KorDaF
Korpora
Deutsch als
Fremdsprache

DATA DRIVEN LEARNING WITH STUDENTS OF GERMAN: A PRACTICAL EXAMPLE OF A SMALL-SCALE CORPUS-BASED RESEARCH PROJECT

Sabrina Link
Università degli Studi di Urbino Carlo Bo

Abstract

Trotz der vielen positiven Effekte, die der Einsatz von Korpora beim Sprachenlernen mit sich bringt, kommen Korpora im Bereich Deutsch als Fremdsprache (DaF) noch relativ begrenzt zum Einsatz, und Veröffentlichungen mit konkreten Beispielen sind selten. Daher wird in diesem Beitrag ein korpusbasiertes Forschungsprojekt vorgestellt, das mit Deutschstudierenden an der Universität Urbino in Italien durchgeführt wurde. In diesem Projekt wurden die Studierenden selbst zu Forschenden und analysierten Häufigkeits- und Kookkurrenzmuster ausgewählter Neologismen im Deutschen Referenzkorpus DeReKo. Das Feedback der Studierenden zeigt, dass sie die Arbeit mit dem Korpus schätzten und dadurch metalinguistisches Wissen sowie verschiedene übertragbare Fähigkeiten entwickeln und teilweise neuen Wortschatz erwerben konnten.

Keywords: linguistische Forschung durch Studierende; projektbasiertes Lernen in DaF; Neologismen; Häufigkeitsanalyse; Kookkurrenzanalyse; DeReKo

Abstract

Despite the numerous benefits of incorporating corpora into language learning, their use in German as a Foreign Language (GFL) remains relatively limited, and publications providing concrete examples are scarce. This paper, therefore, presents a corpus-based research project conducted with German students at the University of Urbino in Italy. As part of this project, students took on the role of researchers, analysing the frequency and co-occurrence patterns of selected neologisms in the German Reference Corpus DeReKo. Student feedback suggests that they valued working with the corpus, as it enabled them to develop metalinguistic knowledge and transferable skills, while also contributing to some degree to vocabulary expansion.

Keywords: student-led linguistic research; project-based learning in GFL; neologisms; frequency analysis; co-occurrence analysis; DeReKo

1. The use of corpora in teaching German as a foreign language (GFL)

The use of corpora – collections of written or spoken utterances in one or more languages in digital form (cf. Lemnitzer / Zinsmeister 2015: 13) – offers not only significant research potential but also considerable pedagogical value, particularly in language teaching and learning. Corpora, as repositories of authentic language material, can be effectively integrated into language instruction to facilitate discovery-based learning. Through the use of corpora, learners can examine various aspects of the target language, develop their vocabulary and grammatical knowledge (cf. e.g., Bernardini 2004; Gaskell / Cobb 2004; Wallner 2008), and resolve linguistic uncertainties through targeted search queries (cf. Farroni 2024: 83). In the classroom, corpora can draw learners' attention to recurring linguistic patterns (cf. Drumbl 2011: 67) and enhance language awareness (cf. Boulton 2010) by encouraging them to engage consciously with and reflect on language rules (cf. Jessner 2008: 276). Furthermore, the discovery-based nature of this learning process fosters learner autonomy (cf. Cobb / Boulton 2015). Consequently, the use of corpora promotes an inductive approach to linguistic phenomena, commonly referred to as data-driven learning (DDL).

> DDL is alleged to enhance cognitive and metacognitive skills, increase sensitivity to authentic language use, provide an interactive approach to constructivist discovery learning, foster motivation especially through individualisation, promote reusable and transferable skills, favour autonomy for life-long learning, and correspond largely to current theories of second language acquisition. (Boulton / Tyne 2013: 99)

Despite clear evidence of the benefits of corpora and DDL, the method remains relatively underutilised in educational contexts (cf. Boulton 2020: XV) and, with few exceptions, continues to be either insufficiently recognised or entirely overlooked in foreign language instruction (cf. Corino 2019: 272). Several factors may explain this limited adoption, including a lack of teacher training and professional development opportunities (cf. Römer 2006; Corino 2019), teacher scepticism, and practical constraints such as time limitations in lesson planning (cf. Flinz 2021: 6) and the challenge of familiarising learners with corpora (cf. Corino / Onesti 2019).

