc/PDF/29i8VTzYfT3M.pdf>]