

AUF DEM WEG ZU EINER GROSSEN DATENBASIS FÜR DEUTSCH ALS L2 – DAS PROJEKT DAKODA

Interview mit **Katrin Wisniewski** und **Matthias Schwendemann**

● **Katrin Wisniewski**
| Universität Leipzig
Matthias Schwendemann
| Universität Leipzig

Ihr Projekt DAKODA hat zum Ziel, eine grosse Anzahl von Lernerkorpora des Deutschen in ein Repositorium zu integrieren und korpusübergreifend zu verknüpfen. Zu nennen wären hier das ZISA-Korpus (Clahsen et al., 1983), die FALKO-Korpusfamilie (Hirschmann et al., 2022), das MERLIN-Korpus (Wisniewski et al., 2013), das deutsch-ungarische Lernerkorpus DULKO (Beeh et al., 2021) oder auch die beiden Schweizer Korpora DiGS (Diehl et al., 2000) und SWIKO (Karges et al., 2022) – die Liste liesse sich fortsetzen. Was sind dabei die grössten Herausforderungen?

Katrin Wisniewski: Ein Mangel an Herausforderungen besteht nicht. Diese betreffen ganz unterschiedliche Ebenen im Projekt.

Zunächst einmal ist die Herausforderung, eine publizierbare Datenbasis zusammenzustellen, nicht zu unterschätzen. In DAKODA bitten wir ja Inhaber:innen

öffentlicher und nicht öffentlicher L2-Korpora darum, uns ihre Daten für unser geplantes Dashboard zur Verfügung zu stellen. Das Dashboard soll eine Auswahl an Suchfunktionen anbieten. Zusätzlich möchten wir andererseits die Daten, wie von Ihnen schon angesprochen, in einem Repositorium bereitstellen und, falls rechtlich möglich, auf verschiedenen Wegen zum Download anbieten.

Nun mangelt es in keiner Weise an der Kooperationsbereitschaft von Kolleg:innen – ganz im Gegenteil! Hier sind wir nahezu ausnahmslos auf sehr positive Reaktionen gestoßen und unseren Kolleg:innen äußerst dankbar für ihre Offenheit. Allerdings ist die Publikation von Lernerkorpora an zahlreiche datenschutz- und lizenzrechtliche Bedingungen geknüpft. Wurden diese Faktoren bei der Korpuserstellung nicht schon mitbedacht und kein entsprechendes Einverständnis der Lernenden eingeholt und keine entsprechende Datenanonymisierung durchgeführt, können die Korpora tatsächlich der Forschungsgemeinschaft nicht mehr zur Verfügung gestellt werden. Bei älteren

Korpora haben wir außerdem einen nicht unerheblichen ‚Schwund‘ festgestellt: Einige Lernerkorpora sind Datenhacks zum Opfer gefallen, regelweise Transkripte wurden bei zu räumenden Büros vernichtet. Viele Korpusdaten wurden nie digitalisiert; in einigen publizierten Korpora fehlen Daten, oder sie sind nicht vollständig anonymisiert usw.

Eine weitere Herausforderung liegt darin, dass sehr viele Lernerkorpora des Deutschen in vielen verschiedenen, uneinheitlichen und nicht kompatiblen Formaten vorliegen und in unterschiedlicher Weise aufbereitet sind. Wenn sich zum Beispiel die Definition für ein Token oder einen Satz in Korpora unterscheidet und man dann Analysen, die auf solchen Einheiten beruhen, über die Korpora hinweg durchführt, muss das zu Folgefehlern führen. Deshalb bringen wir alle Lernerkorpora in ein einheitliches Format und führen bestimmte Vorverarbeitungsschritte neu und für alle Datensätze in gleicher Form durch.

Um ein Dashboard zu erstellen, das die Suche über verschiedene Lernerkorpora hinweg erlaubt, müssen zudem die darin enthaltenen Metadaten aufeinander abgestimmt sein. Metadaten sind «Daten über Daten», also beispielsweise Informationen über das Alter oder die L1 von Lernenden. Auch hier herrscht in der Lernerkorpuslandschaft eine große Vielfalt. Einige Korpora haben sehr wenige Metadaten, andere sehr viele. Auch ist die Definition selbst von scheinbar eindeutigen Variablen wie „L1“ oft unterschiedlich. Wir arbeiten deshalb an einem großen gemeinsamen Metadaten-schema für alle involvierten Korpora.