While corpus linguistics has a well-established tradition, particularly in the field of English as a Foreign Language (EFL), its application within German linguistics and German as a Foreign Language (GFL) developed considerably later (cf. Fandrych / Tschirner 2007: 195). Consequently, in GFL, the approach and its pedagogical potential remain largely underexplored (cf. Flinz et al. 2021: 1). Notably, the first German-language introductions to corpus linguistics only emerged in 2006 (cf. Lemnitzer / Zinsmeister 2006; Scherer 2006), and while the use of corpora in GFL is generally encouraged, relatively few publications provide concrete implementation examples (cf. Chrissou 2011).

This article seeks to contribute to the still limited body of practical applications of corpus use in GFL by presenting a project that integrated the *Deutsches Referenzkorpus* (DeReKo, 'German Reference Corpus') into a second-year German linguistics course (*Lingua Tedesca II*) at the Università degli Studi di Urbino Carlo Bo in Italy. The project primarily aimed to enhance the students' transferable skills, metalinguistic awareness, and German language proficiency while also offering them a more practice-oriented introduction to linguistic research.

2. Corpus-based research project with students of German

2.1 Project idea and objectives

One approach to integrating corpora into language teaching is to enable learners to conduct exploratory, autonomous research (cf. Flinz 2021: 4). This method allows students to engage with corpora independently, positioning them as researchers in their own right (cf. Johns 1988: 14) or as linguistic "Sherlock Holmes" (Johns 1997: 101). This was precisely the objective of the present project: to empower students to assume the role of linguists and conduct small-group linguistic research. In alignment with the course content, the project focused on the temporal development and semantics of COVID-related neologisms.

The students first selected relevant neologisms, including both established terms that had been incorporated into the online version of the *Duden*[1] (and thus, to some extent, lost their neologism status) and unestablished ones. They then conducted both qualitative and quantitative analyses of these terms, examining their frequency since 2020 as well as their most common co-occurrences. The research aimed to address two central questions:

1. How has the frequency of COVID-related neologisms evolved over the past three years?
2. How and in what contexts are these neologisms predominantly used?

---

[1] https://www.duden.de/ (08.11.2024).

To answer these questions, the students interpreted their data, integrating hypotheses that the frequency of COVID-related neologisms would have declined between 2020 and 2023 (the year of the project) and that established neologisms would exhibit higher frequencies than unestablished ones.

For this project, the corpus was used directly in class within student groups, serving both as the foundation for their research and as a source of authentic language data, thereby exposing learners to natural language usage (cf. Lemnitzer / Zinsmeister 2015: 189). After collecting and analysing their data, the students presented their findings and wrote a critical review of the results.

In corpus-based language teaching, three primary approaches can be distinguished along the axes of empiricism/rationalism and deduction/induction (cf. Flinz 2021: 28): corpus-based quantitative, corpus-based quantitative and qualitative, and corpus-supported (cf. Lemnitzer / Zinsmeister 2015: 34). Given that this project combined frequency-based analysis with co-occurrence analysis and subsequent data interpretation, it followed a corpus-based quantitative and qualitative approach – widely regarded as the most suitable method for foreign language teaching (cf. Lemnitzer / Zinsmeister 2015: 37). This research-oriented methodology enabled learners to engage with language at a meta-level while providing an additional channel for linguistic input, complementing the examination of authentic texts (cf. Wallner 2013).

Aligned with data-driven learning (DDL; cf. Section 1), the project sought to enhance a range of transferable skills relevant to employability. According to the International Labour Organisation (ILO), individuals are most employable when they possess broad-based education and training, alongside essential portable skills such as teamwork, problem-solving, information and communications technology (ICT) proficiency, and strong communication and language skills (cf. Brewer 2013: iii). Other key employability skills include planning and organisation, self-management, autonomous working, and critical thinking (cf. ibid.: 1). Accordingly, the project was designed to develop all these competencies, with a particular focus on enhancing students' metalinguistic awareness and practical German language skills through engagement with authentic German text data and an exploration of contemporary linguistic trends.

By positioning students as independent linguistic researchers, the project granted them considerable autonomy. While they received some guidance – such as pre-formulated research questions and references to useful resources – they were ultimately responsible for managing their time effectively, planning and organising their research, and ensuring timely completion. This autonomy was also intended to enhance learner motivation, as autonomy is recognised as one of the so-called *ten commandments* for fostering motivation in language learning (cf. Dörnyei / Csizér 1998).

