

Semantisch-lexikographische Aspekte deverbaler *-ung-* Nominalisierungen: Grundlagen und Beschreibungsmöglichkeiten

Iván Arias-Arias, *Abteilung für Anglistik
und Germanistik, Instituto da Lingua Galega,
Universidade de Santiago de Compostela, Spanien*
(ivanarias.arias@usc.gal)
(<https://orcid.org/0000-0003-2673-0899>)

Zusammenfassung: Zeitgenössische digitale Wörterbücher weisen erhebliche Defizite in der systematischen Darstellung morpho-lexikalischer Informationen auf. Diese Studie untersucht anhand deverbaler *-ung-*Nomina, wie korpusgestützte Argumentstruktur-Analysen deren lexikographische Darstellung verbessern können. Basierend auf einem semantisch annotierten Korpus wurden 200 *-ung-*Nominalisierungen quantitativ und qualitativ analysiert. Die Ergebnisse führen zu zwei zentralen Befunden: (a) Reguläre Polysemie: Die morphologischen Bildungsmechanismen erklären die systematische Polysemie zwischen Prozess- (*nomen actionis*) und Resultatlesarten (*nomen acti*), deren Zusammenhang in bestehenden monolingualen Wörterbüchern nicht in Verbindung mit dem Affix *-ung* präsentiert wird; (b) Stabile Argumentstruktur: Die semantische Argumentstruktur (vom *nomen actionis*) ist durch ein stabiles „Minimalprogramm“ geprägt (das Argument für das Patiens/Thema und das Argument für die Ursache der Veränderung). Die Studie schlägt daraufhin ein optimiertes Modell vor, das diese expliziten Argumentstruktur-Muster und ihre semantischen Funktionen integriert. Dieser systematische Ansatz ist entscheidend für fortgeschrittene DaF-Lernende, da er die produktive Bildung von Nominalisierungen ermöglicht. Die Arbeit demonstriert somit, dass eine systematische morpho-lexikalische Aufbereitung den wachsenden Bedarf an strukturierter digitaler Lexikographie bedient und den Fremdspracherwerb unterstützt.

Stichwörter: DEVERBALE NOMINA, *-UNG-*NOMINALISIERUNGEN, ARGUMENTSTRUKTUR, DIGITALE LEXIKOGRAPHIE, BEDÜRFNISSE DER BENUTZER:INNEN, KORPUSGESTÜTZTE VERFAHREN, REGULÄRE POLYSEMIE

Abstract: Semantic-Lexicographic Aspects of Deverbal *-ung-*Nominalizations in German: Theoretical Foundations and Descriptive Possibilities. Contemporary digital dictionaries exhibit significant shortcomings in the systematic representation of morpho-lexical information. This study uses German deverbal *-ung-*nominalisations to examine how corpus-based argument structure analyses can improve their lexicographic documentation. Based on a semantically annotated corpus, 200 *-ung-*nominalisations were analysed both quantitatively and qualitatively. The results lead to two central findings: (a) Regular polysemy: The morphological formation mechanisms explain the systematic polysemy between process (*nomen actionis*)

and result (*nomen acti*) readings, a relationship whose connection to the *-ung* affix is not presented in existing monolingual dictionaries; (b) Stable argument structure: The semantic argument structure (of the *nomen actionis*) is characterised by a stable "minimal program" (the argument for the patient/theme and the argument for the cause of change). The study consequently proposes an optimised model that integrates these explicit argument structure patterns and their semantic functions. This systematic approach is crucial for advanced learners of German (DaF), as it enables the productive formation of nominalisations. The work thus demonstrates that a systematic morpho-lexical presentation addresses the growing need for structured digital lexicography and supports foreign language acquisition.

Keywords: DEVERBAL NOUNS, *-UNG*-NOMINALISATIONS, ARGUMENT STRUCTURE, DIGITAL LEXICOGRAPHY, USERS' NEEDS, CORPUS-BASED METHODS, REGULAR POLYSEMY

1. Einleitung

Die lexikographische Praxis der Gegenwart sieht sich weiterhin mit grundlegenden Herausforderungen der digitalen Ära konfrontiert, wie in den Villa-Vigoni-Thesen prägnant formuliert wurde (vgl. Schierholz 2021). Zwei zentrale Problemfelder, die eine intensive Diskussion unter Expert:innen erfordern, betreffen zum einen „die geordnete Zusammenführung von automatisch aus Textkorpora erzeugten und gezielt aufbereiteten Daten sowie deren benutzerorientierten Präsentation“ und zum anderen die Ausrichtung „an den spezifischen Bedürfnissen der Benutzer:innen [...] und deren sprachlichen Handlungen bzw. kommunikativen Absichten“ (Schierholz 2021: 305).

Während die Implementierung KI-basierter Systeme in lexikographische Prozesse derzeit im Mittelpunkt des wissenschaftlichen Interesses steht, erhalten diese beiden benutzerorientierten fundamentalen Aspekte paradoxerweise weniger Beachtung. Zahlreiche wissenschaftliche Arbeiten untersuchen, wie künstliche Intelligenz die aufwändige Arbeit von Lexikograph:innen effizienter gestalten und wie entsprechende Daten automatisch gewonnen werden können (vgl. z. B. Lew 2024; de Schryver 2023 oder Fuertes-Olivera 2024). Obwohl diese Untersuchungen von höchstem Interesse sind, da sie zentrale Bereiche wie die Produktionssteigerung in der lexikographischen Arbeit oder die Integration von Chatbots in verschiedene linguistische Anwendungsfelder behandeln, in denen traditionell Wörterbücher verwendet werden, bleibt die kontinuierliche Verbesserung digitaler lexikologischer Informationssysteme eine dringende Notwendigkeit. Wörterbücher spielen nach wie vor eine wesentliche Rolle in der heutigen Gesellschaft und werden keineswegs obsolet (vgl. Arias-Arias et al. 2025 oder Kosem et al. 2019).

Ein wesentlicher Aspekt, der in Wörterbüchern — im Einklang mit den oben erwähnten Villa-Vigoni-Thesen (Schierholz 2021) — noch weiterentwickelt und optimiert werden muss, ist die adäquate Präsentation von Daten aus der morphologisch-lexikalischen Schnittstelle. Deverbale Nomina, wie beispielsweise die auf *-ung*, sind im Deutschen hochproduktiv (Zifonun et al. 1997) und

stellen eine erhebliche Herausforderung im DaF-Unterricht dar. Sowohl die Darstellung semantischer Informationen in Bezug auf lemmatisierte Affixe als auch die Integration von Informationen zur Argumentstruktur bei den entsprechenden lemmatisierten Nomina bedürfen einer eingehenden Diskussion.

Diese Studie widmet sich daher zwei Leitfragen:

1. Inwiefern kann die automatische Extraktion von Daten aus semantisch annotierten Korpora bei der Erfassung lexikologischer Informationen für *-ung*-Nomina unterstützend wirken?
2. Lässt sich eine benutzerfreundliche Organisation der Informationen vorschlagen, die den Bedürfnissen potenzieller Nutzer:innen in diesem Bereich gerecht wird?

Der vorliegende Beitrag gliedert sich wie folgt: In Abschnitt 2 wird analysiert, wie morpho-lexikalische Informationen in monolingualen digitalen Wörterbüchern derzeit dargestellt werden, und es wird begründet, weshalb ihre systematische Integration für DaF-Lernende wesentlich ist. Abschnitt 3 bietet eine theoretische Grundlage zum valenziellen Verhalten von *-ung*-Nominalisierungen und erläutert, wie ihre semantische Struktur bei der lexikographischen Beschreibung berücksichtigt werden sollte. Abschnitt 4 stellt die methodologische Herangehensweise dar: eine korpusgestützte Methode zur (semi)automatischen Extraktion von Argumentstrukturen, deren Validität durch statistische Tests überprüft wird. Abschnitt 5 präsentiert die empirischen Ergebnisse der linguistischen Analyse, die in Abschnitt 6 in einen Vorschlag eines verbesserten mikrostrukturellen Wörterbuchartikel-Modells übersetzt werden.

2. Morphologisch-lexikologische Informationen in monolingualen Wörterbüchern: wozu?

Wortbildungsprozesse verdienen besondere Beachtung, wenn es um deren Darstellung in digitalen Wörterbüchern geht. Eine umfassende Auseinandersetzung mit dieser Problematik findet sich bei Klosa-Kückelhaus (2013a), die betont, dass eine systematische Erfassung morphologischer Phänomene wie Komposition oder Derivation die Vernetzungen im Wortschatz aufzeigt und die lexikalische Beschreibung präzisiert. Die Integration solcher grammatischen Regeln wird in vielen Wörterbüchern durch Lemmatisierung und durch Elemente der Mediostruktur realisiert (vgl. Ulsamer 2013: 22).

Angaben zu Wortbildungsprozessen sowie die Lemmatisierung von Wortelementen fördern den Aufbau von Vernetzungen im mentalen Lexikon sowohl bei Muttersprachler:innen als auch bei Lernenden. Für DaF-Lernende ist dieser Aspekt von besonderer Relevanz, da das Phänomen der Nominalisierung — die Bildung von Nomina ausgehend von Verben durch morphologische Derivationsprozesse — sowie der Nominalstil auf höheren Kompetenzstufen intensiv behandelt werden (vgl. z. B. Buscha et al. 2016 oder Füllemann 2023). Dabei werden

sowohl grammatische Regularitäten als auch semantische Kernkomponenten der Argumentstruktur thematisiert, insbesondere die Vererbung (oder Nichtvererbung) von Argumentstellen vom verbalen in den nominalen Bereich.

Monolinguale Wörterbücher werden von DaF-Lernenden potenziell in drei verschiedenen Situationen zu Rate gezogen, wenn sie auf Informationen bezüglich morphologischer Prozesse zugreifen möchten: bei der Rezeption, Produktion und Wissensbildung (vgl. Bergenholtz und Gouws 2013). Was die Rezeption anbelangt, können sich nach Bergenholtz und Gouws (2013: 71) lediglich erfahrene Benutzer:innen über lemmatisierte Affixe Informationen erschließen, die allen durch einen bestimmten Wortbildungsprozess entstandenen Wörtern gemeinsam sind. Beim Lesen eines Textes müssten DaF-Lernende in einer derartigen Situation zwei aufeinanderfolgende mentale Prozesse vollziehen: (a) zum einen die Identifikation der Komponenten des Derivats, beispielsweise die Erkennung des Stamms und der Affixe (wie etwa $\{Er\}\{\text{leicht}\{er\}\}\{ung\}$), und (b) zum anderen die Erschließung der Stammbedeutung (in dem Beispiel *leicht* bzw. *erleichtern*).

In Bezug auf die Produktion in der Fremdsprache stellen Bergenholtz und Gouws (2013: 73) fest, dass idealerweise alle Wortbildungsprodukte als Lemmata im lexikologischen Informationssystem verzeichnet werden sollten. Wenn ausreichende Informationen im Wörterbuch sowohl aus grammatischer als auch aus lexikologischer Perspektive dargestellt werden, können die Nutzer:innen Affixe zur Bildung neuer Wortschöpfungen in ihren eigenen Texten verwenden. In ähnlicher Weise können Angaben zu Wortbildungsprozessen kognitiv-gesteuerte Situationen unterstützen, obwohl diese Informationen eher in Referenzgrammatiken gesucht werden, da die Voraussetzungen der Nutzer:innen je nach Vorkenntnissen sehr unterschiedlich sein können — seien es Linguistik-Studierende oder Laien (vgl. Bergenholtz und Gouws 2013: 74).

Im Mittelpunkt dieses Abschnittes steht die kritische Auseinandersetzung mit allgemeinen monolingualen Wörterbüchern in Hinblick auf das beschriebene Problem. Konkret wird der Frage nachgegangen, wie die deverbalen Nomina auf *-ung* in monolingualen lexikologischen Informationssystemen dargestellt werden. Das *Digitale Wörterbuch der Deutschen Sprache* (DWDS) lemmatisiert das Suffix *-ung* und verzeichnet Informationen zur Grammatik der entsprechenden Wortbildung, wie das Genus, die durch den Derivationsprozess resultierende Wortart und die Pluralform. Außerdem wird als mediostrukturelle Verbindung zu den entsprechenden Wörterbuchartikeln eine Liste von lexikalischen Einheiten dargeboten, die auf *-ung* enden (siehe Abbildung 1). Dennoch mangelt es dem Wörterbuchartikel zu dem Suffix an semantischen Aspekten, wie einer Bedeutungsparaphrase oder Informationen zum semantischen Verhalten und zu Mustern, in denen Prädikate auf *-ung* eingebettet sein können. Auch im Fall von Affixen müssen Informationen einbezogen werden, die von potenziellen Nutzer:innen benötigt werden können (vgl. Fuertes-Olivera 2025: 162).

