



Thilo Weber  & Sandra Hansen 

Schwankungen zwischen schwacher und starker Substantivflexion

Abstract The present corpus-study investigates fluctuations between the so called “weak” and “strong” nominal inflection classes in German. Corroborating previous studies (Köpcke 1995, Schäfer 2019), we show that the tendency for traditionally weak masculine nouns to join the strong pattern is strongest for nouns displaying phonotactic and semantic properties that are atypical of weak nouns. Conversely, we show that among strong masculine nouns attested in a weak form at least once, nouns with phonotactic and semantic properties typical of weak nouns are overrepresented compared to masculine nouns not attested in weak forms. Using logistic regression, we show, among other things, that non-canonical forms are more likely to occur in informal texts (represented by internet forum discussions) than in (more formal) newspaper texts. While this is fully expected for the traditionally weak nouns, where the use of the non-canonical (strong) forms involves a loss of case suffixes (viz. the loss of *-(e)n* in the accusative and dative singular), it is perhaps more noteworthy with respect to the traditionally strong nouns, where, conversely, the use of the non-canonical (weak) forms leads to additional case morphology. Moreover, we find that the shift from strong to weak differs from the shift from weak to strong with respect to grammatical case (accusative/dative vs. genitive). We propose that is connected to the status of the genitive as a marker of formal/written style.

Keywords morphology, inflection, declension, noun, case, strong, weak

Open Data Der dem Kapitel zugrundeliegende Datensatz „Schwache Maskulina“ ist öffentlich zugänglich (DOI: 10.14618/schwachemaskulinadb).

1 Einleitung

Substantive weisen im Deutschen ein inhärentes Genus auf und flektieren nach Kasus und Numerus. Dabei unterscheiden sie sich gruppenweise dahingehend, in welcher Paradigmenzelle sie welche Form annehmen; d. h., sie lassen sich unterschiedlichen **Flexionsklassen** zuordnen (z. B. Pavlov 1995, Thieroff 2000: 243–336, Nübling 2008, Dudengrammatik 2016: 194–220). Die meisten Nicht-Feminina des

Deutschen folgen der sog. „starken“ Flexion. Das heißt, sie tragen im Genitiv Singular typischerweise das Suffix *-(e)s*¹, während sie im Dativ und Akkusativ genauso wie im Nominativ endungslos bleiben – mit Ausnahme des mehr oder weniger reliktisch erhaltenen Dativ *-e* (zu Letzterem siehe z.B. Konopka 2012) (vgl. *der Tag, des Tages, dem Tag(e), den Tag*). Eine vergleichsweise kleine Anzahl von Maskulina mit mehrheitlich menschlichem bzw. zumindest belebtem Denotat² folgt indes dem sog. „schwachen“ Muster, bei dem im Genitiv, Dativ und Akkusativ Singular (sowie im Plural) das Suffix *-(e)n* auftritt (vgl. *der Matrose vs. des/dem/den Matrosen, der Bär vs. des/dem/den Bären*). Im vorliegenden Kapitel untersuchen wir Schwankungen zwischen diesen beiden Mustern. Vorherige korpusbasierte Untersuchungen zu diesem Gegenstandsbereich leisteten u. a. Kusová (2013, 2014), Paulfranz (2013) und Schäfer (2019).

Wie allgemein bekannt, besteht eine Tendenz dazu, das seltenere, schwache Muster analogisch an das häufigere, starke Muster (das „Standardmuster“ für Maskulina) anzugleichen (vgl. Dudengrammatik 2016: 213). Schwankungen finden sich daher vor allem bei den traditionell schwach flektierenden Substantiven. Betroffen sind insbesondere der Akkusativ (1) und Dativ (2) (hierzu mehr in Abschnitt 4.3.2), wo neben Formen auf *-(e)n* (1a, 2a) somit auch endungslose Formen begegnen (1b, 2b). In geringerem Maße ist im Genitiv eine Ersetzung von *-(e)n* (3a) durch *-(e)s* (3b) zu beobachten.