Working in small groups required students to communicate and collaborate effectively to achieve successful outcomes. Thus, the project not only aimed to improve teamwork and communication skills but also sought to promote social learning. Additionally, by analysing and critically reviewing their findings, students were expected to further refine their critical thinking skills.

To present their data, students needed to familiarise themselves with Excel and PowerPoint-tools with which many had little to no prior experience, particularly in the context of creating figures and visualising data. Thus, beyond enhancing digital competencies, the project required students to engage with a new technological tool – a corpus – which was unfamiliar to all participants. This process helped develop their technological proficiency and independent learning skills. Furthermore, the presentation component of the project aimed to improve students' public speaking abilities. Confronted with multiple new tools and concepts throughout the project, students also strengthened their problem-solving skills by navigating unfamiliar software and analytical methods.

Another crucial employability skill fostered by the project was linguistic proficiency. In addition to working with authentic German text data, students encountered new vocabulary and lexical fields while searching for relevant neologisms. During the frequency analysis, they had to determine

whether corpus search results genuinely reflected neologism usage. For example, the word *Sicherheitsabstand* ('safety distance') existed before the COVID-19 pandemic – commonly referring to the distance between vehicles – and thus had to be excluded when it did not appear in a pandemic-related context. This process required students to critically analyse corpus data, differentiating between various contextual meanings of the same word.

Similarly, conducting a co-occurrence analysis further enhanced students' contextual understanding of the selected neologisms. Through these tasks, they developed a more nuanced awareness of the semantic diversity of certain German words. Finally, by engaging with contemporary linguistic phenomena, students also gained insight into recent developments in the German language.

2.2 Corpus selection

A wide range of corpora is available for German. Some of the largest and most well-known for written German include the *Deutsches Referenzkorpus* (DeReKo, 'German Reference Corpus')[2] and the corpus from the *Digitale Wörterbuch der deutschen Sprache* (DWDS, 'Digital Dictionary of the German Language')[3]. For spoken German, the *Datenbank für Gesprochenes Deutsch* (DGD, 'Database for Spoken German')[4] provides a comprehensive collection of data (for a more detailed overview of German corpora, see Flinz 2021: 9-28).

While the use of customised corpora in German as a Foreign Language (GFL) didactics offers advantages – most notably, allowing learners to work with smaller datasets, thereby reducing the risk of being overwhelmed by an excessive volume of data and failing to achieve meaningful results (cf. Flinz 2021: 12) – it was decided, given the linguistic focus of the course (cf. Section 2.3), not to create a customised corpus for this project. Instead, the *Deutsches Referenzkorpus* (DeReKo) was selected. As the world's largest collection of German-language corpora, DeReKo comprises approximately 57.6 billion words (as of 9 January 2024; cf. IDS 2024). It encompasses a wide variety of texts, including fiction, academic and popular science works, newspaper articles, and other materials chosen based on their scope, quality, and topical relevance, with the majority originating from the present and recent past (cf. Flinz 2021: 10). The corpus is accessed via COSMAS II (Corpus Search, Management and Analysis System), an application used for searching and analysing the dataset (COSMAS II). DeReKo is continuously expanded, growing by approximately 3.1 million words per year (cf. Kupietz et al. 2018: 4353). Given its scale and complexity, it is primarily designed for linguistic research rather than didactic purposes (cf. Flinz 2021: 11). However, this does not imply that it cannot or should not be used in language teaching (see, for example, Nolting / Radke 2019). Nevertheless, a certain amount of time is required to familiarise students with the corpus to ensure its effective use in educational contexts (cf. Steyer 2008: 189).

Although the introduction to the corpus is time-intensive, DeReKo was deemed appropriate for this project for several reasons. Firstly, the ability to navigate a new and complex tool is a crucial skill in the contemporary workplace (cf. Section 2.1). Thus, rather than perceiving the corpus's complexity and unfamiliarity as obstacles, they were viewed as opportunities for students to develop transferable skills. Secondly, DeReKo provides an extensive and up-to-date dataset, which is essential for researching neologisms. Owing to its broad range of texts, students could investigate neologisms across different genres and registers. Additionally, the corpus allows users to filter results by decade, year, text source, topic, and other parameters – an essential feature for the temporal frequency analysis

---

[2] https://www.ids-mannheim.de/digspra/kl/projekte/korpora/ (08.11.2024).
[3] https://www.dwds.de/ (08.11.2024).
[4] https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome (08.11.2024).

the students were required to undertake. Finally, DeReKo supports both frequency and co-occurrence analysis, which were central components of the project. Therefore, despite the time investment required for corpus training, DeReKo was selected as the most suitable resource.