Der *Duden* inkludiert das Suffix *-ung* ebenso in seiner Lemmaliste, bietet aber im Wörterbuchartikel weniger Informationen zur Grammatik der sich daraus ergebenden Wortbildungen als das DWDS. Ein Beispiel für ein Derivat wird

genannt, aber eine exhaustive Liste von Wortbildungen oder Angaben zum semantischen Verhalten von Derivaten auf *-ung* werden nicht angeboten.

Das lexikographische Portal *lexiko* verzeichnet das Suffix nicht als Lemma, verweist dennoch auf das *Grammatische Informationssystem (grammis)* des Instituts für Deutsche Sprache (vgl. Klosa-Kückelhaus 2013b: 180), wo ausführlichere Informationen zu dem Suffix dargeboten werden. Hier werden semantische Aspekte aus kategoriemäßiger Perspektive berücksichtigt, denn es wird festgestellt, dass durch *-ung*-Nomina Tätigkeiten beschrieben werden und dass daraus teilweise Bezeichnungen für Sachen und Sachverhalte entstehen.

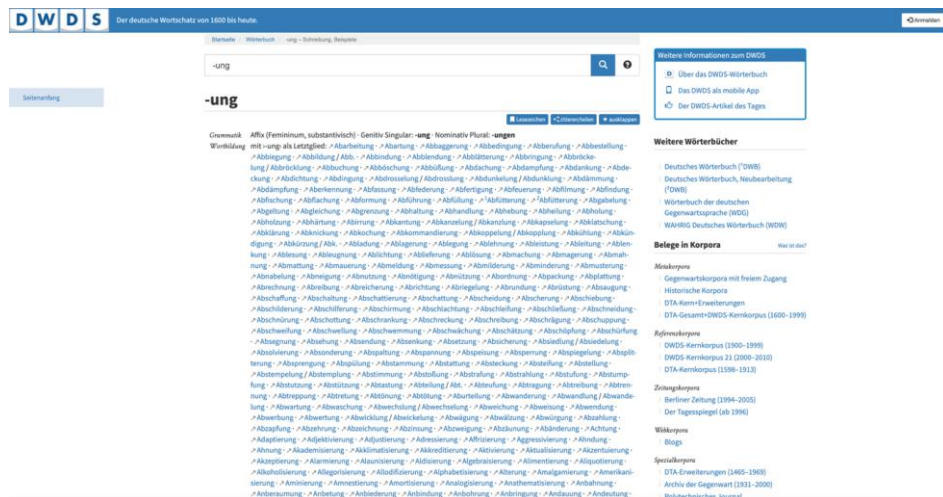


Abbildung 1: Wörterbuchartikel zum Suffix *-ung* in DWDS

Die angeführte Analyse der drei deutschsprachigen Wörterbuchsysteme offenbart ein strukturelles Defizit, das nicht auf diese Ressourcen beschränkt ist, sondern eine systematische Lücke in der lexikographischen Praxis widerspiegelt. Parallel zur allgemeinen Wörterbuchtradition hat sich eine eigenständige Valenzlexikographie entwickelt, die sich explizit der Beschreibung von Argumentstrukturen widmet. Im deutschsprachigen Raum markiert das *Wörterbuch zur Valenz und Distribution deutscher Substantive* von Sommerfeldt und Schreiber (1977) den Ausgangspunkt dieser Tradition für den nominalen Bereich. Diese hat bis heute maßgebliche Weiterentwicklungen erfahren: Die *Erlangen Valency Patternbank* (Herbst und Uhrig 2009) bezieht deutsche Nomina mit Präpositivergänzungen systematisch ein, und multilinguale Ressourcen wie *Portlex* (Domínguez Vázquez und Valcárcel Riveiro 2019) erweitern den Ansatz auf weitere Sprachen¹.

Diese Valenzressourcen beschreiben Argumentstrukturen jedoch ausschließlich auf der Ebene des einzelnen Lemmas und abstrahieren nicht über den morphologischen Bildungsprozess. Ein Eintrag für *Verhaftung* in einem Valenz-

wörterbuch dokumentiert dessen spezifische Argumentstellen, ohne dass diese Information mit dem produktiven Muster des *-ung*-Suffixes in Verbindung gebracht wird. Der morphologische Mechanismus, der die Argumentstruktur systematisch motiviert, bleibt damit lexikographisch unsichtbar. Umgekehrt lemmatisieren allgemeine Wörterbücher wie das DWDS oder der *Duden* das Suffix *-ung*, ohne die damit verbundenen Argumentstrukturmuster explizit zu machen. Erschwerend kommt hinzu, dass Valenzwörterbücher nicht immer benutzerfreundlich gestaltet sind: Ihre Informationsarchitektur setzt oft metalinguistische Vorkenntnisse voraus, und die Darstellung der Argumentstruktur ist für Nicht-Spezialist:innen — darunter fortgeschrittene DaF-Lernende — nicht immer intuitiv zugänglich (vgl. Domínguez Vázquez und Valcárcel Riveiro 2019).

Es ist jedoch bemerkenswert, dass einige monolinguale Wörterbücher bereits einen fortschrittlichen Ansatz zur Darstellung morphologischer Informationen verfolgen. Ein besonders gelungenes Beispiel hierfür stellt die digitale Version des *Den Danske Ordbog* dar (siehe Abbildung 2), die als Vorbild für die lexikographische Behandlung produktiver Suffixe dienen kann². In diesem dänischen Wörterbuch wird unter dem Eintrag für das nominale Suffix *-else* — das funktionale Äquivalent zum deutschen *-ung* — eine umfassende Informationsarchitektur präsentiert, die weit über die Auflistung grammatischer Eigenschaften hinausgeht und systematisch semantische Dimensionen integriert. Konkret werden die verschiedenen Bedeutungsebenen des Suffixes expliziert: zum einen die Bezeichnung von Vorgängen oder Prozessen, zum anderen die Bezeichnung von Ergebnissen oder Resultaten der entsprechenden Handlungen. Damit demonstriert das *Den Danske Ordbog* exemplarisch, was der vorliegende Beitrag für das Deutsche anstrebt: die systematische Verbindung morphologischer, semantischer und argumentstruktureller Information in allgemeinen einsprachigen digitalen Wörterbüchern.

Die Verfügbarkeit solcher morphologisch-semantischen Informationen erweist sich als besonders wertvoll für Nutzer:innen in zwei spezifischen Anwendungskontexten: (a) Wenn fortgeschrittene DaF-Lernende aktiv Sprache produzieren müssen, ermöglicht das Verständnis für Affixmuster eine systematische und regelgeleitete Wortbildung. Anstatt jede Nominalisierung einzeln memorieren zu müssen, können sie auf produktive morphologische Prozesse zurückgreifen; (b) Für Nutzer:innen, die die zugrundeliegenden Strukturprinzipien der Wortbildung verstehen möchten, bietet diese Information eine theoretische Grundlage für die Entwicklung morphologischer Bewusstheit. Hier wird ein fundamentaler kognitiver Prozess im fortgeschrittenen Spracherwerb angesprochen: Lernende entwickeln auf der Basis ihrer Sprachbeobachtungen Hypothesen über die Produktivität morphologischer Prozesse (zur Hypothesenbildung bei der Wörterbuchbenutzung von DaF-Lernenden vgl. Müller-Spitzer et al. 2018 oder Wolfer et al. 2018). Sie vermuten beispielsweise, dass bestimmte Suffixe systematisch zur Bildung von Nomina aus Verben eingesetzt werden können. Diese Hypothesenbildung stellt jedoch nur den ersten Schritt dar. Entscheidend ist die Möglichkeit zur Hypothesenvalidierung durch den Zugang zu verlässlichen

lexikographischen Ressourcen. Nur wenn Wörterbücher explizite Informationen über die semantischen und morphologischen Eigenschaften von Suffixen bereitstellen, können Lernende ihre intuitiven Annahmen überprüfen und gegebenenfalls korrigieren.

Der Hinweis, dass monolinguale Wörterbücher (vgl. Domínguez Vázquez und Valcárcel Riveiro 2015 oder Kosem et al. 2019) relativ häufig zu Rate gezogen werden, unterstreicht die praktische Relevanz der Integration morphologischer Informationen in lexikographische digitale Werke. Dies deutet auf eine Nutzer:innenpräferenz für integrierte Informationssysteme hin, die morphologische, semantische und syntaktische Aspekte in kohärenter Weise verbinden, anstatt diese Informationen auf verschiedene Ressourcentypen zu verteilen. Das dänische Modell demonstriert die Nützlichkeit einer systematischen Integration morphologischer Informationen in allgemeine monolinguale Wörterbücher und könnte als Vorlage für entsprechende Entwicklungen in der deutschen Lexikographie dienen.

Du er her: Forside / Den Danske Ordbog / Ordbog

Vis forkortet

-else suffiks

BØJNING i substantiv fælleskøn: -n, -r, -rne | i substantiv intetkøn: -t, -r, -rne

UDTALE [-ølse]

OPRINDELSE sideform til **-sel**, udviklet af ældre middeldansk *-ilse*, omdannet af oldsaksisk *-isli*, *-islo*

Betydninger

1. bruges til at danne substantiver der betegner en handling eller proces

SYNONYMER -sel | -ing | -tion

GRAMMATIK fx befrielse, skabelse

1.a betegner resultatet heraf eller ting forbundet hermed

GRAMMATIK fx kendelse, tilladelse

Orddannelser

AFLEDNINGER afgørelse sb. | bevægelse sb. | fornøjelse sb. | følelse sb. | indflydelse sb. | misundelse sb. | opfattelse sb. | spøgelse sb. | uddannelse sb. | udvidelse sb.

Rapportér et problem fra Den Danske Ordbog

Abbildung 2: Wörterbuchartikel zum Suffix *-else* in *Den Danske Ordbog*

Die Analyse der lexikographischen Beschreibung von *-ung*-Nomina legt erhebliche Inkonsistenzen zwischen den vorher erwähnten deutschsprachigen Wörterbuchsystemen dar. Am Beispiel des Lemmas *Erleichterung* lassen sich Unterschiede in der Informationsstrukturierung und -darstellung beobachten. Das DWDS privilegiert in seiner Bedeutungsstrukturierung eine der Lesarten des Verbs *erleichtern*, auf die direkt referiert wird. Dieser Ansatz behandelt die derivationelle

Beziehung zwischen Verb und Nomen als primären Organisationsfaktor. Der *Duden* etabliert das sich aus dem beschriebenen Vorgang ergebende Gefühl als primäre Bedeutung und die prozessuale Dimension des Verbs rückt in den Vordergrund. Das lexikographische Portal *lexiko* beschränkt sich auf die Bereitstellung grammatischer Informationen und authentischer Korpusbelege und verzichtet auf explizite Bedeutungsparaphrasen oder semantische Strukturierungen.

Betreffend die Darstellung der Kombinationsmöglichkeiten von *Erleichterung* zeigen die drei untersuchten lexikologischen Informationssysteme unterschiedliche Ansätze. Das DWDS bietet die detailliertesten Informationen zur kontextuellen Verwendung von *Erleichterung*. Diese Informationen werden durch bedeutungsbezogene Beispielangaben vermittelt, wobei die Argumentstruktur des Lemmas veranschaulicht wird: *die Erleichterung der Prüfung, des Wettkampfes durch günstigere Bedingungen*. Der *Duden* gibt lediglich ein einziges Beispiel zur Erschließung der Argumentstruktur (*Erleichterung der Schmerzen*), wobei weitere semantische Kombinationsmöglichkeiten unerwähnt bleiben. *lexiko* zeigt ausschließlich authentische Textstellen aus Korpora, wobei Nutzer:innen Verwendungsmuster selbstständig erkennen müssen, da die Information weniger systematisch aufbereitet ist. Auf diese Weise werden weder die sprach- noch die wissensbezogenen Kompetenzen der Lernenden ausreichend gefördert, da systematische Informationen zur Argumentstruktur und zu semantischen Kombinationsmustern fehlen. Im Folgenden (siehe Abschnitt 3) wird auf die systematische Erfassung der Argumentstruktur dieser nominalen Derivationen eingegangen.

