

Franz Unterholzner, Ilka T. Fladung & Marcel Illetschko

Einsicht in sprachstrukturelle Zusammenhänge und deren Modellierung in Sprachbetrachtungs-Testaufgaben der iKM^{PLUS}

Einblicke in Aufgabenkonstruktion und Antwortverhalten

Zusammenfassung

Der Beitrag gibt Einblick in Testaufgaben zur Sprachbetrachtung aus der österreichischen nationalen Kompetenzmessung iKM^{PLUS}. Im Fokus stehen Überlegungen zur Aufgabenkonstruktion und Analysen des Antwortverhaltens von Schüler:innen bei drei für den Echteinsatz nicht validierten Aufgaben. Die Daten stammen aus Pilotierungsstudien auf Primar- und Sekundarstufe ($n_{\text{Prim}} = 623$; $n_{\text{Sek}} = 822$). Zwei Aufgaben beziehen sich auf orthographisches Wissen (satzinterne Großschreibung und Kommasetzung), eine auf die Unterscheidung von Verben und Nomen/nominalisierten Verben. Die Analysen deuten auf ein gering ausgeprägtes Bewusstsein der Proband:innen für die (morpho)syntaktische Determiniertheit der untersuchten Phänomene hin. Sie zeigen zudem die Schwierigkeiten bei der Konstruktion psychometrisch robuster Aufgaben im Spannungsfeld zwischen Testgüte und unterrichtlich geprägten Vorstellungen. Die Darstellung des Wechselspiels von Konstruktionsüberlegungen und Ergebnisinterpretation trägt zur Diskussion der Frage bei, wie didaktisch fundierte Konstrukte in standardisierte Testformate im Bereich Sprachbewusstheit überführt werden können.

Schlagwörter: Sprachbetrachtung • Sprachbewusstheit • Sprachdidaktik • Kompetenzmessung • Orthographiedidaktik

Didaktik Deutsch

Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur

31. Jahrgang 2026. Heft 61. S. 57–79

DOI: 10.21248/dideu.993

Copyright Dieser Artikel wird unter der Creative-Commons-Lizenz CC BY-NC-ND 4.0 veröffentlicht: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Abstract

Modeling Linguistic Structure in Language Awareness Assessment: Insights into Item Design and Student Response Behavior within the Austrian Competence Assessment iKM^{PLUS}

This paper examines items in the “language awareness” domain of the Austrian competence assessment iKM^{PLUS}. It focuses on item design and student response behavior. Three items that were excluded during the validation process are analysed using data from one primary-level pilot study (n = 623) and one secondary-level pilot study (n = 822). Two items target orthography-related knowledge of written structure and strategies (capitalisation cues and punctuation), the third item focuses on the distinction between verbs and nouns/nominalised verbs. Items were developed to improve the assessment of procedural and metacognitive knowledge rather than declarative knowledge. Analyses highlight design challenges and response patterns. Results demonstrate the difficulty of operationalising strategic metacognitive competence in test items. Student responses indicate limited awareness of (morpho-)syntactic determination. Findings contribute to discussions on translating didactically informed constructs into standardised test items in the field of language awareness.

Keywords: language awareness • language didactics • grammar teaching • competence testing • spelling

1 Einleitung und Problemaufriss

Im deutschdidaktischen Diskurs wird mitunter darauf hingewiesen, dass der schulische Orthographie- und Grammatikunterricht deutlich von der wissenschaftlichen Auseinandersetzung mit den zugehörigen Lerngegenständen wie auch von aktuellen didaktischen Modellierungen entfernt sei. Obwohl einige Modelle für den Orthographie- und Grammatikunterricht seit Längerem in der Wissenschaft als potenziell wirkungsvoll diskutiert würden, hätten diese bisher selten Eingang in die schulische Realität gefunden – so z. B. die Nominalphrase, ein valenzorientiertes syntaktisches Satzverständnis oder das topologische Feldermodell als Visualisierungshilfe (Bredel, 2021; Christ, 2017).

Überlegungen zu deren didaktischem Potenzial werden unter anderem aus der Konsistenz der linguistischen Modellierung abgeleitet, was mitunter als „Lerngegenstandsangemessenheit“ (Müller, 2025, S. 22) o. Ä. bezeichnet wird. Erste Vorschläge zu den hier thematisierten Modellierungen gehen bereits auf die 1990er-Jahre zurück (vgl. Abschnitt 2.1). Insgesamt muss aber – trotz einiger Studien zur satzinternen Großschreibung (s. u.) – noch ein Mangel an empirischer Evidenz bezüglich ihrer Wirksamkeit konstatiert werden (vgl. Betzel, 2024, S. 32–33).

Vor diesem Hintergrund erhebt sich die Frage, ob sich Aufgaben konstruieren lassen, die Hinweise darauf geben, inwieweit sich die genannten didaktischen Modellierungen im Antwortverhalten von Schüler:innen widerspiegeln. Zu diesem Zweck wurden im Rahmen der österreichischen „individuellen Kompetenzmessung PLUS“ (iKM^{PLUS}; Stauber et al., 2024) Aufgaben aus den Kompetenzbereichen *Einsicht in Sprache durch Sprachbetrachtung* (Primarstufe) und *Sprachbewusstsein* (Sekundarstufe) entwickelt. Der vorliegende Beitrag analysiert ausgewählte Aufgaben sowie das zugehörige Antwortverhalten unter Bezug auf die angesprochenen didaktischen Modellierungen. Dabei wird diskutiert, inwieweit sich bestimmte Antwortmuster als Ausdruck dieser Modellierungen interpretieren lassen und welche Konsequenzen sich daraus für die Konstruktion diagnostischer Aufgaben ergeben.

Zwei der Aufgaben stammen aus der Sekundar-, eine aus der Primarstufe. In diesem Beitrag können sie deshalb präsentiert werden, weil sie im Rahmen des Validierungsprozesses aus weiter unten diskutierten Gründen vom Echteinsatz in der Kompetenzmessung ausgeschlossen wurden, was allerdings nicht bedeutet, dass sie für Analysen in Bezug auf Bedingungen der Aufgabenkonstruktion im Wechselspiel mit dem Antwortverhalten von Schüler:innen wertlos wären. Alle Aufgaben behandeln (morpho)syntaktisch determinierte Phänomene. Aufgabe 1 berührt die Thematik der Nominalphrase im Verhältnis zur satzinternen Großschreibung, Aufgabe 2 ein ansatzweise valenzorientiertes Satzverständnis in Bezug auf die Kommasetzung und Aufgabe 3 (die in zwei Varianten präsentiert wird) das visualisierende Potenzial des topologischen Feldermodells für die Unterscheidung der Wortarten Verb und Nomen. Die ersten beiden Aufgaben zielen zwar letztlich auf orthographische Fragestellungen, sind jedoch dem Kompetenzbereich Sprachbetrachtung zuzuordnen, da es bei ihnen nicht um Rechtschreibentscheidungen an sich, sondern um das ihnen zugrunde liegende schriftstrukturelle Wissen geht. Die dritte Aufgabe gehört mit ihrer traditionell schulgrammatischen Ausrichtung auf Wortarten ohnehin zum Kern dieses Kompetenzbereichs.

2 Sprachdidaktisch-theoretischer Hintergrund

Viele rezente orthographiedidaktische Modellierungen eint, dass sie auf der „kognitiven Durchdringung“ (Betzel 2024, S. 32) sprachstruktureller Zusammenhänge basieren, somit, sofern eine gute didaktische Vorstrukturierung gegeben ist, auch entdeckendes Lernen ermöglichen und zur Ver-

fügbare von Strategien beitragen sollen (vgl. Müller, 2024, S. 90–96). In diesem Sinne sind Sprachbetrachtung und Orthographie eng miteinander verknüpfte Bereiche.

Dabei steht als theoretische Annahme hinter der Konstruktion der in diesem Beitrag vorgestellten Aufgaben die modellhafte Vorstellung, wonach deklarativ-grammatisches Wissen sich nicht auf direktem Weg in prozedurale Rechtschreibfähigkeiten überführen lässt (Funke, 2022, S. 440). Demnach kann deklaratives metasprachliches Wissen, kombiniert mit dem entsprechenden metakognitiven Wissen zu Überprüfungsprozeduren etc., also vor allem über den „Umweg“ der Selbstregulation wirksam werden (vgl. die Gemeinsamkeiten der verschiedenen Interface-Positionen in Kohl-Dietrich & Maiberger, 2022, S. 327). So könnte etwa am Beginn eines sprachbewussten Denkvorgangs, z. B. während des Verschriftens, ein erster, nicht zwingend verbalisierbarer sprachaufmerksamer Moment (im Sinne von Noticing; Schmidt, 1990, S. 132) stehen. Darauf könnte ein Innehalten, also ein Moment der „Deautomatisierung“ (Bredel, 2013, S. 24–25) entsprechend der Vorstellung der Aktivierung des „inneren Monitors“ (Ossner, 2007), folgen. Insgesamt wurde damit als leitend für die im Folgenden dargestellte Aufgabenkonstruktion angenommen, für das integrierte Wirksamwerden von Sprachbewusstheit brauche es, über geeignetes metasprachliches Wissen hinaus, auch metakognitiv-strategisches Wissen um anknüpfende Überprüfungsprozeduren (Unterholzner & Müller, 2023).

Zu den Gegenständen didaktischer Modellierungen metasprachlichen Wissens mit Fokus auf sprachstrukturelle Einsicht und Lerngegenstandsangemessenheit gehören die Nominalphrase, ein syntaktischer Satzbegriff, der im Kern der Valenztheorie entnommen ist (Hennig, 2023), sowie das topologische Feldermodell (bspw. Rothstein, 2022).

Die Aufgaben, die im weiteren Verlauf dieses Beitrags vorgestellt werden, greifen solche Modellierungen auf:

- In Aufgabe 1 (Jahrgangsstufe 7 und 8) geht es um Strategiewissen zur satzinternen Großschreibung (u. a. um die auf das Erkennen des Wortes im rechten Nominalfeld zielende Erweiterungsprobe durch ein kongruent flektiertes Adjektiv im Mittelfeld der Nominalphrase).
- Gegenstand von Aufgabe 2 (Jahrgangsstufe 7 und 8) ist das Erkennen von Teilsätzen an ihrem „tragenden“ Element, der Verbgruppe, als Strategie für Entscheidungen über Kommasetzung.
- Aufgabe 3 (Jahrgangsstufe 3) zielt auf das Erkennen unterschiedlicher kategorialer Wortartzuweisungen von äußerlich identischen Wörtern in verschiedenen Satzkontexten – mit und ohne Zuhilfenahme einer Visualisierung durch das topologische Feldermodell.