2.3 Project implementation

The project was conducted over two consecutive academic years (2022/23 and 2023/24) as part of the *Lingua Tedesca II* course at the Università degli Studi di Urbino Carlo Bo in Italy. Second-year German students in Urbino follow a language course and a linguistics course for their German language component. They are required to pass four different exams annually. Related to the language course they have to pass a written exam (which consists of either a grammar test combined with writing and reading comprehension or an essay, depending on their proficiency level), a listening comprehension test, and an oral examination. For the linguistics course they need to pass a German linguistics exam. *Lingua Tedesca II* constitutes the linguistics component for all second-year German students. As such, its primary aim is to enhance students' metalinguistic awareness rather than their practical language skills. This focus is also reflected in the course's instructional language: while authentic German materials (e.g., videos and PowerPoint slides) are incorporated, the course is predominantly taught in Italian. The rationale for this approach lies in the diverse proficiency levels among students, who may range from A2 to C1 level upon entering their second year, depending on the level they started with in the first year. Consequently, while some second-year students operate at an A2 level, others may already have reached C1, making exclusive instruction in German impractical.

Attendance is not mandatory for the course, though passing the exam is a prerequisite for progression. Students have the option to participate in the project or to take an oral examination at the end of the term. Despite the non-compulsory nature of attendance, the majority of students opted to frequent the course regularly. In 2022/23, 23 students participated in the project, while in 2023/24, 34 students participated. The course takes place in either the first or second term and consists of 30 teaching hours, each lasting 45 minutes, in accordance with the academic quarter system.

The course content is determined by the lecturer. In the two years covered by this study, the course focused on neologisms and was structured in two parts: a theoretical introduction and a practical component centred around the project. The theoretical section was divided into three units: an introduction to neologisms (including definitions, linguistic context, and examples), lexicography and neologisms, and psycholinguistics and neologisms, examining how new words become entrenched and institutionalised. Within the lexicography segment, students explored the decision-making process used by *Duden* to include new words in the dictionary, including the role of corpus-based analysis. They were subsequently presented with a selection of neologisms and, working in groups, evaluated whether each term should be added to *Duden*, based on the editorial criteria. As frequency of occurrence is one of the key selection criteria, students then chose one neologism and conducted a preliminary frequency analysis using DeReKo.

Before engaging in independent corpus analysis, students were introduced to DeReKo and COSMAS II through a step-by-step registration process, guided by the lecturer. Since certain aspects of the platform were not immediately intuitive, this support ensured that all students could access the corpus without difficulty. Once registered, students were introduced to the different corpus archives and were informed that the W – *Archiv für geschriebene Sprache* would be used for research purposes. They then conducted a frequency analysis on a given neologism alongside the lecturer, following a modelling approach to develop familiarity with the tool. During this phase, they were also made

aware of key considerations in corpus searches, including query formulation, data reliability, and data presentation.

Following this initial exercise, students transferred the corpus data into Excel and created visual representations of the results, again in parallel with the lecturer. They then compared their figures to those produced by the lecturer to refine their analytical skills. Once confident in their abilities, students conducted their first independent frequency analysis on a previously discussed neologism, with the lecturer providing guidance as needed.

The second phase of the course introduced the project, outlining its research questions, hypotheses (cf. Section 2.1), methodological steps, and group organisation. The project aimed to answer two research questions:

1. How has the frequency of COVID-related neologisms changed over the past three years?
2. How and in what contexts are these neologisms predominantly used?

Two hypotheses were formulated:

1. The frequency of COVID-related neologisms has declined between 2020 and 2023.
2. Established neologisms appear more frequently than unestablished ones.

