



Péter Kappel, Bernadett Modrián-Horváth
(Universität Szeged, Institut für Germanistik)

Das Deutsch-ungarische Lernerkorpus (Dulko)

Projektgruppe am Lehrstuhl für Germanistische Linguistik in Szeged

Projektleitung: Dr. habil. Ewa Drewnowska-Vargáné

Verantwortliche MitarbeiterInnen: Dr. habil. Andreas Nolda (bis 2019), Dr. Bernadett Modrián-Horváth und Dr. Péter Kappel (ab 2019)

ProjektmitarbeiterInnen: Christoph Beeh, Dr. Orsolya Rauzs, Dr. György Scheibl

Datenbasis

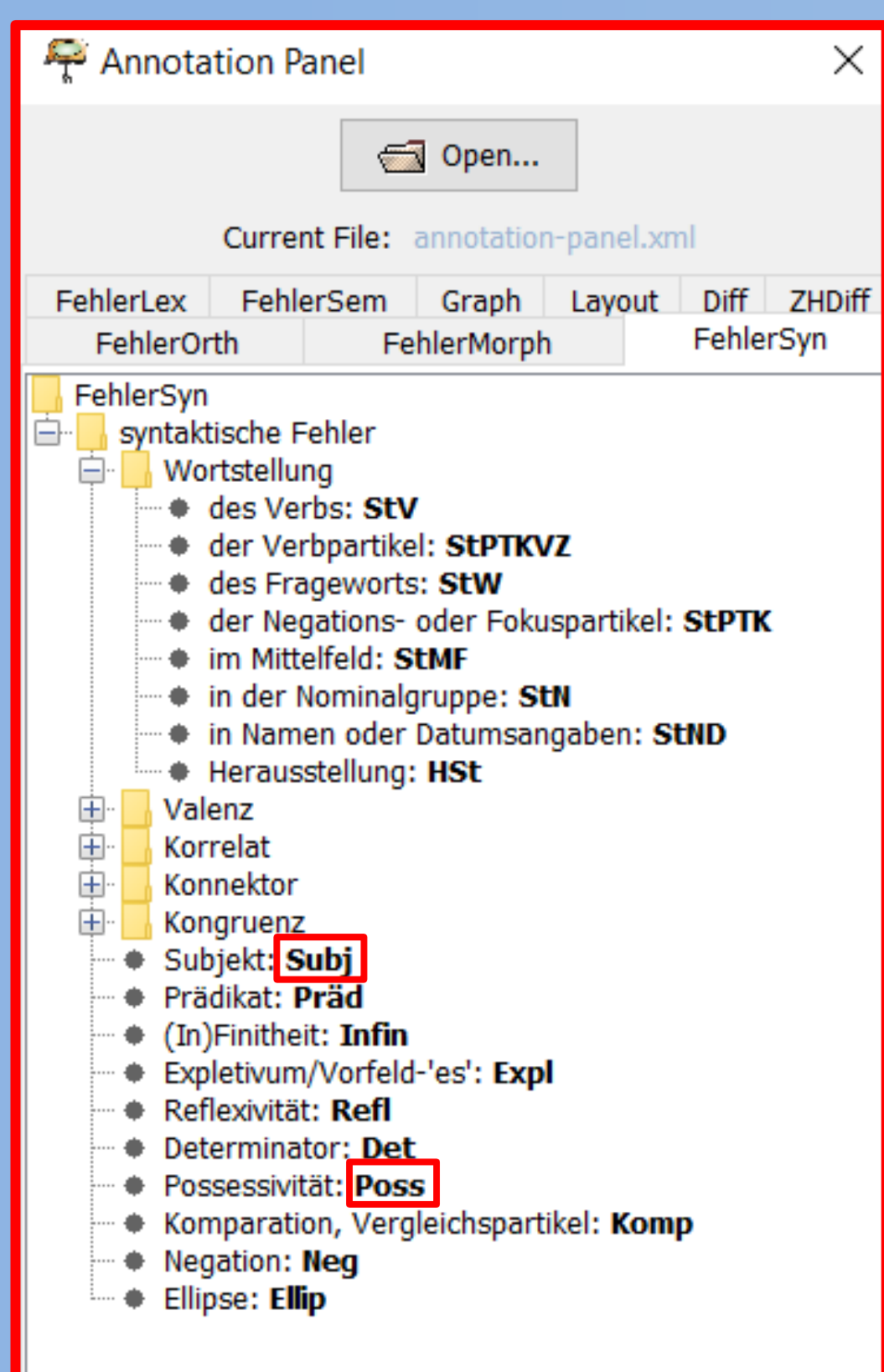
schriftliche Sprachprodukte fortgeschrittener ungarischer Deutschlernender:

- kontrolliert erhobene deutschsprachige Essays (34 Texte) und Übersetzungen (30 Texte)
- keine Hilfsmittel
- Gesamtumfang: 64 Texte, 21.010 Tokens
- ProbandInnen: Studierende am Institut für Germanistik in Szeged mit Ungarisch als L1 und Deutsch als L2 oder L3 (Sprachniveau: mindestens B2 gemäß GER/CEFR)
- Übereinstimmung der Essay-Themen von Dulko mit denen von Falko: Entlohnung, Feminismus, Kriminalität (vgl. Reznicek et al. 2012)
- Datenerhebung: handschriftlich verfasste Texte + Metadaten
- Anonymisierung, Digitalisierung und halbautomatische Annotation (inkl. Lemmatisierung, Satzspannen, pos-Tagging bzw. ZH- und Fehlerannotation)

Spezifika des Korpus

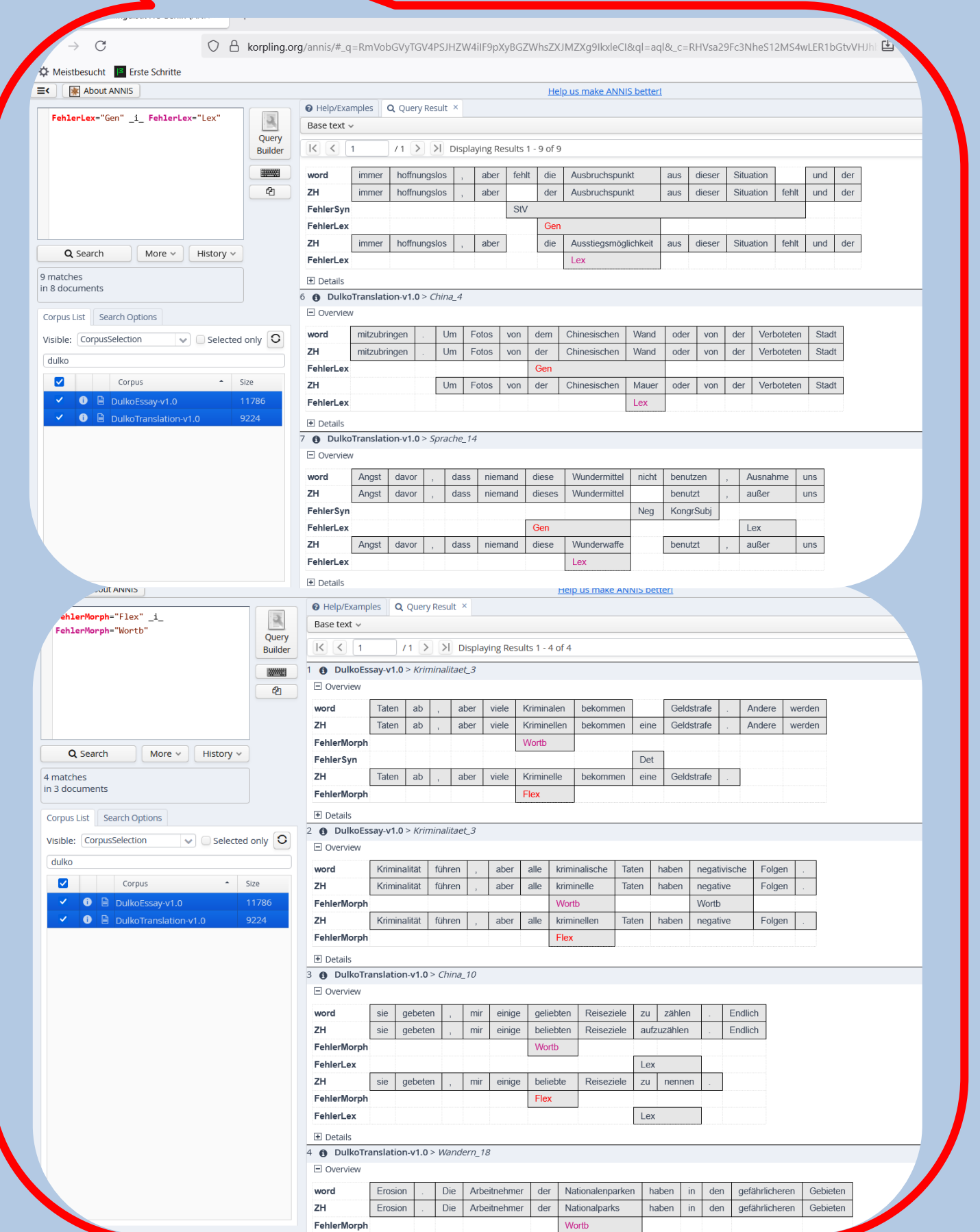
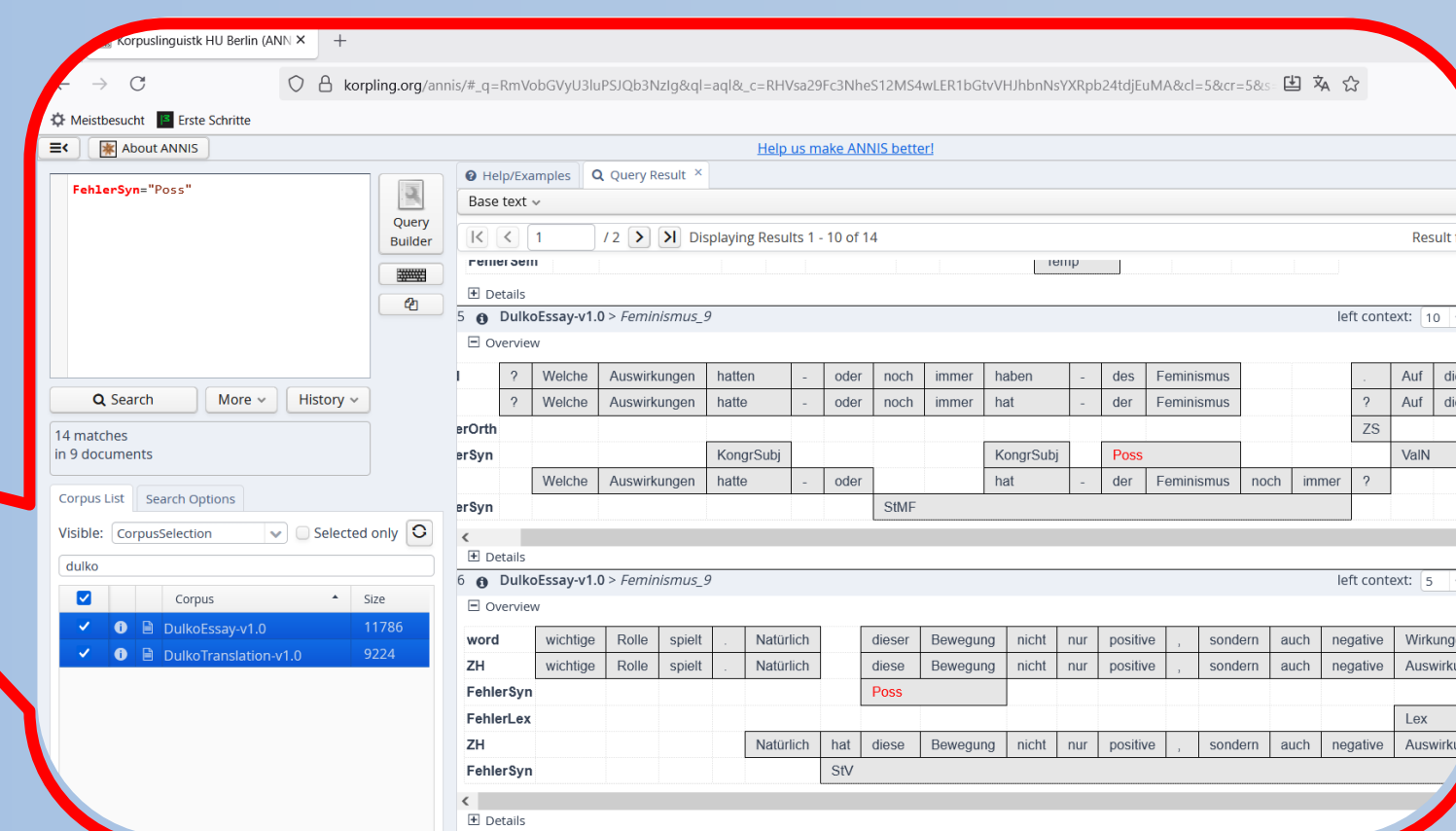
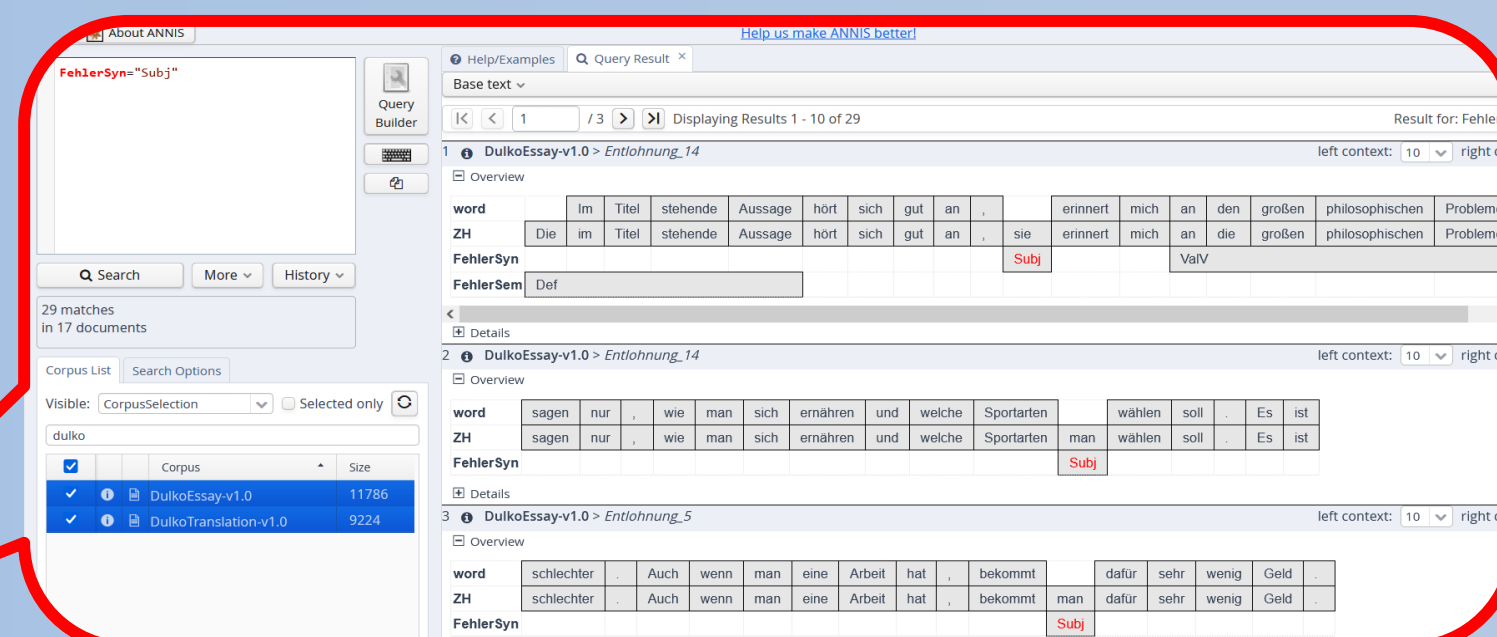