Die bisher genannten Herausforderungen betreffen sämtlich Fragen aus dem Bereich der Open Science. Immer häufiger wird – sinnvollerweise – von Fördergebern in Projekten ein Forschungsdatenmanagement verlangt, das die sogenannten FAIR-Prinzipien berücksichtigt. Forschungsdaten sollten demnach gut auffindbar (*findable*), zugänglich (*accessible*), kompatibel (*interoperable*) und nachhaltig bzw. wiederverwendbar (*reusable*) sein. Hier hat die Lernerkorpuslandschaft des Deutschen noch etliche Herausforderungen zu meistern. Ein Problem sehen wir in der geberseitigen Forderung (aber auch dem forschungsseitigen Wunsch) nach nachhaltigen Daten, ohne dass

«Wir bringen alle Lernerkorpora in DAKODA in ein einheitliches Format und stimmen die Metadaten aufeinander ab. So ermöglichen wir korpusübergreifende Analysen mit großen Datenmengen.»

gleichzeitig eine dauerhafte Förderung bzw. eine stabile übergeordnete, institutionell angebundene Infrastruktur für Lernerkorpora besteht.

Die verschiedenen Korpora wurden auch generiert, um Daten für unterschiedliche Forschungsfragen zu liefern, und können demnach stark variieren. Wie geht das Projekt DAKODA damit um?

Katrin Wisniewski: Auf inhaltlicher Ebene geht es im Kern des DAKODA-Projekts darum, auszuprobieren, wie gut eine automatische Analyse von Verbstellungsphänomenen in Lernaltersprache funktioniert, genauer gesagt eine Analyse der sogenannten Erwerbsstufen, wie sie in der *Processability Theory* beschrieben wurden (Pienemann, 1998, 2005). Automatische Analysen von Lernaltersprache sind aber sehr anspruchsvoll, vor allem bei beginnenden Lernenden. Deshalb trägt DAKODA auch dezidiert explorativen Charakter: Eine „perfekte“ Annotation ist computerbasiert nicht möglich. Dennoch interessiert es uns herauszufinden, welche Dinge mit dem Computer besser oder aber weniger gut funktionieren, auch um zukünftigen Entwicklungen hier vielleicht ein wenig den Weg zu bahnen.

Eine Voraussetzung für das Funktionieren (automatischer) Annotationen ist, dass man die interessierenden Phänomene (bei uns: Erwerbsstufen) sehr klar und detailliert linguistisch definiert. Mit Vagheit konfrontiert man automatische Analysesysteme nämlich besser nicht. Dabei ist uns aufgefallen, dass trotz der enormen Verbreitung der syntaktischen Erwerbsstufen teils noch nie genau beschrieben wurde, wie sich die Stufen eigentlich im Detail definieren. Von daher geht der automatisierten Annotation derzeit eine umfassende linguistische Arbeit voraus.



© Anke Steinberg/Universität Leipzig

Matthias Schwendemann ist wissenschaftlicher Mitarbeiter in den Bereichen Linguistik und Angewandte Linguistik am Herder-Institut der Universität Leipzig. Seine Arbeitsschwerpunkte in Forschung und Lehre liegen in den Bereichen Lexikologie, Wissenschaftssprache und Erwerb und Entwicklung des Deutschen als Fremd- und Zweitsprache sowie der Analyse von Lernaltersprache. Derzeit ist er zudem Mitarbeiter im BMBF-geförderten Drittmittelprojekt DAKODA.



© Sven Reichhold/Universität Leipzig

Katrin Wisniewski hat die Gerhard-Helbig-Professur für Deutsch als Fremd- und Zweitsprache am Herder-Institut der Universität Leipzig inne und forscht zu Grammatikerwerb und Sprachdiagnostik. Sie leitet das DAKODA-Projekt und war auch für das MERLIN-, das DISKO- und das MIKO-Korpus verantwortlich.