Working in small groups (three to four students), participants selected ten COVID-related neologisms – five already listed in the online *Duden* and five not. They used the DWDS *Themenglossar zur COVID-19 Pandemie*[5] and OWID *Neuer Wortschatz rund um die Coronapandemie*[6] to identify relevant terms. Each group then conducted frequency analyses (2020-2023) in COSMAS II and created visualisations using Excel. They subsequently performed a co-occurrence analysis, interpreting and presenting their data autonomously but being offered support from the lecturer, when needed.

The first cohort (2022/23) delivered their findings through in-class presentations, while the second cohort (2023/24) created short video reports[7]. Presentations could be given in either German or Italian, depending on the students' proficiency levels. The project concluded with a short critical written discussion of their own results for the 2022/23 cohort, and a written critical review of another video for the 2023/24 cohort.

2.4 Students' perception and feedback

At the conclusion of the course, students were invited to complete a feedback questionnaire to provide insights into their perceptions of the course. While the original questionnaire included questions on both the theoretical and practical components, the results presented here pertain exclusively to the practical aspect of the corpus project.

In 2022/23, 19 of the participating students completed the feedback form, while in 2023/24, 15 students did so. Initially, students were asked which part of the course they found most instructive – the theoretical or the practical component – and were requested to justify their choice. Of the 34 students who provided feedback across both years, the majority (20 students) considered the practical

---

[5] https://www.dwds.de/themenglossar/Corona (08.11.2024).

[6] https://www.ids-mannheim.de/neologismen-in-der-coronapandemie/ (08.11.2024).

[7] It was decided to change the methodology so that the students did not only critically reflect on their data but also on someone else's data. Hence creating a video and reviewing it afterwards was considered a more appropriate method to increase the students' active and critical engagement with the data.

component to be more instructive, while seven students favoured the theoretical part, and seven found both components equally beneficial.

Students particularly appreciated that the practical component enabled them to apply the theoretical knowledge they had acquired, affording them a more hands-on experience that facilitated a deeper understanding of the subject matter (see student responses 1-3). Additionally, they valued the interactive and collaborative nature of the practical tasks (see student responses 4-5), as well as the opportunity to develop new practical skills (see student responses 6-7).

(1) We were able to put into practice what we had just studied.
(2) Practical, because I fully understood the theory explained.
(3) Both were useful, but I think the practical part was more instructive, you learn a lot by working first hand.
(4) Practical, because by working in groups we had more comparisons with each other and were also able to help each other better.
(5) Practical because it gave us the opportunity to work in groups and practise what was taught.
(6) Practical, because it allowed me to develop skills in analysing data and graphs and because it allowed me to put into practice what had been explained in class and learn how to use different programmes, such as Cosmas.
(7) For different reasons, I think that the two parts were instructive in equal measure: the theoretical part allowed me to learn more about the subject of neologisms from a more conceptual point of view (e.g. theories, definitions), the practical part allowed me to acquire or improve knowledge of a more technical and methodological nature (e.g. learning how to know and make the best use of the constituent elements of a corpus, completing a project together with a group, setting up a presentation containing graphs and tables as clearly as possible).

Students were subsequently asked to respond to a series of five-point Likert-scale questions regarding the practical component, with 5 indicating strong agreement and 1 indicating strong disagreement. These questions aimed to assess how demanding, motivating, engaging, and useful students found this aspect of the course. As illustrated in Figure 1, students generally perceived the practical component as demanding and challenging – 14 out of 34 students (strongly) agreed, 16 remained neutral, and 4 (strongly) disagreed. However, the majority found it motivating (24 out of 34 (strongly) agreed, with only 2 disagreeing), useful (23 out of 34 (strongly) agreed, with only 4 (strongly) disagreeing), and interesting (26 out of 34 (strongly) agreed, with only 1 disagreeing).

