

Lena Heine\*, Madeleine Domenech, Lisa Otto, Astrid Neumann, Michael Krelle, Dominik Leiss, Dietmar Höttecke, Timo Ehmke und Knut Schwippert

## Modellierung sprachlicher Anforderungen in Testaufgaben verschiedener Unterrichtsfächer: Theoretische und empirische Grundlagen

Modelling linguistic demands in assessment tasks for various school subjects: Theoretical and empirical foundations

<https://doi.org/10.1515/zfal-2018-0017>

**Abstract:** Recent years have witnessed a growing interest in the relationship between academic language registers and school success in the German-speaking education system. However, we still know very little about the actual effects that academic language has on the academic performance of students, for instance, in how far the extent to which academic language is used in subject tasks actually makes these tasks more difficult. It is therefore highly vital that any operationalization of difficulty-inducing linguistic features of tasks is made on solid theo-

---

**\*Korrespondenzautor:** Lena Heine, Ruhr-Universität Bochum, Fakultät für Philologie, Germanistisches Institut, 44780 Bochum, E-mail: lena.heine@rub.de

**Madeleine Domenech**, Universität Paderborn, Institut für Germanistik und Vergleichende Literaturwissenschaft, Deutsch als Zweitsprache/Deutsch als Fremdsprache, 33098 Paderborn, E-mail: madeleine.domenech@upb.de

**Lisa Otto**, Ruhr-Universität Bochum, Fakultät für Philologie, Germanistisches Institut, 44780 Bochum, E-mail: lisa.otto@rub.de

**Astrid Neumann**, Leuphana Universität Lüneburg, Institut für Deutsche Sprache und Literatur und ihre Didaktik, 21335 Lüneburg, E-mail: astrid.neumann@leuphana.de

**Michael Krelle**, Technische Universität Chemnitz, Zentrum für Lehrerbildung, 09111 Chemnitz, E-mail: michael.krelle@zlb.tu-chemnitz.de

**Dominik Leiss**, Leuphana Universität Lüneburg, Institut für Mathematik und ihre Didaktik, 21335 Lüneburg, E-mail: leiss@leuphana.de

**Dietmar Höttecke**, Universität Hamburg, Fakultät für Erziehungswissenschaft, AG Didaktik der Physik, 20146 Hamburg, E-mail: dietmar.hoettecke@uni-hamburg.de

**Timo Ehmke**, Leuphana Universität Lüneburg, Institut für Bildungswissenschaft, 21335 Lüneburg, E-mail: tehmke@leuphana.de

**Knut Schwippert**, Universität Hamburg, Fakultät für Erziehungswissenschaft, von-Melle-Park 8, 20146 Hamburg, E-Mail: knut.schwippert@uni-hamburg.de

retical and empirical grounds. The purpose of this article is thus to present the linguistic foundation used in an interdisciplinary empirical study in which 1.346 7<sup>th</sup> and 8<sup>th</sup> graders solved a set of subject-oriented tasks from Maths, Physics, German, PE and Music, while the degree of linguistic demands in the tasks was systematically varied. First, the theoretical and empirical research on linguistic difficulty from a range of research discourses is discussed. The findings are merged into a model of linguistic demands. Its operationalization is then illustrated in three linguistically varied versions of the subject-specific tasks. Finally, an outlook on preliminary results of the empirical study is given, which indicate that the categories used in the model actually do produce differences in subject-task difficulty, even though there are a number of effects that need further investigation.

**Keywords:** Academic language, task difficulty, linguistic difficulty, language in subject-specific tasks

## 1 Einleitung

Schulbezogene Sprachkompetenzen werden in einem breit aufgestellten, interdisziplinären Wissenschaftsfeld mit unterschiedlichen Begriffen beschrieben, bspw. als „Cognitive Academic Language Proficiency“ (Cummins 2008) oder als „Language of Schooling“ (Schleppegrell 2014). Mitunter sind auch allgemeine sprachwissenschaftliche Ansätze in der Diskussion, z.B. wenn Aspekte von „Distanzsprache“ bzw. von „konzeptioneller Schriftlichkeit“ (Koch und Oesterreicher 1985) diskutiert oder wenn „elaborierte Codes“ (Bernstein 1964) zur Abgrenzung herangezogen werden. Zuletzt stand besonders der zur „Leitvokabel“ (Feilke 2012: 4) erhobene Begriff der „Bildungssprache“ im Mittelpunkt (Becker-Mrotzek et al. 2013; Gogolin und Lange 2010; Morek und Heller 2012; Tschirner et al. 2016). Diese Begriffe werden mit mehr oder weniger starker Schärfung und unterschiedlichen Fokussierungen verwendet (vgl. Pohl 2016). Allen unterliegt jedoch die Annahme, dass Schule die Verwendung von solchen Sprachregistern erfordert, die sich von denen des alltäglichen Diskurses (z.B. im Familien- und Freundeskreis) unterscheiden, und dass schulbezogene Sprache im Vergleich zu Alltagssprache in der einen oder anderen Weise erhöhte Anforderungen an ihre Benutzer\*innen stellt. Besonderes Augenmerk wird dabei oft auf Lerner\*innen mit Deutsch als Zweitsprache gelegt, für die schulbezogene Sprachregister besonders herausfordernd seien (Gogolin und Duarte, 2016) und in einem „double disadvantage“ (Lüdemann und Schwerdt 2013; Heppt et al. 2015) resultierten. Sie stünden nämlich nicht nur vor der Aufgabe, Deutsch für den Alltagssprachgebrauch zu

lernen, sondern müssen gleichzeitig auch schulbezogene Sprachkompetenzen aufbauen, die zudem meist nicht explizit vermittelt werden (Vollmer und Thürmann 2010). In der Diskussion um ungleiche Bildungschancen und Schulerfolg werden daher normative bzw. qualifizierende und selektierende Dimensionen bildungssprachlicher Fähigkeiten hervorgehoben (z.B. Gogolin und Lange 2011; Oleschko 2017). Sprache kann dabei unterschiedliche Funktionen einnehmen, so dass es sich bei der Frage nach dem Erwerb der notwendigen sprachlichen Kompetenzen nicht nur um die kommunikative bzw. didaktische Funktion von Sprache in Kontexten der Wissensvermittlung (z.B. Feilke 2012) handelt. Vielmehr können sich über sie auch sozial-identitätsstiftende Merkmale ausdrücken (z.B. Morek und Heller 2012) sowie epistemische Funktionen ausgeübt werden, also gleichzeitig mit sprachlichen Strukturen auch neue Denkmuster ausgebildet werden (z.B. Pohl 2016), wodurch sich klare Verbindungen zwischen Fachlernen und Sprachenlernen, etwa im Sinne einer durchgängigen Sprachbildung, auf tun (vgl. Ahrenholz 2017; Reich 2013).

Der Kenntnisstand darüber, ob und inwiefern schulbezogene Sprache jedoch tatsächlich als (schulische) Barriere wirkt, und wenn ja, für welche Schüler\*innen besonders, muss momentan als noch sehr undifferenziert betrachtet werden. Diese fehlenden Grundlagen haben Konsequenzen für die schulische Praxis:

[W]ir [tun] eigentlich den zweiten vor dem notwendigen ersten Schritt: Wir greifen mit entsprechenden Fördermaßnahmen in einen Spracherwerbsprozess ein, den wir zuvor weder exakt *beobachtet* und präzise *beschrieben* noch hinsichtlich seiner internen Wirkfaktoren *analysiert* und *verstanden* haben (Pohl 2016: 73, Herv. i. O.).

Der vorliegende Beitrag nimmt vor diesem Hintergrund einen spezifischen Blickwinkel ein, der sich auf der Schnittstelle von Fach und Sprache mit der Gestaltung von Textaufgaben für konstruktvalide und faire Leistungsmessungen beschäftigt. Er ist dabei an einem der „internen Wirkfaktoren“ genauer interessiert, nämlich den Effekten, die unterschiedliche sprachliche Ausgestaltungen von Textaufgaben auf die Aufgabenschwierigkeit haben.<sup>1</sup>

---

<sup>1</sup> Der Beitrag ist im Rahmen umfänglicherer Forschungsarbeiten der aus Erziehungswissenschaftler\*innen, Sachfach- und Sprachdidaktiker\*innen und Bildungsforscher\*innen bestehenden Arbeitsgruppe *Fach und Sprache* (AG FuS) entstanden. Diese Arbeitsgruppe befasst sich seit 2011 mit der Erforschung verschiedener Dimensionen von Fachaufgaben, wobei die sprachliche Dimension eine hervorgehobene Rolle einnimmt. Dies erklärt, weshalb im vorliegenden Beitrag textlastige Aufgaben (und nicht etwa Informationstexte) im Blickpunkt stehen. Die AG FuS besteht aus folgenden Mitgliedern: Sprecher: Dominik Leiss (Mathematikdidaktik, Lüneburg), Knut Schwippert (Empirische Bildungsforschung, Hamburg); Beteiligte: Deutschdidaktik – Astrid Neumann (Lüneburg) und Michael Krelle (Chemnitz); Physikdidaktik – Dietmar Höttecke (Hamburg); Sportdidaktik – Claus Krieger (Hamburg) und André Gogoll (Maggingen);

Die Grundannahme ist hierbei, dass Fachaufgaben, die nicht nur fachliche, sondern auch solche sprachliche Kompetenzen messen, die nicht bei allen Schüler\*innen gleichermaßen vorausgesetzt werden können, zur Chancengleichheit beitragen, indem sie das eigentliche fachliche Kompetenzniveau verschleiern. Dies wiederum basiert auf der Prämisse, dass bestimmte sprachliche Strukturen tatsächlich mit erhöhter Verstehensschwierigkeit einhergehen. Zu untersuchen ist somit in Bezug auf schriftlich vorgelegte, textlastige Testaufgaben, ob die Verwendung solcher sprachlicher Mittel in Aufgabenstellungen tatsächlich die fachliche Leistungsfähigkeit von Schüler\*innen beeinträchtigt. Im Diskurs zur schulischen Sprachförderung und -bildung findet sich eine Fülle von Vorschlägen, welche sprachlichen Oberflächenmerkmale typisch für schulische Diskurse seien und besonderer Unterstützung bedürfen, weil sie sich eher nicht spontan im Alltagssprachgebrauch entwickeln können. Genannt werden hier in der Regel (siehe für entsprechende Zusammenstellungen z.B. Berendes et al. 2013; Gantefort 2013; Gogolin und Duarte 2016; Leisen 2013; Morek und Heller 2012; Stahns 2016; für Entsprechungen im Englischen vgl. Überblick in Abedi et al. 1997):

- auf Ebene des Wortschatzes u.a. fachlich bzw. inhaltlich spezifische Lexeme, Fremdwörter und mehrdeutige Ausdrücke, welche als semantisch anspruchsvoll gelten, aber auch strukturell komplexe Elemente wie diskontinuierliche Ausdrücke (z.B. Präfixverben), Nominalisierungen oder Komposita,
- bestimmte morphosyntaktische Merkmale wie bspw. Genitiv- und Passivkonstruktionen, unpersönliche Wendungen, die Verwendung des Konjunktivs,
- komplexe syntaktische Strukturen wie Partizipialkonstruktionen oder Nebensatzgefüge
- sowie satzübergreifende kohäsionsstiftende Mittel wie komplexe Konjunktionen oder textstrukturelle Vor- und Rückverweise.