3. Überlegungen zur nominalen Valenz von *-ung*-Derivaten

Deverbale Nomina auf *-ung* stellen ein zentrales Phänomen an der Schnittstelle von Morphologie, Syntax und Semantik dar. Ihre Analyse offenbart die komplexen Beziehungen innerhalb der Nominalphrase und wirft grundlegende Fragen zur Argumentstruktur auf. Einerseits erben diese Nomina die semantische Valenz des Basisverbs, andererseits unterliegen sie den syntaktischen Realisierungsbedingungen, die für den nominalen Bereich gelten. Aus dieser Asymmetrie zwischen konzeptueller Struktur und syntaktischer Form resultieren vielschichtige Prozesse der Argumentvererbung, -transformation und -tilgung.

In dieser Arbeit wird die Annahme vertreten, dass die Valenz von Nomina als ein System *sui generis* fungiert und dass eine direkte Ableitung verbaler Strukturen nicht immer möglich ist (vgl. Teubert 2003; Domínguez Vázquez 2011). Die nominale Valenz etabliert eigene syntaktische Muster: Ein markantes Beispiel hierfür ist die Genitivphrase, die als typische nominale Ergänzung gilt (vgl. Engel 2004; Domínguez Vázquez 2024) und semantische Relationen ausdrückt, die in der verbalen Sphäre oft durch Subjekt- oder Objektkasus realisiert werden. Dies unterstreicht, dass die semantische Ebene die fundamentale Ebene der Valenz darstellt, auf der die morphosyntaktischen Realisierungen aufbauen.

Der Übergang von einer verbalen zu einer nominalen Konstruktion führt zu einer systematischen Transformation der Argumentrealisierung. Während die Valenzlexikographie diese Strukturen auf Lemmaebene dokumentiert (vgl. Sommerfeldt und Schreiber 1977), bleibt ihre systematische Rückführung auf den morphologischen Bildungsmechanismus — wie im vorigen Abschnitt dargelegt — ein Desiderat. Dies lässt sich anhand der folgenden Beispiele veranschaulichen, die den Transformationsprozess vom Aktivsatz über den Passivsatz zur Nominalkonstruktion nachzeichnen:

- (1) Günstigere Bedingungen erleichtern den Wettkampf. (Aktivsatz)
- (2) Der Wettkampf wird durch günstigere Bedingungen erleichtert. (Passivsatz)
- (3) Die Erleichterung des Wettkampfes durch günstigere Bedingungen. (Nominalphrase)

Im aktiven Satz (1) fungiert *günstigere Bedingungen* als Subjekt und *den Wettkampf* als Akkusativobjekt. Im Passivsatz (2) wird die Perspektive verschoben: Das vormalige Objekt rückt in die Subjektposition, während das Agens durch eine Präpositionalphrase realisiert wird. In der nominalen Konstruktion (3) werden diese Argumente weiter transformiert: Das vormalige Objekt wird zu einer Genitivphrase (*des Wettkampfes*), während das vormalige Subjekt durch eine Präpositionalphrase (*durch günstigere Bedingungen*) aktualisiert wird. Diese nominale Struktur ermöglicht eine höhere Informationsdichte, indem eine komplexe Proposition kompakt kodiert wird (Hennig 2019; Zifonun et al. 1997).

Ein zentrales Merkmal von Nominalisierungen ist die flexible Realisierung der Argumentstruktur. Im Gegensatz zum Verbalsatz, wo bestimmte Ergänzungen obligatorisch sind, wird bei der Nominalisierung oft nur ein „Minimalprogramm“ der Argumente realisiert, während andere Elemente optional werden oder gänzlich entfallen können (vgl. Schumacher 1999). Die Entscheidung, welche Argumente syntaktisch ausgedrückt werden, ist dabei nicht willkürlich, sondern häufig diskurspragmatisch motiviert (vgl. Hennig 2019). Je nach Fokus und Kontext werden nur die relevantesten Teilnehmenden des Ereignisses genannt.

Dies lässt sich gut am Beispiel der Konstruktionen *Wallander verhaftet den Dieb* und *die Verhaftung des Diebs* (vgl. Welke 2011: 253) verdeutlichen. Obwohl das Nomen *Verhaftung* und das Verb *verhaften* die gleiche semantische Basis haben, unterscheidet sich die syntaktische Notwendigkeit ihrer Argumente. In der Nominalphrase bleibt die semantische Rolle des Agens (*Wallander*) zwar konzeptuell erhalten, ihre syntaktische Realisierung als Präpositionalphrase (*durch Wallander*) ist jedoch fakultativ. Der Fokus liegt in diesem Zusammenhang auf der Handlung und dem Patiens. Dieser Kontrast macht deutlich, dass die semantische Ebene die grundlegende Ebene der Valenz ist. Auf dieser semantischen Basis operieren jedoch die flexibleren morphosyntaktischen Regeln der Nominalphrase, die im Gegensatz zum Verbalsatz häufig die Tilgung von Argumenten erlauben.

Die Produktivität der *-ung*-Nominalisierungen ist jedoch nicht uneingeschränkt. Die entscheidende Restriktion liegt in der internen syntaktisch-semanticen Struktur des Basisverbs. Eine auf der lexikalischen Syntax und Semantik fußende Theorie (vgl. Roßdeutscher 2010; Roßdeutscher und Kamp 2010) postuliert, dass die Bildung einer *-ung*-Nominalisierung davon abhängt, ob das Verb *bi-eventiv* aufgebaut ist. Die zentrale Hypothese von Roßdeutscher (2010) und Roßdeutscher und Kamp (2010) besagt, dass eine verbale Konstruktion genau dann eine *-ung*-Nominalisierung besitzt, wenn das Verb eine *bi-eventive* Struktur aufweist. Der semantische Kern dieser Anforderung ist, dass der *-ung*-Operator eine Struktur benötigt, deren semantische Repräsentation eine Kausalitätsbedingung enthält, wobei ein Ereignis einen Zustand verursacht. Die Verben mit *bi-eventiven* Strukturen enthalten zwei Sub-Ereignisse: ein auslösendes Ereignis und einen resultierenden Zustand. Die Kausalbeziehung zwischen beiden wird explizit in der Struktur der Verbalphrase kodiert und lizenziert somit die Anbindung des *-ung*-Operators (vgl. Roßdeutscher und Kamp 2010).

Vor diesem theoretischen Hintergrund wird deutlich, dass nur Verben mit einer inhärenten Kausal-Resultats-Struktur für die *-ung*-Nominalisierung verfügbar sind. Bei der tatsächlichen Realisierung in der Nominalphrase wird oft nur ein „Minimalprogramm“ der Argumente (vgl. Schumacher 1999) umgesetzt, wobei die Auswahl diskurspragmatisch gesteuert ist. Die zugrundeliegende *bi-eventive* Struktur erklärt die semantische Dualität vieler *-ung*-Nomina, denn sie können sich sowohl auf die Handlung (*nomen actionis*) als auch auf deren Ergebnis (*nomen acti*) beziehen (vgl. Engel 2004; Welke 2011). Welke (2011) weist auf den perfektiven Charakter der *-ung*-Nominalisierungen hin und stellt fest, dass sie normalerweise telische Handlungen bezeichnen.

Aus ontologischer Perspektive ist in *GermaNet* (Hamp und Feldweg 1997; Hinrichs et al. 2020; siehe Abbildung 3) zu beobachten, dass die *-ung*-Derivate sehr oft als Geschehen definiert werden und teilweise auch anderen semantischen Klassen zugeordnet werden, denn die Tatsache, dass *-ung*-Nominalisierungen das Resultat der Handlung beschreiben, führt dazu, dass diese Handlungen auch Objekte oder Personen bezeichnen können (vgl. Wöllstein und Dudenredaktion 2022). Dies zeigt sich beispielsweise bei der Vokabel *Bebauung*, die gleichzeitig den Prozess des Bauens (*nomen actionis*) als auch die Gesamtheit der Gebäude auf einer Fläche (*nomen acti*) beschreiben kann und somit als Gruppe klassifiziert wird. Der Fokus wird in dieser Studie auf die erste Lesart beschränkt, d. h. auf die Lesart, die mit einer Zustandsveränderung zusammenhängt: von einem Zustand A erfolgt ein Übergang zu einem Zustand B. Abbildung 3 veranschaulicht, wie die ontologische Vererbung bei einer *-ung*-Nominalisierung in *GermaNet* dargestellt wird. Der Fokus liegt hierbei auf Nomina, die Prozesse oder Handlungen beschreiben und die mindestens in ihrer inhärenten semantischen Argumentstruktur zwei Argumente implizit enthalten, während metaphorische oder übertragene Bedeutungen außer Acht bleiben.

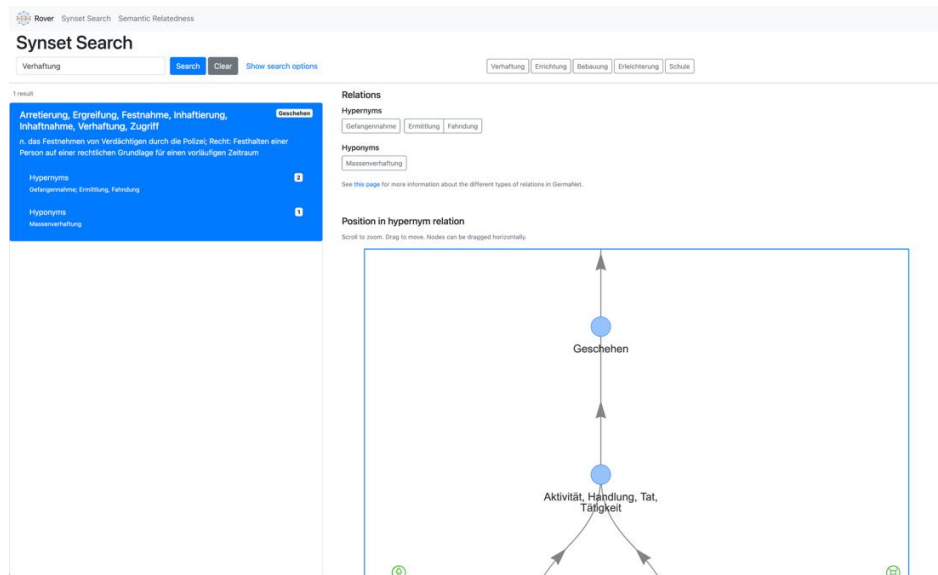


Abbildung 3: Ontologische Klassifikation von *Errichtung* in *GermaNet Rover*

4. Datenerhebung zur Analyse der *-ung*-Nominalisierungen

Der vorliegende Abschnitt setzt sich zum Ziel, die methodologische Grundlage zur Beantwortung der in der Einleitung gestellten Fragen zu schaffen. Da der hier angestrebte Ansatz auf Korpusdaten basiert, wird zunächst auf die Korpusbeschreibung eingegangen, anschließend auf die Extraktionsverfahren und schließlich auf die statistischen Variablen, die für die Analyse und den Vorschlag einer Verbesserung der lexikographischen Informationen in lexikologischen Informationssystemen herangezogen werden.

4.1 Korpusbeschreibung

Um die nachfolgende Datenanalyse nachvollziehbar zu gestalten, wird zunächst das zugrundeliegende Korpus beschrieben. Für die vorliegende Untersuchung wurde ein ad hoc erstelltes Korpus deutschsprachiger *Wikipedia*-Artikel zum Einsatz gebracht, welches zwecks eines umfassenderen Forschungsvorhabens³ zusammengestellt wurde. Die Wahl der *Wikipedia* als Datenquelle wird in Arias-Arias (o. D.) ausführlich erklärt und begründet sich insbesondere in ihrer Repräsentativität für den Nominalstil (Hennig 2019), was sie zu einem besonders geeigneten Untersuchungsgegenstand für die Analyse nominaler Strukturen und verdichteter Ausdrucksweisen macht. *Wikipedia*-Artikel weisen charakteristische Merkmale des elaborierten Schriftsprachgebrauchs auf, darunter komplexe

Nominalphrasen, Funktionsverbgefüge und attributive Verdichtungen (vgl. Hennig 2019), die für (populär)wissenschaftliche und enzyklopädische Textsorten typisch sind. Zudem gewährleistet die thematische Vielfalt der Artikel eine breite lexikalische Abdeckung des Deutschen.