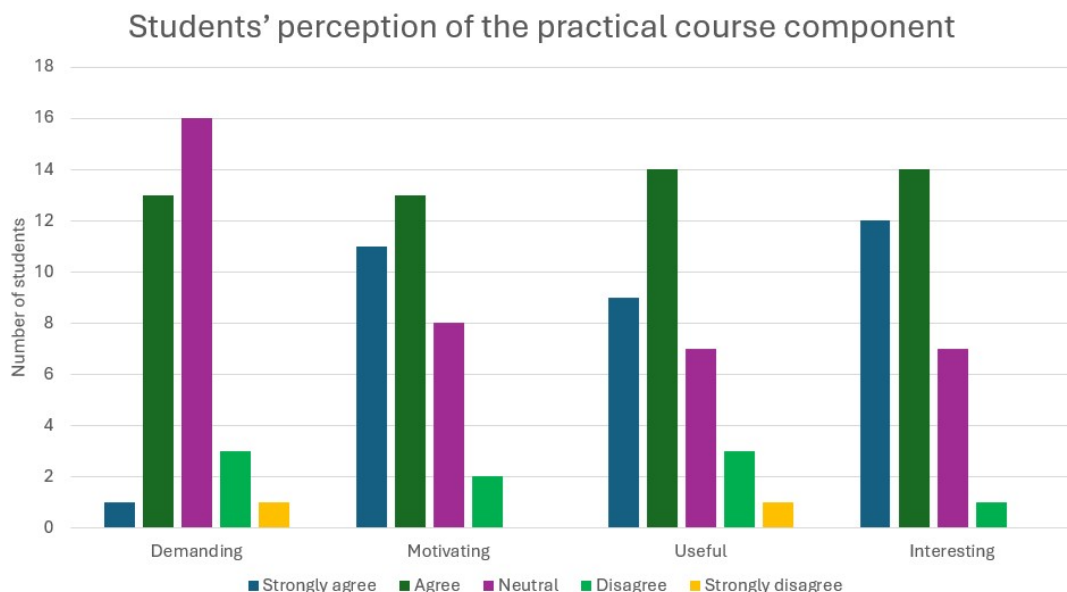Students' perception of the practical course component

Figure 1
Students' perception of the practical course component

As the corpus project was designed not only to enhance students' transferable skills and motivation but also to support their language development, students were asked whether working with the corpus had contributed to their metalinguistic knowledge and practical German skills. While the 2022/23 cohort was asked a single yes-or-no question on whether the practical component had improved their metalinguistic awareness and German proficiency, the questionnaire for the 2023/24 cohort was refined to include two separate five-point Likert-scale questions: one concerning metalinguistic knowledge and the other concerning language proficiency.

In the first cohort, 17 out of 19 students reported that working with the corpus had enhanced their theoretical and practical German skills. In the second cohort, 10 out of 15 students (strongly) agreed that the project contributed to their metalinguistic knowledge (with 5 remaining neutral), while 7 out of 15 (strongly) agreed that it had improved their German proficiency, 7 were neutral, and 1 disagreed.

Students were asked to elaborate on how the practical component had contributed to their linguistic development. Many reported improvements in their vocabulary (see student responses 8-10) and highlighted that working with the corpus allowed them to engage with authentic German usage (see student responses 10-11).

(8) In this part I learnt several new words.
(9) It helped me because we saw how and how often words are used, the various meanings and related meanings, and words that do or do not belong to today's language.
(10) The practical part allowed us to really work with the language, testing the use of words, discovering new ones, checking how often they are actually used in Germany.
(11) It brought me closer to the actual spoken language than the German studied in class.

However, not all students perceived a tangible improvement in their practical German skills, particularly given that Italian was the dominant language in the classroom (see student response 12).

(12) Certainly my theoretical and practical skills have improved because we have worked as linguists, but I do not feel that my language skills have improved because I have not applied German as much as Italian.

In line with this, one student suggested in the feedback section, "I would increase the use of German in the course". Meanwhile, another student commented, "Simplify the materials provided, especially the videos", – all of which were in German. This contrasted yet with another remark: "Reduce the number of times theoretical videos are watched in class". These responses indicate differing preferences among students regarding the balance of German and Italian in instruction. While the majority of students found the corpus project beneficial in developing various skills, particularly metalinguistic knowledge (which is the primary objective of the course), only half of the 2023/24 cohort reported a perceived improvement in their practical German proficiency.

3. Conclusion

This project has demonstrated that students of German can derive significant benefits from working with corpora and engaging in data analysis. Based on student feedback, it is evident that the majority enjoyed the project and valued the opportunity to apply theoretical knowledge in a practical context. The responses further suggest that most of the project's objectives were successfully met.

Students reported improvements in their digital competencies and appreciated the opportunity to engage with a new tool – in this case, the COSMAS II web application from DeReKo. Most students found the project to be engaging, interactive, and motivating, despite its relative complexity. The fact that all groups successfully completed the largely autonomous project suggests that they effectively applied problem-solving, planning, and organisational skills. Additionally, the high quality of the data presentations and critical reflections indicates that students developed strong analytical competencies, including the ability to collect, interpret, and critically evaluate linguistic data.