Während aus der Textverstehens- und Textverständlichkeitsforschung (Christmann und Groeben 2001; Langer et al. 1974) bekannt ist, dass Gliederungs- und Ordnungsprinzipien der Informationsdarbietung in Texten (Absätze, Überschriften, Hervorhebungen, Redundanzen) sowie die textliche Fokussierung auf das Wesentliche als zentral für die Schwierigkeit eines Textes betrachtet werden können, fällt auf, dass die oben genannten Merkmale vor allem auf lexikalisch-syntaktischer Ebene angesiedelt sind. Zwar liegen erste empirisch basierte Merk-

---

Musikdidaktik – Michael Ahlers (Lüneburg); Sprachbildung und Mehrsprachigkeit – Lena Heine (Bochum); empirische Bildungsforschung – Timo Ehmke (Lüneburg) und Schulpädagogik – Marc Kleinknecht (Lüneburg) sowie sieben dazugehörigen Mitarbeiter\*innen, Näheres unter <http://www.fach-und-sprache.de/>.

malsbeschreibungen vor, die zeigen, dass diese Kategorien (zumindest teilweise) tatsächlich in typischen schulfachbezogenen Diskursen vorkommen (z.B. Schramm et al. 2013 für Sachunterricht; Hartung 2013 für Geschichte; Niederhaus et al. 2016 für Mathematik; Berendes et al. 2017 für Erdkunde; Rincke 2010 für Physik). Ob diese Strukturen aber tatsächlich zu Verstehensschwierigkeiten führen, ist bislang noch weitgehend ungeklärt. So verzeichnen in fachbezogenen Kontexten bei Sachtexten und textlastigen Fachaufgaben beispielsweise Deppner (1989) und Starauschek et al. (2006) bei der Vereinfachung auf sprachstrukturell-semanticischer Ebene bessere Verstehensleistungen von Schüler\*innen, während Kohnen et al. (2017), Haag (2015) und Heppt et al. (2015) keine oder nur geringe Auswirkungen von sprachlich-semanticen Textanpassungen messen. Härtig und Kohnen (2017) weisen für hohe Textkohäsion sogar größere Schwierigkeiten bei Schüler\*innen nach, die über hohes Vorwissen verfügen, während bei Schmitz und Gräsel (2016) eine höhere Kohäsionsdichte Schüler\*innen mit mehr thematischem Vorwissen das Verstehen erleichtert. Vergleichbar uneindeutig ist die Lage im englischsprachigen Raum: Bird und Welford (1995), Mesmer, Cunningham und Hiebert (2012) und Prophet und Badede (1995) zeigen Item-Bias-Effekte der sprachstrukturellen Ausgestaltung von Sachtexten und -aufgaben, während die Metastudie von Kieffer et al. (2009) zum Schluss gelangt, die Vereinfachung der Testsprache sei weitgehend wirkungslos.

Wir werden daher im Folgenden den Forschungsstand zu empirisch beschreibbaren sprachlichen Merkmalen hinsichtlich ihrer schwierigkeitsgenerierenden Effekte für die Textrezeption aus verschiedenen relevanten Teildisziplinen aufarbeiten, um zu prüfen, für welche sprachlichen Oberflächenmerkmale tatsächlich Hinweise für ihren schwierigkeitsgenerierenden Charakter vorliegen. Dabei arbeiten wir heraus, dass die entsprechenden sprachlichen Strukturen (z.B. auf Lexik- oder Syntaxebene) sich nach drei unterliegenden schwierigkeitsgenerierenden Prinzipien ordnen lassen (Kapitel 2). Diese bilden den Ausgangspunkt für die Operationalisierung dieser Prinzipien in unterschiedlichen sprachlichen Kategorien (Kapitel 3), deren Umsetzung anhand von Ausschnitten aus Physik- und Mathematikaufgaben illustriert wird. Auf das schwierigkeitsgenerierende Potenzial dieser Prinzipien wird abschließend anhand von ausgewählten Daten aus einer empirischen Studie ein Ausblick gegeben. Die detaillierte Darstellung der Ergebnisse dieser Studie erscheint in einem eigenständigen Beitrag (vgl. Schwippert et al. eingereicht); hier soll es vorrangig um die linguistische Fundierung des Designs gehen.

## 2 Ansätze zur Erfassung von sprachlicher Schwierigkeit bei der Dekodierung von Texten

Das Lesen eines (Aufgaben-)Textes wird in allen bekannten Modellvorstellungen als ein interaktiver Vorgang gefasst, der sich zwischen dem Lesenden mit seinem Vorwissen und weiteren Personenmerkmalen, dem Text und seinen Merkmalen, der Leseaktivität (z. B. ihrer Zielsetzung) in einem weiteren soziokulturellen Kontext abspielt (vgl. u. a. Christmann und Groeben 2001; Goeperich 2008; Guthrie et al. 2004; Langer et al. 1974; RAND Reading Study Group 2002; Van Dijk und Kintsch 1983; Verhoeven und Perfetti 2008; Zwaan und Radvansky 1998 u. a.). Einen Text lesend verstehen zu können, wird also als Integrationsfähigkeit von text- und wissensbasierten Prozessen gefasst und umfasst somit deutlich mehr als nur die Dekodierung sprachlicher Formen. Somit ist die Schwierigkeit eines Textes nur unter Einbezug sowohl von Textmerkmalen als auch Leser- und Kontextmerkmalen zu bestimmen. Dennoch versuchen verschiedene Ansätze, Textschwierigkeit weitgehend rezipientenunabhängig bzw. vor dem Hintergrund „prototypischer“ Leser\*innen über beschreibbare linguistische Merkmale zu ermitteln. Während frühe Ansätze aus der Lesbarkeitsforschung (z. B. Gunning 1969; Kincaid et al. 1975) lediglich Wortlänge und Silbenanzahl sowie Satzlänge zur Bestimmung von Textschwierigkeit messen, greifen neuere Zugänge auf eine Vielzahl an Einzelmaßen zurück, darunter Wort- und Kollokationsfrequenz, lexikalische Breite, morphologische und syntaktische Textmerkmale, Kohäsions- und Kohärenzmarker sowie auch semantische Merkmale, wie die Anzahl an Propositionen im Satz. Sie weisen gute Übereinstimmungen mit Ergebnissen von Verstehenstests auf (vgl. Graesser et al. 2004; Schwarm und Ostendorf 2005; Vajjala und Meurers 2012; s. für einen guten Überblick Beinborn 2016). Vor der Brück et al. (2008) und Hancke et al. (2012) stellen Ansätze vor, die gezielt auf das Deutsche bezogen sind und auch die morphologischen Besonderheiten des Deutschen berücksichtigen.<sup>2</sup> Bislang liegen allerdings noch keine überprüften Anwen-

---

<sup>2</sup> Auch die Diskussion um Übersetzungen von Gebrauchstexten in Leichte Sprache ist in diesem Kontext relevant. Leichte Sprache entspricht jedoch unserer Testlogik nicht, weil sie Partizipations-, Lern- und Brückenfunktionen verfolgt (Bredel und Maas 2016: 10). Die ethischen Leitlinien der Leichten Sprache zeigen auf, dass es sich bei ihr um eine Varietät des Deutschen handelt, die der Inklusion von Menschen mit kognitiven oder sensorischen Beeinträchtigungen und funktionalen Analphabeten mithilfe einer im Bereich Satzbau, Wortschatz und mit Bezug auf das vorauszusetzende Weltwissen systematisch reduzierten und in besonderer visuell aufbereiteten äußeren Form dient (Maas 2015: 11–12). Diese Aufgabe ist besonders zu würdigen und von entsprechend geschulten Übersetzer\*innen auszuführen. Zudem sind Texte in Leichter Sprache als *Ergänzungen* und nicht alleinige Testformate einsetzbar: „Leichte-Sprache-Texte ersetzen

dungen für das Deutsche vor, die für unsere Zwecke (insbesondere hinsichtlich der schulischen Adressatengruppe) direkt verwendbar wären. Zusätzlich erscheint es mit Hinblick auf didaktische Zielsetzungen relevant, schwierigkeitsgenerierende Prinzipien auch ohne computerbasierte Auswertungsmöglichkeiten transparent zu machen und zu verdeutlichen, wie sich diese auf sprachlicher Oberfläche anhand unterschiedlicher Phänomene manifestieren können. Im Folgenden geben wir deshalb einen Überblick über empirische Hinweise darauf, dass die in der Literatur als typisch für schulbezogene Texte identifizierten Einzelmaße tatsächlich zu erhöhter Verstehensschwierigkeit führen. Nicht alle Einzelphänomene sind für das Deutsche bereits empirisch untersucht worden. Die Belege, die uns vorliegen, legen es jedoch nahe, dass sich drei Prinzipien für eine Erfassung von sprachlicher Schwierigkeit anbieten: Strukturelle Komplexität, Eindeutigkeit von Form-Bedeutung-Beziehungen und Frequenz. Wir ordnen den folgenden Überblick daher nach diesen Kategorien.

## 2.1 Strukturelle Komplexität

Sprachliche Schwierigkeit von Texten lässt sich u.a. darüber bestimmen, wie hoch die Kosten der kognitiven Verarbeitung bestimmter sprachlicher Strukturen sind. Diese hängen mit dem Grad sprachlicher Komplexität der jeweiligen Äußerung zusammen, welche in der Psycholinguistik bspw. definiert wird als

a property or quality of a phenomenon or entity in terms of (1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationships between the constituent components (Bulté und Housen 2012: 22).

Die Schwierigkeit ist in diesem Sinne also einerseits durch die Menge der zu verarbeitenden Einzelelemente bestimmt, andererseits durch die Anzahl und die Qualität der einzelnen Prozesse, die für die Verarbeitung sprachlicher Information angenommen werden und die den Grad der kognitiven Belastung ausmachen. Diese Prozesse werden je nach sprachtheoretischer Verankerung unterschiedlich modelliert: Auf Grundlage eines generativen Ansatzes wird dabei etwa angenommen, dass verschiedene Prozeduren (z.B. Rechtsbewegung, Deletion oder Zwischenstrukturen) auf einer basalen Struktur durchgeführt werden müssen, um die entsprechende Struktur zu produzieren bzw. rückgängig gemacht werden müs-

---

nicht den Ausgangstext. Sie ergänzen ihn und stellen ein Zusatzangebot dar. Ausgangs- und Zieltext werden teilweise flankierend gelesen, [...]“ (Maas 2015: 139; vgl. für eine kritische Diskussion dieser Sichtweise jedoch Bock et al. 2017).

sen, um sie zu entschlüsseln. Dies macht die Verarbeitung einer solchen Struktur komplexer, erfordert mehr Schritte in der Sprachverarbeitung und bindet damit mehr kognitive Ressourcen. Eher kognitiv-linguistisch orientierte Studien nehmen stattdessen allgemeine kognitive Prozesse als Maßstab, hier v.a. den Grad der Belastung des Arbeitsgedächtnisses, die angesichts begrenzter Verarbeitungskapazitäten z.B. durch die Verarbeitung bestimmter syntaktischer Strukturen verursacht wird (Housen 2014). Es wird davon ausgegangen, dass einerseits bereits aufgebaute Strukturen, die „Speicherkosten“ verursachen, aktiv gehalten werden müssen, andererseits „Integrationskosten“ anfallen, wenn neue Elemente in bereits aufgebaute syntaktische Strukturen integriert werden müssen. Dabei wird angenommen:

[...] (1) the longer a predicted category must be kept in memory before the prediction is satisfied, the greater is the cost for maintaining that prediction; and (2) the greater the distance between an incoming word and the most local head or dependent to which it attaches, the greater the integration cost (Gibson 1998: 1).