**An wen werden sich dieses
Repositorium und das von Ihnen
beschriebene Dashboard haupt-
sächlich richten? Und für welche
Verwendungszwecke kann man
diese jeweils konsultieren?**

Matthias Schwendemann: Kurz gesagt: Aus unserer Sicht wird bei DAKODA für alle etwas Interessantes dabei sein. Wir haben tatsächlich ein sehr breites und diverses Zielpublikum im Blick, also beispielsweise Spracherwerbsforscher:innen oder Wissenschaftler:innen, die sich grundsätzlich mit Lerner Sprache und/oder mit sprachlicher Variation auseinandersetzen, aber auch und nicht zuletzt Sprachdidaktiker:innen. Zusätzlich kann DAKODA aufgrund der für das Deutsche als L2 relativ großen Datenmenge auch für Computerlinguist:innen relevante Einsichten bereithalten. Eine weitere interessante Forschungsperspektive ergibt sich zudem dadurch, dass

der verschiedenen DAKODA-Werkzeuge deutlich: dem Repositorium einerseits und dem interaktiven Dashboard andererseits. Einige der Funktionalitäten des Dashboards (z.B. ein Filtern der Korpusdaten nach sprachlichen Strukturen auf bestimmten Erwerbsstufen der Processability Theory) werden natürlich sehr eng an unsere konkreten Forschungsfragen angelehnt sein und daher vor allem für Spracherwerbsforscher:innen relevant sein, die sich für Erwerbsstufen im Bereich der Wortstellung interessieren. Gleichzeitig wird das Dashboard aber die Möglichkeit bieten, ganz unterschiedliche Korpora anhand ihrer Metadaten zu durchsuchen und so besser kennenzulernen und miteinander in Beziehung zu setzen. Die Aufbereitung der zahlreichen Metadaten in den verschiedenen Korpora mit dem Ziel einer größeren Einheitlichkeit ist gleichzeitig auch ein sehr umfangreiches Teilprojekt von DAKODA. Nutzer:innen des Dashboards wird es dann möglich sein, in den DAKODA-Daten etwa nach Lernenden mit ganz spezifischen Eigenschaften (Alter, Erstsprache(n), Sprachniveau, Sprachbiografie etc.) zu suchen und etwa die Texte dieser Lernenden vergleichend zu analysieren.

Was das Repositorium angeht, so erfüllt dieses aus unserer Perspektive mehrere sehr wichtige Funktionen. Auf einer grundsätzlichen Ebene sind Daten innerhalb eines Repositoriums öffentlich und damit prinzipiell zugänglich und in vielen Fällen auch downloadbar. Das DAKODA-Repositorium wird so interessierten Wissenschaftlern eine ganze Reihe an Möglichkeiten zur weiteren Arbeit mit den Korpusdaten geben. Gleichzeitig sorgt die Ablage von Daten in einem Repositorium im besten Fall dafür, dass Daten an mehreren Orten gesichert sind und so die Gefahr von Datenverlusten reduziert wird.

Letztlich erhoffen wir uns, dass durch DAKODA von vielen Wissenschaftler:innen aus unterschiedlichen Forschungsgebieten bestimmte Forschungsfragen nochmal oder auch ganz neu an eine große Datenbasis gestellt werden können und so die empirische Wissensbasis der Spracherwerbsforschung erweitert werden kann. Dies ist besonders auch deshalb der Fall, weil im Rahmen von DAKODA einige Korpora zum allerersten Mal überhaupt veröffentlicht und zur Verfügung

«Wir erhoffen uns, dass durch DAKODA bestimmte Forschungsfragen nochmal oder auch ganz neu an eine große Datenbasis gestellt werden können und so die empirische Wissensbasis der Zweitspracherwerbsforschung erweitert werden kann.»

in DAKODA nicht nur Daten von L2-Lernenden, sondern auch Daten von Menschen mit Deutsch als Erstsprache zur Verfügung gestellt werden. Und zuletzt ist es natürlich so, dass die vielen Korpora aus DAKODA zu sehr unterschiedlichen Zwecken und vor dem Hintergrund sehr vielfältiger Erkenntnisinteressen erhoben wurden. Wer sich also bereits in der Vergangenheit aus bestimmten Gründen für spezifische Korpora interessiert hat, die jetzt auch in DAKODA zu finden sind, wird nun neue Zugangswege zu diesen Daten erhalten und vor allem auch die Möglichkeit, diese Daten mit anderen Lernerkorpusdaten in Verbindung zu setzen und zu vergleichen.

Diese sehr vielfältige Ausrichtung zeigt sich in der auch von Katrin Wisniewski gerade angesprochenen Konzeption

gestellt werden (z.B. das MULTILIT-Korpus (Schroeder/Schellhardt, 2015) oder auch das umfangreiche Chinesische Deutschlernerkorpus (Wu/Li, 2022)).