Diese drei Phänomenbereiche werden im Folgenden näher betrachtet.

2.1 Nützliche Strategien für die satzinterne Groß- und Kleinschreibung erkennen (Aufgabe 1)

Für viele Kinder und Jugendliche liegt im Bereich der satzinternen Großschreibung ein Fehler Schwerpunkt, wobei Nominalisierungen durch syntaktische Konversion besondere Schwierigkeiten bereiten (Rautenberg & Wahl, 2024, S. 2). An dieser Stelle wird die „Verzahnung von Sprachreflexion und Orthographie“ (Bredel, 2022, S. 258) deutlich. Denn ohne einen syntaktischen Blick auf den Satz und die Wortgruppe (insgesamt vgl. Hübl & Steinbach, 2015), bspw. indem man eine „vom Satzkontext unabhängige und vereinfachende Wortartbestimmung“ (Bangel et al., 2020, S. 55) durchführt, lässt sich keine informierte Entscheidung über Groß- oder Kleinschreibung treffen. Daher behandeln syntaktisch fundierte Ansätze (Funke, 1995; Röber-Siekmeyer, 1999) die attributiven Adjektive im Mittelfeld der Nominalphrase als den entscheidenden Indikator für die Großschreibung

des Wortes im rechten Feld der Nominalphrase.¹ Denn durch stufenweises Verlängern des Mittelfelds mit Hilfe von Attributen kann das sich immer weiter nach rechts verschiebende, großzuschreibende Wort im rechten Nominalfeld sichtbar gemacht werden (Röber-Siekmeyer, 1999). Darüber hinaus kann der nominale Charakter des rechts stehenden Wortes an der Notwendigkeit der Kongruenz bzw., daraus resultierend, an der Flexion des hinzugestellten Adjektivs abgelesen werden (Funke, 1995).

Der Einfluss der Einsicht in diese syntaktisch(-morphologischen) Eigenschaften der Nominalphrase auf die satzinterne Großschreibung wurde in mehreren Studien untersucht (vgl. Bangel et al., 2020, S. 62–67), zuletzt bspw. bei Rautenberg und Wahl (2019), Bangel (2022) sowie Rautenberg und Wahl (2024). Insgesamt ergibt sich aus diesen Studien, dass syntaxbasierte didaktische Ansätze als vielversprechend eingeordnet werden können (Rautenberg & Wahl, 2024, S. 20). In Bezug auf vom Satzkontext isolierte Modellierungen zeigen Studien demgegenüber (und durchaus in das Gesamtbild passend), dass sich das vom Satzkontext unabhängige semantische Kriterium des Bezugs auf einen Gegenstand (Konkretum) oder eine Idee (Abstraktum) als „Großschreibung auslösendes Kriterium“ (Bredel, 2010, S. 218) im weiteren Verlauf des Erwerbs der satzinternen Großschreibung hartnäckig hält. Dies führt zu Problemen (vgl. Betzel, 2015) und bereitet Schüler:innen in der Sekundarstufe I Schwierigkeiten, syntaktische Analysen für die Richtigschreibung zu leisten (Prosch, 2016, S. 241).

2.2 Über Strategien zur Kommasetzung zwischen Teilsätzen entscheiden (Aufgabe 2)

Das Komma als fehlerträchtiger Lerngegenstand ist gut untersucht (für einen Überblick zur Studienlage vgl. Fürstenberg, 2023, S. 22–30). In der schulischen Praxis sind bei der Vermittlung der Kommasetzung Ansätze weit verbreitet, die sich auf eine große Zahl an Regeln stützen, wodurch die klaren und verhältnismäßig einfachen – weil eindeutig syntaktisch bestimmten (Günther, 2000, S. 285) – Grundverhältnisse eher verschleiert als einsichtig gemacht werden (vgl. bspw. Noack, 2018, S. 5). Dies gilt insbesondere für das Satzgrenzenkomma: Die weit verbreitete Signalwortdidaktik, die den Blick fast ausschließlich auf Junktionen in Sätzen lenkt, scheint einen syntaktisch determinierten Blick auf das Satzgrenzenkomma zu erschweren (Bredel, 2015a, S. 135). In der Sprachdidaktik wird folglich seit längerer Zeit die im Vergleich zu den mit Signalwörtern verbundenen zahlreichen Regeln und Unterregeln didaktisch reduzierte Orientierung am Satz nach valenztheoretischen Prinzipien gefordert (Esslinger & Noack, 2020, S. 47–52). Dabei ist hier aus didaktischer Sicht mit ‚valenztheoretisch‘ vor allem der folgende, im syntaktischen Sinne satzdefinierende Aspekt gemeint: „Sätze sind Verben mit ihren zugehörigen Wortgruppen“ (Lindauer & Sutter, 2005, S. 29). Es reicht demnach für die Trennung von Teilsätzen durch Komma für einen hohen Anteil an Einzelfällen aus, nach den Verben Ausschau zu halten.

Bis dato existiert nur eine auf dem valenzorientierten Vermittlungsansatz basierende Wirkungsstudie. Sie misst für eine sehr auf Reduktion und Einfachheit zielende Variante eine mittlere Effektstärke (Unterholzner, 2021, S. 198). Während dies zwar einen allgemeinen Wirksamkeitsnachweis für die Vermittlungsmethode darstellt, kann daraus aber kein spezifischer Vorteil des valenztheoretischen Ansatzes abgeleitet werden, da der Vergleich mit der Kontrollgruppe einem anderen Aspekt

¹ Wegen der uneinheitlichen, auf unterschiedlichen theoretischen Konzeptualisierungen basierenden Verwendung der Ausdrücke *Kern* und *Kopf* in Zusammenhang mit der Nominalphrase werden hier nach dem Vorbild von Laser und Riegler (2025, S. 49–51) die Bezeichnungen *linkes Nominalfeld* (durch Artikelwort besetzt) und *rechtes Nominalfeld* (durch Nomen oder nominal verwendetes, großzuschreibendes Wort besetzt) verwendet.

der didaktischen Intervention diene (metakognitiv orientierte Didaktisierung), das Treatment der Kontrollgruppe folglich ebenso als valenzorientierter Ansatz durchgeführt wurde.

2.3 Verben und Nomen/nominalisierte Verben unterscheiden (Aufgabe 3)

Für das Unterscheiden von Verben und Nomen im konkreten Satzkontext kann auch das topologische Feldermodell nützlich sein (vgl. bspw. Metzger, 2017, S. 79–80). Es zielt auf Visualisierung ab, um den kompetenten Umgang mit typischen Satzstrukturen des Deutschen zu erleichtern (zum Visualisierungsaspekt vgl. Uhl, 2019). Von besonderem Interesse für die vorliegende Aufgabe ist die Eigenschaft des Feldermodells, deutlich zu veranschaulichen, dass Verben nicht einfach beliebige, sondern sehr systematisch bestimmte Positionen im Satz besetzen (Laser & Riegler, 2025, S. 50–53). Daraus ergibt sich eine Dreiteilung in Verberst-, Verbzweit- und Verbletztsatz (Granzow-Emden, 2022, S. 214–225), und die Nichtverschiebbarkeit der Verbgruppe wird deutlich sichtbar (Granzow-Emden, 2022, S. 212).² Das folgende Beispiel eines Verbzweitsatzes aus Granzow-Emden (2019, S. 12) zeigt, dass das Feldermodell sowohl für das Identifizieren von Verben als auch für das Erkennen von Nominalisierungen durch syntaktische Konversion eine unterstützende Visualisierung darstellen kann.

Vorfeld	linkes Verbfeld	Mittelfeld	rechtes Verbfeld	Nachfeld
Nach dem Schwimmen	sonne	ich mich auf der Liegewiese		.
Ich	sonne	mich nach dem Schwimmen auf der Liegewiese		.
Auf der Liegewiese	sonne	ich mich nach dem Schwimmen		.

Tab. 1: Syntaktische Konversion im topologischen Feldermodell

Im Fall einer Fragestellung wie „Schreibt man die Wörter SCHWIMMEN und SONNE hier groß oder klein?“ wäre man bei vom Satzkontext isolierter Betrachtung verleitet, *schwimmen* klein- und *Sonne* großzuschreiben, was damit korrespondiert, dass *schwimmen* aus lexikalischer Perspektive ein Verb und *Sonne* ein Nomen darstellt. Eine zu überprüfende Annahme könnte nun sein, dass es Schüler:innen, die daran gewöhnt wären, Sätze mit Hilfe des topologischen Feldermodells zu betrachten,³ leichter fallen könnte, *sonne* hier syntaktisch-relational als Verb zu identifizieren, weil es stabil im linken Verbfeld auftaucht. Gleiches gilt mit anderen Vorzeichen für das am rechten Rand der Nominalphrase stehende *Schwimmen*. Großzuschreibende Wörter treten also gehäuft am Ende von Satzgliedern – und im Fall der Vorfeldprobe gehäuft als letztes Wort vor dem linken Verbfeld – auf. Von Interesse ist hier auch, inwiefern das topologische Feldermodell Schüler:innen nicht nur bei der expliziten Analyse von Sätzen, sondern durch die grafische Vorstrukturierung und somit Verdeutlichung auch bei impliziten Lernprozessen (implizite Formfokussierung; Klages & Pagonis, 2025, S. 13–15) unterstützt – bei solchen unbewussten Lernprozessen also, in denen das prozedurale System bei gehäufte Wahrnehmung bestimmter wiederkehrender Muster automatisierte, dem Bewusstsein nicht zugängliche Reaktionen auf die Wahrnehmung bestimmter struktureller Merkmale aufbaut.

² Dadurch wird nahegelegt, das Prädikat bzw. die Verbgruppe schulgrammatisch nicht als Satzglied zu behandeln. Auf den Diskurs um die Notwendigkeit des Prädikatsbegriffs soll hier nicht näher eingegangen werden (vgl. dazu bspw. Granzow-Emden, 2021, S. 26–27, oder Christ, 2019).

³ Böhler und Cristante (2024) verwenden diese Modellierung bereits für die Primarstufe.

Für das Feldermodell existieren schulbezogene Publikationen (bspw. Wöllstein & Schmadel, 2015; hier insbesondere Gallmann, 2015; Granzow-Emden, 2022; Rothstein, 2022), Modellierungen für die Sekundarstufe (Granzow, 2019; Metzger, 2017) und die Primarstufe (Böhler & Cristante, 2024), theoretisch hergeleitete Annahmen zu dessen Überlegenheit im Hinblick auf den Rechtschreibunterricht (Bredel, 2015b) und eine Interviewstudie zum Blick von Lehrkräften auf das Feldermodell (Betzel & Kleinbub, 2025).