- (1) a. Man muss ihn nicht mögen, **den Berliner Bären**. (die tageszeitung, 25.2.2005, S. 25)
- b. Sie sprach trotzdem auf der Gedenkveranstaltung für **den erschossenen Bär** Bruno auf dem Münchner Marienplatz [...]. (Süddeutsche Zeitung, 21.9.2007, S. 35)
- (2) a. Stimmen die Merkmale überein, handelt es sich **beim Bären** aus dem Münstertal um ein eineinhalb Jahre altes Jungtier. (Saale-Zeitung, 20.11.2013, S. 6)
- b. Da bei Kronstadt die höchste Braunbärdichte weltweit mit **einem Bär** pro Quadratkilometer herrscht, reagieren die Menschen gelassen. (Die Südostschweiz, 15.8.2005, o.S.)

1 Siehe Bubenhofer, Hansen-Morath & Konopka (2014) zur Gesamtheit der Varianten der starken Genitivmarkierung und zu deren Verteilung.

2 Wir verwenden im vorliegenden Kapitel auf die Lexemebene bezogen den Begriff „Denotat“. Geht es hingegen darum, auf wen oder was sich ein konkreter Beleg eines Lexems bezieht, sprechen wir von „Referent“.

- (3) a. Der polnische Präsident Walesa mahnt einen raschen Beitritt an, weil der Erfolg Schirinowskijs die Gefährlichkeit **des russischen Bären** zeige. (Süddeutsche Zeitung, 5.1.1994, S. 4)
 b. Die Heimat **des größten Bärs** der Welt sind die Harburger Berge. (Der Prignitzer, 25.2.2013, S. 25)

Seltener lassen sich auch Schwankungen in die umgekehrte Richtung beobachten. Sie betreffen dann also Substantive, die (zumindest im Singular) traditionell stark flektieren. Beispiele dafür sind Substantive auf *-or* (Kusová 2013); der wohl bekannteste Fall ist das Lexem *Autor*, vgl. (4–6).³

- (4) a. Ich weiß nicht, ob wir **einen weiteren Autor** beauftragen werden, um die Serie zu beenden. (<http://www.starwars-union.de/forum/board15-das-erweiterte-universum-eu/board16-literatur/board34-b-cher/15860-imperial-commando-2>)
 b. Von Spenden für die CDU ist die Rede, von einem Geldbetrag für **einen Autoren** eines Wulff-Buches. (Die Rheinpfalz, 10.4.2013, o.S.)
- (5) a. Und dann vielleicht mit **einem zweiten Autor** als Unterstützung... (<http://www.tabletopwelt.de/forum/archive/index.php/t-91621.html>)
 b. Die Ausführung stammt zum großen Teil vom amerikanischen **Autoren** Eric Hufschmid [...]. (<http://www.panzerfreund.de/forum/bothread-6767.html>)
- (6) a. Das 420 Seiten starke Buch stammt aus der Feder **des Autors** Adam Victor. (<http://www.elvispresley-community.de/archive/index.php/t-4150.html>)
 b. Im Mittelpunkt stehen in diesem Jahr die Märchen **des dänischen Autoren** Hans Christian Andersen. (Hamburger Morgenpost, 10.11.2005, S. 14)

Kusová (2013, 2014) untersucht die Flexionsschwankungen im Deutschen Referenzkorpus (DeReKo), Paulfranz (2013) in deutschen Lokal- und Regionalzeitungen. Schäfer (2019) verwendet das DECOW16B-Webkorpus (Schäfer & Bildhauer 2012), wobei er ein multivariates Modell verwendet. Im vorliegenden Kapitel wenden wir Schäfers (2019) multivariaten Ansatz auf Stichproben aus dem auf DeReKo

3 Da Lexeme wie *Autor* ohnehin nur im Singular stark flektieren, während sie im Plural mit *-en* markiert werden, werden sie insgesamt der „gemischten“ Flexion zugeordnet. Wenn wir sie hier als „stark“ bezeichnen, dann beziehen wir uns dabei auf ihr Verhalten im Singular.