Ein in der Psycholinguistik vergleichsweise breit untersuchtes Feld ist hierbei das der „Einschubphänomene“, insbesondere bei Partikelverben oder Relativsätzen (z.B. Levy et al. 2012; Levy und Keller 2013; Gabler 2013). Verwendet werden typischerweise Grammatikalitäts- und Schwierigkeitseinschätzungen von Muttersprachler\*innen (siehe zusammenfassend Gibson 1998) oder Blickbewegungsmessungen (Kliegl et al. 2004), die immer häufiger auch mit EEG-Messungen komplettiert werden (z.B. Bornkessel-Schlesewsky et al. 2011; Gabler 2013), aber auch Lesezeitmessungen (z.B. King und Just 1991; Konieczny 2000). Längere Lesezeiten ergeben sich beispielsweise, je weiter ein Relativsatz vom Bezugsnomen entfernt ist (z.B. Konieczny 2000, dessen Ergebnisse darauf hinweisen, dass die Zuordnung des Relativpronomens zum weiter entfernt stehenden Kopfnomen größere kognitive Anstrengung erfordert als zu nahestehenden). U.a. Hawkins' (2004) Untersuchung deutet an, dass neben Menge und Art von Verarbeitungsprozessen (hier in Bezug auf die Distanz der Extraposition von Relativsätzen) und der Anzahl der zu integrierenden Einzelelemente (Länge des Relativsatzes) auch semantische Aspekte die Verarbeitungsanforderungen bestimmen. Dies zeigt sich ebenfalls in einer Studie von Lahmann (2011, dargestellt in Hulstijn 2015: 72ff.). Erhoben wurden u.a. Grammatikalitätsurteile von L1-Sprecher\*innen in Bezug auf unterschiedlich komplexe Satzstrukturen im Deutschen. Einfache Stimuli beinhalteten Hauptsätze in der am häufigsten vorkommenden Konstituentenreihenfolge (S-V-O, auch mit Satzklammer), komplexe Sätze bestanden aus stark markierten Strukturen wie HS mit indirektem Objekt + direktem Objekt + Subjekt-NO oder Sätze mit Objekt Cleft Clauses oder Mit Objekt-WH-Fragen (*David fragt sich, wen am Freitag der Chauffeur abgeholt hat*). Bei der Beurteilung der

Grammatikalität derartiger Sätze erkannten alle Probanden die einfachen Strukturen als richtig, allerdings nur 87 % die hoch komplexen Strukturen.

Auch für Passivkonstruktionen konnte gezeigt werden, dass sie höhere Anforderungen an Verstehensprozesse stellen als ihre aktiven Entsprechungen (z.B. Berndt et al. 2004; Kharkwal und Stromswold 2014). Sie werden vergleichsweise spät im L1-Erwerb erworben (Diessel 2004) und gehören auch bei L1-Sprecher\*innen nicht zur linguistischen Kernkompetenz im Sinne einer von allen L1-Sprecher\*innen beherrschten *Basic Language Cognition* (Hulstijn 2015; Street und Dabrowska 2010).

Auch wenn diese Aspekte (Anzahl und Beschaffenheit der einzelnen Elemente bzw. deren Relationen untereinander) in der Forschung zumeist auf syntaktischer Ebene untersucht worden sind, lassen sie sich auch auf Wort- und Phrasenebene beziehen. Auf Wortebene finden sich Hinweise darauf, dass sich Wortlänge und morphologische Komplexität (z.B. Nominalisierungen) auf die Verarbeitungsanforderungen auswirken, u.a. korrelieren sie in Eye-Tracking-Studien mit der Fixationsdauer (Kliegl et al. 2004; Solomyak und Marantz 2010); auf Phrasenebene konnte gezeigt werden, dass komplexe Nominalphrasen schwerer zu verarbeiten sind als einfache (Fang et al. 2006; Schmidt 1993).

## 2.2 Eindeutigkeit von Form-Bedeutung-Beziehungen

Ein anderer Blickwinkel rückt die Transparenz auf semantischer Ebene, also die Eindeutigkeit von Form-Bedeutung-Beziehungen, als Maß für sprachliche Schwierigkeit in den Fokus. Gerade im Kontext der L2-Forschung ist die Assoziation von unterschiedlichen Graden an semantischer Plausibilität und unterschiedlichen Anforderungen an Verstehensprozesse mehrfach belegt (vgl. Štekauer 2005 oder Fusté-Herrman 2008; für idiomatische Ausdrücke Irujo 1993; für semantisch implausible Passivsätze Street und Dabrowska 2010). Daneben gibt es starke Hinweise auf die Relevanz semantischer Transparenz bzw. Eindeutigkeit von Sprachstrukturen auch für den schulischen Kontext. So identifizieren Köhne et al. (2015: 77, Hervorhebungen im Original) im Rahmen des Projekts BiSpra folgende Elemente als schwierigkeitsgenerierend:

Ausdrücke, die, lexikalisch isoliert betrachtet, ambig oder mehrdeutig (polysem) sind bzw. genauer: die eine konkrete, alltägliche, und eine (oder mehrere) abstraktere Bedeutungen aufweisen (z.B. *abgeben*, *angeben*, *durchgehen*, *liegen*, *scheinen*, *abgeschlossen*) sowie solche Präfixverben, die sich aus einem Präfix wie *er-*, *ver-* oder *be-* und einem alltäglich gebräuchlichen Verb wie *setzen*, *stellen* oder *legen* zusammensetzen.

Dies wird durch die Befunde anderer Projekte zum schulbezogenen Textverständnis für verschiedene Altersstufen und Fächer bestätigt – bspw. in weiteren Teilprojekten in BiSpra (Runge 2013) oder des Projekts NaTexI für das Verständnis naturwissenschaftlicher Texte in der Sekundarstufe I (Kohnen et al. 2017).

Die Explizierung von Bedeutungsrelationen ist bei der textuellen Gestaltung von Bedeutung; insbesondere gibt es Hinweise darauf, dass ein Text umso einfacher verständlich wird, je mehr kohäsionsstiftende Mittel er enthält und je transparenter er kohäsive Bezüge macht (Breindl und Waßner 2006; Schmitz und Gräsel 2016). Enthält ein Text allerdings viele pronominale Referenzen, die für ein Verständnis zunächst eine semantische Interpretation erfordern, kann sich seine Schwierigkeit erhöhen (Bryant und Noschka 2015).

### 2.3 Frequenz

Beide bisher besprochenen Prinzipien, strukturelle Komplexität und Eindeutigkeit von Form-Bedeutung-Beziehungen, korrelieren oft mit deren Vorkommenshäufigkeit. In gebrauchsbasierten Ansätzen, die vor allem mit Korpusdaten arbeiten, wird daher oft angenommen, dass sich letztlich jegliche Dimension kognitiver Verarbeitungsschwierigkeit mit der Frequenz im sprachlichen Input erklären lässt (z.B. Ellis 2002; Scontras et al. 2014) – insbesondere, da frequente Strukturen in Zusammenhang mit dem Wissensabruf von automatisiertem Sprachwissen und Chunks stehen (Ellis und Shintani 2014) und somit Auswirkungen auf die Effizienz der Sprachverarbeitungsprozesse haben. Derartige erwartungsbasierte Untersuchungen sind zu einer Reihe von Sprachen durchgeführt worden (für die deutsche Sprache etwa Konieczny 2000; Konieczny und Döring 2003; Levy und Keller 2013). Auch aus der Sprachtestforschung (Kniffka und Linnemann 2016) gibt es Hinweise darauf, dass bestimmte Strukturen (im verwendeten C-Test v.a. paarige Konjunkturen wie *sowohl ... als auch; zwar, ... aber*, ebd.: 172ff.), aber auch Kollokationen und Phraseologismen wie *den Geldreichtum mehren; dem Ziel dienen; xy liegt die Annahme zugrunde*, sowie Ellipsen für die untersuchte Schüler\*innengruppe besonders schwierigkeitsgenerierend waren. Diese Gefüge sind verhältnismäßig wenig frequent, was die Autoren mit geringerer Vertrautheit und geringerer Automatisierung in Verbindung bringen. Auch auf morphosyntaktischer Ebene lassen sich ähnliche Beobachtungen anschließen: So ist beispielsweise die zum Bezugsnomen adjazente Position von Relativsätzen nicht nur mit unterliegenden komplexeren Verarbeitungsmechanismen erklärbar, sondern sie ist auch in deutlich mehr Kontexten möglich und damit auch insgesamt häufiger als die extraponierte Position (zusammenfassend Gabler 2013: 13). Ebenso sind Aktivkonstruktionen

ungleich häufiger im sprachlichen Input als entsprechende Passivkonstruktionen (Cornelis 1996).