- Anpassung an vorhandene Erhebungsverfahren, Aufgabenstellung, Textsorte (Essay), Themen, Annotationsgrundlage – Falko; Vergleichbarkeit
- tiefgreifende, teils manuelle Annotation (Zielhypothesen, Fehlertagging)
- EXMARaLDA (Dulko) von Andreas Nolda: frei verfügbare Toolsammlung für den Partitur-Editor von EXMARaLDA (Schmidt / Wörner 2014); 2018 mit dem Innovationspreis der Universität Szeged ausgezeichnet; seit 2023 integrierter Bestandteil von EXMARaLDA
- Workflow intersubjektiv zuverlässig durch Teamarbeit (u.a. mit MuttersprachlerInnen)
- Übersetzungstexte (inkl. muttersprachliche Vorlagen)
- explizites, mehrdimensionales Fehler-Tagging mit selbst erarbeiteten Fehlerkategorien inkl. Fehlerbereiche (Nolda/Hirschmann 2019)
- beliebig viele, kumulativ zu interpretierende Zielhypothesen für überlappende Fehler

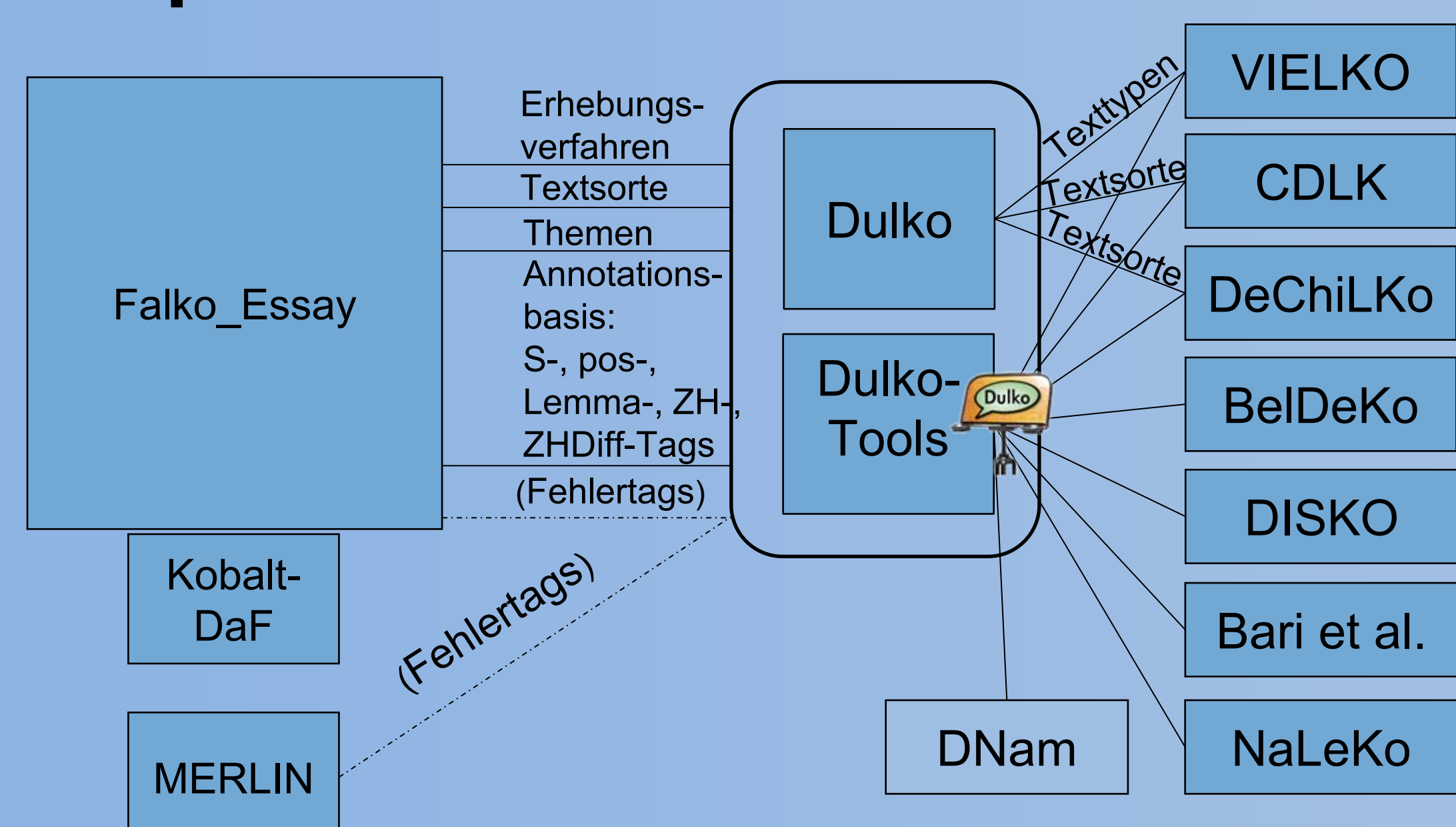


Fehlerannotation

- Datenabdeckung
- Vergleichbarkeit schaffen, aber
- typologische Eigentümlichkeiten nicht verdecken, z.B.
- **Subj**: fehlendes oder überflüssiges Subjekt (neben **ValIV**) – Ungarisch als „Pro-Drop“-Sprache
- **Poss**: fehlerhafte Realisierung des Possessors (im Ungarischen kein Genitiv und kein *habeo*-Verb vorhanden)

