

KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

METHODENMESSE

Mittwoch, 16. März 2022, 15:45 Uhr bis 17:45 Uhr

Chinesisches Deutschler-Korpus (CDLK)

Yuan LI und Zekun WU

Die deutsche Sprache erlebt in China seit der Jahrtausendwende einen Boom. Der neuesten Erhebung 2020 zufolge gibt es in allen Bereichen, also Schule, Hochschule und Erwachsenenbildung, 20% mehr Deutschler als noch vor 5 Jahren. Die bestehenden Forschungsprojekte zur Lerner Sprache chinesischer Deutschler verfolgten meist einen qualitativen und induktiven Ansatz und stützten sich selten auf umfangreiche Korpora. Weltweit gesehen enthalten nur einige Lerner Korpora (Falko, AleSKo, Kobalt und BFSU-Korpus) schriftliche Texte der chinesischen Deutschler. Die ersten drei wurden in Deutschland erstellt, enthält nur wenige Texte und nur von fortgeschrittenen Studierenden (39, 43 bzw. 20 Texte auf dem Niveau B2). Das BFSU-Korpus aus China besteht aus Prüfungsaufsätzen chinesischer DeutschstudentInnen, ist aber nicht öffentlich zugänglich. Es ist zu erkennen, dass es bei den aktuellen Korpora nur begrenzte Textsorten und nicht umfassende Deutschniveaus sowie Lernstufen gibt.

Angesichts dieses Desiderats wird ein umfangreiches „Chinesische Deutschler-Korpus“ (CDLK) unter der Leitung von Prof. Dr. Yuan Li an der Zhejiang Universität aufgebaut. Das ist ein multidimensional annotiertes Lernerkorpus, das alle Niveaus bzw. Lernstufen abdeckt und die Texte sowohl von SchülerInnen, als auch von StudentInnen enthält. Darüber hinaus werden die Texte von denselben Lernern innerhalb von drei Jahren regelmäßig gesammelt.

Im CDLK werden Texte mit unterschiedlichen Themen und aus verschiedenen Genres (argumentative, bildbeschreibende, deskriptive und erzählende Texte) im Unterricht von Deutschlernern aus unterschiedlichen Jahrgängen von Schulen und Universitäten in verschiedenen Regionen handschriftlich innerhalb von 30 Minuten ohne Hilfsmittel produziert. Ein wichtiger Teil des CDLK ist die Mehrebenenannotation, inklusive Fehlerannotation und Abhängigkeits-Annotation. Die Fehlerkategorien wurden auf Basis Hirschmann und Noldas differenziert und bereichert, um Äußerungen chinesischer Deutschler kultursensibel und -spezifisch anzupassen. Die Abweichungen von den Originaläußerungen werden nicht nur auf der Normalisierungsebene (Orthografie, Morphosyntax, Morphologie, Lexik, Syntax, Semantik), sondern auch auf der höheren Ebene (Logik, Stil, Transfer) annotiert. Die Annotation wurde mit EXMARaLDA vorgenommen. Damit wurden die Zielhypothesen automatisch mit POS-, lemma annotiert. Dann wurden alle Abweichungen (ZHDiff) zwischen den Zielhypothesen und der Originaläußerungen automatisch annotiert, und manuell auf die jeweiligen Fehlerkategorien positioniert. Jeder Fehler wurde von drei Forschern gesondert identifiziert und kategorisiert. Bei Abhängigkeits-Annotation wird Abhängigkeitsrelation zuerst automatisch mit SpaCy annotiert, dann manuell überprüft.

Auf der Grundlage des CDLK wurden bereits einige Studien durchgeführt. In Anlehnung an der Theorie der dynamischen Systeme und mit dem CAF-Rahmen wurde die Lernaltersprache querschnittlich oder längsschnittlich aus einer bestimmten Perspektive vergleichend durchgeführt: Auf der lexikalischen Ebene wurden die Verwendung von Konnektoren, Verben auf Basis der Valenzgrammatik, Präpositionen, Chunk sowie lexikalische Fehler untersucht; auf der syntaktischen Ebene konzentrieren wir uns derzeit auf die Entwicklung der syntaktischen Komplexität, die Wortstellung der Verben und die syntaktische Vielfalt; auf der textuellen Ebene wird die Diskurskohärenz thematisiert; darüber hinaus werden Faktorenkorrelationen erforscht, wie z.B. die Beziehung zwischen dem Einstiegsalter der Lerner und der Lerngeschwindigkeit sowie die Korrelation zwischen dem Einstiegsalter und der Korrektheit der schriftlichen Produktion.

Literatur:

- Lüdeling, Anke & Merja Kytö (Hrsg.) (2009): *Corpus Linguistics. An International Handbook. Vol 2.* (Reihe Handbücher zur Sprach- und Kommunikationswissenschaft). Mouton de Gruyter, Berlin.
- Nolda, Andreas (2019): *Annotation von Lernerdaten mit EXMARaLDA (Dulko).* Manuscript. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin.
- Reznicek, Marc; Anke Lüdeling; Cedric Krummes; Franziska Schwantuschke; Maik Walter; Karin Schmidt; Hagen Hirschmann & Torsten Andreas (2012): *Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.01.*
- Reznicek, Marc; Anke Lüdeling & Hagen Hirschmann (2013): *Competing target hypotheses in the Falko corpus: A flexible multi-layer corpus architecture.* In: Ana Díaz-Negrillo, Nicolas Ballier & Paul Thompson (Hrsg.) *Automatic Treatment and Analysis of Learner Corpus Data.* John Benjamins, 101-124, Amsterdam / Philadelphia.
- Yuan Li u.a. (2021): *Das Handbuch CDLK. Version 2.0.* (unveröffentlichtes Manuskript)
- 文秋芳, 梁茂成 & 晏小琴 编著(2008): *中国学生英语口语语料库(2.0版).* 外语教学与研究出版社, 北京. (Wen, Qiufang; Maocheng Liang & Xiaoqin Yan (Hrsg.) (2008): *Spoken and Written English Corpus of Chinese Learners. Version 2.0.* Foreign Language Teaching and Research Press, Beijing.)