3 Methode

3.1 Hintergrund: Testaufgaben im Kompetenzbereich Sprachbetrachtung in der iKM^{PLUS}

Die vorgestellten Aufgaben stammen aus der iKM^{PLUS}. Die *individuelle Kompetenzmessung PLUS* ist ein in Österreich flächendeckend eingesetztes Large-Scale-Instrument zur Kompetenzmessung, das Individualförderung, Unterrichts- und Schulentwicklung, Schulqualitätsmanagement und Bildungsmonitoring in einem Format vereint. Es wurde ab 2020 durch das *Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen* (IQS) implementiert und befindet sich seit Herbst 2025 im zweiten Erhebungszyklus (Stauber et al., 2024). Im Rahmen der iKM^{PLUS} liegt ein Schwerpunkt darauf, Lehrkräften nicht nur ein niederschwelliges Diagnostiktool zur Evaluierung ihres Unterrichts anzubieten, das kompetenzbereichsspezifische Rückmeldungen zum Lernstand der Schüler:innen geben kann, sondern auch Hinweise zur weiteren Förderung bereitzustellen. Dies gilt insbesondere für die sogenannten Bonusmodule, die in den Kompetenzbereichen *Einsicht in Sprache durch Sprachbetrachtung* (Primarstufe) und *Sprachbewusstsein* (Sekundarstufe) freiwillig eingesetzt werden können.

3.2 Eingesetzte Aufgaben

Die Aufgaben, die für die iKM^{PLUS} pilotiert wurden, waren entweder kurz und für sich stehend oder in einem sogenannten Testlet mit anderen Aufgaben verbunden. Als Testlet bezeichnet man eine Gruppe von Aufgaben, die auf einen gemeinsamen Text oder multimodalen Stimulus bezogen sind (Pospeschill, 2022, S. 45). Der Einsatz von Testlets zielt auf das Desiderat, das Konzept des integrierten Grammatikunterrichts in der Kompetenzmessung abzubilden (vgl. Krelle, 2016, S. 217).

Ein zentrales Anliegen der Aufgabenkonstruktion bestand darin, die Kompetenzfacetten des prozeduralen Wissens (im Sinne nicht verbalisierbarer Fähigkeiten) und des metakognitiv-strategischen Wissens (Wenn-dann-Szenarien zu Überprüfungsprozeduren) gegenüber auf deklaratives Wissen abzielenden Anteilen aufzuwerten.⁴ Mit prozeduralem Wissen sei hier im Sinne von Squire & Dede (2015, S. 2) nicht ein Wissen *über* Abläufe bzw. Prozeduren gemeint, sondern die Kodierungsform gänzlich impliziter, hochgradig automatisierter Handlungsschemata, die nur ausgeführt, also demonstriert, aber nicht verbalisiert werden können.⁵

Die im Folgenden vorgestellten Aufgaben wurden im Anschluss an die Pilotierung aufgrund verschiedener Mängel im Validierungsprozess ausgeschlossen. Die Gründe hierfür werden im Verlauf

⁴ Für die Definition der Wissensbegriffe folgen wir der Konzeptualisierung der Verhältnisse zwischen prozeduralem, deklarativem und metakognitivem Wissen nach Unterholzner und Müller (2023) sowie der Unterscheidung von Aufmerksamkeit und Bewusstheit nach Müller und Unterholzner (2022).

⁵ Im Gegenzug wird episodisches Wissen neben semantischem Wissen als zweite Art deklarativen Wissens – in diesem Fall über Abläufe – verstanden. Das kann den Ablauf der Bedienung einer Kaffeemaschine ebenso betreffen wie metakognitiv-strategische Vorstellungen über den Schritt-für-Schritt-Ablauf grammatischer Proben (Squire & Dede, 2015, S. 3).

der Aufgabenanalysen erläutert. Alle Aufgaben haben ein Multiple-Choice-Format mit jeweils vier Antwortoptionen, von denen genau eine korrekt ist. Die korrekten Antwortoptionen werden in der Folge als Attraktoren, die inkorrekten als Distraktoren bezeichnet.

Die besondere Herausforderung bei der Formulierung von Multiple-Choice-Aufgaben besteht darin, dass der Attraktor zweifelsfrei richtig und alle Distraktoren zweifelsfrei falsch sein müssen. Im Idealfall sollten sich die Antwortoptionen gegenseitig ausschließen, in etwa ähnlich lang und möglichst ähnlich formuliert sein (Pospeschill, 2022, S. 49–50). Grundsätzlich ergibt sich daraus die Schwierigkeit, drei plausible sowie disjunkte Distraktoren zu konstruieren, die einerseits ausreichend ökologisch valide, andererseits aber trotzdem eindeutig falsch sind, ohne aber wiederum „Test-Wiseness“ begünstigende Indikatoren zu enthalten (Millman et al., 1965, S. 707). Auf Seiten der Schüler:innen bedeutet dies, dass sie im Fall aller vier Antwortoptionen das Gefühl haben sollten, diese könnten auch tatsächlich mit dem durch die Aufgabe gestellten Problem zu tun haben, und keine auf Basis konstruktirrelevanter Signale entstehende dominante Intuition zur Aufgabenlösung entwickeln. So fungieren neben sprachlichen Hinweisen wie „immer“ oder „nie“ (Pospeschill, 2022, S. 65) auch – oft nur implizit wahrgenommene – Alleinstellungsmerkmale einzelner Antwortoptionen als Hinweisreize, sich für die betreffende Antwortoption zu entscheiden. Das könnte beispielsweise dann der Fall sein, wenn drei von vier Antwortoptionen bei einer Wortartentscheidung auf semantische Unterscheidungskriterien anspielen, die vierte Antwortoption aber auf ein syntaktisches Kriterium verweist. Erschwerend kommt hinzu, dass die vier Antwortoptionen für Schüler:innen auch einfach formuliert sein sollen, was durch unterschiedlich komplexe Zusammenhänge dahinter ebenfalls erschwert wird und wiederum zur Vermeidung eindeutiger sprachlicher Indikatoren im Widerspruch steht. Diese Gründe begünstigen die Konstruktion einfacher, auf deklaratives Wissen zielender Aufgaben wie „Um welche Wortart handelt es sich bei dem markierten Wort?“ mit den Antwortmöglichkeiten „Nomen/Verb/Adjektiv/Pronomen“. Nicht zuletzt deshalb stellt die Konstruktion von Aufgaben, die auf Kompetenz (und nicht ausschließlich auf deklaratives Wissen) abzielen, eine große Herausforderung dar.

3.3 Stichprobe

Die vorgestellten Aufgaben stammen aus zwei verschiedenen Pilotierungsstudien: die ersten beiden Aufgaben aus der Pilotierung von 117 Aufgaben für die Sekundarstufe I (Januar 2024, 7. und 8. Jahrgangsstufe, Online-Erhebung), die dritte Aufgabe aus der Pilotierung von 145 Aufgaben für die Primarstufe (Frühjahr 2023, Ende 3. Jahrgangsstufe, Paper-Pencil-Erhebung). Jede Schülerin und jeder Schüler bekam jeweils einen Online-Test bzw. ein Testheft mit 32 oder 33 Aufgaben zur Bearbeitung. Weil es jeweils mehrere Testhefte gab, in denen unterschiedliche Kombinationen von Aufgaben enthalten waren, variieren die Fallzahlen für die Bearbeitung der einzelnen Aufgaben.

Die bereinigte Stichprobe in der Sekundarstufe umfasste insgesamt $n = 561$ Schüler:innen der 7. Jahrgangsstufe und $n = 261$ in der 8. Jahrgangsstufe ($n_{7+8} = 822$). Die einzelnen Aufgaben wurden von mindestens 133 und maximal 428 Schüler:innen bearbeitet. Die Erhebungen wurden an verschiedenen Schulen (30 Schulen im Fall der 7., 15 Schulen im Fall der 8. Jahrgangsstufe) der Schultypen Mittelschule und Gymnasium durchgeführt. Die Schulen lagen teils im ländlichen, teils im städtischen Raum; dabei wurden alle neun Bundesländer Österreichs abgedeckt.

Die bereinigte Stichprobe in der Primarstufe umfasst insgesamt $n = 623$ Schüler:innen aus 33 verschiedenen Schulen. Die einzelnen Aufgaben wurden von mindestens 119 und maximal 497

Schüler:innen bearbeitet. Durchgeführt wurde diese Pilotierung in der 3. Jahrgangsstufe der österreichischen Volksschule; auch hier wurde auf eine Abdeckung des ländlichen und des städtischen Raums sowie aller Bundesländer geachtet.

3.4 Durchführung und Auswertung

Die Durchführung der beiden Testungen verlief standardisiert und die Auswertung (teil)automatisiert. Sie erfolgte anhand mehrerer deskriptiv-statistischer Kennwerte, von denen die für die vorliegende Studie relevanten an dieser Stelle kurz erläutert seien.

Für die Güte ganzer Aufgaben wurden der Schwierigkeitswert p und die Trennschärfe einer Aufgabe r_{pb} betrachtet. Der Schwierigkeitswert p ist dabei ein Dezimalwert, der als Anteil der korrekt gelösten Aufgaben gelesen werden kann. Ein Wert von $p = 0.3$ bedeutet folglich, dass 30 % der Schüler:innen die Aufgabe korrekt lösen konnten. Der Wert für die Trennschärfe gibt die Korrelation zwischen der Beantwortung einer einzelnen Aufgabe und den erreichten Punkten in der gesamten Erhebung wieder. Damit ist die Trennschärfe ein Maß dafür, wie gut eine einzelne Aufgabe zwischen Schüler:innen mit hoher bzw. niedriger Performanz differenzieren kann. Für die Interpretation des Werts kann gelten: Je näher der Wert an +1 liegt, desto besser stimmt die Aufgabe mit dem im Gesamttest erhobenen Leistungseindruck überein – die einzelne Aufgabe differenziert auf ähnliche Weise wie der Gesamttest (Moosbrugger & Kelava, 2012, S. 86). Es dürfen grundsätzlich folgende Grenzwerte als Interpretationshilfe angenommen werden: stark: $r_{pb} > 0.5$; mittel: $r_{pb} > 0.3$ und schwach: $r_{pb} > 0.1$ (Döring, 2022, S. 473). Ein negativer Wert – also eine negative Korrelation zwischen Aufgabe und gesamter Testung – verweist darauf, dass Schüler:innen mit einer insgesamt schwachen Gesamtleistung die Aufgabe tendenziell besser lösen als Schüler:innen mit einer guten Gesamtleistung.