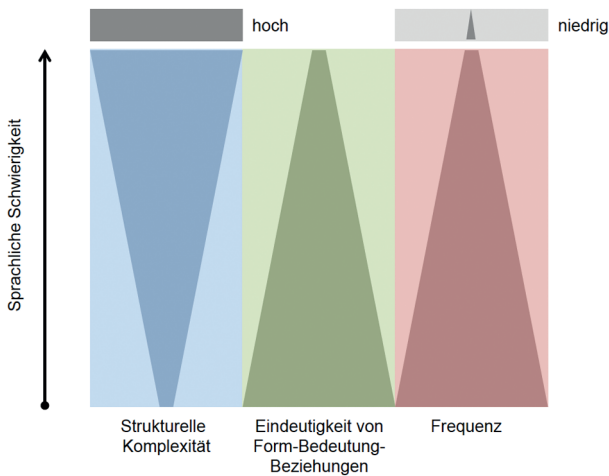
## 2.4 Erkenntnisse aus der L2-Forschung

Die Frage nach dem Umgang mit sprachlicher Schwierigkeit stellt sich insbesondere, wenn es sich um Schüler\*innen handelt, die Deutsch zusätzlich zu einer oder mehreren bereits beherrschten Sprachen erlernen. Sie müssen sich in schulischen Kontexten (v.a. sprachlich vermittelte) fachspezifische Kompetenzen aneignen, während sie gleichzeitig noch grammatikalische und lexikalische Grundlagen des Deutschen erlernen. Hier ist die Frage zu stellen, ob sich für L2-Lernende spezifische Bereiche identifizieren lassen, die in der Rezeption von Sprache besondere Schwierigkeiten darstellen (de Graaff und Housen 2009; Long und Robinson 1998). Sie ist vor allem in der L2-Testforschung dahingehend gestellt worden, ob Lesen in der L2 eher als „reading problem or a language problem“ anzusehen ist (Alderson 1984) – also, ob begrenzte Kompetenzen in der Zielsprache eine herausgehobene Rolle spielen oder ob Performanz eher durch allgemeine kognitiv-sprachliche oder metakognitive Prozesse beeinflusst wird. Meta-Analysen (Grabe 2009; Jeon und Yamashita 2014; auch Alderson et al. 2014) zeigen, dass Grammatik-, aber insbesondere Wortschatzwissen von L2-Probanden die am stärksten mit Leseverstehen korrelierenden Variablen sind. Auf textueller Oberflächenebene werden, wie auch in L1-Kontexten, der Grad der Komplexität von grammatikalischen Strukturen (Freedle und Kostin 1999: 27; Grotjahn 2000: 47), aber auch die „Zahl der schwierigen und/oder unvertrauten (seltenen) Wörter“ (Grotjahn 2000: 47) als zentrale Schwierigkeitsvariablen für die Textrezeption in der neuen Sprache genannt, wodurch auch hier das Prinzip „Frequenz“ in den Vordergrund rückt. In L2-Kontexten wird „Schwierigkeit“ jedoch vor allem unter Erwerbsaspekten mit Phasen der L2-Aneignung in Verbindung gebracht. Verhältnismäßig „schwieriger“ sind somit Konstruktionen, die später im L2-Erwerbsprozess erworben werden. Verhältnismäßig gut gesichert sind Erkenntnisse aus sogenannten Morphemstudien, die zeigen, dass bestimmte grammatikalische Muster lernerübergreifend in derselben Reihenfolge erworben werden. Beispielsweise werden im Verbalbereich Infinitive vor regelmäßigen vor unregelmäßigen Präsensformen erworben, für den Satzbau: Hauptsatz vor W-Fragen vor Verbal-klammer vor Inversion, Nominativ vor übrigen Kasusformen (Pienemann 1989; Diehl 1999; s. a. Dimroth und Schimke 2012; Grieshaber 2005). Neben Frequenz findet sich auch hier der Aspekt der strukturellen Komplexität wieder. Einschränkend ist hier zu sagen, dass die Erkenntnisse zu den Morphemstudien aus Sprachproduktionsdaten hergeleitet wurden und unklar ist, inwiefern später erworbene

Strukturen tatsächlich mit erhöhten Verstehenschwierigkeiten in Zusammenhang stehen. Weiterhin sind an dieser Stelle noch typische Schwierigkeitsbereiche des Deutschen wie bspw. Pronomen, (komplexe) Nominalphrasen, Verbformen etc. zu nennen, die besonders immer dann als „schwieriger“ eingeschätzt werden, wenn sich die Strukturen aus der L1 besonders vom Deutschen unterscheiden.

## 2.5 Zwischenfazit und einschränkende Bemerkungen

Die tatsächlichen Anforderungen beim Leseverstehen ergeben sich aus einem Komplex an individuellen und situativen Faktoren in dynamischer Interaktion mit einem Textprodukt. Die eindeutige Bestimmung des sprachlichen Schwierigkeitsgrades einer textlastigen Aufgabe ist für authentische Situationen somit kaum möglich. Dennoch lassen sich bestimmte sprachliche Dimensionen nennen, die potenziell eine Auswirkung auf die Verständlichkeit beim Lesen haben: So deuten die vorangegangenen Ausführungen darauf hin, dass das Vorkommen strukturell einfacher, semantisch klarer und hochfrequenter Konstruktionen im Text das Sprachverstehen erleichtert (Abbildung 1).



**Abbildung 1:** Drei Dimensionen sprachlicher Textschwierigkeit beim Lesen

Diese Dimensionen passen weitgehend auf solche Phänomene, die in der Literatur wiederkehrend als charakteristisch für Texte in der Schule genannt werden (z.B. semantisch oder strukturell intransparente Lexik bzw. Morphosyntax, komplexe Syntax, hohe Satz- und Phrasenlänge mit hoher Propositionsdichte, fehlende se-

mantische Redundanz sowie unklare Verweisstrukturen, vergleichsweise seltene Lexik und Satzstrukturen etc.). Die L2-Forschung liefert außerdem Anhaltspunkte dafür, dass bereits auf niedrigen Erwerbsstufen beherrschte grammatikalische Strukturen mit größerer Wahrscheinlichkeit bei allen Schüler\*innen als verständlich vorausgesetzt werden können als solche, die erst später erworben werden.

Einschränkend ist anzumerken, dass sich viele der genannten Studien nicht auf das Deutsche beziehen, sondern entsprechende Phänomene in anderen Sprachen, v. a. im Englischen, untersucht wurden. Zusätzlich ist zu bedenken, dass die Kategorisierungen der entsprechenden Phänomene letztlich auch von den verwendeten theoretischen Rahmungen abhängen – so argumentieren Vertreter\*innen gebrauchsbasierter Ansätze sehr viel stärker über den Faktor Frequenz als Forscher\*innen, die sich innerhalb eines modular-symbolverarbeitenden Paradigmas verorten. Mit dem Wissen um diese Einschränkungen lässt der Literaturüberblick dennoch eine Eingrenzung auf die vorgeschlagenen allgemeinen Prinzipien zu (Abbildung 1), die sich wiederkehrend als bedeutsam für Effekte von erhöhter Anforderung an die sprachliche Verarbeitung erwiesen haben.

### 3 Operationalisierung der Prinzipien in Testaufgaben

Will man unterschiedliche Grade sprachlicher Schwierigkeit in (Aufgaben-)Texten für ein Forschungsdesign in eine konkrete Form umsetzen, so bedarf es der Operationalisierung der drei Prinzipien in sprachlichen Elementen. In den aktuellen Untersuchungen der AG FuS sollten Fachaufgaben in drei verschiedenen sprachlichen Versionen zum Einsatz kommen. Ziel ist es, die Auswirkungen solcher Modellierungen auf Leistungen von Schülergruppen in den Blick zu nehmen. Aus diesem Grund haben wir die in Abbildung 1 dargestellten Kontinua in drei Stufungen unterteilt und mit typischen Indikatorphänomenen gefüllt, an denen sich die Prinzipien niederschlagen bzw. die sich anbieten, auf drei Ebenen variiert zu werden (Tabelle 1).

Bewusst versucht das Kriterienraster in Tabelle 1 dabei nicht, eine vollständige Abbildung aller potentiell möglichen sprachlichen Variationsbereiche zu liefern, sondern es benennt mit den hervorgehobenen Kategorien gezielt nur besonders saliente und damit leicht zu fokussierende Merkmale. Die Cut-off-Points zwischen den Versionen sind dabei zunächst reine Setzungen. Basierend auf vergleichbaren Designs (Leiss et al. 2017; Haag 2015) erschien es wichtig, die einzelnen Versionen in einem möglichst breiten Spektrum an sprachlicher Spannweite bei gleichbleibender Fachsprachlichkeit anzusiedeln, damit in ihnen auch

wirklich Unterschiede zu Tage treten können. Dennoch sollten sie – trotz ihrer Künstlichkeit – jeweils real stattfindender Sprachverwendung nahekommen. Außerdem war uns wichtig, eine Anwendbarkeit in den Fachdidaktiken zu ermöglichen, die kein sprachtheoretisches Vertiefungswissen erforderte.

**Tabelle 1:** Operationalisierung des Modells in konkreten Kriterien auf sprachlicher Ebene

	Kriterien	Version 1	Version 2	Version 3
Verben	Frequenz	hochfrequent ( <i>haben, sein, machen, ...</i> )	mittelfrequent ( <i>verursachen, verbleiben, ...</i> )	wenig frequent ( <i>erstehen, illustrieren, ...</i> )
	Semantik	semantisch transparente Ausdrücke (z.B. <i>tragen, halten, kaufen, ...</i> )	mehrdeutige bzw. semantisch intransparente Ausdrücke ( <i>anhalten, einhalten, anfallen, ...</i> )	idiomatische Ausdrücke ( <i>mit seinem Latein am Ende sein, zu Buche schlagen, ...</i> )
	Morphologie	Simplex	Präfix und Partikel	Präfix und Partikel, Nominalisierungen
	Flexion	Indikativ	Konjunktiv	Passiv, dreiteilige verbale Ausdrücke (z.B. <i>gemacht haben wollen</i> )
Nomen	Frequenz	hochfrequent ( <i>Haus, Baum, Fußball, ...</i> )	mittelfrequent ( <i>Gebäude, Material, ...</i> )	wenig frequent ( <i>Gebälk, Verzehr, ...</i> )
	Morphologie	Komposita möglichst vermeiden (bis auf hochfrequente)	transparente Komposita (z.B. <i>Nom+Nom wie Herbsttag, Wohnort, Arbeitgeber</i> , Derivation mit <i>-keit, -heit</i> etc.)	semantisch und grammatikalisch komplexe Komposita (z.B. <i>Neuanschaffungskosten</i> )
	Kasus	im Nominativ und Akkusativ	im Dativ	im Genitiv
Adjektive	Frequenz	hochfrequent ( <i>gut, lang, wenig, ...</i> )	mittelfrequent ( <i>wertvoll, gering, ...</i> )	wenig frequent ( <i>existentiell, ansässig, kostspielig, betagt, ...</i> )
	Semantik	semantisch transparente Ausdrücke ( <i>groß, breit, teuer, ...</i> )	mehrdeutige bzw. semantisch intransparente Ausdrücke ( <i>kostbar, ...</i> )	semantisch intransparente und idiomatische Ausdrücke ( <i>erschwinglich, in die Jahre gekommen, ...</i> )

Tabelle 1: (fortgesetzt)

	Kriterien	Version 1	Version 2	Version 3
Pronomen	Kasus	im Nominativ und Akkusativ	im Dativ	im Genitiv
	Unpersönliche Ausdrücke	keine	<i>jemand, man</i> + Aktiv	<i>man</i> + Passiv
Syntax	Fragen	nur direkte Fragen	auch indirekte Fragen	auch indirekte Fragen
	Satzgefüge	hauptsächlich Hauptsätze mit kanonischer Wortstellung und Koordinationen, Nebensätze mit <i>weil, dass, deshalb</i>	Hauptsätze und Nebensätze auch mit allen übrigen Konnektoren mit kanonischer Wortstellung	hauptsächlich Nebensätze, auch mit Topikalisierungen etc.
	Einschübe in Satzklammer	vermeiden/nur möglichst kurze	Satzglieder	Sätze
Textebene	Kohäsion	v. a. explizite Wiederaufnahme, implizite Wiederaufnahme durch Pronomen, wenn das Bezugsnomen eindeutig und nah ist	implizite Wiederaufnahme durch Pronomen, wenn Bezugsnomen weiter entfernt steht, Wiederaufnahme durch Synonyme	Ellipsen

Version 1 repräsentiert dabei eine sprachliche Ausformung, in der – soweit fachlich möglich – hochfrequenter und stark kollozierender Alltagswortschatz, kanonische Strukturen und möglichst explizite Bezüge zum Einsatz kommen. Version 2 dagegen realisiert u. a. typisch bildungssprachliche Wendungen, bleibt allerdings in Hinblick auf die zu verarbeitenden Elemente noch im Bereich vergleichsweise häufig auftretender und kanonischer Strukturen. Version 3 dagegen geht gezielt auf die Ebene von hoher Informationsdichte in lexikalischen und syntaktischen Strukturen, die z. T. hohe Inferenzleistungen erfordern, auf seltene und wenig antizipierbare Strukturen abzielen und damit einen breiten Wortschatz erfordern und die Verarbeitungslast erhöhen.