Werden auch nach Abschluss des Projekts neue Korpora in die Korpussammlung integriert?

Katrin Wisniewski: Wie oben schon gesagt, müssen Korpora je erst in ein einheitliches Format überführt werden, bevor sie in die DAKODA-Datenbasis aufgenommen werden können. Auch eine Übernahme in das Metadatenschema ist nötig. Dafür sind jedoch personelle Ressourcen nötig, die nach Projektende nicht mehr zur Verfügung stehen werden. So bedauerlich das ist: Dieser Situation ließe sich nur abhelfen, wenn es eine institutionell angebundene, übergreifende Lernerkorpus-Infrastruktur und eine langfristige strukturelle Förderung gäbe.

Welche Ratschläge würden Sie Forschenden geben, die vorhaben, ein Korpus zu erstellen?

Matthias Schwendemann: Hier hilft es vielleicht, zunächst rückwärts und vor allem FAIR (*Findability, Accessibility, Interoperability, Reusability*) zu denken. Die nützlichsten Korpusdaten für Wissenschaftler:innen und Sprachdidaktiker:innen sind Daten, die öffentlich und mit möglichst geringen (oder gar keinen) Zugangsvoraussetzungen auffindbar, verfügbar und zugänglich sind. Die erste Frage, die man sich bei der Konzeption eines Lernerkorpus stellen könnte, ist also: Wie müsste meine Einverständniserklärung und meine Datenschutzerklärung aussehen, damit ich die Daten später möglichst ohne Einschränkungen veröffentlichen kann? Hier sind einige grundlegende rechtliche Fragen zu beachten, die in den letzten Jahren enorm an Bedeutung gewonnen haben. Sehr wichtig ist hier vor allem die Tatsache, dass sowohl eine Datenschutzerklärung erhoben werden muss, die vor allem klärt, wie die persönlichen Daten der Lernenden bzw. Teilnehmenden geschützt werden, als auch eine Einverständniserklärung, in der es darum geht, wie die erhobenen Daten denn tatsächlich verwendet und vor allem veröffentlicht werden. Hier sind in jedem einzelnen Fall unter Umständen eine Reihe von spezifischen Fragen und

«Bei der Erhebung von Korpusdaten sollte die potenzielle Nachnutzbarkeit dieser Daten immer schon mitgedacht werden.»

Kontexten zu bedenken und wir würden allen raten, sich hier schon früh und proaktiv mit den Rechtsabteilungen bzw. den Justiziaten der eigenen Institutionen zu besprechen und gemeinsame Lösungen zu erarbeiten. Einverständniserklärungen sollten im besten Fall so abgefasst sein, dass sie eine Veröffentlichung unter freien Creative Commons-Lizenzen (CC) oder die Veröffentlichung mit möglichst geringen Einschränkungen und Zugriffshürden erlauben. Sofern die dann erhobenen Einverständnis- und Datenschutzerklärungen dies zulassen, sollte ab diesem Zeitpunkt ein Open-Access- und ein Open-Science-Ansatz verfolgt werden. Daten könnten dann beispielsweise auf Repositorien, wie zum Beispiel dem [IRIS-Repositorium](#) oder auch dem DAKODA-Repositorium, gespeichert, veröffentlicht und verfügbar gemacht werden, was die Nachnutzung durch andere Wissenschaftler:innen begünstigen würde.

Gleichzeitig, und jetzt springe ich nochmal ganz an den Anfang, sollten sich Wissenschaftler:innen bei der Erstellung von Korpora natürlich auch fragen, wie die erhobenen Daten die konkrete Forschungsfrage, für welche das jeweilige Korpus erhoben werden soll, beantworten können. Aber darüber hinaus sollte die Nachnutzbarkeit der erhobenen Daten direkt mitgedacht werden. Bereits erhobene Daten können dann nachhaltig von anderen Wissenschaftler:innen und in Lehr- und Lernkontexten genutzt werden, wenn zum Beispiel umfangreiche Metadaten erhoben wurden, die eine Vielzahl von potenziellen Forschungsfragen beantworten können. Metadaten sind gewissermaßen Daten über die im Korpus verfügbaren Sprachdaten. Dies können Informationen über die Erhebung des Korpus, über die Hintergründe und Eigenschaften der Lernenden oder auch über die Art und Weise der verwendeten Elizitierungsverfahren sein. Auch wenn bei der Erhebung von Metadaten