Das Korpus umfasst mehr als 40 Millionen Tokens und bietet damit eine empirische Grundlage für quantitative und qualitative Analysen. Die linguistische Annotation der Korpusdaten erfolgte mittels UDPipe (Straka 2018), einem Tool zur automatischen morphosyntaktischen Annotation, das auf den *Universal Dependencies* basiert und eine konsistente Auszeichnung von Wortarten, Abhängigkeitsrelationen und morphologischen Merkmalen ermöglicht. Darüber hinaus wurden die Daten mit semantischen Informationen aus *GermaNet* (Hamp und Feldweg 1997) angereichert, um lexikalisch-semantische Analysen zu ermöglichen. Die Kombination aus morphosyntaktischer und semantischer Annotation schafft die Voraussetzung für eine mehrdimensionale Korpusanalyse, die sowohl strukturelle als auch semantische Aspekte der zu untersuchenden sprachlichen Phänomene erfassen kann.

4.2 Zur Datenerhebung

Was die Datenerhebung anbelangt, wird nun ein mehrstufiges Verfahren angewandt, das auf der automatischen Extraktion linguistisch annotierter Daten beruht. Ziel dieser Arbeit ist es, zu beobachten, inwiefern die automatische Extraktion annotierter Korpusdaten zur systematischen Beschreibung der Argumentstrukturen von *-ung*-Derivaten beitragen kann. Die Extraktion erfolgt durch einen ad hoc erstellten Python-Algorithmus⁴, der mithilfe von Regeln operiert, die den nachfolgend beschriebenen linguistischen Kriterien entsprechen. Die Extraktion erfolgt nach mehreren aufeinander aufbauenden Kriterien, die sowohl morphologische, semantische als auch syntaktische Ebenen berücksichtigen:

- (i) Identifikation des Regens. In einem ersten Schritt werden alle Nomina identifiziert, die als Regens einer Nominalphrase fungieren und zwei zentrale Voraussetzungen erfüllen: Sie müssen erstens auf *-ung* enden und zweitens im semantischen Annotationsmodul mindestens als {Geschehen} klassifiziert sein. Die semantische Annotation erfolgt auf Grundlage der *GermaNet*-Daten (Hamp und Feldweg 1997), wobei zum gegenwärtigen Zeitpunkt keine automatische Disambiguierung polysemer Formen vorgenommen wird. Dies bedeutet, dass ein Nomen wie *Zeitung*, das zwar formal auf *-ung* endet, jedoch nicht ereignisbezogen ist, aufgrund fehlender {Geschehen}-Annotation nicht in die Auswahl gelangt, während ein Nomen wie *Vergiftung*, das als Ereignis- oder Handlungsnomen semantisch markiert ist, extrahiert wird.
- (ii) Syntagmatische Realisierung der Argumente. In einem zweiten Schritt wird die syntagmatische Umgebung des Regens untersucht, um potenzielle Argumente zu identifizieren. Hierbei werden Konstituenten des

rechten Feldes in Betracht gezogen, da Argumente im Deutschen typischerweise postnominal realisiert werden (vgl. Engel 2004). Im Fokus stehen dabei insbesondere Genitivphrasen (z. B. *die Zerstörung der Stadt*) sowie Präpositionalphrasen (z. B. *die Entscheidung für eine Lösung*), die als prototypische Realisierungsformen nominaler Argumente gelten (vgl. Domínguez Vázquez 2011).

- (iii) Syntaktisch-dependenzielle Relation. Auf syntaktischer Ebene werden nur solche Phrasen extrahiert, die in der Dependenzannotation als Modifikatoren des Kopfnomens getaggt sind. Dies schließt vor allem die Dependenzrelation *nmod* ein, die im *Universal Dependencies*-Framework für die Annotation von Argumenten und Adjunkten bei Nomina verwendet wird. Durch diese Einschränkung wird sichergestellt, dass nur syntaktisch an das Nomen gebundene Konstituenten erfasst werden, während satzübergreifende Elemente im Prinzip ausgeschlossen bleiben.
- (iv) Kontextfenster. Um die Analyse handhabbar zu halten und gleichzeitig einen ausreichenden Kontext zu gewährleisten, wird ein Fenster von 5 Tokens rechts vom Kopfnomen definiert. Innerhalb dieses Fensters werden die oben genannten syntaktischen und morphologischen Kriterien angewandt.

Es ist anzumerken, dass dieses automatische Extraktionsverfahren zwangsläufig auch falsch-positive Treffer liefert, etwa wenn eine Präpositionalphrase als Adjunkt fungiert, aber fälschlicherweise als Argument interpretiert wird, oder wenn semantische Mehrdeutigkeiten nicht aufgelöst werden. Dennoch liefert gerade die Analyse dieser Fälle wertvolle Einsichten in die Grenzfälle zwischen Argumenten und Adjunkten sowie in die Variabilität der Argumentrealisierung bei *-ung*-Derivaten.

4.3 Zur Analyse der Daten

Dieser Abschnitt widmet sich dem Analyseverfahren der extrahierten Korpusdaten, das sowohl qualitative als auch quantitative Aspekte umfasst. Ziel ist es, einerseits zu ermitteln, welche der automatisch extrahierten Konstruktionen tatsächlich Argumentstrukturen realisieren und welche nicht, und andererseits die strukturellen und semantischen Eigenschaften dieser Argumentrealisierungen systematisch zu beschreiben.

Die qualitative Analyse konzentriert sich auf die Validierung der extrahierten Daten und die Identifikation echter Argumentstrukturen. Im Zentrum steht dabei die folgende Frage: Welche der extrahierten postnominalen Phrasen sind tatsächlich Argumente des jeweiligen *-ung*-Nomens und lösen somit eine Argumentstruktur aus? Die manuelle Annotation der Belege erlaubt es, falsch-positive Fälle zu identifizieren und die Genauigkeit des Extraktionsverfahrens zu evaluieren. Darüber hinaus liefert die qualitative Betrachtung Aufschluss über

Grenzfälle und ambige Konstruktionen, die für die theoretische Diskussion der Argumentstruktur von Interesse sein können.

Die quantitative Analyse zielt darauf ab, Häufigkeitsmuster und strukturelle Präferenzen in der Argumentrealisierung von *-ung*-Derivaten zu ermitteln. Zunächst wird erhoben, wie viele der extrahierten Konstruktionen tatsächlich Argumentstrukturen entsprechen, um die Effektivität des Extraktionsverfahrens zu beurteilen. Anschließend werden die validierten Argumentrealisierungen in Hinblick auf folgende Variablen analysiert:

- (i) Semantische Klassifikation der Prädikate und der Argumente: Da die Korpusdaten mit *GermaNet* (Hamp und Feldweg 1997) annotiert sind, können im Fall der Prädikate die Synsets zur granularen semantischen Klassifikation herangezogen werden. Was die Argumente anbetrifft, wird hier in Betracht gezogen, zu welcher semantischen Kategorie (z. B. {Artefakt} oder {Substanz}) das jeweilige Argument gehört.
- (ii) Art der Ergänzung: Welche Rolle erfüllt das Argument in der Argumentstruktur?
- (iii) Syntaktische Realisierung der Ergänzung: Wie wird das Argument oberflächlich realisiert?

Für die Analyse wird ein randomisiertes Sample von 200 Belegen aus dem Arbeitskorpus extrahiert. Diese Stichprobengröße ist für eine kombinierte qualitativ-quantitative Pilotanalyse, wie sie hier angestrebt wird, methodisch angemessen (vgl. Levshina 2015): Sie erlaubt einerseits eine manuelle Validierung sämtlicher Belege und bietet andererseits eine hinreichende empirische Basis für die angewandten statistischen Verfahren. Die Ergebnisse sind dabei als Ausgangspunkt für weiterführende Untersuchungen an größeren Datensätzen zu verstehen.

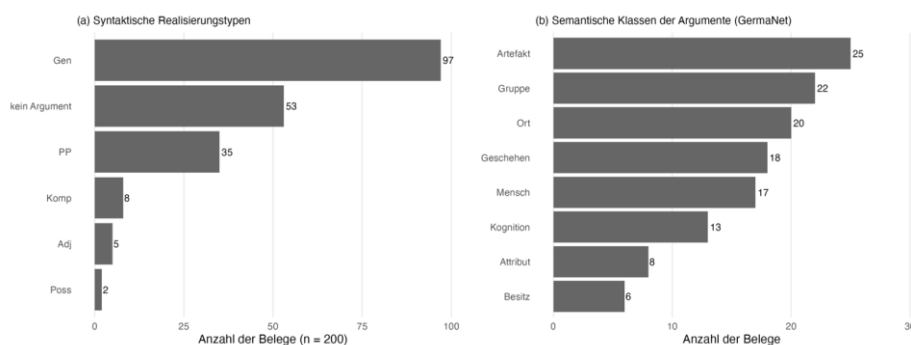


Abbildung 4: Deskriptive Analyse der extrahierten Argumentstrukturen

Die deskriptive Auswertung des Samples gibt zunächst einen Überblick über die Verteilung der extrahierten Argumentstrukturen (siehe Abbildung 4). Hinsichtlich der syntaktischen Realisierung zeigt sich eine klare Präferenz für die Genitivphrase als primären Realisierungstyp, gefolgt von Präpositionalphrasen und — in deutlich geringerer Anzahl — Komposita und anderen Typen. Was die semantische Klassifikation der Argumente gemäß *GermaNet*-Annotation betrifft, verteilen sich die häufigsten Klassen auf {Artefakt}, {Gruppe}, {Ort} und {Mensch}, wobei {Artefakt} und {Gruppe} die frequentesten Kombinationspartner der *-ung*-Prädikate darstellen. Diese deskriptiven Befunde bilden die empirische Grundlage für die in Abschnitt 5 vorgenommene qualitative und quantitative Analyse der Argumentrealisierungsmuster.

Zur Korrelationsanalyse werden Chi-Quadrat-Tests (χ^2) und Cramérs V herangezogen (vgl. Baayen 2008; Levshina 2015), da diese Verfahren die (Un)abhängigkeit zwischen zwei nominalen Variablen messen. Der Chi-Quadrat-Test prüft dabei, ob die beobachteten Häufigkeiten in einer Kontingenztafel signifikant von den unter Unabhängigkeit erwarteten Häufigkeiten abweichen. Cramérs V quantifiziert die Stärke dieses Zusammenhangs auf einer Skala von 0 (kein Zusammenhang) bis 1 (perfekter Zusammenhang) und ermöglicht so eine inhaltliche Interpretation jenseits der bloßen Signifikanzaussage (vgl. Levshina 2015: 211-215). Ein Wert von Cramérs V $\geq 0,5$ gilt dabei als Hinweis auf einen starken Zusammenhang (vgl. Cohen 1988).

Aufgrund der teilweise niedrigen Frequenzen in den Kreuztabellen werden zusätzlich Resamplingverfahren eingesetzt. Der Rückgriff auf Resampling anstelle des Fisher-Exakt-Tests ergibt sich dabei aus der Dimensionalität der verwendeten Kontingenztabellen: da der Fisher-Exakt-Test bei größeren Tabellenstrukturen rechnerisch aufwendig ist und interpretatorisch eingeschränkt bleibt, erlauben Resamplingverfahren eine flexible und robuste Schätzung des *p*-Werts auch bei mehrdimensionalen Kreuztabellen mit geringen Zellbesetzungen (vgl. Levshina 2015: 218-221). Dies gewährleistet die statistische Validität der Ergebnisse auch dort, wo einzelne Zellen der Kontingenztafel nur wenige Belege aufweisen.

Darüber hinaus ermöglicht die Kombination aus qualitativer Validierung und quantitativer Mustererkennung eine empirisch fundierte Beschreibung der Argumentrealisierung bei *-ung*-Derivaten und liefert Erkenntnisse darüber, inwieweit automatische Extraktionsmethoden zur linguistischen Korpusanalyse beitragen können.

5. Analyse der Resultate

Des Weiteren werden die extrahierten Belege qualitativ und quantitativ ausgewertet, und die Ergebnisse werden in diesem Abschnitt eingehend diskutiert. Die Analyse gliedert sich in drei Hauptbereiche: Zunächst wird auf die Validierung des Extraktionsverfahrens eingegangen, wobei besondere Aufmerksamkeit der Unterscheidung zwischen echten Argumentstrukturen und Nicht-Argumentstrukturen geschenkt wird. Anschließend werden Grenzfälle thematisiert, in denen

Elemente zwar zum verbalen Rahmen des Hauptverbs im Satz gehören, aber auf der nominalen Ebene in unveränderter Form realisiert werden können. Abschließend wird die quantitative Analyse präsentiert, die sowohl Häufigkeitsverteilungen als auch strukturelle Realisierungsmuster und Abhängigkeiten zwischen verschiedenen Variablen in den Blick nimmt.