Wie Tabelle 1 illustriert, konkretisieren sich in den Kriterien die Prinzipien „von einfach zu komplex“, „von prototypisch bzw. eindeutig zu ungewöhnlich“ und „von hochfrequent zu selten“. Diese werden zunächst für Verben, Nomen und Adjektive anhand von konkreten Merkmalen verdeutlicht. Als Anhaltspunkt

für die Frequenz und Auftretenswahrscheinlichkeit wurde das Leipziger Korpus ([http://corpora.uni-leipzig.de/de?corpusId=deu\\_newscrawl\\_2011](http://corpora.uni-leipzig.de/de?corpusId=deu_newscrawl_2011)) zu Grunde gelegt, wobei Version 1 ca. bis zur Frequenzstufe 9 des Korpus<sup>3</sup>, Version 2 von 10 bis 13 und Version 3 ab Stufe 14 aufwärts gesetzt wurde.<sup>3</sup>

In der Verb- und Adjektivsemantik kommt in Version 1 semantisch transparenter Wortschatz zum Einsatz, also Verben mit einer verhältnismäßig klar begrenzbaren prototypischen Bedeutung wie z.B. *haben*, *tragen*, *gehen*, *kaufen* oder Adjektive wie z.B. *gut*, *lang*, *wenig*. Version 2 erlaubt hier Verben, die von der prototypischen Bedeutung abweichen oder die aufgrund bestimmter Bestandteile leicht mit anderen Bedeutungen verbunden werden könnten, wie es z.B. häufig durch Präfix- und Partikelverben geschieht (*verursachen*, *umfassen*, *vermachen*, etc., Köhne et al. 2015), oder Adjektive wie *kostbar*. Bei Version 3 setzt sich der Einsatz semantisch intransparenten Wortschatzes fort, integriert aber auch gezielt Idiome (*mit seinem Latein am Ende sein*, *blauäugig*).

Dieses Kriterium hat seine Entsprechung auch auf der Ebene der Verbmorphologie, indem hier auf einer Spanne von Simplexformen bis hin zu Präfix- und Partikelverben gedacht wird.

Hinsichtlich der Verbflexion sind in Version 1 Indikativ und Infinitiv verortet; Version 2 operiert zusätzlich mit Konjunktivformen, die einerseits im sprachlichen Input insgesamt weniger frequent und andererseits auch in ihrer Bedeutung komplexer zusammengesetzt sind. Bei Version 3 wird gezielt versucht, komplexe Verbgefüge zu bilden und Passivkonstruktionen einzubauen. In der Nominalmorphologie hat das Prinzip „von einfach zu komplex“ seine Entsprechung auf der Ebene der Komposita.

Pronomen werden im Kriterienraster noch einmal gesondert in den Blick genommen, weil durch die Verwendung von *man* bzw. *jemand* angenommen wird, dass das Ausfüllen der semantisch unspezifischen und impliziten Bedeutungseinheit mit einem höheren kognitiven Aufwand verbunden ist als die explizite Nennung eines konkreten Akteurs.

Auf syntaktischer Ebene werden die Modellprinzipien in verschiedenen Satzmustern umgesetzt: Bei Version 1 sind dies vor allem Hauptsätze in kanonischer Form; sind sie verbunden, dann durch Koordination oder durch die hochfrequenten Konnektoren *weil*, *wenn*, *dass* und *deshalb*, die laut Europäischem Referenzrahmen auch von Deutschlerner\*innen bereits auf dem Niveau A2 beherrscht

---

<sup>3</sup> Hier besteht eine Einschränkung für die Nutzung darin, dass es sich beim Leipziger Korpus um ein Zeitungskorpus handelt und damit bestimmte Lexeme genrespezifisch überrepräsentiert werden. Die hier erzielten Richtwerte wurden daher im Einzelfall durch Einschätzungen kompetenter Sprecher\*innen abgeglichen, denn es gibt gute Anhaltspunkte dafür, dass diese dann als zuverlässig betrachtet werden können (Überblick in Ellis 2002: 145ff.).

werden. Version 2 komplettiert mit allen übrigen Konnektoren und verwendet typischerweise einfache Hauptsatz-Nebensatz-Konstruktionen, bleibt aber bei der kanonischen Wortstellung. Bei Version 3 wird dies aufgebrochen: Hier finden gezielt ungewöhnliche Wortstellungen wie Topikalisierungen Verwendung und der Einsatz mehrerer untergeordneter Sätze ist die Regel.

### 3.1 Die Umsetzung der Kriterien in sprachlich variierten Fachaufgaben

Für das Untersuchungsdesign der AG FuS wurden von den Fachdidaktiker-Teams Aufgaben mit fachlichem Fokus entwickelt. Diese werden an anderer Stelle in ihrer fachlichen Einbettung eingehender beschrieben (Leiss et al. 2017). Ihnen allen ist gemeinsam, dass sie im Aufgabenstamm jeweils ein Szenario in Textform präsentieren und dass dieser Text für den Aufbau einer Problemrepräsentation und damit als Voraussetzung zur Lösung der Aufgabe verstehend gelesen werden muss.

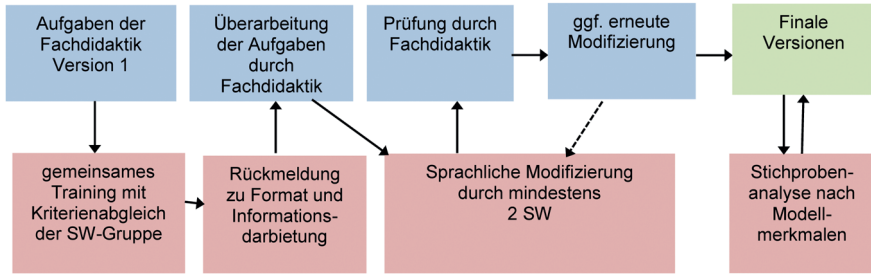
Die von den Fachdidaktiker\*innen erstellten Aufgaben wurden anschließend durch die fünf am Projekt beteiligten Sprachwissenschaftler\*innen (SW), die zur Diskussion und anschließenden Festlegung einiger basaler Kriterien zum Vorgehen zuvor ein gemeinsames Training absolvierten, sprachlich modifiziert. Die sprachlichen Versionen durchliefen während der Konstruktionsphase mehrere Aushandlungsrunden, einerseits innerhalb der sprachwissenschaftlichen Gruppe, andererseits zwischen Sprachwissenschaftler\*innen und Fachdidaktiker\*innen, sodass gewährleistet werden konnte, dass sowohl auf sprachlicher als auch auf fachlicher Ebene alle Kriterien Berücksichtigung finden (Abbildung 2). Als konsensuelles Ergebnis wurden jeweils drei sprachliche Versionen unter Rückbezug auf die Modellvorstellung erarbeitet. In den mehrtägigen Sitzungen wurde dabei gleichzeitig ein Abgleich in der Interpretation des Kriterienrasters erzielt.

Zuletzt erfolgte mittels MAXQDA<sup>4</sup> eine stichprobenhafte Auswertung der finalen Aufgabenversionen über die sprachlichen Modellmerkmale hinweg, um zu gewährleisten, dass die verschiedenen Ebenen tatsächlich zufriedenstellend umgesetzt wurden.

Die fachlichen Dimensionen wurden über die drei sprachlichen Versionen hinweg konstant gehalten. Das bedeutet, dass über die drei Versionen hinweg darauf geachtet wurde, dass keine zusätzlichen oder fehlenden Propositionen z. B. durch ausschmückende Zusätze oder Weglassungen, wesentliche Umstellungen

---

4 <https://www.maxqda.de/> (zuletzt abgerufen am 14.12.2017).



**Abbildung 2:** Beispielhafter Ablauf einer Aufgabenvariation

der Darbietungsreihenfolge, durch die die Repräsentation des Szenarios beeinflusst würde, Änderungen des Textformats (Auflistungen gegenüber Fließtext o.ä.) auftraten. Dies war u.a. deshalb wichtig, um über die Versionen hinweg in etwa vergleichbarem Maße eine gute Textgliederung und Kohäsivität beizubehalten, sowie die situative Einbettung der eigentlichen Aufgabenstellung konstant zu halten (oben, Kap. 4.2). Tabelle 2 illustriert ausschnittsweise, wie dies für die drei Versionen aussah:

**Tabelle 2:** Illustration der sprachlichen Ausformung in den drei Versionen anhand eines Ausschnittes aus Aufgabe P1

Version 1	Version 2	Version 3
Tim will, dass eine Glühlampe leuchtet. Er nimmt vier Dinge: eine Batterie, eine Glühlampe, ein kurzes Kabel und ein langes Kabel. Er verbindet alles zu einem Stromkreis. [...]	Tim hat vor, eine Glühlampe zum Leuchten zu bringen. Er baut einen Stromkreis auf, für den er vier Gegenstände miteinander verbindet: eine Batterie, eine Glühlampe sowie ein kurzes und ein langes Kabel. [...]	Mit dem Ziel, eine Glühlampe zum Leuchten zu bringen, verbindet Tim vier Einzelbauteile – eine Batterie, eine Glühlampe, ein kurzes und ein langes Kabel – zu einem Stromkreis. [...]

