

Schriften des Instituts für Dokumentologie und Editorik — Band 16

# **Digitale Edition in Österreich**

## **Digital Scholarly Edition in Austria**

---

herausgegeben von | edited by  
Roman Bleier, Helmut W. Klug

2023

BoD, Norderstedt

**Bibliografische Information der Deutschen Nationalbibliothek:**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

**Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 29. April 2023.**

2023

Herstellung und Verlag: Books on Demand GmbH, Norderstedt

ISBN: 978-3-743-102-842

Einbandgestaltung: Stefan Dumont; Coverbild: wurde von Roman Bleier und Helmut Klug für ein KONDE-Poster (DHa 2017) erstellt

Satz: Roman Bleier und Lua $\TeX$

# Zweisprachiger Spracherwerb: Longitudinalstudie anhand schriftlicher Texte der Hermagoras-Volksschule – Digitalisierung, Auszeichnung, Auswertung

Ursula Doleschal, Lisa Rieger

<b>Mitarbeiterinnen:</b>	Ursula Doleschal (Projektleitung), Lisa Rieger
<b>Institutionen:</b>	Universität Klagenfurt, Institut für Slawistik
<b>Fördergeber:</b>	Fakultät für Kulturwissenschaften der Alpen-Adria-Universität Klagenfurt
<b>Website:</b>	<a href="https://campus.aau.at/cris/project/0f4de0c9610ea75c01611937ce72037f">https://campus.aau.at/cris/project/0f4de0c9610ea75c01611937ce72037f</a>

Dieses Projekt hat zum Ziel, mit Methoden der Digital Humanities den Fortschritt in Sprach- und Textkompetenz von Schülerinnen und Schülern der zweisprachigen *Mohorjeva ljudska šola/Hermagoras Volksschule* in Klagenfurt mit dualem deutsch-slowenischem Immersionsunterricht auf Basis von ihnen verfasster, authentischer Texte zu analysieren. Dafür wurde von 2015 bis 2018 eine Klasse von der zweiten bis zur vierten Schulstufe begleitet. In diesem Zeitraum wurden von 11 Schülerinnen und Schülern 234 Texte, zusammengefasst in 24 Textsammlungen, digitalisiert, transkribiert, ausgezeichnet und mit einem Auswertungsprogramm<sup>1</sup> hinsichtlich quantitativer Parameter analysiert. Nach einer Gesamtauswertung erfolgte eine Auswertung nach Erstsprachen.

Die Transkription erfolgte manuell in *Transkribus*. Dabei wurde nur das, was von den Schülerinnen und Schülern selbst verfasst wurde, im vorliegenden Wortlaut transkribiert. Anhand der Kopien ergaben sich folgende Probleme:

1. An den Seiten abgeschnittene Wörter: Diese wurden nur so weit transkribiert, wie die Buchstaben noch eindeutig erkennbar waren, ansonsten wurde mit dem POS-Tag „gap“ auf die Unvollständigkeit hingewiesen.
2. Schwer leserliche und schlecht erkennbare Wörter: Diese wurden so gut wie möglich rekonstruiert, um das Wort im Weiteren lemmatisieren zu können.

---

<sup>1</sup> Das Auswertungsprogramm wurde nach gemeinsamer Absprache von der Computerlinguistin am Institut für Germanistik der AAU Klagenfurt, Elisabeth Eder, MA, speziell für den Zweck des Projekts erstellt.

3. Abgeschnittene oder unleserliche längere Textpassagen: Diese führten dazu, dass der Text aus der Analyse ausgeschieden wurde.