Ein Beispiel: Für eine Aufgabe mit $p = 0.3$ und $r_{pb} = 0.61$ gilt, dass es sich um eine eher schwierige Aufgabe handelt, weil es nur 30 % der Schüler:innen gelungen ist, sie zu lösen. Es gibt aber einen stark ausgeprägten positiven Zusammenhang mit dem Gesamtergebnis, sodass man die Aufgabe als gut geeignet für den Einsatz in Bezug auf das jeweils dahinterstehende Kompetenzmodell einordnen kann (für die Primarstufe vgl. Unterholzner & Illetschko, 2025a, S. 1–3; für die Sekundarstufe Unterholzner & Illetschko, 2025b, S. 1–4).

Für weitere Analysen wird zudem die Antwortverteilung $p_A, p_B \dots$ der einzelnen Antwortoptionen sowie die Trennschärfe für einzelne Antwortoptionen r_{pbA}, r_{pbB} usw. betrachtet. Dabei gilt: Attraktoren sollten eine positive Trennschärfe aufweisen, wohingegen die Trennschärfe für Distraktoren negativ sein sollte. Sind in den Analysen für Kennwerte zusätzliche Indexzahlen angegeben, so geben diese die Jahrgangsstufe an, für die der Wert gilt. Ein Beispiel: r_{pbD8} steht für die Trennschärfe von Antwortoption D für die 8. Jahrgangsstufe.

4 Analysen

In der Folge werden drei Aufgaben vorgestellt, die den drei Thematiken Großschreibung des Wortes im rechten Nominalfeld, Kommasetzung und valenzorientierter Satzbegriff sowie Wortarterkennen im topologischen Feldermodell zuzuordnen sind. Zunächst wird jeweils die Aufgabenkonstruktion beschrieben, dann werden die Ergebnisse vorgestellt und in einem dritten Schritt erfolgt eine knappe Diskussion.

4.1 Aufgabe „Kehren“: Strategien zur Entscheidung über satzinterne Groß- und Kleinschreibung

Vorstellung der Aufgabe

Die Aufgabe „Kehren“ wurde als Einzelaufgabe in der Pilotierung der Sekundarstufe (13- und 14-jährige Schüler:innen) gestellt und zielt auf das Erkennen eines durch syntaktische Konversion verursachten nominalisierten Gebrauchs von Verben ab (Müller, 2024, S. 72). Hier muss keine Rechtschreibentscheidung im eigentlichen Sinne getroffen werden. Vielmehr wird danach gefragt, welche Strategie zielführend ist, um eine Rechtschreibentscheidung zu treffen, weshalb ein Einsatz im Kompetenzbereich Sprachbetrachtung gerechtfertigt ist. Die Aufgabe ist metasprachlich, weil hier Wissen über sprachstrukturelle Zusammenhänge nützt, um entscheiden zu können, welche Strategie zielführend ist; und sie ist metakognitiv, weil sie eine Situation simuliert, in der man zum Zweck einer orthographischen Entscheidung aus einer sprachlichen Handlung aussteigt, um strategisch das eigene Handeln zu überdenken und anzupassen.

NACH DEM ESSEN KEHRE ICH UNTER DEM TISCH DIE ESSENSRESTE ZUSAMMEN. MIR IST DAS KEHREN MIT DEM BESEN VIEL LIEBER ALS DAS STAUBSAUGEN.

Schreibt man **KEHREN** in den beiden Sätzen klein oder groß? Was ist nützlich, um das herauszufinden?

- ☐ Rückwärts lesen: Den Satz von hinten nach vorne lesen.
- ☐ Wörter zählen: Das wievielte Wort im Satz ist **KEHREN**?
- ☐ Ergänzen: Kann man *SCHNELLE* zu **KEHREN** hinzufügen?
- ☐ Nachschlagen: Im Wörterbuch unter **KEHREN** nachlesen.

Abb. 1: Aufgabe „Kehren“

Der Attraktor C zielt auf die Erweiterungsprobe ab, wonach das funktionierende Einsetzen eines adjektivischen Attributs ins Mittelfeld der Nominalphrase, das in Kongruenz gebracht werden muss, einen Hinweis auf die Großschreibung darstellt. Demnach ist die nicht funktionierende Einsetzung **Nach dem Essen schnelle Kehre ich ...* o. Ä. ein Hinweis darauf, dass es sich hierbei nicht um eine Nominalisierung handelt, während *Mir ist das schnelle Kehren ...* die Großschreibung nahelegt.

Der Distraktor A stellt den Versuch dar, einen Ratschlag abzubilden, der im Deutschunterricht als Revisionsstrategie mitunter gegeben wird. Es kann nicht festgestellt werden, ob dieser Ratschlag auch in Zusammenhang mit satzinterner Großschreibung unterrichtlich auftritt. Denkbar ist allerdings, dass Schüler:innen diesen Zusammenhang herstellen. Mit B ist ein Distraktor gegeben, der Schüler:innen vermutlich selten bis gar nicht im Zusammenhang mit Fragen der Großschreibung begegnet und die Aufgabenschwierigkeit demnach etwas absenken sollte. Distraktor D stellt wohl eine Strategie dar, die Schüler:innen bei Rechtschreibfragen (aber nicht bei Fragen der Nominalisierung) häufig empfohlen wird. Es ist tendenziell problematisch, dass die Antwortoptionen A bis C auf Handlungswissen abzielen, das direkt mit sprachbezogenen Prozeduren assoziiert ist, während D auf eine „Metaprozedur“ (Wörterbuchnutzung) anspielt. Verschiebungen im Antwortverhalten hin zu Distrak-

tor D aufgrund seiner (explizit oder implizit) wahrgenommenen Andersartigkeit können also nicht ausgeschlossen werden.

Da im Fall aller drei in den Distraktoren verarbeiteten Strategien die Syntaktik eines Satzes keine Rolle spielt, steht eine inkorrekte Antwort auch für die Nichtaktivierung eines Bewusstseins für die Relevanz der syntaktischen Zusammenhänge im Kontext der Aufgabe.

Ausgeklammert für die Konstruktion der weiteren Antwortoptionen wurde der weit verbreitete, aber nicht unproblematische Artikeltest (Fuhrhop, 2021, S. 42; Rautenberg & Wahl, 2019, S. 84). Da eine Antwortoption mit Artikeltest gleichwohl nicht als zur Gänze falsch beurteilt werden kann, war es nicht möglich, den Artikeltest in einem Distraktor neben der Erweiterungsprobe als Inhalt des Attraktors zu verwenden.

Ergebnisse

Die vorliegende Aufgabe ($n_7 = 290$; $n_8 = 261$) wurde von über 30 % der Schüler:innen korrekt mit der Antwortoption C, in deren Kern die Erweiterungsprobe steht, beantwortet ($p_7 = .36$; $p_8 = .31$), über 40 % der Schüler:innen wählten jeweils den Distraktor D; die Distraktoren A und B wurden sehr selten gewählt (vgl. Tab. 2), was wahrscheinlich darauf zurückzuführen ist, dass diese Antwortoptionen wenig realistisch und somit wenig ökologisch valide sind. Diese Verteilung auf die Antwortoptionen ist testtheoretisch problematisch, weil ein Distraktor, der zu selten ausgewählt wird, der also auch für Personen ohne entsprechendes Wissen direkt auszuschließen wäre, für eine ‚versteckte‘ Formatänderung sorgt: Aus dem ursprünglichen 3-zu-1-Verhältnis von Distraktoren und Attraktor wird dann ein 2-zu-1-Verhältnis (oder ausgeglichener) und damit auch eine Verschiebung der Ratewahrscheinlichkeit auf einen höheren Wert als .25. Die Trennschärfe der gesamten Aufgabe liegt wohl auch deshalb bei niedrigen .20 (r_{pb7}) bzw. .28 (r_{pb8}).

7. Jahrgangsstufe			8. Jahrgangsstufe		
<i>n</i>	<i>P</i>	<i>r_{pb}</i>	<i>n</i>	<i>p</i>	<i>r_{pb}</i>
290	.36	.20	261	.31	.28
AO	<i>P</i>	<i>r_{pb}</i>	AO	<i>p</i>	<i>r_{pb}</i>
A	.06	−.03	A	.07	.01
B	.14	−.20	B	.17	−.44
C (korrekt)	.36	.20	C (korrekt)	.31	.38
D	.42	.01	D	.44	−.08

Tab. 2: Deskriptive Statistik der Aufgabe „Kehren“

Diskussion

Für den Einsatz im Rahmen der iKM^{PLUS} ist diese Aufgabe nach der Pilotierung vor allem deswegen als ungeeignet bewertet worden, weil die Distraktoren A und B zu wenig plausibel waren und zu selten gewählt wurden. Damit in Zusammenhang steht auch der geringe Trennschärfewert von Distraktor A und die insgesamt niedrige Trennschärfe der gesamten Aufgabe ($r_{pb7} = .20$; $r_{pb8} = .28$). Neben dieser Problematik muss zudem festgestellt werden, dass die gewählte Frageformulierung „Schreibt man

KEHREN in den beiden Sätzen klein oder groß?“ für manche Schüler:innen möglicherweise irreführend war, weil man sie so verstehen könnte, als müsste das in Frage stehende Wort in beiden Fällen auf dieselbe Weise verschriftet werden. Trotzdem lohnt es sich, das Antwortverhalten der Proband:innen auf Vertrautheit mit der Erweiterungsprobe bzw. Spuren syntaktischer Bewusstheit zu prüfen.