Wie sich in diesem Ausschnitt zeigt, realisieren alle drei Versionen dieselben Sinneinheiten.<sup>5</sup> Bestimmtes Vokabular wird dabei über die Versionen hinweg konstant gehalten: *Glühlampe*, *Batterie*, (*kurzes und langes*) *Kabel*, *Stromkreis* und auch *verbinden* sind für das fachliche Problemlöszenario notwendiges Fachvo-

<sup>5</sup> Freilich verändern sich der konnotative Gehalt und andere Bedeutungsdimensionen immer zu einem gewissen Grad, wenn andere Formen verwendet werden; so steckt beispielsweise in „etwas vorhaben“ (Version 2) noch sehr viel stärker ein Moment des Planens als in der Entsprechung „wollen“ (Version 1). Vollständige Synonymie lässt sich nicht erzielen.

kabular und werden daher bei allen drei Versionen identisch realisiert (siehe Kohnen et al. 2017 für ein vergleichbares Vorgehen in Bezug auf Fachbegriffe). Während Version 1 allerdings mit *wollen*, *nehmen* und *Dinge* hochfrequente und morphologisch einfache Lexeme realisiert, finden sich auf Version 2 als Entsprechung Partikelverben (*vorhaben*, *aufbauen*) und mit *Gegenstände* ein typisch bildungssprachliches Lexem mit mittlerer Häufigkeit. Version 3 realisiert vergleichbares Vokabular wie Version 2, tauscht aber *Gegenstände* gegen seltene und morphologisch komplexe *Einzelbauteile* aus; vor allem aber setzt es auf syntaktischer Ebene eine komplexe Konstruktion um: Was in Version 1 in kurzen Einzelsätzen mit Standardwortstellung und in Version 2 mit einfachen und verhältnismäßig kurzen Nebensatzkonstruktionen umgesetzt wird, kommt in Version 3 in einem einzigen Satzgefüge zum Ausdruck. Dieses weicht von der kanonischen Wortstellung ab (so wird das Subjekt *Tim* erst nach dem Infinitivsatz mit vorangestelltem Subjunktiv *mit dem Ziel* in der Mitte der Einheit realisiert, was u. a. mit Subjekt-Verb-Inversion einhergeht) und erfordert durch den langen Einschub in den Gedankenstrichen eine verhältnismäßig große Gedächtnisleistung.

**Tabelle 3:** Illustration der sprachlichen Ausformung in den drei Versionen anhand der Aufgabe Ma3

Version 1	Version 2	Version 3
Emil und Tarek wohnen im selben Haus. Emil war mit seinem Auto 21 Tage im Urlaub in Schweden. Tarek war zur selben Zeit mit seinem Auto in Spanien. Beide Autos sind jetzt sehr schmutzig. [...]	Emil und Tarek bewohnen dasselbe Haus. Emil hat auf einer Urlaubsreise 21 Tage lang mit seinem Auto Schweden bereist, während Tarek zur selben Zeit mit seinem Fahrzeug in Spanien unterwegs war. Daher sind beide Autos nun sehr schmutzig, [...]	Emil und Tarek sind im selben Mehrfamilienhaus wohnhaft. Nachdem Emil mit seinem Auto eine 21-tägige Urlaubsreise nach Schweden und Tarek im selben Zeitraum mit seinem Auto eine Fahrt nach Spanien unternommen hat, sind beide Fahrzeuge derzeit stark verunreinigt. [...]
Es gibt die Putzstation oder die Waschstraße. Zur Putzstation kostet die Fahrt hin und zurück 2 Euro. Sie ist 6 km entfernt. An der Putzstation machen alle Leute ihre Autos selber sauber. Das kostet nur 0,50 Euro pro Minute. [...]	Sie können die Putzstation aufsuchen oder die Waschstraße benutzen. Für die Hin- und Rückfahrt zur Putzstation fallen Fahrtkosten in Höhe von 2 Euro an – sie liegt in 6 km Entfernung. Man kann dort das Auto selber reinigen, was nur 0,50 Euro pro Minute kostet. [...]	Dazu besteht die Möglichkeit, die Putzstation in Anspruch zu nehmen oder die Waschstraße aufzusuchen. Bei erstgenannter sind Kosten für die Hin- und Rückfahrt in Höhe von 2 Euro einzuplanen. Die Anfahrtstrecke dorthin beträgt 6 km. Bei der Putzstation hat man die Möglichkeit, das Fahrzeug in Eigenarbeit zu säubern, wobei lediglich Kosten von 0,50 Euro pro Minute entstünden. [...]

Auch hier werden in Version 1 die Lexeme *wohnen*, *sein* und *schmutzig* entsprechend der Version 2 Ausdrücke *bewohnen* und *bereisen* verwendet. Diese wiederum finden Entsprechung in den Wendungen *wohnhaft sein*, *eine Reise unternehmen* und *verunreinigt sein*.

Ebenfalls zeigen sich in Version 2, aber vor allem Version 3 Wendungen, die oft mit Nominalisierungen, indirekter Kohärenz und unpersönlichen Ausdrücken einhergehen. Sowohl *entfernt sein* (Version 1) als auch *benutzen* und *selber (reignen)* (Version 2) haben in Version 2 bzw. 3 eine Entsprechung in nominalisierter Form: *Entfernung*, *in Anspruch nehmen* und *in Eigenarbeit (zu säubern)*. Sie zeigen sich häufig als Folge, wenn man gezielt versucht, nichtkanonischen Satzbau und seltene Wendungen einzubinden. Dieses Beispiel kann auch als Illustration für die elliptische Form der Wiederaufnahme in Version 3 dienen: Während in Version 1 Kohärenz durch die explizite Nennung *die Putzstation – sie erfolgt*, wird in Version 3 eine indirektere Form der Wiederaufnahme vollzogen. Hier müssen die Referenten gedanklich ergänzt werden: *Dazu besteht die Möglichkeit, die [Putzstation] in Anspruch zu nehmen oder die Waschstraße aufzusuchen. Bei [erstgenannter] [...]*.

Ein weiteres Merkmal, das erst in den Versionen 2 und 3 auftritt und welches sich in dieser Aufgabe widerspiegelt, sind unpersönliche Ausdrücke, wie das Beispiel *man*.

## 4 Abschließende Bemerkungen zur empirischen Validierung und Ausblick

Das in diesem Beitrag dargestellte Modell und seine Operationalisierungen sind als das Ergebnis theoretischer Überlegungen und als heuristische Zusammenstellung von empirischen Belegen entstanden und sollte hier auch in erster Linie als solches vorgestellt werden. Mit ihm haben wir drei sprachlich unterschiedlich anspruchsvolle Versionen derselben Fachaufgaben erstellt. Natürlich schließt sich die Frage an, inwiefern sich die verschiedenen Versionen im praktischen Einsatz tatsächlich auf die Fähigkeit von Schüler\*innen auswirken, Aufgaben zu lösen. Lösen sie die sprachlich schwierigen Aufgabenversionen weniger häufig richtig als die sprachlich einfachen Versionen?

Aufbauend auf einer ersten Pilotuntersuchung mit Aufgabenlösungen von 601 Schüler\*innen der Jahrgangsstufen 7 und 8 unterschiedlicher Schulformen, die zu einer Überarbeitung und Schärfung des Designs führte (dargestellt in Leiss et al. 2017), wurden auf der Grundlage des in diesem Beitrag dargestellten Modells und seiner Operationalisierungsdimensionen für die Fächer Mathematik, Physik,

Musik, Sport und Deutsch jeweils fünf Fachaufgaben erstellt. Diese Aufgaben griffen einzelne fachliche Inhaltsbereiche heraus und hatten über die Fächer hinweg unterschiedliche fachliche Fokussierungen (genauere Informationen dazu und zu ihrer fachlichen Einbettung finden sich in Ahlers und Seifert 2015 a, b; Gogoll 2014; Plath und Leiss 2017; Plath et al. 2014). Ihnen war jedoch gemeinsam, dass sie als Textaufgaben ausgeformt waren. Die Deutschaufgaben erforderten als Lösung einen schriftlichen Text; alle anderen Fachaufgaben waren als Single-Choice-Formate konzipiert, bei denen die Antwortalternativen über die drei Versionen hinweg identisch formuliert waren. Für jede Aufgabe wurden drei sprachliche Versionen nach der vorstehend präsentierten Modellvorstellung ausgeformt. Im Rahmen einer weiteren Testung wurden diese Aufgaben in einem Rotationsdesign (sechs verschiedene Testhefte) in 7. und 8. Klassen verschiedener Schulformen (Gymnasium, Gesamtschulen, Stadtteilschulen) und Bundesländer erprobt (N = 1.346 Schüler\*innen aus insgesamt 78 Klassen, Durchschnittsalter 14,0 Jahre). Zusätzlich zu den Testaufgaben absolvierte jede\*r Schüler\*in einen C-Test zur Erfassung der allgemeinen Sprachkompetenz im Deutschen und füllte einen Hintergrundfragebogen aus. Die Studie wird im Detail in Schwippert et al. (eingereicht) dargestellt, so dass wir an dieser Stelle auf eine kleinschrittige Darstellung verzichten.

Für die Bewertung der Schwierigkeit der einzelnen Aufgaben wurde die Lösungswahrscheinlichkeit genutzt. Durch das Rotationsdesign wurde jede Aufgabenversion von ca. 220 Schüler\*innen bearbeitet, wobei sich die bearbeitenden Schüler\*innengruppen in Bezug auf ihre Hintergrundvariablen und ihre allgemeinen sprachlichen Fähigkeiten im Deutschen nicht unterschieden.

Insgesamt zeigte sich, dass über die verschiedenen Aufgaben hinweg die Anteile korrekt gelöster Aufgaben je Fachaufgabe von Sprachversion 1 zu Sprachversion 3 in der Tendenz sinken. Die mittlere Lösungshäufigkeit über alle Items und Fächer hinweg liegt für die Sprachversion 1 bei 41 Prozent und für die Sprachversion 3 bei 38 Prozent<sup>6</sup>. Auch wenn dieses Muster nicht ganz stabil ist und sich bei einzelnen Aufgaben nicht zeigt, manifestiert sich bei den meisten Aufgaben die erwartete Abstufung hinsichtlich korrekt gelöster Aufgaben zwischen Version 1 und 3. Interessanterweise war jedoch nur bei etwas mehr als der Hälfte Version 2 auch tatsächlich „mittelschwer“.

Insgesamt bestätigen die Ergebnisse, dass die nach dem in diesem Beitrag vorgestellten Modell vorgenommene sprachliche Modellierung von fachlichen Testaufgaben durchaus unterschiedlich schwierige Aufgabenversionen hervor-

---

<sup>6</sup> Aufgrund des besonderen Formats der Deutschaufgaben beziehen sich diese Berechnungen nur auf die Mathematik-, Physik-, Sport- und Musik-Items.

bringt, indem sie gestufte Effekte auf die Lösungshäufigkeiten bewirkt. Allerdings deuten die Ergebnisse auch darauf hin, dass noch weitere Merkmale neben der sprachlichstrukturellen Ausformung bestimmend für die Aufgabenschwierigkeit sind. Um sich einem besseren Verständnis des Zusammenhangs zwischen sprachlicher Ausformung und verschiedenen fachlichen Dimensionen genauer nähern zu können, laufen derzeit Vorarbeiten für Folgeuntersuchungen der Arbeitsgruppe (<http://www.fach-und-sprache.de>).