5.1 Validierung und Identifikation von Argumentstrukturen

Zur Validierung der extrahierten Daten musste eine manuelle Verarbeitung und Überprüfung der Belege vorgenommen werden, um festzustellen, welche polysemen lexikalischen Einheiten⁵ im jeweiligen Kontext nicht die Bedeutung {Geschehen} aufweisen. Besonders problematisch erwiesen sich dabei Einheiten mit regulärer Polysemie (vgl. Pustejovsky 1995 oder Pustejovsky und Batiukova 2019), die simultan zwei Bedeutungen aktivieren können (siehe 5.2). Die folgenden Belege veranschaulichen diese Problematik:

- (4) Zur **Ausstattung** der Männer gehören oft Waffen [...] {Artefakt}
- (5) Der Verein hat rund 500 Mitglieder, darunter eine große **Abteilung** für Jugendsport [...] {Gruppe}
- (6) Die europäische **Bevölkerung** in der Stadt war zu portugiesischer Zeit gering. {Gruppe}

Diese drei Beispiele beschreiben keine Geschehen, sondern Resultate von Prozessen, die durch die entsprechenden *-ung*-Nomina bezeichnet werden könnten. In Beispiel (4) bezeichnet *Ausstattung* das Resultat des Ausstattens (die konkreten Gegenstände), in (5) bezeichnet *Abteilung* die resultierende organisatorische Einheit, und in (6) meint *Bevölkerung* die Gruppe von Menschen als Resultat des Bevölkerungens. Da diese resultativen Lesarten nicht die ereignishaftige Bedeutungskomponente als primäre Lesart auslösen, wurden sie aus dem Datensatz ausgeschlossen. Dies betrifft 21% der Daten, konkret wurden 42 Belege entfernt.

Ein besonders interessanter Fall ist das Wort *Weitsprung*, das als {Geschehen} getaggt wurde und auf *-ung* endet, jedoch keine *-ung*-Ableitung vom zugrunde liegenden Verb *springen* darstellt, sondern ein implizites Derivat durch internen Ablaut (vgl. Zifonun et al. 1997). Solche Fälle erfordern besondere Aufmerksamkeit bei der automatischen Verarbeitung.

Darüber hinaus wurden Einheiten identifiziert, bei denen das *nmod* zwar syntaktisch korrekt annotiert wurde, aber keine direkte Aktivierung einer Argumentstelle des nominalen Prädikats auslöst. Hier lassen sich zwei Falltypen unterscheiden:

- (i) Fälle wie Beispiel (7), in denen das *nmod* keine semantische Ergänzung des Kopfnomens darstellt:
- (7) Die **Steuerung** ist ähnlich der Vorgänger.

Das Nomen *Vorgänger* wird vom Parser als *nmod* annotiert, erfüllt aber keine eindeutige semantische Argumentfunktion des Kopfnomens, denn es

liegt in diesem Fall eine syntaktische Irregularität vor: Die Ähnlichkeitskonstruktion ist elliptisch aufgebaut (*Die Steuerung ist ähnlich [der Steuerung] der Vorgänger*). Derartige Fälle werden aus der weiteren Analyse ausgeschlossen.

- (ii) Fälle wie Beispiel (8), in denen das als *nmod* getaggte Element (*Regierungszeit*) lediglich eine temporale Angabe darstellt, während die tatsächliche Argumentergänzung vom System übersehen wird. Dies geschieht häufig, wenn mehrere nominale Einheiten im rechten Feld der Nominalphrase auftreten und alle als *nmod* annotiert werden:
- (8) Die **Wiedereinsetzung** bürgerlicher Rechte in seiner Regierungszeit markiert für Uruguay den Beginn einer neuen Ära, von dem alle gesellschaftlichen Bereiche profitierten.

In diesem Beispiel wäre *bürgerlicher Rechte* die eigentliche Ergänzung von *Wiedereinsetzung*, während *Regierungszeit* nur eine temporale Angabe darstellt. Die syntaktische Annotation differenziert jedoch nicht zwischen obligatorischen Argumenten und fakultativen Angaben.

Fälle wie (8) werden nicht aus dem Datensatz ausgeschlossen, da sie tatsächlich eine Ergänzung aufweisen. Im Rahmen der manuellen Verarbeitung der Daten wird jedoch darauf hingewiesen, welches *nmod* nicht korrekt als Argument identifiziert worden ist. Diese Annotationen dienen als Baseline für künftige Ansätze zur automatischen Disambiguierung zwischen argumentalen und adjunktalen *nmod*-Relationen.

Nach der manuellen Überprüfung zeigt sich, dass 73,5% der verbleibenden Belege tatsächlich eine Argumentstruktur realisieren. Die meisten dieser nominalen Prädikate weisen mono- oder biargumentale Strukturen auf, was die syntaktisch-semantische Komplexität der *-ung*-Nominalisierungen verdeutlicht. Dieser Befund deutet darauf hin, dass *-ung*-Derivate sich zur Entwicklung argumentstrukturbasierter Disambiguierungsverfahren (z. B. zwischen den Lesarten {Geschehen} und {Artefakt}, wie bei *Ausstattung* in (4)) eignen. Die nachfolgende Analyse geht zunächst auf diese Grenzfälle ein und konzentriert sich anschließend auf die validierten Strukturen.

5.2 Analyse von Grenzfällen

Die Analyse der exkludierten Belege gibt Aufschluss über Muster von regulärer Polysemie in den *-ung*-Nomina. Eine systematische Auswertung der Daten lässt feststellen, dass die häufigste Kategorie die Polysemie {Geschehen}-{Gruppe} ist, gefolgt von {Geschehen}-{Artefakt} und weiteren semantischen Alternationen. Die quantitative Analyse zeigt, dass die Kategorie {Gruppe} mit 30,24% am stärksten vertreten ist, gefolgt von {Artefakt} mit 20,93% und {Kognition} mit 11,63%. Die Kategorien {Attribut}, {Ort} und {Relation} weisen jeweils 6,98% auf, während die übrigen Kategorien jeweils unter 3% liegen. Diese

Verteilung deutet darauf hin, dass *-ung*-Nominalisierungen besonders produktiv in der Bezeichnung von Kollektiven, materiellen Objekten und kognitiven Zuständen als von der {Geschehen}-Bedeutung abgeleitete Lesart sind.

Das häufigste Polysemiemuster zeigt sich in der Alternation zwischen Prozess- und Kollektivlesart. Ein exemplarischer Beleg hierfür findet sich in (9):

- (9) Seligmann befand die Firma lebensfähig und wurde quasi als Lohn für seine solide Arbeit vom Aufsichtsrat der Firma, dem auch Magnus angehörte, gebeten, in die **Leitung** der Firma einzutreten.

Hier zeigt das Nomen *Leitung* zwei semantische Lesarten: einerseits den Prozess des Leitens als {Geschehen} und andererseits das Leitungsgremium als {Gruppe}. Weitere Beispiele für dieses Muster finden sich in Nomina wie *Verwaltung*, *Regierung* oder *Vertretung*, die sowohl den Handlungsprozess als auch die ausführende Personengruppe bezeichnen können.

Die zweithäufigste Polysemie betrifft die Alternation zwischen Prozess und resultierender Sache. Die richtige Interpretation und semantische Annotation solcher polysemer Lesarten kann oft erst nur durch kontextuelle Analyse geklärt werden. Besonders aussagekräftig sind die Realisierungsmuster der Argumentstruktur und die syntaktische Einbettung (vgl. Domínguez Vázquez 2024): Der unbestimmte Artikel in *eine Bemalung* (siehe (11)) deutet auf eine konkrete Sache hin (Resultat-Lesart), während eine prozessuale Lesart wahrscheinlich durch andere Konstruktionen signalisiert wird. Diese Alternation zeigt sich besonders deutlich in dem Beleg (10):

- (10) 1508 wurde mit der Puente del Obispo über den Guadalquivir die **Verbindung** zur Stadt Baeza geschaffen.

Das Nomen *Verbindung* weist in diesem Beleg eine charakteristische Doppelreferenz auf: Es bezeichnet gleichzeitig den Prozess des Verbindens ({Geschehen}) und — in diesem Kontext als primäre Lesart — das daraus resultierende materielle Objekt ({Artefakt}). Beide Lesarten sind nicht als alternative Interpretationen zu verstehen, sondern als komplementäre Facetten einer einzigen konzeptuellen Einheit im Sinne der regulären Polysemie (vgl. Pustejovsky und Batiukova 2019): Der Kontext *wurde [...] geschaffen* aktiviert zwar die Resultat-Lesart als dominant, schließt die Prozess-Lesart jedoch nicht aus. Ein zum Teil analoges Muster zeigt sich bei *Bemalung* in Beispiel (11):

- (11) Nose art ist eine **Bemalung** oder Zeichnung auf dem Rumpf eines Luftfahrzeugs in der Nähe der Flugzeugnase.

Bemalung alterniert zwischen zwei Lesarten: dem Prozess des Bemalens und dem bemalten Objekt als konkretes Resultat.

Die {Kognition}-Lesart wird in Beispiel (12) exemplifiziert:

- (12) [...] in der sie ihre emanzipatorische **Vorstellung** von der Gleichheit beider Geschlechter festhielt.

Das Nomen *Vorstellung* wird hier als {Kognition} annotiert, da es sich auf eine Idee oder ein Konzept bezieht — ein mentales Konstrukt, das durch einen kognitiven Prozess (*sich vorstellen*) entstanden ist. In diesem Zusammenhang wird {Geschehen} in eine mentale Entität transformiert.

Die Analyse bestätigt, dass {Geschehen} als Basis-Lesart deverbaler *-ung*-Nominalisierungen fungiert und systematisch mit anderen semantischen Kategorien alterniert. Diese reguläre Polysemie ist charakteristisch für deverbale Nominalisierungen (vgl. Pustejovsky und Batiukova 2019) und spiegelt die konzeptuelle Nähe zwischen Handlungen, handelnden Akteuren und Handlungsergebnissen wider.

Es ist dabei anzumerken, dass die Charakterisierung der {Geschehen}-Lesart als Zustandsveränderung nicht für alle *-ung*-Derivate in gleicher Weise zutrifft. Während Nomina wie *Verhaftung* oder *Erleichterung* prototypisch ein kausativ-resultatives Muster aufweisen — ein Element wird von einem Ausgangszustand A in einen Resultatzustand B überführt —, bezeichnen andere *-ung*-Nomina telische Prozesse ohne eigentliche Zustandsveränderung des Arguments. Ein Beispiel hierfür ist *Messung*: Das gemessene Objekt erfährt keine ontologische Veränderung, sondern wird lediglich einer Beobachtungshandlung unterzogen. Diese Fälle gehören zwar ebenfalls zur {Geschehen}-Lesart im Sinne von *GermaNet*, weisen aber eine abgestufte Kausalstruktur auf (vgl. Roßdeutscher und Kamp 2010). Für die lexikographische Beschreibung bedeutet dies, dass das vorgeschlagene Modell nicht als starres Schema, sondern als prototypisches Muster zu verstehen ist, das je nach lexikalischer Semantik des Basisverbs in seiner Ausprägung variieren kann.

5.3 Argumentstrukturen von *-ung*-Nominalisierungen mit {Geschehen}-Lesart

Bei der Analyse der *-ung*-Argumentstrukturen in der {Geschehen}-Lesart zeigt sich, dass 79% der untersuchten Belege eine Argumentstruktur aufweisen. Dieses Ergebnis unterstreicht die Aussagekraft der angewandten Methode zur (semi)automatischen Extraktion von Daten und dementsprechend zur (semi)automatischen Erstellung von (Valenz)wörterbüchern. Die Beobachtung bestätigt, dass das „Minimalprogramm“ solcher lexikalischen Einheiten auf dem biargumentalen Phrasenmuster A1 + A2 basiert (vgl. Domínguez Vázquez 2011). Diese beiden Argumente, die im Folgenden erläutert werden, sind dem Prädikat semantisch inhärent und realisieren sich teilweise nur oberflächlich, um die vorher erwähnte *bi-eventive* Struktur (vgl. Roßdeutscher und Kamp 2010) transparenter zu machen.

