



Abschlussbericht *Sanders' Sprachschatz Digital* im Rahmen Lehre@LMU von Sebastian Göttel

Antragsteller: Sebastian Göttel
Betreuerin: PD Dr. Christiane Wanzeck
Projektname: Sanders' Sprachschatz Digital

Geförderte Mittel: Anschaffung des zweibändigen Wörterbuchs von Daniel Sanders: *Deutscher Sprachschatz geordnet nach Begriffen zur leichten Auffindung und Auswahl des passenden Ausdrucks. Ein stilistisches Hilfsbuch für jeden Deutsch Schreibenden.* Hoffmann & Campe: Hamburg, 1873.
Buchbinderarbeiten: Heraustrennen des Textblocks aus dem Einband durch eine Buchbinderei, um optimale Ergebnisse beim Einscannen zu gewährleisten. Ziel des Projekts war es, das Einscannen des Wörterbuchs in höchster Qualität zu ermöglichen.

Untersuchungsgegenstand: Das onomasiologische Wörterbuch von Daniel Sanders ist ein einmaliges zeitgenössisches Dokument über den Wortschatz des Deutschen, das in seiner Qualität bis heute unübertroffen ist.

Im Sommersemester 2013 wurde meine Bachelorarbeit zu Sanders' *Sprachschatz* von dem Fachbereich der Germanistischen Linguistik an der LMU-München angenommen. Zur Erstellung meiner Arbeit wäre ein angemessen digital aufbereitetes Material eine große Hilfe gewesen, leider musste ich feststellen, dass die Online-Versionen des Wörterbuchs (etwa archive.org oder der Bodleian Library, University of Oxford) für linguistische Untersuchungen ungeeignet sind. Eine immer wiederkehrende Problematik eingescannter Texte, ist die mangelnde Berücksichtigung eines zu engen Seitenfalzes. Dadurch geht der Text an diesen Stellen verloren. Daher ist das Anliegen dieses Projekts, das selbsterworbene Wörterbuch von der Buchbinderei Schiedeck in München „öffnen“ zu lassen, indem der Buchblock aus dem Einband herausgetrennt wurde. Damit war es möglich, jede einzelne Seite vollständig mit dem Scanner zu erfassen.

Der Digitalisierungsprozess wurde durchgeführt mittels Adobe Acrobat XI Pro. Zur Vorbereitung auf den anschließenden OCR-Prozess wurden die insgesamt 1072 Seiten in Graustufen bei 600 dpi eingescannt. Bei der anschließenden Erfassung des Texts konnte das original Seitenlayout beibehalten werden. Der Erfassungsprozess wurde somit auf professionellem Niveau durchgeführt.

Die beiden in Frage kommenden OCR-Programme, Adobe Acrobat Pro und Abbyy FineReader wurde von mir einem Testdurchlauf unterzogen. Dabei hat sich gezeigt, dass Adobe in seiner Vielfältigkeit für mein Projekt, das geeignetere Programm ist. Für die Nutzer liegt meine Datenerhebung als Word-, JPG- Datei und als PDF-Dokument vor.

Ein Ausschnitt des erfassten Wörterbuchs füge ich in Form einer Word-Datei bei. Bei Interesse ist die Bereitstellung der vollständigen Version selbstverständlich möglich.

Nicht zuletzt möchte ich mich für die Unterstützung, auch im Namen von Frau Dr. Wanzeck, im Rahmen der Lehre@LMU bedanken.

Sebastian Göttel