Attraktor C wurde im Verhältnis zu A und B mit $p_7 = .36$ bzw. $p_8 = .31$ deutlich häufiger gewählt und weist auch eine erwartbare schwache ($r_{pbC7} = .20$) bis mittlere ($r_{pbC8} = .38$) positive Korrelation mit einem guten Gesamtergebnis auf. Allerdings wurde der Wörterbuch-Distraktor D von jeweils über 40 % der Schüler:innen, also noch häufiger, gewählt (bei geringen bzw. negativen Trennschärfen von $r_{pbD7} = .01$, $r_{pbD8} = -.08$). Dass es im Fall von Nominalisierungen nicht sinnvoll ist, einzelne Wörter im Wörterbuch nachzuschlagen, scheint für diese Schüler:innen weniger die Entscheidungsgrundlage darzustellen als vielleicht die Verlockung, mit der Antwortoption Wörterbuch eine Strategie zu wählen, die ihnen in der Schule mutmaßlich oft begegnet. Möglicherweise spielt hier auch eine Rolle, dass sich D insofern von den drei anderen Antwortoptionen alleinstellend unterscheidet, als es eine „Metaprozedur“ darstellt. Jedenfalls zeigen Schüler:innen, die sich für A, B oder D entschieden haben, im Rahmen der Testung kein Bewusstsein für die syntaktischen Bedingungen der satzinternen Großschreibung. Hinzu kommt, dass freilich auch die Wahl von Attraktor C nicht zwingend durch syntaktische Bewusstheit oder Bekanntheit mit der Erweiterungsprobe verursacht sein muss, sondern auch von anderen Gründen wie wenig plausiblen Distraktoren (A und B) verursacht sein kann. Insgesamt lässt dies einerseits den Rückschluss zu, dass die syntaktische Determiniertheit der Großschreibung den meisten Proband:innen nicht bewusst war; andererseits veranschaulicht diese Aufgabe die vielen Fallen, die im Prozess der Aufgabenkonstruktion lauern.

4.2 Aufgabe „Vorfeld“: Kommasetzung zwischen (vermeintlichen) Teilsätzen

Vorstellung der Aufgabe

Die Aufgabe „Vorfeld“ ist eine Multiple-Choice-Aufgabe, die als dritte von insgesamt vier Aufgaben innerhalb eines Testlets gestellt wurde. Der zugehörige achtzeilige Stimulustext sowie die anderen drei Aufgaben erreichten gute Kennwerte und werden deswegen zukünftig in der iKM^{PLUS} eingesetzt. Deshalb kann der Stimulustext an dieser Stelle nicht gezeigt werden. Um den Kontext der Aufgabe trotzdem nachvollziehbar zu machen, soll stattdessen auf ein strukturgleiches, lexikalisch verfremdetes Satzmaterial als Ausgangsbeispiel hingewiesen werden: „... im Dunkel auftauchen. Anders als bei den besonders gut ausgeleuchteten Straßen gibt es für diese Wege noch keine Verbesserungsvorschläge. Insgesamt ist ...“

Wie kannst du herausfinden, ob du nach dem unterstrichenen Teil des Satzes einen Beistrich setzen musst?

Du könntest prüfen, ob...

- ☐ es im unterstrichenen Teil ein Verb gibt.
- ☐ du beim Vorlesen eine Pause machst.
- ☐ ein typisches Signalwort vorkommt.
- ☐ du hier auch einen Punkt setzen könntest.

Abb. 2: Aufgabe „Vorfeld“

Konzeptionell wurden alle vier Aufgaben zum Stimulus unterschiedlich gestaltet. Für die hier gezeigte Aufgabe steht mit der Kommasetzung erneut ein von syntaktischen Zusammenhängen bestimmtes Phänomen im Mittelpunkt – konkret das Erkennen der Anzahl von Teilsätzen an der Anzahl der Verben/Verbgruppen. Auch im Fall dieser zweiten Aufgabe gilt wiederum, dass nicht eigentlich eine Interpunktionsfrage zu lösen war, vielmehr nach einer geeigneten Bearbeitungsstrategie gefragt wurde. Konkret wurde die unterstrichene Textpassage, auf die sich diese Aufgabe bezieht, mit Rücksicht auf das Phänomen des Vorfeldkommata gestaltet (Esslinger & Noack, 2020, S. 16).

Anders als im Fall der ersten Aufgabe ist eine syntaktische Betrachtungsweise in der Didaktik der Kommasetzung wohl weit verbreitet – und zwar in Zusammenhang mit der Vielzahl an vermittelten Regeln und Unterregeln, die mit der Unterscheidung verschiedener Satztypen zusammenhängt. Der valenzorientierte Aspekt von Attraktor A kommt im Wesentlichen dadurch zum Ausdruck, dass darin (ohne genauere Differenzierung) auf die Verbgruppe als dasjenige Element angespielt wird, an dem der Satz sozusagen aufgehängt ist, und eben nicht auf eine elaborierte Satztypenlehre – die „vom Kernproblem ablenkt und die meisten Schüler und Schülerinnen verwirrt und überfordert“ (Lindauer & Sutter, 2005, S. 28).

Distraktor B kann einer womöglich nicht unterrichtlich vermittelten (und bei manchen Schüler:innen auch nur implizit vorhandenen) Intonations-/Pausenstrategie, wonach eine „Segmentierungspause als Grenze zwischen intonatorischen Einheiten“ (Müller, 2007, S. 74) wahrgenommen wird, zugewiesen werden, Distraktor C der Signalwortstrategie (Müller, 2007, S. 76–79). Distraktor D spiegelt den Umstand, dass es oft schwierig ist, neben dem Attraktor drei plausible Zusammenhänge in Distraktoren zu verpacken. Das macht D wiederum tendenziell problematisch. Die Bezeichnung ‚Beistrich‘ statt Komma im Aufgabentext ist den nationalen Gewohnheiten geschuldet.

Ergebnisse

Die vorliegende Aufgabe ($n_7 = 133$; $n_8 = 126$) wurde von weniger als 20 % der Schüler:innen – also mit einem Wert unter der Ratewahrscheinlichkeit von .25 – korrekt beantwortet ($p_7 = .17$; $p_8 = .14$), über 30 % der Schüler:innen wählten jeweils die Distraktoren B und C. Die Trennschärfe der Aufgabe ist in beiden Jahrgangsstufen negativ ($r_{pb7} = -.11$; $r_{pb8} = -.16$), was eine schwache, aber systematische Tendenz anzeigt, wonach Schüler:innen, die insgesamt in der Erhebung besser abschnitten, mit höherer Wahrscheinlichkeit zu einer falschen Antwort tendierten (vgl. Tab. 3). In dieses Bild passt, dass die

Distraktoren B und C (mit leicht positiven Trennschärfewerten) und der Attraktor A (mit leicht negativem Trennschärfewert) jeweils invertiert sind.

7. Jahrgangsstufe			8. Jahrgangsstufe		
<i>n</i>	<i>p</i>	<i>r_{pb}</i>	<i>n</i>	<i>p</i>	<i>r_{pb}</i>
133	.17	-.11	126	.14	-.16
AO	<i>p</i>	<i>r_{pb}</i>	AO	<i>p</i>	<i>r_{pb}</i>
A (korrekt)	.17	-.11	A (korrekt)	.14	-.16
B	.42	.11	B	.31	.04
C	.32	.01	C	.36	.11
D	.10	-.06	D	.19	-.04

Tab. 3: Deskriptive Statistik der Aufgabe „Vorfeld“

Diskussion

Die Distraktoren B und C zur Pausen- und Signalwortstrategie wurden mit Werten zwischen 31 % und 42 % durchwegs am häufigsten gewählt: Schüler:innen der 7. Jahrgangsstufe neigten am häufigsten zur Pausenstrategie B, und diese Tendenz verstärkte sich bei den besseren Schüler:innen noch. Schüler:innen der 8. Jahrgangsstufe tendierten mit ebendieser Tendenz am häufigsten zur Signalwortstrategie C. Beide Distraktoren weisen einen positiven Trennschärfewert auf, während der wenig gewählte Attraktor A einen negativen Trennschärfewert hat. Bei allen drei Antwortoptionen ist die Trennschärfe im Vergleich zu den testtheoretischen Erwartungen also leicht invertiert. Das bedeutet, dass Proband:innen mit besserem Testergebnis eher die falschen Antwortoptionen wählten als die schwächeren, und umgekehrt. Dass die falschen Annahmen bei den lernschwächeren Proband:innen gegen die Erwartungen einen tendenziell weniger starken Überhang haben, könnte ein Hinweis auf die Dysfunktionalität der womöglich unterrichtlich vermittelten Kommastrategien „segmentierende Pausen“ und „Signalwörter“ sein, dem weiter nachzugehen wäre.

Neben der grundsätzlichen psychometrischen Problematik der invertierten Trennschärfe, die zum Ausscheiden der Aufgabe führte, weist dies allerdings auf eine Problematik hin, der Kompetenzmessungen allgemein unterworfen sind: Aufgaben können nicht rein theoretisch-kriterial und unabhängig von der Population validiert werden. In Fällen, in denen ein dem Lerngegenstand aus theoretischer Perspektive angemessenes Entscheidungskriterium für eine Aufgabe den Lernerfahrungen der Schüler:innen weitestgehend widerspricht – zum Beispiel wegen einer potenziell weit verbreiteten unterrichtlichen Vermittlung dysfunktionaler Strategien –, führt die übliche Pilotierungspraxis, die passende Schwierigkeits- und Trennschärfewerte anstrebt, aber zwangsläufig zum Aussortieren einer Aufgabe, auch wenn sie theoretisch-kriterial durchaus konsistent ist. Dieser Punkt ist nicht neu: Bereits Goldstein (1979, S. 216) hat darauf hingewiesen, dass eine derartige Praxis der Itemselektion dazu führt, dass die gemessene Leistung letztlich durch ebendie Items definiert wird, die sich modellkonform verhalten – während gerade die abweichenden Items, deren Schwierigkeit nicht für alle Lernenden gleich ausfällt, inhaltlich von besonderem Interesse sein könnten. Dass auch in der deutschdidaktischen Kompetenzmessung ein Bewusstsein für diese Problematik besteht, zeigt etwa der Verweis von Bremerich-Vos und Böhme (2009, S. 384) auf Goldstein (1979).

Eine invertierte Trennschärfe kann in diesem Sinne also auch als ein Indiz für ein systematisches Problem in der Unterrichtsrealität, das dem Antwortverhalten der Schüler:innen zugrunde liegt, diskutiert werden. Zum Beispiel wäre es denkbar, dass leistungsstarke Schüler:innen unterrichtlich vermittelte dysfunktionale Strategien williger und schneller übernehmen und sie sich daher tendenziell eher für einen damit zusammenhängenden Distraktor entscheiden. Eine Art „Vermittlungsresistenz“ schwächerer Schüler:innen wäre in dem Fall sozusagen ein Schutz gegen die Übernahme einer im Unterricht angebotenen dysfunktionalen Strategie. Dass dieses Szenario den Entstehungshintergrund für die Kennwerte darstellt, soll für die vorliegende Aufgabe gar nicht behauptet werden. Aber diese Ergebnisse stehen zu solch einer Annahme jedenfalls nicht im Widerspruch und werfen eher Fragen grundsätzlicher Art auf.