Das erste Argument (A1) beschreibt normalerweise das Element, das die Zustandsänderung erfährt. Im Beispiel *die Erleichterung der Schmerzen* erfährt das Nomen *Schmerzen* eine Veränderung vom Ausgangszustand (*-erleichtert*) zum Resultatzustand (*+erleichtert*). Morphosyntaktisch wird A1 bevorzugt durch eine Genitivphrase realisiert, wobei auch alternative Realisierungsmöglichkeiten bestehen, darunter lexikalische Komposita (wie *Strukturänderung*) oder Präpositionalphrasen mit genitivischer Bedeutung (wie *Landung von Leut-*

nant Basil Hall). Aus statistischer Perspektive besteht eine signifikante Korrelation ($p = 0,0004998$) zwischen der Aktivierung des Arguments A1 und seiner Realisierung als Genitivphrase⁶ ($\chi^2 = 449,99$; Cramérs $V = 0,7095$).

Aus semantischer Sicht ist das frequenteste Element ($n = 15$) das Muster [Prädikat (-ung)] + [A1 (Genitiv) + {Artefakt}], wie in *Errichtung einer Kuppel*. Daneben erscheint das Muster [Prädikat (-ung)] + [A1 (Genitiv) + {Gruppe}], exemplifiziert durch *Bildung einer gemeinsamen Großloge aller Bundesstaaten* oder *Ausgliederung der Profimannschaft*. Qualitativ betrachtet zeigt sich, dass die Sprache in diesem Kontext verschiedene Realisierungsmöglichkeiten bereitstellt: Bei der Phrase *Schallausbreitung tiefer Frequenzen* wird die Konstruktion eines Kompositums für das A1 lizenziert, das dann durch eine weitere Spezifizierung mittels einer Genitivphrase präzisiert wird (*Schalle tiefer Frequenzen werden ausgebreitet*). Wie erwähnt, offenbart sich die valenzielle Struktur von A1 durch eine verbale Paraphrase mittels des dem -ung-Nomen zugrunde liegenden Verbs: *die Bearbeitung der Aufgaben > die Aufgaben werden bearbeitet*. Das Argument A1 wird in 89,11% der Belege realisiert, sei es als monoargumentale Struktur oder auch in Kombination mit anderen Argumenten auf der Oberfläche.

Eine eingehendere Auseinandersetzung mit der kombinatorischen Semantik aus statistischer Sicht erbringt Hinweise dafür, dass Prädikate bestimmter semantischer Unterklassen⁷ tendenziell mit einigen semantischen Klassen in den Argumentslots kombinieren. Aus statistischer Perspektive können diese jedoch lediglich als unabhängige Variablen betrachtet werden. Der χ^2 -Test ist in diesem Zusammenhang nicht signifikant ($p = 0,09345327$) und kann daher nach dem Resampling-Verfahren nicht als angemessen beurteilt werden. Die Kombinationen sind — wie oben dargelegt — unabhängig von der Unterklasse des Prädikats, d. h., {Geschehen} kombiniert im Allgemeinen gleichermaßen mit {Artefakt} (siehe (13)), {Gruppe} (siehe (14)), {Mensch} (siehe (15)) oder {Kognition} (siehe (16)):

- (13) An der **Entwicklung von Bild und Tonsystemen** für das digitale Kino der Zukunft sind viele Firmen interessiert [...].
- (14) 1981 begann die **Auslieferung der Baureihe 111** [...].
- (15) Dem Zyklus kommt damit auch eine besondere Bedeutung für die **Entwicklung der Figur des Dagobert Duck** zu.
- (16) Bei der **Entwicklung der Systemtheorie** durch Physiker, Mathematiker und Techniker gingen diese immer wieder auf Analogien [...].

Eine Analyse des zweiten Arguments zeigt, dass es typischerweise die semantische Rolle der Ursache oder des Mittels für die im Nomen ausgedrückte Zustandsveränderung übernimmt. Auf diese Weise wird nach wie vor die Annahme untermauert, dass in den -ung-Strukturen eine Konstellation vorliegt, in der ein Element (A1) eine Veränderung bzw. einen Prozess erfährt, die von einem anderen Element hervorgerufen wird (A2). Ein Beispiel ist *die Erleichterung der Schmerzen durch Tabletten*, wobei die *Tabletten* die Ursache für die *Erleichterung* sind.

Statistisch gesehen ist die explizite Nennung dieses Arguments selten; es tritt nur in 6,12% der Belege auf und erscheint dabei fast nie allein, sondern in Kombination mit anderen Argumenten. Obwohl es theoretisch möglich ist, dass es als einziges Argument realisiert wird (z. B. in Fällen wie *eine Ernährung durch Filtration*), ist dies unter den analysierten Belegen kaum zu beobachten. Syntaktisch wird dieses Argument meist durch eine Präpositionalphrase mit *durch*, seltener auch mit *mit* oder *von*, realisiert. Seine kausale Rolle lässt sich durch eine verbale Paraphrase bestätigen, bei der das Element die Subjektrolle einnimmt: Aus *die Ernährung [der Pflanzen] durch Filtration* wird der Satz *die Filtration ernährt [die Pflanzen]*. Gerade aus diesem Grund folgen die frequentesten Strukturen (n = 2) dem Muster [Prädikat (*-ung*)] + [A2 (Präposition)] + {Substanz]}:

(17) Durch **Behandlung der Schmelze mit Magnesium** [...].

Statistische Tests sind im vorliegenden Fall nicht aussagekräftig, da die Muster nur sehr selten gemeinsam auftreten (d. h. eine geringe Kookkurrenz aufweisen). Eine valide statistische Analyse ist bei einer derart geringen Fallzahl nicht gewährleistet.

Die Realisierung des dritten Arguments ist an spezifische Bedingungen geknüpft und erfolgt nicht standardmäßig. Seine Projektion lässt sich auf zwei primäre Mechanismen zurückführen:

- (i) Valenzvererbung (vgl. Hölzner 2007, Teubert 2003): Das Argument wird direkt vom zugrunde liegenden Verb vererbt. Das Nomen übernimmt also die Valenzstruktur des Verbs, von dem es abgeleitet ist, und lizenziert dadurch dasselbe Argument, wie in *einer baldigen Versetzung ins Ausland* oder *Anpassung urbaner Räume an den Klimawandel*.
- (ii) Inhärente Semantik: Alternativ kann das Argument auch durch die eigenständige, lexikalische Bedeutung des Stammnomens motiviert sein. Dies ist insbesondere dann der Fall, wenn man davon ausgeht, dass das erste (A1) und zweite Argument (A2) die Kernargumente der morphologischen *-ung*-Konstruktion darstellen und das dritte Argument eine zusätzliche, semantisch geforderte Ergänzung ist, wie in *Freisetzung von Chemikalien in ihre Nachbarschaft*.

Die Realisierung des dritten Arguments ist direkt von der Semantik des jeweiligen Prädikats abhängig. Das bedeutet, dass seine Realisierung und Funktion nicht pauschal für alle Nominalisierungen auf *-ung* beschrieben werden können. Aufgrund dieser starken lexikalischen Abhängigkeit entzieht es sich einer systematischen Generalisierung über die gesamte Klasse der *-ung*-Derivate hinweg. Eine ad hoc produzierte Struktur mit verdichteter Syntax kann alle drei Argumente gleichzeitig lizenzieren: *die Beteiligung der Studierenden durch die zuständige Abteilung an den organisierten Aktivitäten*.

Als viertes Argument wurde in einigen analysierten Belegen eine Adjunktorphrase (vgl. Wöllstein und Dudenredaktion 2022) als Nominalergänzung (Domínguez Vázquez 2011) identifiziert. Dieses Argument lässt sich leicht syste-

matisieren, da es ohne Restriktion bei vielen Nomina möglich ist — wie etwa die Verwendung von *Gusseisen als Baumaterial*.

Im Gegensatz zu A1 und A2, die das „Minimalprogramm“ der Argumentstruktur von *-ung*-Nominalisierungen bilden, werden weitere Argumente nur unter spezifischen Bedingungen projiziert: A3 wird als lexikalisch-spezifisch betrachtet, während A4 als unbeschränkt produktiv bei Nomina definiert wird. DaF-Lernende würden bereits profitieren, wenn lexikologische Informationssysteme die A1- und A2-Muster der {Geschehen}-Lesart im Wörterbuch als semantische Information explizit darstellen. Auf dieser Grundlage können sie produktiv Hypothesen bilden, wie neue *-ung*-Nominalisierungen im Deutschen strukturiert sind. Diese Hypothesen können dann durch Wörterbuchrecherche validiert oder korrigiert werden: Wenn die gebildete Hypothese mit den lexikographischen Daten übereinstimmt, wird sie bestätigt; wenn nicht, können Lernende ihre Annahmen anpassen. Im Folgenden wird vorgeschlagen, wie diese Erkenntnisse in einer verbesserten lexikographischen Darstellung der *-ung*-Lemmata praktisch unterstützend wirken können.

6. Vorschlag zur Verbesserung lexikographischer Darstellung

Eine kritische Auseinandersetzung zwischen der lexikographischen Darstellung von *-ung*-Derivaten in monolingualen Wörterbüchern (siehe Abschnitt 2, Abbildung 1) und den empirischen Daten aus dieser Studie (siehe Abschnitt 5) zeigt ein erhebliches Defizit: Bestehende Wörterbücher bieten keine Informationen, die DaF-Lernende bei der Verifizierung oder Korrektur ihrer Ausgangshypothesen unterstützen. Müller-Spitzer et al. (2018: 309-310) beschreiben diesen Prozess präzise:

The students begin by formulating an initial hypothesis [...], based either on intuition before initiating a search process or on hypotheses formulated on the basis of a specific search action, such as the search for the meaning or translation equivalent of a word. From this point onwards, the whole search process focuses exclusively on the attempt to confirm this hypothesis [...].

Fortgeschrittene Lernende (auf der C-Stufe gemäß des Gemeinsamen Europäischen Referenzrahmens) formulieren Hypothesen über die Argumentstruktur von *-ung*-Nomina und versuchen, diese mit einem monolingualen Wörterbuch zu bestätigen — finden dort jedoch keine relevanten Informationen. Die vorliegende Studie zeigt, dass lexikologische Informationssysteme um folgende Aspekte erweitert werden sollten: (a) semantische Beschreibung der Basis-Lesart: *-ung*-Nominalisierungen drücken in ihrer {Geschehen}-Lesart typischerweise Prozesse oder Zustandsveränderungen aus; (b) Argumentstruktur-Muster: Explizite Informationen, wie A1 und A2 generell realisiert werden; (c) Produktive Anwendungsbeispiele: konkrete Beispiele für vollständig realisierte Argumentstrukturen, die Lernenden zeigen, wie die Muster funktionieren; (d) Einbezug regulärer Polysemie: Hinweis darauf, dass *-ung*-Nominalisierungen oft zwischen {Geschehen}

und beispielsweise {Gruppe}, {Artefakt} oder {Kognition} alternieren. Solche lexikographischen Angaben würden DaF-Lernenden ermöglichen, ihre Hypothesen nicht nur zu überprüfen, sondern auch eigenständig neue Nominalisierungen produktiv zu bilden.

In der Mikrostruktur monolingualer Wörterbücher sollten erweiterte Angaben unter dem Lemma *-ung* hinzugefügt werden, wie am Beispiel des Online-Wörterbuchs *Den Danske Ordbog*. Im semantischen Teil (vgl. Engelberg und Lemnitzer 2009) sollte zunächst erläutert werden, dass *-ung* ein hochproduktives Suffix im Deutschen ist, das primär aus Verben abgeleitete Nominalisierungen bildet.

Diese Nomina bezeichnen in ihrer prototypischen {Geschehen}-Lesart Prozesse oder Zustandsveränderungen. Wie in Abschnitt 5.2 dargelegt, ist dabei zwischen kausativ-resultativen Mustern und telischen Prozessen ohne eigentliche Zustandsveränderung des internen Arguments zu unterscheiden. Aus lexikographischer Perspektive ist diese Unterscheidung jedoch nicht als Einschränkung des vorgeschlagenen Modells zu verstehen: Wie die Phrase *die Beobachtung der Phänomene durch die Wissenschaftler zeigt*, lizenzieren auch telische *-ung*-Nominalisierungen eine biargumentale Struktur mit A1 als Thema der Handlung und A2 als deren Agens oder Mittel. Der Unterschied liegt nicht in der Argumentstruktur selbst, sondern in der semantischen Rolle von A1: Bei resultativen Nominalisierungen trägt A1 die Rolle des Patiens, das eine Zustandsveränderung erfährt; bei telischen Nominalisierungen trägt A1 die Rolle des Themas, auf das die Handlung gerichtet ist, ohne dass eine Transformation stattfindet. Das lexikographische Modell kann beide Fälle unter einem einheitlichen Rahmen erfassen, sofern die semantische Rolle von A1 je nach Verbklasse entsprechend spezifiziert wird.