While students reported an increase in their metalinguistic awareness and some valued working with authentic German texts, only approximately 50% of the 2023/24 cohort perceived an improvement in their practical German skills. The reported improvements primarily concerned lexical development, which aligns with the focus on neologisms. Thus, while corpus analysis supported students' theoretical linguistic competence, most did not perceive significant gains in their overall German proficiency. Although the primary aim of the course was linguistic rather than language acquisition, a greater perceived improvement in practical language skills would have been desirable. However, this limitation may not be attributable to the corpus project itself but rather to the broader institutional structure of the course—specifically, the varying levels of German proficiency among students and, consequently, differing preferences regarding the use of German versus Italian in instruction.

In sum, integrating corpus analysis into German studies presents considerable pedagogical potential. Not only does it enhance students' metalinguistic knowledge and provide insight into authentic language use, but it also fosters the development of transferable skills essential for their future professional endeavours.

References and Resources

Bernardini, Silvia (2004): Corpora in the classroom. An overview and some reflections on future developments. In: Sinclair, John (ed.): *How to use corpora in language teaching*. Amsterdam / Philadelphia: John Benjamins Publishing, 15-36.

Boulton, Alex / Tyne, Henry (2013): Corpus linguistics and data-driven learning: A critical overview. In: *Bulletin suisse de Linguistique appliquée*. Neuchâtel: Institut de linguistique de l'Université. Rôle des pratiques langagières dans la constitution des espaces sociaux pluriels d'aujourd'hui Actes du colloque VALS-ASLA (Lausanne, 1-3 février 2012), 97-118.

Boulton, Alex (2010): Consultation de corpus et styles d'apprentissage. In: *Cahiers de l'APLIUT* 29: 1, 98-115.

Boulton, Alex (2020): Data-Driven Learning for Younger Learners: Obstacles and Optimism. In: Crosthwaite Peter (ed.): *Data-Driven Learning for the Next Generation*. New York: Routledge, XIV-XX.

Brewer, Laura (2013): Enhancing youth employability: what? why? and how?: guide to core work skills, International Labour Office, Geneva. http://www.ilo.org/skills/pubs/WCMS_213452/lang--en/index.htm (08.11.2024).

Chrissou, Marios (2011): *Mit Textkorpora im Unterricht arbeiten und Sprachförderung initiieren.* https://www.uni-due.de/imperia/md/content/prodaz/chrissou_korpusarbeit.pdf (08.11.2024).

Cobb, Tom / Boulton, Alex (2015): Classroom applications of corpus analysis. In: Biber, Douglas / Reppen, Randi (eds.): *Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, 478-497.

Corino, Elisa / Onesti, Cristina (2019): Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning. In: *Frontiers in Education*, 4: 7, 1-12. https://doi.org/10.3389/feduc.2019.00007.

Corino, Elisa (2019): Data-driven Learning: tra lingue straniere e CLIL, tra ricerca e didattica. In: *EL.LE* 8: 2, 271-286.

COSMAS II: http://www2.ids-mannheim.de/cosmas2/uebersicht.html (08.11.2024).

Dörnyei, Zoltan / Csizér, Kata (1998): Ten commandments for motivating language learners: results of an empirical study. In: *Language Teaching Research*, 2: 3, 203-229. https://doi.org/10.1177/136216889800200303.

Drumbl, Hans (2011): Sprachdidaktik und Korpora. In: Abel, Andrea / Zanin, Renata (Hrsg.): *Korpora in Lehre und Forschung*. Bolzano: Bolzano University Press, 55-99.

Fandrych, Christian / Tschirner, Erwin (2007): Korpuslinguistik und Deutsch als Fremdsprache. Ein Perspektivenwechsel. In: *Deutsch als Fremdsprache* 44: 4, 195-204.

Farroni, Cristina (2024): Korpora im DaF-Unterricht: Einsatz im Schulbereich. In: *Linguistik Online*, 126: 2, 83-98. https://doi.org/10.13092/lo.126.11044.