natürlich Fragen der Datensparsamkeit und auch potenziell ethische Fragen zunächst handlungsleitend sein sollten, gibt es im Bereich der Lernerkorpusforschung immer stärkere Bemühungen, sich auf eine Reihe von grundlegenden Metadaten zu einigen, die in jedem Fall erhoben werden sollten (vgl. das [Core Metadata Schema for Learner Corpora](#) von Paquot et al., 2023). Vorlagen wie das Schema von Paquot et al. (2023) stellen dabei sehr hilfreiche Vorlagen bei der Erstellung eigener Korpora dar. Neben den Metadaten wäre ein weiterer vielversprechender Punkt, die Vergleichbarkeit mit anderen Lernerkorpora des Deutschen als L2 direkt mitzudenken. Dies könnten Wissenschaftler:innen etwa dadurch erreichen, dass ähnliche Aufgabenformate wie in bereits bestehenden Korpora zugrunde gelegt werden, oder auch dadurch, dass ähnliche Gruppen von Lernenden untersucht werden. Nicht zuletzt sind erstsprachliche Vergleichsgruppen, die beispielsweise dieselben Aufgabenstellungen bearbeiten, eine sehr hilfreiche Ergänzung, um Lernerkorpusdaten weiter anzureichern. Ob es für den jeweiligen Forschungskontext sinnvoll ist, L1-Gruppen einzubeziehen, sollte bei der Konzeption neuer Korpora direkt mitbedacht werden. In der Regel werden solche Vergleiche die vorhandenen Daten und die daraus ableitbaren Schlussfolgerungen allerdings deutlich bereichern.

Und nicht zuletzt möchten wir (natürlich mit einem kleinen Augenzwinkern) auch mal dazu motivieren, ab und zu

groß zu denken und sich nicht von der vielen, vielen Arbeit, die das Erheben und Erstellen von Lernerkorpora bedeutet, einschüchtern oder gar abschrecken zu lassen. Tatsächlich rechtfertigen die fertigen Lernerkorpora wirklich alle Mühen auf dem Weg dorthin und werden uns als Fachcommunity in den nächsten Jahren dabei helfen, gemeinsam viele Fragen des Fremd- und Zweitspracherwerbs nochmal ganz neu zu stellen und hoffentlich auch beantworten zu können.

Was soll ihrer Meinung nach zum Thema ergänzt werden?

Katrin Wisniewski & Matthias Schwendemann: Um die oben genannten (und weitere) Herausforderungen der Lernerkorpusarbeit anzugehen, sind eine gute Vernetzung und weiterhin kooperative und interdisziplinäre Projekte nötig. Vor allem in der Zusammenarbeit mit der Computerlinguistik sehen wir großes Potenzial.

Die Arbeit mit Lernerkorpora ist für Praktiker:innen oft mit Hindernissen verbunden. An viele Korpora kommen etwa Lehrende gar nicht heran, und meist ist eine einigermaßen aufwändige Einarbeitung in die jeweilige Korpus-Suchsprache erforderlich. Wir hoffen, dass in Zukunft mehr niedrigschwellige Angebote zur Lernerkorpusnutzung entwickelt werden (vielleicht vergleichbar dem [ZuMult-Projekt](#), Fandrych et al. 2023). Vielleicht könnten KI-Anwendungen hier spannende Entwicklungen ermöglichen.

«Ein fertiges Lernerkorpus rechtfertigt alle Mühen auf dem Weg dorthin.»

DAKODA – Datenkompetenzen für DaF/DaZ

DAKODA ist ein interdisziplinäres Projekt mit dem übergeordneten Ziel, die Datenkompetenzen des wissenschaftlichen DaF/DaZ-Nachwuchses im Bereich Lernerkorpusforschung voranzutreiben. Im Projekt werden sprachtechnologische Ressourcen für erwerbsbezogene Fragestellungen entwickelt und auf Basis einer breiten Datengrundlage erprobt. DAKODA soll somit Möglichkeiten und Grenzen der Anwendung computerlinguistischer Verfahren für Lerner-sprachenanalysen explorativ ausloten. Dazu kooperiert das Projektteam am Herder-Institut der Universität Leipzig (Leitung: Prof. Dr. Katrin Wisniewski) mit dem Language Technology Lab der FernUniversität in Hagen (Leitung: Prof. Dr.-Ing. Torsten Zesch).