Der Einbezug dieser Informationen würde den produktiven Prozess von DaF-Lernenden unterstützen, wie in der Abbildung 5 veranschaulicht wird: Von einem bereits bekannten Verb können sie ad hoc ein Nomen bilden. Wenn die Argumentstruktur deutlich in monolingualen Wörterbüchern dargestellt ist, können sie diese Muster replizieren und neue Konstruktionen erstellen — vorausgesetzt, ihnen ist klar, welche die zugrunde liegenden Annahmen sind und dass nicht alle Verben diese Nominalisierung gestatten.

Es ist anzumerken, dass die beschriebenen Informationen auch idealerweise in der Mikrostruktur der einzelnen *-ung*-Lemmata dokumentiert werden sollten. Allerdings trägt ihre systematische Integration in das Affix-Lemma *-ung* bereits wesentlich dazu bei, dass: (a) verschiedene Fälle abgedeckt werden, denn Benutzer:innen können von einem Wörterbuchartikel aus auf die Argumentstruktur-Muster zugreifen, ohne jedes einzelne Nomen nachschlagen zu müssen; (b) produktive Wortbildung unterstützt wird, denn durch das Verständnis der allgemeinen Muster können Benutzer:innen neue Nominalisierungen eigenständig bilden; (c) Wissensbildung gefördert wird, denn Benutzer:innen entwickeln ein metalinguistisches Bewusstsein für die Produktivität morphologischer Prozesse.

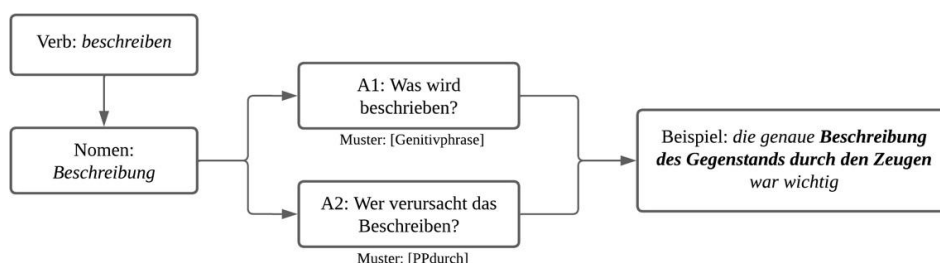


Abbildung 5: Produktionsprozess und Hypothesenbildung mithilfe verbesserter lexikologischer Informationssysteme

Die hier vorgeschlagene Beschreibungsperspektive unterscheidet sich von der klassischen Valenzlexikographie in einem grundlegenden Punkt: Während Valenzwörterbücher wie Sommerfeldt und Schreiber (1977) oder *Portlex* (Domínguez Vázquez und Valcárcel Riveiro 2019) die Argumentstruktur vom einzelnen Lemma aus beschreiben, geht der vorliegende Ansatz von einer übergeordneten morphologischen Einheit aus: dem Suffix *-ung* als Organisationsprinzip einer Familie von Argumentstrukturmustern. Diese Perspektive findet eine Entsprechung in der konstruktionsgrammatisch orientierten Arbeit von Proost (2023), die zeigt, dass Argumentstrukturmuster als Abstraktionen über individuelle Prädikat-Argumentstrukturen beschrieben werden können, die eine stabile Form-Bedeutungs-Paarung aufweisen und in Musterfamilien organisiert sind.

Der vorliegende Beitrag geht dabei in zweierlei Hinsicht über bestehende Ansätze hinaus. Erstens knüpft er an die lexikographische Praxis des DWDS an, das unter dem Lemma *-heit* bereits semantische Kategorialinformation bereitstellt, und erweitert diese Logik für *-ung* um die syntaktisch-semantische Schnittstelle: nicht nur was das Suffix bedeutet, sondern welche Argumentstellen es systematisch lizenziert und wie diese realisiert werden. Zweitens ist diese Information aus einer produktiven Perspektive entscheidend: Sie erlaubt fortgeschrittenen DaF-Lernenden genau jene Muster zu validieren, die sie für die eigenständige Bildung neuer *-ung*-Nominalisierungen benötigen. Die praktische Umsetzung eines solchen Modells fügt sich dabei in den gegenwärtigen Wandel des lexikographischen Prozesses ein: Klosa-Kückelhaus und Tiberius (2025) zeigen, dass die Entwicklung zentralisierter lexikographischer Datenbanken neue Möglichkeiten eröffnen, strukturierte (morpho-lexikalische) Information in digitale Wörterbücher zu integrieren. Die hier vorgeschlagenen Argumentstrukturmuster könnten in einem solchen Rahmen semi-automatisch aus annotierten Korpora gewonnen und in bestehende Wörterbuchinfrastrukturen eingespeist werden. Dieser Ansatz ist über den vorliegenden Einzelfall hinaus auf weitere derivationell produktive Wortklassen übertragbar.

7. Schlussfolgerung

Die vorliegende Studie ist von einer doppelten lexikographischen Lücke ausgegangen: Allgemeine Wörterbücher beschreiben das Suffix *-ung* morphologisch, ohne die damit verbundenen Argumentstrukturmuster explizit zu machen; Valenzwörterbücher beschreiben Argumentstrukturen auf Lemmaebene, ohne den morphologischen Mechanismus, der diese Strukturen systematisch motiviert, lexikographisch sichtbar zu machen. Der vorliegende Beitrag hat untersucht, wie eine korpusgestützte Argumentstrukturanalyse diese Lücke schließen kann.

Die Analyse der 200 Belege zeigt, dass die (semi)automatische Extraktion von Argumentstrukturen aus einem semantisch annotierten Korpus eine zuverlässige empirische Grundlage für die lexikographische Beschreibung deverbaler *-ung*-Nominalisierungen liefert. Aus linguistischer Perspektive bestätigt sich eine stabile biargumentale Grundstruktur, die das semantische Kernmuster der {Geschehen}-Lesart konstituiert. Dabei zeigt die Studie, dass automatische Extraktion durch menschliche Expertise begleitet werden muss: Fälle regulärer Polysemie und falsch-positive Treffer erfordern eine manuelle Validierung, die zugleich als Grundlage für künftige automatische Disambiguierungsverfahren dienen kann. Damit wird die erste Leitfrage der Untersuchung beantwortet.

Die Studie schlägt vor, die Mikrostruktur des Suffix-Lemmas *-ung* in allgemeinen monolingualen digitalen Wörterbüchern systematisch um syntaktisch-semantische Schnittstelleninformation zu erweitern. Dieser Schritt geht über das hinaus, was das DWDS bereits für *-heit* leistet, und orientiert sich an dem von Proost (2023) beschriebenen Prinzip, wonach morphosyntaktische Einheiten Familien von Argumentstrukturmustern mit stabiler Form-Bedeutungs-Paarung lizenzieren. Das vorgeschlagene Modell ist dabei nicht als starres Schema zu verstehen, sondern als prototypisches Muster, das je nach lexikalischer Semantik des Basisverbs variiert. Aus der Perspektive fortgeschrittener DaF-Lernender ist diese Information entscheidend: Sie erlaubt die Validierung von Hypothesen über neue Nominalisierungen und unterstützt deren produktive Bildung. Damit wird die zweite Leitfrage sowie deren praktische Implikationen beantwortet.

Für zukünftige Forschung wäre es wünschenswert, das vorgeschlagene lexikographische Modell prototypisch zu implementieren und in Usability-Studien mit DaF-Lernenden zu evaluieren. Darüber hinaus ist der methodische Ansatz — korpusgestützte Extraktion, semantische Klassifikation mittels einer Ontologie und statistische Validierung — auf weitere produktive Derivationsuffixe des Deutschen sowie prinzipiell auf andere Sprachen mit vergleichbaren Ressourcen übertragbar. Die Integration argumentstruktureller Information in die Mikrostruktur digitaler Wörterbücher stellt damit kein Einzelfalldesiderat dar, sondern ein generelles Prinzip der morphologisch motivierten Beschreibung derivationell produktiver Wortklassen.

Danksagung

Dieser Aufsatz ist Teil des Projekts PID2022-137170OB-I00, finanziert durch MCIN/AEI/10.13039/501100011033/FEDER, EU, wobei PID2022-137170OB-I00 die in der Bewilligungsentscheidung genannte Referenz ist; MCIN das Akronym des Ministeriums für Wissenschaft und Innovation der spanischen Bundesregierung; AEI das Akronym der Staatlichen Forschungsagentur; 10.13039/501100011033 der DOI (Digital Object Identifier) der Agentur; und FEDER das Akronym des Europäischen Fonds für regionale Entwicklung. Die Durchführung dieser Studie war z.T. dank des Programms zur Ausbildung von Universitätslehrenden (Förderkennzeichen FPU21/00188) vom spanischen Ministerium für Universitäten möglich.

Endnoten

1. Für das Tschechische ist in diesem Zusammenhang *NomVallex* von Kolářová und Vernerová (2022) zu nennen.
2. Erwähnenswert ist in dieser Hinsicht, dass das *Den Danske Ordbog* mit Daten aus dem dänischen *FrameNet* angereichert wurde (vgl. Nimb et al. 2017), wodurch semantisch-strukturelle Informationen über Argumentstrukturmuster systematisch in die lexikographische Beschreibung integriert werden konnten.
3. Der Korpusaufbau, die Korpusbeschreibung und die Korpusannotation sind grundlegender Teil einer sich in Erarbeitung befindenden Doktorarbeit. Auf Anfrage können Daten zur Verfügung gestellt werden.
4. Auf Anfrage wird der Algorithmus bereitgestellt.
5. Das Korpus enthält zum gegenwärtigen Zeitpunkt kein Disambiguierungsverfahren, da zunächst der Versuch unternommen wird, ausgehend von der Argumentstruktur Richtlinien zur Bedeutungsdisambiguierung vorzuschlagen. Dieser methodische Ansatz zielt darauf ab, syntaktisch-semantische Korrelationen zu identifizieren, die als Disambiguierungskriterien dienen können.
6. Die anderen untersuchten Korrelationen, beispielsweise zwischen der semantischen Klasse des Prädikats und der semantischen Klasse des Arguments, erwiesen sich als statistisch nicht signifikant und werden daher hier nicht weiter betrachtet.
7. Die semantische Klasse des Prädikats ist auf {Geschehen} festgelegt, da dies für die Analyse vordefiniert wurde. Die Annotation mit *GermaNet* ermöglicht jedoch eine granularere Analyse, indem auf Präfixe der Synsets zurückgegriffen wird. Theoretisch werden unter der Bezeichnung {Geschehen=s1.*} Einheiten subsumiert, die eine zumindest teilweise ähnliche Bedeutung in jeglicher Hinsicht aufweisen.