Ebenfalls zur invertierten Trennschärfe von Attraktoren passend konnte auch Müller eine Dominanz der Pausen- und Signalworthypothese zu Ungunsten eines syntaktischen Zusammenhangs finden (Müller, 2007, S. 241–248). Im Fall der Pausenhypothese hält Müller es allerdings zumindest für fragwürdig, aus diesen Ergebnissen auf das Unterrichtsgeschehen zu schließen: Es sei „[n]icht wahrscheinlich [...], dass die Orientierung an der Intonation auf Instruktion zurückzuführen ist“, so Müller (2007, S. 242). Die Signalworthypothese betreffend kann man mit Bredel (2015a, S. 135) möglicherweise annehmen, das Antwortverhalten im Fall der vorliegenden Aufgabe sei unterrichtlich bedingt, wofür auch spricht, dass es mit einem Jahr mehr Schulunterricht, also in der 8. Jahrgangsstufe, zu einer Verschiebung hin zur Signalworthypothese kommt.

4.3 Aufgabe „Fischen“: Wortartenkategoriale Unterschiede mit und ohne Feldermodell erkennen

Vorstellung der Aufgabe

Die Primarstufenaufgabe „Fischen“ thematisiert die Kategorisierung von Wortarten. Anders als im Fall der ersten beiden Aufgaben geschieht dies nicht in Form sich daraus ergebender orthographischer Konsequenzen, sondern ausschließlich als für sich stehende sprachbetrachtende Fragestellung. Hier wird also nicht danach gefragt, ob die jeweilige Strategie für die Beantwortung nützlich sein könnte, sondern direkt nach einer Lösung aus dem Bereich der Wortartenkategorisierung. Konzeptionell ist diese Aufgabe an die Odd-one-out-Aufgaben von Funke (2005, S. 189–192) angelehnt. Das heißt, es geht darin um das Erkennen einer unterschiedlichen Kategorienzuweisung⁶ der (bei Großschreibung) äußerlich identischen Wörter *fischen* (Verb) und *Fischen* (Nomen) – und zwar ausschließlich in Form der basalen Frage danach, in welchem der vier Beispiele es sich bei FISCHEN um eine von den anderen Beispielen abweichende Wortart handelt, ohne genauer nach der Bestimmung der Wortart zu fragen.

⁶ Die im Aufgabenstamm verwendete Formulierung „gehört zu einer anderen Wortart“ sollte die Problematik abschwächen, dass Proband:innen, die unterrichtlich an eine lexikalisch fundierte Wortartenkategorisierung (eine lexikalisch an sich festgelegte Wortart „wird als eine andere Wortart verwendet“), und Proband:innen, die an eine syntaktische Perspektive („ist eine andere Wortart“ im gegebenen syntaktischen Kontext) gewöhnt sind, diese Aufgabe als unterschiedlich schlüssig empfinden.

In welchem Satz gehört FISCHEN zu einer anderen Wortart als in den anderen Sätzen?

☐	Viele von ihnen	FISCHEN	derzeit noch im Trüben.		
☐	Die meisten Angler	FISCHEN	gerne in Fließgewässern.		
☐	Diese zwei Firmen	FISCHEN	nach Daten von Personen.		
☐	Manche	können	mit FISCHEN nichts anfangen.		

Abb. 3: Aufgabe „Fischen“, Variante mit Feldermodell

In welchem Satz gehört FISCHEN zu einer anderen Wortart als in den anderen Sätzen?

☐ Viele von ihnen FISCHEN derzeit noch im Trüben.

☐ Die meisten Angler FISCHEN gerne in Fließgewässern.

☐ Diese zwei Firmen FISCHEN nach Daten von Personen.

☐ Manche können mit FISCHEN nichts anfangen.

Abb. 4: Aufgabe „Fischen“, Variante ohne Feldermodell

Diese und weitere hier nicht vorgestellte Aufgaben wurden in der Testung in einer quasi-experimentellen Parallelführung je einmal mit (Abb. 3) und einmal ohne unterstützende Visualisierung (Abb. 4) durch ein topologisches Feldermodell ohne Überschriften angeboten. Diese Entscheidung folgte dem Erkenntnisinteresse daran, ob die graphische Strukturierung durch das Feldermodell unterstützend wirke (vgl. Uhl, 2019, S. 361; Rödel, 2023). Die leitende Annahme war hier, dass die Visualisierung nicht nur den Schüler:innen, die die Strukturanalyse mit dem Feldermodell bereits im Unterricht kennengelernt haben, eine Entscheidungshilfe bietet, sondern es auch den nicht mit dem Feldermodell vertrauten Schüler:innen ermöglicht, die Rolle des erfragten Wortes dadurch schneller zu erfassen. Es wurde darauf geachtet, Parallelversionen mit und ohne Feldermodell jeweils nicht im selben Testheft unterzubringen.

Für alle Feldermodell-Versionen war bereits im Vorfeld der Pilotierung entschieden worden, sie nur in der Pilotierungsstudie einzusetzen, nicht aber in den Echteinsatz zu bringen. Dies geschah einerseits aus Gründen der ökologischen Validität vor dem Hintergrund eines Unterrichtsgeschehens, in dem das Feldermodell mutmaßlich wenig verbreitet ist, und andererseits, weil es testtheoretisch problematisch ist, wenn die Visualisierungshilfe das sprachliche Material so vorstrukturiert, dass diese optische Strukturierung die korrekte Antwortoption durch ihre Andersartigkeit als möglichen Attraktor ausweist. Deshalb fiel auch die Entscheidung dahingehend, die Feldermodell-Tabelle nicht mit den Überschriften „Vorfeld“, „linkes Verbfeld“ (oder „linke Satzklammer“), „Mittelfeld“, „rechtes Verbfeld“ (oder „rechte Satzklammer“) und „Nachfeld“ zu versehen, da insbesondere die Verbfeld-

Bezeichnungen testtheoretisch unsauber einen Teil der Lösung mitgeliefert hätten. Dies reduzierte zwar die Zirkularitäts-Problematik, die durch die Integration der didaktischen Vermittlung / des Lernprozesses (durch das Feldermodell repräsentiert) in die Leistungsfeststellung entstand. Trotzdem blieb dieses Problem in Form der graphischen Vorstrukturierung inhärent.

Ergebnisse

Die Aufgabe „Fischen“ ($n_3 = 123$ bzw. 122) wurde in der Version ohne Feldermodell von einem Viertel ($p = .25$), in der Version mit Feldermodell von einem Drittel ($p = .33$) der Schüler:innen korrekt beantwortet. Die Trennschärfe der Version ohne Feldermodell liegt bei $r_{pb} = .31$, diejenige der Version mit Feldermodell bei $r_{pb} = .35$.

„Fischen“ (3. Jahrgangsstufe)					
ohne Feldermodell			mit Feldermodell		
<i>n</i>	<i>p</i>	<i>r_{pb}</i>	<i>n</i>	<i>p</i>	<i>r_{pb}</i>
122	.25	.31	123	.33	.35
AO	<i>p</i>	<i>r_{pb}</i>	AO	<i>p</i>	<i>r_{pb}</i>
A	.08	-.02	A	.07	-.05
B	.14	-.15	B	.07	-.01
C	.26	.16	C	.10	.05
D (korrekt)	.25	.31	D (korrekt)	.33	.35

Tab. 4: Deskriptive Statistik der Aufgabe „Fischen“, Varianten mit und ohne Feldermodell

Diskussion

Die Version mit Feldermodell erwies sich für die Proband:innen als etwas einfacher. Ihre höheren p -Werte im Fall der korrekten Antwortoption sowie die niedrigeren im Fall der Distraktoren deuten auf die hinweisgebende Funktion der Visualisierung hin. Der im Vergleich zu den Distraktoren A und B höhere Wert bei Distraktor C in der Version ohne Feldermodell ($p_{C\text{-ohne}} = .26$) legt darüber hinaus die Versuchung für die Proband:innen nahe, diese Antwortoption aufgrund der anderen Semantik (*fischen* ist nur hier auf Daten, nicht auf Fische bezogen) als abweichend und damit als angenommen korrekt auszuwählen. Die Visualisierung durch das Feldermodell bewirkt hier eine Verschiebung im Antwortverhalten weg von der (vermutlich teils impliziten) Stützung auf Semantik ($p_{C\text{-ohne}} = .26$ zu $p_{C\text{-mit}} = .10$) hin zur syntaktischen Unterscheidungsvariante ($p_{D\text{-ohne}} = .25$ zu $p_{D\text{-mit}} = .33$).

Möglicherweise ist dies ursächlich zu unbestimmbarem Anteil auf die schiere visuelle Andersartigkeit des Attraktors zurückzuführen. Darüber hinaus könnte in unbestimmtem Ausmaß auch eine Rolle spielen, dass die unterschiedlichen semantischen Interpretationen von *Fischen* im Attraktor D (die Handlung des Fischens gegenüber den Tieren) mit und ohne Visualisierung durch das Feldermodell unterschiedlich gewichtet wahrgenommen werden – was wiederum mit der Wahrnehmung von *Fischen* als nominalisiertem Verb (Tätigkeit) oder als eigentlichem Nomen (Tiere) korrespondieren könnte. Die Werteververschiebung hin zu D in der Version mit Feldermodell stellt für sich also noch keinen hinreichenden Hinweis auf eine durch das Feldermodell beeinflusste Erhöhung der syntaktischen

Einsicht dar. Sie deutet allerdings schon an, dass der visuelle Aufbau des Feldermodells hinweisgebend kongruent mit den syntaktischen Zusammenhängen ist, die es visualisieren möchte. Das mag ein Indiz für den Wert des Feldermodells aus der Perspektive der Vermittlung sein, aus der Perspektive der Testung besteht aber das Problem der Zirkularität. So etwa, wenn man mit einem Visualisierungsansatz wie dem Feldermodell eine Leistung testen möchte, aber damit gleichzeitig ein instruierendes/unterstützendes Element in die Testung eingebaut hat, was zu dem zirkulären Effekt führt, dass ebendie Tendenzen in den Ergebnissen gestärkt werden, die durch das instruierende Element eingebracht werden. Anders gesagt: Es kommt heraus, was man selbst hineingesteckt hat. Ein Fazit daraus ist, dass solch eine quasiexperimentelle vergleichende Messung möglicherweise ein schwacher Hinweis auf die geeignete Darstellungsform der Visualisierung ist, letztlich aber nicht in der Kompetenzmessung eingesetzt werden kann und selbst für die Gewinnung von Zusatz Erkenntnissen nur eingeschränkten Wert hat.