Das Projekt wird von 10.2022 – 9.2025 in der Förderline «Datenkompetenzen des wissenschaftlichen Nachwuchses» vom BMBF gefördert (https://www.bildung-forschung.digital/digitalezukunft/de/wissen/Datenkompetenzen/datenkompetenzen_wissenschaftlichen_nachwuchs/datenkompetenzen_wiss_nachwuchs.html).

www.dakoda.org



Literatur

Beeh, Christoph / Drewnowska-Vargáné, Ewa / Kappel, Péter / Modrián-Horváth, Bernadett / Nolda, Andreas / Rauzs, Orsolya / Scheibl, György (2021): *Dulko-Handbuch*. Szeged, Hungary: Institut für Germanistik der Universität Szeged. <https://doi.org/10.14232/dulko-handbuch-v1.0>.

Clahsen, Harald / Meisel, Jürgen / Pienemann, Manfred (1983): *Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter*. Tübingen: Narr. <http://doi.org/10.25592/uhhfdm.1463>.

Diehl, Erika / Christen, Helen / Leuenberger, Sandra / Pelvat, Isabelle / Studer, Thérèse (2000): *Grammatikunterricht: Alles für der Katz? Untersuchungen zum Zweitspracherwerb Deutsch*. Tübingen: Niemeyer Reihe germanistische Linguistik, 220. <https://www.unige.ch/lettres/alman/de/recherche/abgeschlossene-projekte/digs/digs-korpus> [12.03.2024].

Fandrych, Christian / Schmidt, Thomas / Wallner, Franziska / Wörner, Kai (2023): Zugänge zu mündlichen Korpora für DaF und DaZ: Das ZuMult-Projekt. In: *Korpora Deutsch als Fremdsprache*, 3(1).

Hirschmann, Hagen / Lüdeling, Anke / Shadrova, Anna / Bobeck, Dominique / Klotz, Martin / Akbari, Roodabeh / Schneider, Sarah / Wan, Shujun (2022): FALKO. Eine Familie vielseitig annotierter Lernerkorpora des Deutschen als Fremdsprache. In: *Korpora Deutsch als Fremdsprache* 2(2), 139–148, <https://doi.org/10.48694/kordaf.3552>.

Karges, Katharina / Studer, Thomas / Hicks, Nina (2022): Lerner-sprache, Aufgabe und Modalität: Beobachtungen zu Texten aus dem Schweizer Lernerkorpus SWIKO. In: *Zeitschrift für germanistische Linguistik* 50: 1, 104–130. <https://doi.org/10.1515/zgl-2022-2050>.

Pienemann, Manfred (Hg.) (2005): *Cross-Linguistic Aspects of Processability Theory*. Amsterdam: John Benjamins.

Pienemann, Manfred (1998): *Language processing and second language development*. Amsterdam: John Benjamins.

Schroeder, Christoph / Schellhardt, Christin (2015): Nominalphrasen in deutschen und türkischen Texten mehrsprachiger SchülerInnen. In: Ziegler, Arne / Köpcke, Klaus-Michael (Hrsg.): *Deutsche Grammatik in Kontakt*. Berlin: De Gruyter, 241–262. <https://doi.org/10.1515/9783110367171-011>.

Wisniewski, Katrin / Schöne, Karin / Nicolas, Lionel / Vettori, Chiara / Boyd, Adriane / Meurers, Detmar / Abel, Andrea / Hana, Jirka (2013): MERLIN. An Online Trilingual Learner Corpus Empirically Grounding the European Reference Levels in Authentic Learner Data. In: *ICT for Language Learning 2013. Conference Proceedings*. https://conference.pixel-online.net/conferences/ICT4LL2013/common/download/Paper_pdf/322-CEF03-FP-Wisniewski-ICT2013.pdf [12.03.2024].

Wu, Zekun / Li, Yuan (2022): Zur syntaktischen Komplexität des Schriftdeutschen chinesischer Deutschlerner/-innen – Eine korpusbasierte Profilanalyse. In: *Deutsch als Fremdsprache* 4. <https://doi.org/10.37307/j.2198-2430.2022.04.04>.