Literaturverzeichnis

- Arias-Arias, Iván, María José Domínguez Vázquez und Carlos Valcárcel Riveiro. 2025. People, Machines and Dictionaries: Is Artificial Intelligence Killing Dictionaries? Dias, Idalete, Rufus H. Gouws, Anja Lobenstein-Reichmann und Stefan J. Schierholz (Hrsg.). 2025. *Regarding the Significance of Lexicography: Europe and Australia*: 235-258. Berlin/Boston: De Gruyter.
<https://doi.org/10.1515/9783112216644-012>

- Arias-Arias, Iván.** o. D. (vsl. 2026). *(Semi)automatische semantische Annotation nominaler Argumentstrukturen im Deutschen: korpusbasierte Methoden und lexikographische Anwendungen* [Doktorarbeit]. Santiago de Compostela: Universidade de Santiago de Compostela.
- Baayen, R. Harald.** 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Bergenholtz, Henning und Rufus H. Gouws.** 2013. The Presentation of Word Formation in General Monolingual Dictionaries. *Lexikos* 23: 59-76.
<https://doi.org/10.5788/23-1-1204>
- Buscha, Anne, Susanne Raven und Szilvia Szita.** 2016. *Erkundungen C1. Deutsch als Fremdsprache (Integriertes Kurs- und Arbeitsbuch C1)*. Leipzig: Schubert.
- Cohen, Jacob.** 1988. *Statistical Power Analysis for the Behavioral Sciences*. Second edition. New York: Lawrence Erlbaum Associates.
- Den Danske Ordbog.** <https://ordnet.dk/ddo> [15. Oktober 2025].
- de Schryver, Gilles-Maurice.** 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387.
doi: 10.1093/ijl/ecad021
- Digitales Wörterbuch der deutschen Sprache (DWDS).** <https://www.dwds.de/> [15. Oktober 2025].
- Domínguez Vázquez, María José.** 2011. *Kontrastive Grammatik und Lexikographie: spanisch-deutsches Wörterbuch zur Valenz des Nomens*. München: Iudicum.
- Domínguez Vázquez, María José.** 2024. Analysing, Compiling, and Representing Argument Pattern Structures: from Form to Meaning and Back. Giacomini, Laura und Valentina Piunno (Hrsg.). 2024. *Patterns of Meaning in Lexicography and Lexicology*: 297-316. Berlin/Boston: De Gruyter.
<https://doi.org/10.1515/9783111545943>
- Domínguez Vázquez, María José und Carlos Valcárcel Riveiro.** 2015. Hábitos de uso de los diccionarios entre los estudiantes universitarios europeos: ¿nuevas tendencias? Domínguez Vázquez, María José, Xavier Gómez Guinovart und Carlos Valcárcel Riveiro (Hrsg.). 2015. *Lexicografía de las lenguas románicas: Aproximaciones a la lexicografía moderna y contrastiva* (vol. II): 165-190. Berlin/Boston: De Gruyter.
<https://doi.org/10.1515/9783110310337.165>
- Domínguez Vázquez, María José und Carlos Valcárcel Riveiro.** 2019. PORTLEX as a Multilingual and Cross-lingual Online Dictionary. Domínguez Vázquez, María José, Mónica Mirazo Balsa und Carlos Valcárcel Riveiro (Hrsg.). 2019. *Studies on Multilingual Lexicography*: 135-158. Berlin/Boston: De Gruyter.
<https://doi.org/10.1515/9783110607659-008>
- Duden.** <https://www.duden.de/woerterbuch> [15. Oktober 2025].
- Elexiko.** <https://www.owid.de/docs/elex/start.jsp> [15. Oktober 2025].
- Engel, Ulrich.** 2004. *Deutsche Grammatik. Neubearbeitung*. München: Iudicum.
- Engelberg, Stefan und Lothar Lemnitzer.** 2009. *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg.
- Fuertes-Olivera, Pedro A.** 2024. Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes. *Lexikos* 34(1): 123-140.
<https://doi.org/10.5788/34-1-1883>
- Fuertes-Olivera, Pedro A.** 2025. *A Guide to Practical Online Lexicography*. London: Routledge.
<https://doi.org/10.4324/9781003426998>
- Füllemann, Angelika.** 2023. *Grammatik C1. Das Training*. Stuttgart: Klett.

- GermaNet Rover*. <https://weblicht.sfs.uni-tuebingen.de/rover/> [15. Oktober 2025].
- Grammatisches Informationssystem (grammis)*. <https://grammis.ids-mannheim.de/> [15. Oktober 2025].
- Hamp, Birgit und Helmut Feldweg**. 1997. *GermaNet*. A Lexical-Semantic Net for German. Vossen, Piek, Geert Adriaens, Nicoletta Calzolari, Antonio Sanfilippo und Yorick Wilks (Hrsg.). 1997. *Proceedings of the ACL/EACL-97 Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*: 9-15. Madrid: Association for Computational Linguistics.
<https://aclanthology.org/W97-0802.pdf>
- Hennig, Mathilde**. 2019. *Nominalstil. Möglichkeiten, Grenzen, Perspektiven*. Tübingen: Narr.
- Herbst, Thomas und Peter Uhrig**. 2009. *Erlangen Valency Patternbank. A Corpus-based Research Tool for Work on Valency and Argument Structure Constructions*.
<https://www.erlangen-linguistik-online.uni-erlangen.de/projekte/erlangen-valency-patternbank.shtml>
- Hinrichs, Marie, Richard Lawrence und Erhard Hinrichs**. 2020. Exploring and Visualizing WordNet Data with *GermaNet Rover*. Navaretta, Costanza und Maria Eskevich (Hrsg.). 2020. *CLARIN Annual Conference Proceedings*: 32-36. Virtual Edition.
https://office.clarin.eu/v/CE-2020-1738-CLARIN2020_ConferenceProceedings.pdf
- Hölnzer, Matthias**. 2007. *Substantivvalenz. Korpusgestützte Untersuchungen zu Argumentrealisierungen deutscher Substantive*. Tübingen: Niemeyer.
- Klosa-Kückelhaus, Annette (Hrsg.)**. 2013a. Einleitung. Klosa-Kückelhaus, Annette (Hrsg.). 2013. *Wortbildung im elektronischen Wörterbuch*: 7-12. Tübingen: Narr.
- Klosa-Kückelhaus, Annette**. 2013b. Wortbildung in *lexiko*: Gegenwart und Zukunft. Klosa-Kückelhaus, Annette (Hrsg.). 2013. *Wortbildung im elektronischen Wörterbuch*: 175-196. Tübingen: Narr.
- Klosa-Kückelhaus, Annette und Carole Tiberius**. 2025. The Lexicographic Process Revisited. *International Journal of Lexicography* 38(1): 1-12.
<https://doi.org/10.1093/ijl/ecae016>
- Kolářová, Veronika und Anna Vernerová**. 2022. NomVallex: A Valency Lexicon of Czech Nouns and Adjectives. Calzolari, Nicoletta, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk und Stelios Piperidis (Hrsg.). 2022. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*: 1344-1352. Marseille, France: European Language Resources Association.
<https://aclanthology.org/2022.lrec-1.143/>
- Kosem, Iztok, Robert Lew, Carolin Müller-Spitzer, Maria Ribeiro Silveira, Sascha Wolfer, Amelie Dorn, Antton Gurrutxaga, Klara Ceberio, Elixabete Etxeberria, Marie-Aude Lefer, Dirk Geeraerts, Kristina Štrkalj Despot, Tomislav Stojanov, Nikola Ljubešić, Michal Škrabal, Barbora Štěpánková, Veronika Vodrážková, Henrik Lorentzen, Lars Trap-Jensen, Jelena Kallas, Maria Tuulik, Kristina Koppel, Margit Langemets, Tarja Heinonen, Izabella Thomas, Tinatin Margilitadze, Stella Markantonatou, Voula Giouli, Chris Mulhall, Ilan Kernerman, Yifat Ben-Moshe, Tsvi Sadan, Andrea Abel, Martina Nied Curcio, Lidija Tanturovska, Biljana Nikovska, Carole Tiberius, Oddrun Grønvik, Marit Hovdenak, Sturla Berg-Olsen, Knut E Karlsen, Christian-Emil Smith Ore, Monika Biesaga, Tanara Zingano Kuhn, Joao Silvestre, Elena Isabelle Tamba, Gabriela Haja, Marius-Radu Clim, Madalin-Ionel Patrascu, Toma Tasovac, Snežana Petrović, Špela Arhar Holdt, Carlos Valcarcel Riveiro, María José Domínguez Vázquez, Elena Volodina, Ildikó Pilán, Emma**

- Sköldberg, Louise Holmer und Hilary Nesi.** 2019. The Image of the Monolingual Dictionary across Europe. Results of the European Survey of Dictionary Use and Culture. *International Journal of Lexicography* 32(1): 92-114.
<https://doi.org/10.1093/ijl/ecy022>
- Levshina, Natalia.** 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Lew, Robert.** 2024. Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications* 11: 1-8.
<https://doi.org/10.1057/s41599-024-02889-7>
- Müller-Spitzer, Carolin, María José Domínguez Vázquez, Martina Nied Curcio, Idalete Maria Silva Dias und Sascha Wolfer.** 2018. Correct Hypotheses and Careful Reading Are Essential: Results of an Observational Study on Learners Using Online Language Resources. *Lexikos* 28: 287-315.
<https://doi.org/10.5788/28-1-1466>
- Nimb, Sanni, Anna Braasch, Sussi Olsen, Bolette Sandford Pedersen und Anders Søgaard.** 2017. From Thesaurus to *FrameNet*. Kosem, Iztok, Carole Tiberius, Miloš Jakubiček, Jelena Kallas, Simon Krek und Vít Baisa (Hrsg.). 2017. *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2017 Conference*: 1-22.
<https://elex.link/elex2017/wp-content/uploads/2017/09/paper01.pdf>
- Proost, Kristel.** 2023. Paradigmatische Relationen syntagmatischer Muster: Muster- und Familienbildung von Prädikat-Argumentstrukturen mit skalarer Bedeutung und präpositional eingeleitetem Argument. Mollica, Fabio und Sören Stumpf (Hrsg.). 2023. *Konstruktionsgrammatik IX. Konstruktionsfamilien im Deutschen*: 267-325. Tübingen: Stauffenburg.
- Pustejovsky, James.** 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Pustejovsky, James und Olga Batiukova.** 2019. *The Lexicon*. Cambridge: Cambridge University Press.
- Roßdeutscher, Antjie.** 2010. German *-ung*-Nominalisation. An Explanation of Formation and Interpretation in a Root-Based Account. Olsen, Susan (Hrsg.). 2010. *New Impulses in Word Formation*: 101-132. Hamburg: Buske.
- Roßdeutscher, Antjie und Hans Kamp.** 2010. Syntactic and Semantic Constraints on the Formation and Interpretation of *-ung*-nouns. Rathert, Monika und Artemis Alexiadou (Hrsg.). 2010. *The Semantics of Nominalizations across Languages and Frameworks*: 169-214. Berlin/New York: De Gruyter.
<https://doi.org/10.1515/9783110226546.169>
- Schierholz, Stefan J.** 2021. Die Villa Vigoni Thesen zur Lexikographie. *Lexicographica* 37(1): 303-305.
<https://doi.org/10.1515/lex-2021-0017>
- Sommerfeldt, Karl-Ernst und Herbert Schreiber.** 1977. *Wörterbuch zur Valenz und Distribution der Substantive*. Tübingen: Niemeyer.
- Schumacher, Helmut.** 1999. Von *verletzen* zu *verletzbar* und *Verletzung*. Zu einigen Aspekten des Zusammenhangs der Valenz von Verben, Adjektiven und Substantiven. Skibitzki, Bernd und Barbara Wotjak (Hrsg.). 1999. *Linguistik und Deutsch als Fremdsprache: Festschrift für Gerhard Helbig zum 70. Geburtstag*: 177-188. Tübingen: Niemeyer.
- Straka, Milan.** 2018. UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. Zeman, Daniel und Jan Hajič (Hrsg.). 2018. *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*: 197-207. Brussels, Belgium: Association for Computational Linguistics.
<https://aclanthology.org/K18-2020/>

- Teubert, Wolfgang.** 2003. Die Valenz nichtverbaler Wortarten: das Substantiv. Ágel, Vilmos, Ludwig M. Eichinger, Hans-W. Eroms, Peter Hellwig, Hans Jürgen Heringer und Henning Lobin (Hrsg.). 2003. *Dependenz und Valenz. Ein internationales Handbuch zur zeitgenössischen Forschung. 1. Halbband*: 820-835. Berlin: De Gruyter.
- Ulsamer, Sabina.** 2013. Wortbildung in Wörterbüchern — Zwischen Anspruch und Wirklichkeit. Klosa-Kückelhaus, Annette (Hrsg.). 2013. *Wortbildung im elektronischen Wörterbuch*: 13-62. Tübingen: Narr.
- Welke, Klaus.** 2011. *Valenzgrammatik des Deutschen. Eine Einführung*. Berlin/New York: De Gruyter.
- Wolfer, Sascha, Martina Nied Curcio, Idalete Maria Silva Dias, Carolin Müller-Spitzer und María José Domínguez Vázquez.** 2018. Combining Quantitative and Qualitative Methods in a Study on Dictionary Use. Čibej, Jaka, Vojko Gorjanc, Iztok Kosem und Simon Krek (Hrsg.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*: 101-112. Ljubljana: Ljubljana University Press, Faculty of Arts.
<https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/2981-1>
- Wöllstein, Angelika und Dudenredaktion.** 2022. *Duden — Die Grammatik*. Mannheim: Dudenverlag.
- Zifonun, Gisela, Ludger Hoffmann und Bruno Stecker.** 1997. *Grammatik der deutschen Sprache*. Berlin/Boston: De Gruyter.