Auch beide Aufgabenversionen ohne Feldermodell wurden nach der Pilotierung wegen der ungenügenden Trennschärfewerte als ungeeignet für den Einsatz klassifiziert; alle Aufgabenversionen mit Feldermodell waren, wie oben begründet, ohnehin nie für den Echteininsatz vorgesehen. Ein Schluss aus dem Antwortverhalten der Schüler:innen ist jedenfalls, dass Odd-one-out-Aufgaben, die auf syntaktische Unterschiede zielen, gut daraufhin kontrolliert werden sollten, ob allfällige andere Unterschiede (z. B. semantische) stabil gehalten werden.

5 Zusammenfassende Betrachtungen und Ausblick

In diesem Beitrag wurden drei Aufgaben des Kompetenzbereichs *Einsicht in Sprache durch Sprachbetrachtung* (Primarstufe) bzw. *Sprachbewusstsein* (Sekundarstufe) vorgestellt, die im Rahmen einer Pilotierungsstudie für die iKM^{PLUS} von Schüler:innen aus Österreich bearbeitet wurden. Alle Aufgaben eint, dass sie mit dem Ziel entwickelt wurden, rezente didaktische Modellierungen zu integrieren: Der Attraktor der Aufgabe „Kehren“ zur satzinternen Großschreibung nominalisierter Verben zielte auf die Struktur der Nominalphrase und die damit zusammenhängende Erweiterungsprobe durch ein ins Mittelfeld der Nominalphrase einzusetzendes, in Kongruenz zu bringendes Adjektiv ab. Die Idee hinter dem Attraktor der Komma-Aufgabe „Vorfeld“ war ein in den Grundzügen valenztheoretisch fundiertes, bewusst reduziert gehaltenes Verständnis von ‚(Teil-)Satz‘, wonach am Vorliegen eines Verbs oder einer Verbgruppe erkannt werden kann, ob ein solcher vorliegt. In der Aufgabe „Fischen“ wurde in zwei Varianten untersucht, ob sich das Antwortverhalten von Schüler:innen in Bezug auf die Wortartenkategorisierung verändert, wenn zu bearbeitende Sätze in eine Tabelle nach Form des topologischen Feldermodells eingebettet sind.

An den Darstellungen der Überlegungen zur Aufgabenkonstruktion einerseits und der Diskussion des Antwortverhaltens der Schüler:innen auf Basis der psychometrischen Kennwerte andererseits wurde das Wechselspiel von Konstruktionsüberlegungen und Ergebnisinterpretation ersichtlich. Es wurden die Herausforderungen nachvollziehbar gemacht, die mit der Konstruktion kompetenzorientierter Sprachbetrachtungsaufgaben verbunden sind. Während Testaufgaben, die auf die reine Reproduktion kategorial-terminologischen Fachwissens zielen, verhältnismäßig einfach zu konstruieren sind, ist es deutlich schwieriger, andere Kompetenzfacetten zu testen, etwa metakognitives strategisches Wissen im Sinne von: „Immer, wenn ich x im sprachlichen Material beobachte, hilft mir die Wahl von Strategie y“ in Verbindung mit Überprüfungsprozeduren.

Das Antwortverhalten der Schüler:innen liefert insgesamt nur wenige Hinweise auf eine Einsicht in die syntaktische Determiniertheit der besprochenen Phänomene. Über die Gründe dafür lässt sich aus den vorliegenden Daten nichts ableiten. Jedenfalls steht die Antwortverteilung nicht im Widerspruch zu zwei bekannten Forderungen: erstens einer schriftstrukturell dem Lerngegenstand angemessenen Modellierung und zweitens einer Engführung von Orthographie- und Grammatikunterricht mit Augenmerk auf die kognitive Durchdringung sprachstruktureller Zusammenhänge. Es bleibt der Eindruck, dass der für Sprachbewusstheit zentrale Prozess der Distanzierung von der überwiegend intuitiv verarbeiteten Semantik und Pragmatik zugunsten einer bewussten Auseinandersetzung mit syntaktischen Strukturen alles andere als einfach ist.

Darüber hinaus wurden in diesem Beitrag auch Probleme in der Aufgabenkonstruktion diskutiert, die letztlich zur Nichtvalidierung der Aufgaben beitrugen. Es wurde gezeigt, wie sich die Aufgabenkonstruktion hier in einem Spannungsfeld zwischen kriterialer Angemessenheit einerseits und Testfairness in Abhängigkeit vom anzunehmenden Unterrichtsgeschehen andererseits bewegt. Gerade vor dem Hintergrund der deutschdidaktischen Diskussion über die mögliche Dysfunktionalität verbreiteter schulgrammatischer Modellierungen erscheint es lohnend, psychometrische Kennwerte einzelner Antwortoptionen – insbesondere Schwierigkeitswerte und (invertierte) Trennschärfen – systematisch daraufhin zu untersuchen, inwieweit sie sich für die explorative Identifikation unterrichtlich vermittelter Misskonzeptionen nutzen lassen. Anknüpfen ließe sich dabei an die distraktor-basierte Diagnostik, wie sie Sadler (1998) für naturwissenschaftliche Concept Inventories beschreibt. Übertragen auf den Bereich der Sprachbetrachtung könnten Distraktoren gezielt so konstruiert werden, dass sie konkurrierende sprachbezogene Modellierungen oder typische schulisch erworbene Misskonzeptionen repräsentieren. Die Analyse ihres Antwortverhaltens würde damit nicht nur Aussagen über die Aufgabenqualität, sondern zugleich über mögliche fachliche Denkweisen der Schülerinnen und Schüler erlauben. Auf diese Weise ließe sich die Kompetenzmessung stärker mit sprachdidaktischer Grundlagenforschung verbinden.

Literatur

- Bangel, M. (2022). Potentiale einer syntaxbasierten Vermittlung der satzinternen Großschreibung und mögliche vermittlungsunabhängige Einflussfaktoren. In H. Hlebec & S. Sahel (Hrsg.), *Orthographieerwerb im Übergang* (S. 119–145). Erich Schmidt Verlag.
<https://link.springer.com/book/10.37307/b.978-3-503-20650-6>
- Bangel, M., Rautenberg, I., & Weth, C. (2020). Syntaxorientierte Didaktik der Großschreibung: ein Forschungsüberblick. *Didaktik Deutsch*, 25(48), 55–70.
- Betzel, D. (2015). *Zum weiterführenden Erwerb der satzinternen Großschreibung. Eine leistungsgruppendifferenzierte Längsschnittstudie in der Sekundarstufe I*. Schneider Verlag Hohengehren.
- Betzel, D. (2024). Von der wortbildorientierten zur schriftlinguistisch fundierten Rechtschreibdidaktik. *Der Deutschunterricht*, 76(4), 26–35.
- Betzel, D., & Kleinbub, I. (2025). Das topologische Modell im Sprachunterricht – eine Interviewstudie mit Sekundarstufenlehrkräften. *Research on Language*, 2(1), 34–55.
<https://doi.org/10.26204/KLUEDO/9043>
- Böhler, M., & Cristante, V. (2024). *Grammatik in der Grundschule. Unterrichtsentwürfe für sprachheterogene Klassen*, 1–4. Klett Kallmeyer.

- Bredel, U. (2010). Die satzinterne Großschreibung: System und Erwerb. In U. Bredel, A. Müller & G. Hinney (Hrsg.), *Schriftsystem und Schrifterwerb: linguistisch – didaktisch – empirisch* (S. 217–243). De Gruyter.
- Bredel, U. (2013). *Sprachbetrachtung und Grammatikunterricht* (2., durchges. Aufl.). Schöningh.
- Bredel, U. (2015a). Das Satzgrenzenkomma und seine Didaktisierung – das Verb als Zentrum kommarelevanter Strukturen. In B. Mesch & B. Rothstein (Hrsg.), *Was tun mit dem Verb? Über die Möglichkeit und Notwendigkeit einer didaktischen Neuerschließung des Verbs* (S. 135–149). De Gruyter.
- Bredel, U. (2015b). Topologie und Orthographie. In A. Wöllstein & S. Schmadel (Hrsg.), *Das topologische Modell für die Schule* (S. 205–217). Schneider Verlag Hohengehren.
- Bredel, U. (2021). Grammatikunterricht ohne Grammatik – Fehlkonstruktionen der Schulgrammatik am Beispiel der Satzgliedbestimmung und die Folgen für die Lehrerbildung. In S. Döring & D. Elsner (Hrsg.), *Grammatik für die Schule* (S. 71–84). Buske.
- Bredel, U. (2022). Sprachreflexion und Orthographieunterricht. In H. Gornik & I. Rautenberg (Hrsg.), *Sprachreflexion und Grammatikunterricht* (S. 257–272). Schneider bei wbv.
- Bremerich-Vos, A., & Böhme, K. (2009). Kompetenzdiagnostik im Bereich „Sprache und Sprachgebrauch untersuchen“. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 376–392). Beltz.
- Christ, R. (2017). Niemand braucht das Prädikat. Zur Systematisierung der schulischen Satzgliedanalyse. *Linguistische Berichte*, 49(250), 169–217.
- Christ, R. (2019). Ohne ‚Prädikat‘, mit Topologie: Induktive Satzgliedanalyse in Klasse 8. In A. Betz & A. Firstein (Hrsg.), *Schülerinnen und Schülern Linguistik näher bringen. Perspektiven einer propädeutischen Wissenschaftsdidaktik* (S. 123–147). Schneider Verlag Hohengehren.
- Döring, N. (2022). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.
- Esslinger, G., & Noack, C. (2020). *Das Komma und seine Didaktik*. Universitätsbibliothek der Ruhr-Universität Bochum. <https://doi.org/10.46586/SLLD.146>
- Fuhrhop, N. (2021). Das Schriftsystem als Teilgebiet der Grammatik im Deutschunterricht. *ide*, 45(3), 38–47.
- Funke, R. (1995). Das Heben des Wortartschatzes. Nomen im Kontext sehen. *Praxis Deutsch*, 22(129), 57–60.
- Funke, R. (2005). *Sprachliches im Blickfeld des Wissens. Grammatische Kenntnisse von Schülerinnen und Schülern*. Niemeyer.
- Funke, R. (2022). Grammatikunterricht, grammatisches Wissen und schriftsprachliches Können. In H. Gornik & I. Rautenberg (Hrsg.), *Sprachreflexion und Grammatikunterricht* (S. 430–454). Schneider bei wbv.
- Fürstenberg, M. (2023). *Kommagebrauch im Deutschen. Eine empirische Untersuchung zur Kommasetzung beim Schreiben und Einsetzen*. De Gruyter. <https://doi.org/10.1515/9783111322667>
- Gallmann, P. (2015). Das topologische Modell: Basisartikel. In A. Wöllstein & S. Schmadel (Hrsg.), *Das topologische Modell für die Schule* (S. 1–36). Schneider Verlag Hohengehren.
- Goldstein, H. (1979). Consequences of Using the Rasch Model for Educational Assessment. *British Educational Research Journal*, 5(2), 211–220. <https://doi.org/10.1080/0141192790050207>