(Release 2017-II, vgl. Kupietz et al. 2018) basierenden KoGra-Untersuchungskorpus des Projekts Korpusgrammatik (Bubenhofer et al. 2014) und dem Subkorpus Foren des DECOW-Webkorpus an. Damit zielen wir auf die Erfassung potentieller Unterschiede zwischen redigierten und distanzsprachlichen Texten (KoGra) und eher informellen und nähesprachlichen Texten (Webforen) ab.

Aufbauend auf Köpcke (1995), zeigt Schäfer (2019), dass eine wichtige Rolle spielt, wie prototypisch das jeweilige Lexem für die Klasse der schwachen Maskulina ist. So halten schwache Maskulina mit prototypischen semantischen und phonotaktischen Eigenschaften eher am schwachen Muster fest als Lexeme mit weniger prototypischen Eigenschaften. Im vorliegenden Kapitel erweitern wir Schäfers Analyse um einen um weitere potentiell variationssteuernde Faktoren. Diese umfassen das Auftreten einer explikativen Apposition unmittelbar nach dem Maskulinum (*dem Fabrikant(en) Meier*) sowie das Vorkommen eines vorangehenden Adjektivs (*dem italienischen Architekt(en)*). Zum anderen vertiefen wir die Untersuchung der selteneren Schwankungen von stark zu schwach: Anders als die Schwankung von schwach zu stark untersucht Schäfer (2019) letztere nicht im Rahmen eines multivariaten Modells, sondern auf Grundlage deskriptiver Statistiken für eine Auswahl von *or*-Substantiven. In der vorliegenden Untersuchung betrachten wir beide Schwankungsrichtungen mittels multivariater Modelle, und bei der Schwankung von stark zu schwach berücksichtigen wir auch Substantive, die nicht auf *-or* enden.

Wie sich zeigen wird, erklärt der Prototypenansatz nicht nur die Schwankungen innerhalb der traditionell schwach flektierenden Maskulina. Er liefert auch einen Beitrag zum Verständnis dafür, welche starken Maskulina sich den schwachen anschließen. Darüber hinaus zeigen wir, dass Abweichungen vom jeweils kodifizierten Muster eher in den Webforen als in KoGra auftreten. Für die traditionell schwach flektierenden Substantive ist dies das erwartete Ergebnis: Denn das starke Muster bringt hier – im Akkusativ und Dativ – einen Abbau von Kasusflexion mit sich (Endungslosigkeit statt *-(e)n*). Dies passt zum allgemeineren Trend informeller Varietäten des Deutschen. Wie sich zeigt, verhalten sich die Foren aber auch bei den starken Substantiven in stärkerem Maße Norm-divergent. Dadurch kommt es im Akkusativ und Dativ zu einem Mehr an Kasusmorphologie im Vergleich mit den KoGra-Texten. Zudem finden wir auch spezifische Unterschiede zwischen den beiden Schwankungsrichtungen. Für die Schwankung von schwach zu stark bestätigt unsere Analyse den Gegensatz zwischen Genitiv einerseits und Akkusativ und Dativ andererseits. Für die Schwankung von stark zu schwach finden wir hingegen keinen Effekt für den Faktor Kasus. Wir schlagen vor, dass dies mit dem Status des Genitivs als Marker für Hoch- und Schriftsprachlichkeit zusammenhängt.

2 Hintergrund

Die Begriffe „stark“ und „schwach“ sind im Zusammenhang mit der Substantivflexion mehrdeutig. Sie dienen einerseits der Typisierung von Formen bzw. Allomorphen. Andererseits werden sie als Dachbegriffe für historische Deklinationenklassen verwendet, die sich auf Grundlage dieser Formen bzw. Allomorphe unterscheiden lassen. Im vorliegenden Kapitel spielen beide Verwendungen eine Rolle. Wir untersuchen, inwiefern Substantive, die sich historisch (bzw. hier eher: der kodifizierten Norm gemäß, vgl. Abschnitte 4 und 5) der einen bzw. der anderen Klasse zuordnen lassen, im Sprachgebrauch auch tatsächlich in den entsprechend erwartbaren Formen auftreten oder stattdessen in den Formen der jeweils anderen Klasse.