Flinz, Carolina (2021): KORPORA in DaF und DaZ: Theorie und Praxis. In: *Zeitschrift für interkulturellen Fremdsprachenunterricht*, 26: 1, 1-43.

Flinz, Carolina / Hufeisen, Britta / Luppi, Rita / Mell, Ruth M. (2021): Korpora als Grundlage für das Lehren und Lernen von Deutsch als Fremdsprache. In: *Korpora Deutsch als Fremdsprache*, 1: 2, 1-4.

Gaskell, Delian / Cobb, Thomas (2004): Can learners use concordance feedback for writing errors? In: *System*, 32: 3, 301-319.

Jessner, Ulrike (2008): A DST Model of Multilingualism and the Role of Metalinguistic Awareness. In: *The Modern Language Journal* 92: 2, 270-283.

Johns, Tim (1988): Whence and whither classroom concordancing? In: Bongaerts, Theo / De Haan, Pieter / Lobbe, Sylvia / Wekker, Herman (eds.): *Computer applications in language learning*. Dordrecht: Foris, 9-27.

Johns, Tim (1997): Contexts: The Background, Development and Trialling of a Concordance-Based CALL program. In: Wichmann, Anne / Fligelstone, Steven / McEnery, Tony (eds.): *Teaching and Language Corpora*. Harlow: Addison Wesley Longman, 100-115.

Kupietz, Marc / Lüngen, Harald / Kamocki, Paweł / Witt, Andreas (2018): The German Reference Corpus DeReKo: New Developments – New Opportunities. In: Calzolari, Nicoletta / Choukri, Khalid / Cieri, Christopher / Declerck, Thierry / Hasida, Koiti / Isa-hara, Hitoshi / Maegaard, Bente / Mariani, Joseph / Moreno, Asuncion / Odijk, Jan / Piperidis, Stelios / Tokunaga, Takenobu (eds.): *Proceedings of The Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: European Language Resources Association (ELRA), 4353-4360.

Lemnitzer, Lothar / Zinsmeister, Heike (2006): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.

Lemnitzer, Lothar / Zinsmeister, Heike (2015): *Korpuslinguistik. Eine Einführung*. 3. Auflage. Tübingen: Narr.

Nolting, Antje / Radtke, Nadja (2019): Korpusbasierte Lexikografie. Nutzung von Korpora und Analysewerkzeugen im Unterricht für Deutsch als Muttersprache und Fremdsprache. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 24: 1, 107-126.

Römer, Ute (2006): Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments. In: *ZAA* 54: 2, 121-134.

Scherer, Carmen (2006): *Korpuslinguistik*. Heidelberg: Winter.

Steyer, Kathrin (2008): Kollokationen in deutschen Wörterbüchern und in der deutschen Wörterbuchforschung. In: *Lexicographica: International Annual for Lexicography* 24, 185-207.

Wallner, Franziska (2008): Schriftliche Textproduktion in Deutsch als fremde Wissenschaftssprache – Überlegungen zu Einsatz und Nutzen korpuslinguistischer Kollokationsanalysen im Fach DaF. In: Krumm, Hans-Jürgen / Portmann-Tselikas, Paul (Hrsg.): *Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache*. Innsbruck et al.: Studienverlag, 153-163.

Wallner, Franziska (2013): Korpora im DaF-Unterricht – Potentiale und Perspektiven am Beispiel des DWDS. In: *Revista Nebrija de Lingüistica Aplicada* 13, https://www.nebrija.com/revista-linguistica/korpora-im-daf-unterricht-potentiale-und-perspektiven-am-beispiel-des-dwds.html (08.11.2024).

Biographical note: Sabrina Link is currently a lecturer in German as a Foreign Language (GfL) and German Linguistics at the Università degli Studi di Urbino Carlo Bo in Italy. Her research interests in the field of linguistics particularly focus on gender-fair language use in the media, lexical studies (e.g., the use of neologisms), and pragmatics (e.g., the use of routine formulae). Additionally, she is

engaged in research on language didactics, with a special interest in project- and activity-based teaching and learning methods and the application of linguistics in GfL instruction.

Contact address
Sabrina Link
Scuola di Lingue e Letterature Straniere
Dipartimento di Scienze della Comunicazione, Studi Umanistici e Internazionali (DISCUI)
Piazza Rinascimento, 7
61029 Urbino
Italy
Sabrina.link@uniurb.it