- Granzow-Emden, M. (2019). *Deutsche Grammatik verstehen und unterrichten. Eine Einführung* (3., überarb. und erw. Aufl.). Narr Francke Attempto.
- Granzow-Emden, M. (2021). Alles wie gehabt? Zum neuen Verzeichnis grundlegender grammatischer Fachausdrücke. *Didaktik Deutsch*, 26(50), 24–29.
- Granzow-Emden, M. (2022). Sprachstrukturen verstehen. Die Entwicklung grammatischer Kategorien. In H. Gornik & I. Rautenberg (Hrsg.), *Sprachreflexion und Grammatikunterricht* (S. 213–241). Schneider bei wbv.
- Günther, H. (2000). „... und hält den Verstand an“ – Eine Etüde zur Entwicklung der deutschen Interpunktion 1522–1961. In R. Thieroff, M. Tamrat, N. Fuhrhop & O. Teuber (Hrsg.), *Deutsche Grammatik in Theorie und Praxis* (S. 275–286). De Gruyter.
<https://doi.org/10.1515/9783110933932.275>
- Hennig, M. (2023). Valenz. *Der Deutschunterricht*, 75(3), 49–59.
- Hübl, A., & Steinbach, M. (2015). Die Topologie der NP als Grundlage für den Erwerb der satzinternen Großschreibung. In A. Wöllstein & S. Schmadel (Hrsg.), *Das topologische Modell für die Schule* (S. 219–238). Schneider Verlag Hohengehren.
- Klages, H., & Pagonis, G. (2025). Wer gibt wem was? *Fremdsprache Deutsch*. Vorab-Onlinepublikation. <https://doi.org/10.37307/j.2194-1823.2025.72.04>
- Kohl-Dietrich, D., & Maiberger, E.-L. (2022). Die Interface-Hypothese: Annahmen zum Zusammenspiel von implizitem und explizitem Wissen im Spracherwerb. In K. Madlener-Charpentier & G. Pagonis (Hrsg.), *Aufmerksamkeitslenkung und Bewusstmachung in der Sprachvermittlung. Kognitive und didaktische Perspektiven auf Deutsch als Erst-, Zweit- und Fremdsprache* (S. 315–339). Narr Francke Attempto.
- Kreiter, C. D., & Frisbie, D. A. (1989). Effectiveness of Multiple True-False Items. *Applied Measurement in Education*, 2(3), 207–216. https://doi.org/10.1207/s15324818ame0203_2
- Krelle, M. (2016). ‚Sprache und Sprachgebrauch untersuchen‘ im Licht von Vergleichsarbeiten und Unterrichtsentwicklung in der Grundschule. In M. Krelle & W. Senn (Hrsg.), *Qualitäten von Deutschunterricht* (S. 207–228). Klett Fillibach.
- Laser, B., & Riegler, S. (2025). Wortarten syntaktisch denken. *Research on Language*, 2(2), 43–60.
<https://doi.org/10.26204/KLUEDO/9600>
- Lindauer, T., & Sutter, E. (2005). Könige, Königreiche und Kommaregeln. Eine praxistaugliche Vereinfachung des Zugangs zur Kommasetzung. *Praxis Deutsch* (191), 28–35.
- Metzger, S. (2017). *Grammatik unterrichten mit dem Feldermodell. Didaktische Grundlagen und Aufgaben für die Orientierungsstufe*. Klett Kallmeyer.
- Millman, J., Bishop, C. H., & Ebel, R. (1965). An Analysis of Test-Wisness. *Educational and Psychological Measurement*, 25(3), 707–726. <https://doi.org/10.1177/001316446502500304>
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2., aktual. und überarb. Auflage). Springer. <https://doi.org/10.1007/978-3-642-20072-4>
- Müller, A. (2024). *Rechtschreiben lernen. Die Schriftkultur entdecken – Grundlagen und Übungsvorschläge* (4. Aufl.). Klett Kallmeyer. <https://elibrary.utb.de/doi/book/10.5555/9783772791116>
- Müller, A. (2025). „Wenn man einen Ansatz einführt, muss man immer bedenken, wie die Situation in der Praxis ist“ – Transferforschung als deutschdidaktische Aufgabe und Herausforderung. *Didaktik Deutsch*, 30(58), 18–32. <https://doi.org/10.21248/dideu.775>

- Müller, H.-G. (2007). *Zum „Komma nach Gefühl“. Implizite und explizite Kommakompetenz von Berliner Schülerinnen und Schülern im Vergleich*. Peter Lang.
- Müller, H.-G., & Unterholzner, F. (2022). Sprachbewusstheit und Sprachaufmerksamkeit. Ein Vorschlag für ein empirisch operationalisierbares Minimalprogramm. *Didaktik Deutsch*, 27 (52/53), 38–55. <https://doi.org/10.21248/dideu.93>
- Noack, C. (2018). Kommasetzung ist schwer, oder? *Deutschunterricht*, 71(1), 4–9.
- Ossner, J. (2007). Sprachbewusstheit: Anregung des inneren Monitors. In H. Willenberg (Hrsg.), *Kompetenzhandbuch für den Deutschunterricht* (S. 134–147). Schneider Verlag Hohengehren.
- Pospeschill, M. (2022). *Testtheorie, Testkonstruktion, Testevaluation* (2., aktualis. Aufl.) Reinhardt.
- Prosch, A. (2016). *Entwicklung von Rechtschreibkompetenz. Differentielle Analysen mit NEPS-Daten der Haupterhebungen in den Klassenstufen fünf und sieben sowie der Entwicklungsstudien in den Klassenstufen sechs und sieben*. Logos.
- Rautenberg, I., & Wahl, S. (2019). Der Einfluss der Nominalgruppenstruktur auf die Groß-/Kleinschreibung. Eine empirische Untersuchung im 2. und 6. Schuljahr. *Didaktik Deutsch*, 24(46), 83–101.
- Rautenberg, I., & Wahl, S. (2024). Welche sprachstrukturellen Faktoren beeinflussen die Großschreibung von Schüler*innen im Deutschen? *Zeitschrift für Sprachlich-Literarisches Lernen und Deutschdidaktik*, 4, 1–27. <https://doi.org/10.46586/SLLD.Z.2024.11376>
- Röber-Siekmeyer, C. (1999). *Ein anderer Weg zur Groß- und Kleinschreibung*. Klett.
- Rödel, L. (2023). Grammatikdidaktische Visualisierungen. *Der Deutschunterricht*, 75(3), 5–14.
- Rothstein, B. (2022). Sich Sätze erklären – das Feldermodell. In H. Gornik & I. Rautenberg (Hrsg.), *Sprachreflexion und Grammatikunterricht* (S. 515–524). Schneider bei wbv.
- Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, 35(3), 265–296. [https://doi.org/10.1002/\(SICI\)1098-2736\(199803\)35:3<265::AID-TEA3>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1098-2736(199803)35:3<265::AID-TEA3>3.0.CO;2-P)
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Squire, L. R., & Dede, A. J. O. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor perspectives in biology*, 7(3), 1–15. <https://doi.org/10.1101/cshperspect.a021667>
- Stauber, A., Stahl, J., Bruneforth, M., & Illetschko, M. (2024). Die individuelle Kompetenzmessung PLUS (iKM^{PLUS}) – Pädagogische Diagnostik im Spannungsfeld von Individualförderung und Systemmonitoring. In C. Schreiner, G. Schauer & C. Kraler (Hrsg.), *Pädagogische Diagnostik und Lehrer:innenbildung: Bildungswissenschaftliche und fachdidaktische Perspektiven* (S. 157–172). Klinkhardt.
- Uhl, B. (2019). Topologie und sprachliches Lernen – vier Argumente für das grammatikdidaktische Potenzial des Stellungsfeldermodells. *Mitteilungen des Deutschen Germanistenverbandes*, 66(4), 360–368. <https://doi.org/10.14220/mdge.2019.66.4.360>
- Unterholzner, F. (2021). *Sprachbetrachtung – metakognitiv aktivierend und altersangepasst. Eine Interventionsstudie zur Vermittlung von das/dass-Schreibung und Kommasetzung zwischen (Teil-) Sätzen*. Universitätsbibliothek der Ruhr-Universität Bochum. <https://doi.org/10.46586/SLLD.197>
- Unterholzner, F., & Illetschko, M. (2025a). *Deutsch (Einsicht in Sprache durch Sprachbetrachtung) in der iKM^{PLUS} im Detail (samt Förderempfehlungen): Primarstufe*. IQS – Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen. <https://www.iqs.gv.at/downloads/nationale-kompetenzerhebung/ikm-plus-volksschule/lehrpersonen/deutsch>

- Unterholzner, F., & Illetschko, M. (2025b). *Deutsch (Sprachbewusstsein) in der iKM^{PLUS} im Detail (samt Förderempfehlungen): Sekundarstufe*. IQS – Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen. <https://www.iqs.gv.at/downloads/nationale-kompetenzerhebung/ikm-plus-sekundarstufe/lehrpersonen/deutsch>
- Unterholzner, F., & Müller, H.-G. (2023). Metakognition als Brücke zwischen sprachlichem Wissen und Können. *Didaktik Deutsch*, 28(55), 20–38. <https://doi.org/10.21248/dideu.677>
- Wöllstein, A., & Schmadel, S. (Hrsg.). (2015). *Das topologische Modell für die Schule*. Schneider Verlag Hohengehren.

Anschrift der Verfasser:innen:

Franz Unterholzner, Pädagogische Hochschule Salzburg Stefan Zweig (IQS bis September 2025),
Akademiestraße 23–25, 5020 Salzburg
franz.unterholzner@phsalzburg.at

Ilka T. Fladung, Universität Duisburg-Essen, Universitätsstraße 2, 45141 Essen
ilka.fladung@uni-due.de

Marcel Illetschko, IQS – Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen,
Alpenstraße 121, 5020 Salzburg
marcel.illetschko@iqs.gv.at