Als „schwache Formen“ bezeichnen wir hier Akkusativ-, Dativ- und Genitiv-Singular-Formen mit der „schwachen Endung“ *-(e)n*. Als „starke Formen“ bezeichnen wir Genitiv-Singular-Formen mit der Endung *-(e)s*, Dativ-Singular-Formen ohne Endung oder mit der Endung *-e* sowie Akkusativ-Singular-Formen ohne Endung.⁴

2.1 Typische und untypische schwache Maskulina

Köpcke (1995: 168–176) untersucht die historische Klasse der schwachen Maskulina auf der Lexemebene und zeigt, dass sie sich anhand phonotaktischer und semantischer Merkmale in prototypische und weniger prototypische Vertreter gliedern lässt. In phonotaktischer Hinsicht prototypisch ist etwa finales Schwa, in semantischer Hinsicht prototypisch ist ein menschliches Denotat. Substantive wie z. B. *Matrose* vereinen diese beiden Eigenschaften.

Schäfer (2019) wendet Köpckes Ansatz im Rahmen einer Korpusuntersuchung auf die Tokenebene an, um damit auch Schwankungen bei einzelnen Substantiven erfassen zu können. Er zeigt, dass die Substantive mit prototypischen Merkmalen in stärkerem Maße am schwachen Muster festhalten als Substantive mit weniger prototypischen Merkmalen (2019: 405). Auf der Basis einer Auswahl von 62 *or*-Substantiven demonstriert Schäfer zudem, dass der Prototypen-Ansatz auch zum Verständnis von Schwankungen in die umgekehrte Richtung beiträgt: Es wird deutlich, dass der Anteil schwacher Formen bei Substantiven mit menschlichem Denotat – wie eben *Autor* – höher ist als bei solchen mit nicht-menschlichem Denotat (wie z. B. *Transistor*).

4 Folgt man einer engeren Definition und macht die Bezeichnung „starke Form“ vom Vorliegen eines overtten starken Suffixes abhängig, dann wären nur Genitive auf *-(e)s* und Dative auf *-e* als starke Formen zu bezeichnen.

2.2 Variation im Nominativ Singular und Genitivformen auf *-(e)ns*

Bei einem Teil der historisch schwachen Maskulina, die auf Schwa enden und einen unbelebten Referenten haben, kann *-n* auch im Nominativ Singular auftreten (Thieroff & Vogel 2009: 45–46, Augustin 2012: 77–81, Paulfranz 2013: 90–91). Thieroff & Vogel (2009) zufolge ist bei *Buchstabe(n)*, *Gedanke(n)*, *Name(n)* und *Wille(n)* noch die *n*-lose Form bevorzugt; bei *Funke(n)*, *Glaube(n)* und *Hode(n)* kämen beide Formen gleichermaßen vor; und bei *Friede(n)*, *Gefalle(n)*, *Haufe(n)*, *Same(n)* und *Schade(n)* sei die Form ohne *-n* bereits veraltet. Augustin (2012) führt Korpusanalysen zu einzelnen Lexemen durch und zeigt, dass die Variation auch von der intendierten Lesart abhängt. So neigt z. B. *Funke(n)* im wörtlichen Sinne (‘glühendes Teilchen’) stark zur *n*-losen Variante (z. B. *ein Funke springt*), während bei Verwendungen mit übertragener Bedeutung (‘ein bisschen’, vgl. z. B. *ein Funken Hoffnung*) die *n*-haltige Form überwiegt. Geht man davon aus, dass der Nominativ Singular immer der Stammform entspricht, dann sind die Formen auf *-n* als Stammvarianten anzusehen. Damit wiederum sind Akkusativ- und Dativformen auf *-n* nicht eindeutig analysierbar. Es kann sich um eine endungshaltige Form zu einem *n*-losen Stamm handeln (*einen/einem Funke-n*), oder um eine endungslose Form zu einem *n*-haltigen Stamm (*einen/einem Funken_*).