Die slowenischen Texte wurden mit dem morphosyntaktischen Tagger *Obeliks* (Romih, Krek und Kosem, n.d.), die deutschen mit *WebLicht* (CLARIN-D, 2017a) ausgezeichnet. Die Zuordnung der Kategorien und Eigenschaften erfolgt bei *Obeliks* anhand der Ergebnisse aus dem Projekt *Jeziškoslovno označevanje slovenščine* (JOS) (Inštitut Jožef Stefan, n.d.), in *WebLicht* basiert sie auf dem Stuttgart-Tübingen Tagset (Schiller et al. 1999). Beim Vergleich der Guidelines (Erjavec et al. 2010) wurde beschlossen, aufgrund folgender Phänomene von einem direkten Vergleich zwischen den beiden Sprachen abzusehen:

1. Kategorien, welche nur in einer der beiden Sprachen vorhanden sind,
2. unterschiedliche Zuordnungen zu den jeweiligen Kategorien und
3. unterschiedlicher Umfang bei der Auszeichnung morphologischer und syntaktischer Eigenschaften.

Trotz der statistisch niedrigen Fehlerquote war es notwendig, die Auszeichnungsergebnisse in folgenden Fällen manuell zu berichtigen:

1. abgeschnittene und unleserliche Wörter,
2. Rechtschreibfehler und
3. abgeteilte Wörter.

Das Auswertungsprogramm errechnete aus den XML-Dateien für jede Textsammlung, aufgeteilt nach Schülerin bzw. Schüler, folgende Parameter: durchschnittliche Satzanzahl und Satzlänge, durchschnittliche Wortanzahl sowie die durchschnittliche Anzahl von Verben und Konjunktionen. Nach einer Gesamtauswertung wurden die Ergebnisse nach den Erstsprachengruppen Deutsch, Slowenisch, Deutsch und Slowenisch sowie Slowenisch und BKS-Sprachen zusammengefasst. Die Entwicklung der slowenischen und deutschen Texte wurde zuerst getrennt betrachtet, bevor am Ende eines jeden Unterkapitels die prozentuellen Veränderungen einander gegenübergestellt wurden. Totale Zahlen wurden nicht berücksichtigt, um einen Vergleich erschwerende Faktoren, wie unterschiedliche Sprachstruktur und Auszeichnungsstandards, möglichst zu minimieren.

Durch die quantitative computergestützte Analyse konnten objektive Daten für die Beurteilung der Sprachkompetenz der Schülerinnen und Schüler gewonnen werden. Allerdings stellte sich auch heraus, dass reine Zahlen Spielraum für Interpretationen

lassen und somit in Zukunft nicht als alleiniges Mittel zur Beurteilung von Sprachkompetenz eingesetzt werden sollten, sondern vielmehr als objektive Ergänzung zu einer qualitativen Analyse. Aus technischer Sicht wäre es für die weitere Forschung wünschenswert, durch interdisziplinäre Zusammenarbeit, aufbauend auf den aufgezeigten Problemen einen einheitlichen Leitfaden für die Digitalisierung und Auszeichnung von Texten zu erarbeiten und somit einen Vergleich zwischen ähnlich gelagerten Projekten zu ermöglichen.

## Literatur

- CLARIN-D. 2017a. „WebLicht. Main Page.“ Zugriff: 5. April 2019. [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main\\_Page](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page).
- CLARIN-D. 2017b. „WebLicht. Tools in Detail.“ Zugriff: 5. April 2019. [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Tools\\_in\\_Detail](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Tools_in_Detail).
- Erjavec, Tomaž, et al. 2010. „Oblikoskladenjske specifikacije JOS.“ Zugriff: 5. April 2019. <http://nl.ijs.si/jos/msd/html-sl/josMSD-sl.html>.
- Institut Jožef Stefan. o.J. „Projekt JOS: jezikoslovno označevanje slovenskega jezika.“ Zugriff: 5. April 2019. <http://nl.ijs.si/jos/index-sl.html>.
- Romih, Miro, Simon Krek und Iztok Kosem. o.J. „Sporazumevanje v slovenskem jeziku. Označevalnik.“ Zugriff: 5. April 2019. <http://oznacevalnik.slovenscina.eu/Vsebine/Sl/SpletniServis/SpletniServis.aspx>.
- Schiller, Anne, et al. 1999. „Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset).“ Zugriff: 5. April 2019. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.