## Literatur

- Abedi, J., Lord, C., & Plummer, J. R. 1997. *Final report of language background as a variable in NAEP mathematics performance*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Ahlers, M. & Seifert, A. 2015a. Lingua musica? Zur Erfassung musiksprachlicher Kompetenzen Jugendlicher in textbasierten Testinstrumenten. In W. Auhagen, C. Bullerjahn, H. Höge (Hrsg.), *Angewandte Musikpsychologie*, 222–237. Göttingen: Hogrefe.
- Ahlers, M. & Seifert, A. 2015b. Sprachliche Heterogenität im Musikunterricht. In A. Niessen & J. Knigge (Hrsg.), *Theoretische Rahmung und Theoriebildung in der musikpädagogischen Forschung*, 235–249. Münster: Waxmann.
- Ahrenholz, B. 2017. Sprache in der Wissensvermittlung und Wissensaneignung im schulischen Fachunterricht. In B. Lütke, I. Petersen, T. Tajmel (Hrsg.), *Fachintegrierte Sprachbildung. Forschung, Theoriebildung und Konzepte für die Unterrichtspraxis*, 1–32. Berlin: De Gruyter.
- Alderson, J. C. 1984. Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Hrsg.), *Reading in a Foreign Language*, 1–24. London: Longman.
- Alderson, J. C., Haapakangas, E., Huhta, A., Nieminen, L. & Ullakonja, R. 2014. *The Diagnosis of Reading in a Second or Foreign Language*. New York, NY: Routledge.
- Becker-Mrotzek, M., Schramm, K.; Thürmann, E. & Voller, H. J. 2013. *Sprache im Fach. Sprachlichkeit und fachliches Lernen*. Münster: Waxmann.
- Beinborn, L. 2016. Predicting and Manipulating the difficulty of text-completion exercises for language learning. Dissertation TU, Darmstadt, [http://tuprints.ulb.tu-darmstadt.de/5647/1/DissertationBeinborn\\_publishedVersion\\_September20\\_online.pdf](http://tuprints.ulb.tu-darmstadt.de/5647/1/DissertationBeinborn_publishedVersion_September20_online.pdf) [letzter Zugriff 15.05.2018]
- Berendes, K., Dragon, N., Sabine, W., Heppt, B. & Stanat, P. 2013. Hürde Bildungssprache? Eine Annäherung an das Konzept „Bildungssprache“ unter Einbezug aktueller empirischer Forschungsergebnisse. In A. Redder & S. Weinert (Hrsg.), *Sprachförderung und Sprachdiagnostik: Interdisziplinäre Perspektiven*, 17–41. Münster: Waxmann.
- Berendes, K., Vajjala, S., Meurers, D., Bryant, D., Wagner, W., Chinkina, M., & Trautwein, U. 2017. Reading Demands in Secondary School: Does the Linguistic Complexity of Textbooks Increase With Grade Level and the Academic Orientation of the School Track? *Journal of Educational Psychology*. Advance online publication. doi:10.1037/edu0000225
- Berndt, R. S., Mitchum, C., Burton, M. & Haendiges, A. 2004. Comprehension of reversible sentences in aphasia: The effects of verb meaning. *Cognitive Neuropsychology* 21(2). 229–245.

- Bernstein, B. 1964. Elaborated and restricted codes: Their social origins and some consequences. *American Anthropologist* 66(6/2). 55–69.
- Bird, E. & Welford, G. 1995. The effect of language on the performance of second-language students in science examinations. *International Journal of Science Education* 17(3). 389–397.
- Bock, B.M., Lange, D. & Fix, U. 2017. Das Phänomen “Leichte Sprache” im Spiegel aktueller Forschung – Tendenzen, Fragestellungen und Herangehensweisen. In B. M. Bock, U. Fix, & D. Lange (Hrsg.), „Leichte Sprache“ im Spiegel theoretischer und angewandter Forschung, 11–34. Berlin: Frank & Timme.
- Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L, Genc, S., Philipp, M., Roehm, D., & Schlesewsky, M. 2011. Think Globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language* 117(3). 133–152.
- Bredel, U. & Maas, C. 2016. *Ratgeber Leichte Sprache: Die wichtigsten Regeln und Empfehlungen für die Praxis*. Berlin: Dudenverlag.
- Breindl, E., & Waßner, U. H. 2006. Syndese vs. Asyndese. Konnektoren und andere Wegweiser für die Interpretation semantischer Relationen in Texten. In H. Blühdorn, E. Breindl & U. H. Waßner (Hrsg.), *Text-Verstehen. Grammatik und darüber hinaus*, 46–70. Berlin: Walter de Gruyter.
- Bryant, D. & Noschka, N. 2015. Personal- und Demonstrativpronomen im Sprachverstehensprozess: Untersuchungen zum Erwerb funktionaler Anapherndistribution bei DaM, DaF und DaZ. In H. Klages & G. Pagonis (Hrsg.), *Linguistisch fundierte Sprachförderung und Sprachdidaktik*, 17–46. Berlin: Walter de Gruyter.
- Bulté, B. & Housen, A. 2012. Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Hrsg.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 21–46. Amsterdam: Mouton de Gruyter.
- Christmann, U. & Groeben, N. 2001. Psychologie des Lesens. In B. Franzmann, K. Hasemann, D. Löffler & E. Schön (Hrsg.), *Handbuch Lesen*, 145–223. Baltmannsweiler: Schneider.
- Cornelis, L. 1996. English and Dutch: The passive difference. *Language Sciences* 18. 247–64.
- Cummins, J. 2008. BISC and CALP: Empirical and Theoretical Status of the Distinction. In B. Street & N. H. Hornberger (Hrsg.), *Encyclopedia of Language and Education*, Bd. 2, 71–83. New York: Springer.
- de Graaff, R. & Housen, A. 2009. Investigating the effects and effectiveness of L2 instruction. In M. Long & C. Doughty (Hrsg.), *The handbook of language teaching*, 726–755. Oxford: Blackwell.
- Diehl, E. 1999. Schulischer Grammatikerwerb unter der Lupe. Das Genfer DiGS [Deutsch in Genfer Schulen-Projekt]. *Bulletin suisse de linguistique appliquée* 70. 7–26.
- Diessel, H. 2004. *The Acquisition of Complex Sentences*. Cambridge: Cambridge University Press.
- Dimroth, C. & Schimke, S. 2012. Der Erwerb der Fintheit im Deutschen: Ein Vergleich von kindlichen und erwachsenen L2 Lernern. In B. Ahrenholz & W. Knapp (Hrsg.), *Sprachstand erheben – Spracherwerb erforschen*, 287–306. Freiburg: Fillibach.
- Ellis, R. 2002. The Place of Grammar Instruction in the Second/Foreign Language Curriculum. In E. Hinkel & S. Fotos (Hrsg.), *New Perspectives on Grammar Teaching in Second Language Classrooms*, 14–34. London: Routledge.
- Ellis, R. & Shintani, N. 2014. *Exploring Language Pedagogy through Second Language Acquisition Research*. London, UK: Routledge.
- Fang, Z., Schleppegrell, M. J. & Cox, B. E. 2006. Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research* 38. 247–273.

- Feilke, H. 2012. Bildungssprachliche Kompetenzen – fördern und entwickeln. *Praxis Deutsch* 39 (233). 4–13.
- Freedle, R. & Kostin, I. 1999. Does the Text Matter in a Multiple-Choice Test of Comprehension? The Case for the Construct Validity of TOEFL's Minitalks. *Language Testing* 16(1). 2–32.
- Fusté-Herrmann, B. 2008. *Idiom comprehension in bilingual and monolingual adolescents*. University of South Florida Scholar Commons. <http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1255&context=etd> [letzter Zugriff 26.06.2017].
- Gabler, K. 2013. *Wortstellungsvariation im Deutschen: Psycholinguistische Untersuchungen zur Relativsatzposition*. Dissertation. Philipps-Universität Marburg <http://archiv.ub.uni-marburg.de/diss/z2014/0086/pdf/dkg.pdf> [letzter Zugriff 13.09.2017].
- Gantefort, C. 2013. *Schriftliches Erzählen mehrsprachiger Kinder. Entwicklung und sprachübergreifende Fähigkeiten*. Münster: Waxmann.
- Gibson, E. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76.
- Goepferich, S. 2008. Textverstehen und Textverständlichkeit. In N. Janich (Hrsg.), *Textlinguistik: 15 Einführungen*, 291–312. Tübingen: Narr.
- Gogolin, I. & Lange, I. 2011. Bildungssprache und durchgängige Sprachbildung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit*, 107–127. Wiesbaden: VS Verlag.
- Gogolin, I., Duarte, J. 2016. Bildungssprache. In J. Kilian, B. Brouer, D. Lüttenberg (Hrsg.), *Handbuch Sprache in der Bildung*, 478–499. Berlin & Boston: Mouton de Gruyter.
- Gogoll, A. 2014. Das Modell der sport- und bewegungskulturellen Kompetenz und seine Implikationen für die Aufgabenkultur im Sportunterricht. In M. Pfitzner (Hrsg.), *Aufgaben im Sportunterricht – Konzepte und Forschungsbefunde zum kompetenzorientierten Sportunterricht*, 93–110. Wiesbaden: VS-Verlag.
- Grabe, W. 2009. *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M. & Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods* 36(2). 193–202.
- Grießhaber, W. 2005. *Sprachstandsdiagnose im Zweitspracherwerb: Funktional-pragmatische Fundierung der Profilanalyse*. URL: <http://spzwww.uni-muenster.de/griesha/pub/tprofilanalyse-azm-05.pdf> [letzter Zugriff 13.09.2017].
- Grotjahn, R. 2000. Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TestDaF. In S. Bolton (Hrsg.), *TestDaF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar*, 7–55. Köln: Gilde.
- Gunning, R. (1969). The fog index after twenty years. *Journal of Business Communication* 6(2). 3–13.
- Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., Scaffidi, N. T. & Tonks, S. 2004. Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology* 96(3). 403–423.
- Haag, N. 2015. *Differenzielle Validität von Mathematiktestaufgaben für Kinder mit nicht-deutscher Familiensprache. Welche Rolle spielt die sprachliche Komplexität der Aufgaben?* Dissertation, HU Berlin, <https://edoc.hu-berlin.de/handle/18452/18050> [letzter Zugriff 10.09.2017].
- Hancke, J., Vajjala, S., & Meurers, D. 2012. Readability Classification for German using Lexical, Syntactic, and Morphological Features. *Proceedings of COLING*. 1063–1080.