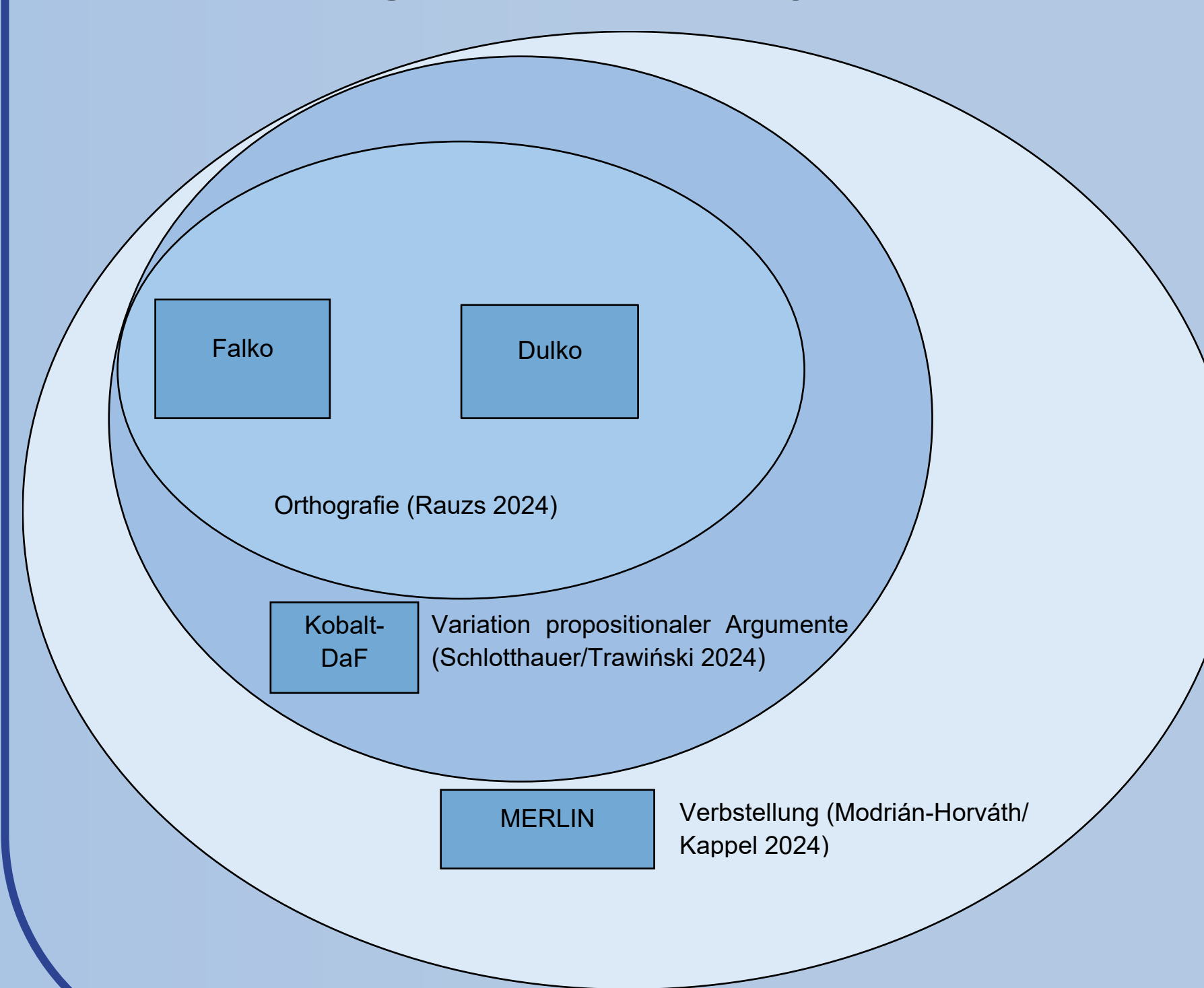


Dulko im Netzwerk von (Lernenden-) Korpora



Anwendungen des Korpus

Vergleichende Analysen



Weitere neuere Anwendungen u.a.:

- Anaphernforschung (Drewnowska-Vargáné 2023)
- Tertiärsprachenforschung (Molnár 2024)
- Verbvalenz (Szerecz 2020)
- studentische Arbeiten im Bereich der Adjektivdeklinaton, der Subjektverwendung, der Wortbildung, der Flexion, des Kasustransfers usw.