- Härtig, H. & Kohnen, N. 2017. Die Rolle der Termini beim Lernen mit Physikschulbüchern. In B. Ahrenholz, B. Hövelbrinks & C. Schmellentin (Hrsg.), *Fachunterricht und Sprache in schulischen Lehr-/Lernprozessen*, 55–72. Tübingen: Narr.
- Hartung, O. 2013. *Geschichte – Schreiben – Lernen. Empirische Erkundungen zum konzeptionellen Schreibhandeln im Geschichtsunterricht*. Münster: Lit Verlag.
- Hawkins, J. A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Heppt, B., Haag, N., Böhme, K. & Stanat, P. 2015. The Role of Academic-Language Features for Reading Comprehension of Language-Minority Students and Students from Low-SES Families. *Reading Research Quarterly* 50(1). 61–82.
- Housen, A. 2014. Difficulty and Complexity of Language Features and Second Language Instruction. In C. A. Chappelle (Hrsg.), *Encyclopedia of Applied Linguistics*, 1–7. Oxford: Blackwell.
- Hulstijn, J. H. 2015. *Language proficiency in native and non-native speakers: Theory and research*. Amsterdam: John Benjamins.
- Irujo, S. 1993. Steering Clear: Avoidance in the Production of Idioms. *International Review of Applied Linguistics in Language Teaching* 31(3). 205–220.
- Jeon, E. H. & Yamahita, J. 2014. L2 Reading Comprehension and Its Correlates: A Meta-Analysis. *Language Learning* 64. 160–212.
- Kharkwal, G. & Stromswold, K. 2014. Good-enough language processing: evidence from sentence-video matching. *Journal of Psycholinguistic Research* 43(1). 27–43.
- Kieffer, M. J., Lesaux, N. K., Rivera, M. & Francis, D. J. 2009. Accommodations for English Language Learners Taking Large-Scale Assessments: A Meta-Analysis on Effectiveness and Validity. *Review of Educational Research* 79(3). 1168–1201.
- King, J. & Just, M. A. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30. 580–602.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L. & Chissom, B. S. 1975. *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel*. Institute for Simulation and Training 56 (<http://stars.libraryucf.edu/istlibrary/56>, letzter Zugriff 28.02.2018)
- Kliegl, R., Grabner, E., Rolfs, M. & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology* 16. 262–284.
- Kniffka, G. & Linnemann M. 2016. Zur Ermittlung der Schwierigkeit (fach-)sprachlicher Strukturen mittels fachlicher C-Tests – Eine explorative Studie. In E. Tschirner, O. Bärenfänger & J. Möhring (Hrsg.), *Deutsch als fremde Bildungssprache. Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik*, 163–178. Tübingen: Stauffenburg.
- Koch, P. & Oesterreicher W. 1985. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36(85). 15–43.
- Köhne, J., Kronenwerth, S., Redder, A., Schuth, E. & Weinert, S. 2015. Bildungssprachlicher Wortschatz – linguistische und psychologische Fundierung und Itementwicklung. In A. Redder, J. Naumann & R. Tracy (Hrsg.), *Forschungsinitiative Sprachdiagnostik und Sprachförderung – Ergebnisse*, 67–92. Münster, New York: Waxmann.
- Kohlen, N., Bernholt, S., Retelsdorf, J. & Härtig, H. 2017. Leichte Sprache im Physikunterricht. In B. M. Bock, U. Fix, & D. Lange (Hrsg.), „Leichte Sprache“ im Spiegel theoretischer und angewandter Forschung, 337–341. Berlin: Frank & Timme.
- Konieczny, L. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29(6). 627–645.

- Konieczny, L. & Döring, P. 2003. Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. *Proceedings of ICCS/ASCS*, 330–335. Sydney: University of New South Wales.
- Langer, I., Schulz von Thun, F. & Tausch, R. 1974. *Verständlichkeit in Schule, Verwaltung, Politik und Wissenschaft*. München, Basel: Reinhardt.
- Leisen, J. 2013. *Handbuch Sprachförderung im Fach*. 2., unv. Aufl. Stuttgart: Klett.
- Leiss, D., Domenech, M., Ehmke, T. & Schwippert, K. 2017. Schwer – schwierig – diffizil: Zum Einfluss sprachlicher Komplexität von Aufgaben auf fachliche Leistungen in der Sekundarstufe I. In D. Leiss, M. Hagen, A. Neumann, K. Schwippert (Hrsg.), *Mathematik und Sprache: Empirischer Forschungsstand und unterrichtliche Herausforderungen*, 99–125. Münster: Waxmann.
- Levy, R., & Keller, F. 2013. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 68(2). 199–222.
- Levy, R., Fedorenko, E., Breen, M. & Gibson, T. 2012. The processing of extraposed structures in English. *Cognition* (122). 12–36.
- Long, M. & Robinson, P. 1998. Focus on form: Theory, research and practice. In C. Doughty & J. Williams (Hrsg.), *Focus on form in classroom second language acquisition*, 15–63. Cambridge: Cambridge University Press.
- Lüdemann, E. & Schwerdt, G. 2013. Migration Background and Educational Tracking. Is there a Double Disadvantage for Second-Generation Immigrants? *Journal of Population Economics* 26(2). 455–481.
- Maas, C. 2015. *Leichte Sprache. Das Regelbuch*. Berlin: LiT-Verlag.
- Mesmer, H. A., Cunningham, J. W. & Hiebert, E. H. 2012. Toward a Theoretical Model of Text Complexity for the Early Grades: Learning From the Past, Anticipating the Future. *Reading Research Quarterly* 47(3). 235–258.
- Morek, M. & Heller, V. 2012. Bildungssprache – kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für angewandte Linguistik* 57(1). 67–101.
- Niederhaus, C., Pöhler, B. & Prediger, S. 2016. Relevante Sprachmittel für mathematische Textaufgaben – Korpuslinguistische Annäherung am Beispiel Prozentrechnung. In E. Tschirner, O. Bärenfänger & J. Möhring (Hrsg.), *Deutsch als fremde Bildungssprache. Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik*, 135–161. Tübingen: Stauffenberg.
- Oleschko, S. 2017. Differenzielle Lernmilieus und Sprachenbildung – zur Bedeutung der Sprachsoziologie für den Diskurs um Sprachfähigkeit in der Schule. In B. Lütke, I. Petersen, & T. Tajmel (Hrsg.), *Fachintegrierte Sprachbildung. Forschung, Theoriebildung und Konzepte für die Unterrichtspraxis*, 51–68. Berlin: De Gruyter.
- Pienemann, M. 1989. Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics* 10. 52–79.
- Plath, J., Leiss, D. 2017. The impact of linguistic complexity on the solution of mathematical modelling tasks. *Zentralblatt der Didaktik der Mathematik*. <https://doi.org/10.1007/s11858-017-0897-x>
- Plath, J., Leiss, D., Schwippert, K. & Neumann, A. 2014. Das verstehe ich nicht! Eine Untersuchung zur Konstruktion des Situationsmodells. *Beiträge zum Mathematikunterricht*. 903–906.
- Pohl, T. 2016. Die Epistemisierung des Unterrichtsdiskurses – ein Forschungsrahmen. In E. Tschirner, O. Bärenfänger, & J. Möhring (Hrsg.), *Deutsch als fremde Bildungssprache. Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik*, 55–80. Tübingen: Stauffenberg.

- Prophet, R. B. & Badede, N. B. 2009. Language and student performance in Junior Secondary Science examinations: The case of second language learners in Botswana. *International Journal of Science and Mathematics Education* 7. 235–251.
- RAND Reading Study Group. 2002. Reading for understanding: Toward an R&D program in reading comprehension. Santa Monica etc.: RAND education.
- Reich, H. 2013. Durchgängige Sprachbildung. In I. Gogolin, I. Lange, U. Michel & H. H. Reich (Hrsg.), *Herausforderung Bildungssprache – und wie man sie meistert*, 55–70. Münster: Waxmann.
- Rincke, K. 2010. Alltagssprache, Fachsprache und ihre besondere Bedeutung für das Lernen. *Zeitschrift für Didaktik der Naturwissenschaften* 16, 235–260.
- Runge, A. 2013. Die Nutzung von (bildungssprachlichen) Verben in naturwissenschaftlichen Aufgabenstellungen bei SchülerInnen der Jahrgangsstufen 4 und 5. In A. Redder und S. Weirner (Hrsg.), *Sprachförderung und Sprachdiagnostik. Interdisziplinäre Perspektiven*, 152–173. Münster, New York: Waxmann.
- Schleppegrell, M. J. 2010. *The Language of Schooling. A Functional Linguistics Perspective*. New York: Routledge.
- Schmidt, J. E. 1993. *Die deutsche Substantivgruppe und die Attribuierungskomplikation*. Tübingen: Niemeyer.
- Schmitz, A. & Gräsel, C. 2016. Bei welchen Lernenden fördert globale Textkohäsion das Verstehen von Sachtexten? Eine Studie zu Wechselwirkungen zwischen globaler Textkohäsion und kognitiven Verständnisvoraussetzungen. *Unterrichtswissenschaft* 3. 267–281
- Schramm, K., Hardy, I. Saalbach, H. & Gadow, A. 2013. Wissenschaftliches Begründen im Sachunterricht. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen*, 295–314. Münster, New York: Waxmann.
- Schwarm, S. E. & Ostendorf, M. 2005. Reading level assessment using support vector machines and statistical language models. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. 523–530. University of Michigan.
- Schwippert, K., Leiss, D., Ehmke, T., Höttecke, D., Heine, L., & Neumann, A. (eingereicht). *Empirische Fundierung der Untersuchung von sprachvariieren Fachaufgaben*.
- Scontras, G., Tsai, E., Mai, K. & Polinsky, M. 2014: Chinese scope: An experimental investigation. In U. Etxeberria, A. Fălăuș, A. Irurtzun & B. Leferman (Hrsg.), *Proceedings of Sinn und Bedeutung* 18. Harvard: Harvard University Press.
- Solomyak, O. & Marantz, A. 2010. Evidence for early morphological decomposition in visual word recognition: A single-trial correlational MEG study. *Journal of Cognitive Neuroscience* 22. 2042–2057.
- Stahns, R. 2016. Bildungssprachliche Merkmale von Texten und Items: Zur Operationalisierung des Konstrukts „Bildungssprache“. *Didaktik Deutsch* 41. 44–55.
- Štekauer, P. 2005. *Meaning Predictability in Word Formation*. Amsterdam, New York: John Benjamins.
- Street, J. A. & Dabrowska, E. 2010. More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua* 120 (8). 2080–2094.
- Staraushek, E. 2006. Der Einfluss von Textkohäsion und gegenständlichen externen piktoralen Repräsentationen auf die Verständlichkeit von Texten zum Physiklernen. *Zeitschrift für Didaktik der Naturwissenschaften* 12. 127–157.
- Sumfleth, E. & Schüttler, S. 1995. Linguistische Textverständlichkeitskriterien. Helfen Sie bei der Darstellung chemischer Inhalte? *Zeitschrift für Didaktik der Naturwissenschaften* 1. 55–72.

- Trim, J., North, B. & Coste, D. 2001. *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. München: Klett-Langenscheidt.
- Tschirner, E.; Bärenfänger, O. & Möhring, J. (Hrsg.). 2016. *Deutsch als fremde Bildungssprache: Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik*. Tübingen: Stauffenburg.
- Vajjala, S., & Meurers, D. 2012. On improving the accuracy of readability classification using insights from second language acquisition. *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, 163–173. Association for Computational Linguistics.
- Van Dijk, T. A. & Kintsch, W. 1983. *Strategies of discourse comprehension*. New York: Academic Press.
- Verhoeven, L. & Perfetti, C. 2008. Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology* 22(3). 293–301.
- Vollmer, H. J. & Thürmann, E. 2010. Zur Sprachlichkeit des Fachlernens: Modellierung eines Referenzrahmens für Deutsch als Zweitsprache. In B. Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache*. 2. Auflage, 107–132. Tübingen: Narr.
- vor der Brück, T., Hartumpf, S. & Helbig, H. 2008. A readability checker with supervised learning using deep indicators. *Informatica*, 32(4). Abdruck ohne Seitenangaben.
- Zwaan, R. A. & Radvansky, G. A. 1998. Situation models in language comprehension and memory. *Psychological bulletin*, 123(2). 162–185.