Literatur

Bari, D. É. et al. (2021): „Mein Oma nicht surfing in die Internet ...“ Eine Pilotstudie zum sprachlichen Management ungarischer Prüfungskandidaten und deren Bewertung. In: Zeitschrift für Mittteleuropäische Germanistik 7(1), 61–77.
Beeh, C. et al.: (2021): Dulko-Handbuch: Aufbau und Annotationsverfahren des deutsch-ungarischen Lernerkorpus. Version 1.0. Szeged: Universität Szeged, Institut für Germanistik. <http://doi.org/10.14232/dulko-handbuch-v1.0>.
Boyd, A. et al. (2014): The MERLIN corpus: Learner language and the CEFR. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). Reykjavik: European Language Resources Association (ELRA), 1281–1288.
Drewnowska-Vargáné, E. (2023): Gebrauch von ‚Anaphern‘ in den Essays des fehlerannotierten deutsch-ungarischen Dulko-Korpus. Vortrag an der Tagung Fehlerannotierte Lernerkorpora des Deutschen: Methodologie und Empirie.
Hien, Dang (2022): Possible Applications of the Vietnamese Learner Corpus for Teaching and Research at the University of Hanoi. In: Nusantara Science and Technology Proceedings, 132–139.
Hirschmann, H. et al. (2023): NaLeKo. Ein komplex annotiertes Lernerkorpus mit schriftlichen Erzähltexten des Deutschen als Erst- und Zweitsprache. In: Korpora Deutsch als Fremdsprache 3(2), 230–238.
Modrián-Horváth, B. / Kappel, P. (2024): Auf der Spur von syntaktischen Fallen in Texten fortgeschrittener Deutschlernender. Zur Verbstellung in Lernendenkorpora. In: Korpora Deutsch als Fremdsprache 4(2), 177–206.
Molnár, K. (2024): Lernerkorpora in der Tertiärsprachenforschung. In: Korpora Deutsch als Fremdsprache 4(2), 77–86.
Muntschick, E. et al. (2020). DISKO (Deutsch im Studium: Lernerkorpus): Handbuch.

Nolda, A. (2024): Die Dulko-Tools des EXMARaLDA-Partitur-Editors. Von einer externen Toolsammlung zum integrierten Bestandteil. In: Korpora Deutsch als Fremdsprache 4(2), 224–235.
Nolda, A. (2025): Annotation von Lernerdaten mit EXMARaLDA. Die Dulko-Tools. Posterpräsentation an der Tagung „Große Lernerkorpora – Möglichkeiten und Grenzen.“
Nolda A. / Hirschmann, H. (2019): Dulko – auf dem Weg zu einem deutsch-ungarischen Lernerkorpus. In: Eichinger, L., Rauzs, O. (2024): Orthografische Fehler in Dulko- und Falko-Texten. In: Korpora Deutsch als Fremdsprache 4(2), 207–223.
Reznicek, M. et al. (2012): Das Falko-Handbuch. Korpusaufbau und Annotationen. Version 2.01. Berlin: Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik – Korpuslinguistik.
Schlotthauer, S. / Trawiński, B. (2024): Variation bei der Realisierung propositionaler Argumente in deutschen Lernertexten. In: Korpora Deutsch als Fremdsprache 4(2), 105–125.
Schmidt, T., Wörner, K. (2014): EXMARaLDA. In: Durand, J. et al. (Hrsg.): The Oxford Handbook of Corpus Phonology. Oxford: Oxford University Press, 402–419.
Strobl, C., Wedig, H. (2023): Beldeko Summary Corpus v1.1.0. Eurac Research CLARIN Centre. Szerecz, S. (2022): Verbvalenzfehler aus dem Übersetzungskorpus von Dulko. In: Acta germanica iuvenum 3, 225–281.
Wu, Z., Li, Y. (2022): Zur syntaktischen Komplexität des Schriftdeutschen chinesischer Deutschler/-innen – Eine korpusbasierte Profilanalyse. In: Deutsch als Fremdsprache 4, 207–217.
Zinsmeister, H. et al. (2012): Das Wissenschaftliche Netzwerk „Kobalt-DaF“. In: Zeitschrift für Germanistische Linguistik 40(3), 457–458.



Fehlerannotation