Im Genitiv Singular kommt bisweilen eine Form auf *-(e)ns* vor (Krischke 2012, vgl. z. B. *des Bärens*, *des Bauerns*, *des Automaten*). Krischke bezeichnet diese als „kasusaugmentierende“ Variante; sie lässt sich als Kombination aus schwachem und starkem Suffix analysieren. Besonders regelmäßig kommt sie bei den o. g. Substantiven vor, die *-n* auch im Nominativ Singular aufweisen (können). Parallel dazu, wie bei diesen Substantiven Akkusativ- und Dativformen auf *-n* auch als endungslose („starke“) Formen zu einem *n*-haltigen Stamm gedeutet werden können, können Genitivformen auf *-ns* hier als reguläre starke Genitivformen auf *-s* gedeutet werden (*eines Funken-s*). Die Ausweitung von *-n* auf den Nominativ sowie die Einfügung des Genitiv *-s* nach dem *-n* stellt somit eine alternative Form der Angleichung des schwachen Musters an das starke Muster dar.

2.3 Syntaktisch bedingte Unterlassung der Kasusflexion

Im Akkusativ und Dativ führt die Angleichung der schwachen an die starke Flexion zu Endungslosigkeit (1b, 2b). Diese Fälle sind zu unterscheiden vom sog. „syntaktisch bedingte[n] Wegfall der Kasusendung *-en*“ (vgl. Dudengrammatik 2016: 213, Müller 2002). Dieser ist im Deutschen ganz regelmäßig dort zu beobachten, wo dem Substantiv weder ein flektierter Determinierer noch ein flektiertes Adjektiv vorangeht (7–8). Oft wäre bei einer Kasus-flektierten Form in solchen

Fällen auch unklar, ob Singular oder Plural intendiert ist (vgl. mit (7): *DAAD-Preis für **Studenten** aus China*).

- (7) DAAD-Preis für **Student** aus China (Dudengrammatik 2016: 213, unsere Hervorhebung)
- (8) Die Bewohner betrachteten den Feuerwehrmann als **Held** (neben: als Helden) (Dudengrammatik 2016: 973, unsere Hervorhebung)

Der syntaktisch bedingte Wegfall von *-(e)n* ist Teil einer allgemeineren Regel der Unterlassung von Kasusflexion (im Singular) in Fällen, in denen dem Substantiv kein flektierter Determinierer oder kein flektiertes Adjektiv vorausgeht (vgl. Dudengrammatik 2016: 973–977, Müller 2002). Er betreffe neben *-(e)n* auch das – allerdings ja ohnehin nur noch relikthaft vorhandene – Dativ-*e*, vgl. *Das ist eine Truhe aus Holz* vs. *Das ist eine Truhe aus hartem Holz(e)*.

Zu beachten ist, dass die syntaktisch bedingte Unterlassung der Kasusflexion auf Akkusativ und Dativ beschränkt ist. Für den Genitiv hat sich mit der sog. „Genitivregel“ (Dudengrammatik 2016: 977–978) eine Sonderregel herausgebildet, die von vornherein genau diejenige syntaktische Umgebung ausschließt, in der es zur syntaktisch bedingten Unterlassung der Kasusflexion gekommen wäre: So kann gemäß dieser Regel eine Nominalphrase „nur dann im Genitiv stehen [...], wenn sie (i) mindestens ein adjektivisch flektiertes Wort und (ii) mindestens ein Wort mit *s-* oder *r-*Endung enthält.“ In Kombination führen die syntaktisch bedingte Endungslosigkeit und die Genitivregel dazu, dass das *-(e)n*-Suffix am Substantiv nur dann auftritt, wenn der jeweilige nicht-nominativische Kasus ohnehin bereits am pränominalen Ausdruck eindeutig markiert ist. In Bezug auf die Markierung von Kasus ist das Suffix also redundant, was seinen Schwund im Akkusativ und Dativ begünstigen dürfte. Sowohl das Phänomen der syntaktisch bedingten Endungslosigkeit als auch die Tendenz zum Schwund des Suffixes selbst in Gegenwart eines Begleiters können als Teil einer allgemeineren langfristigen Entwicklungstendenz des deutschen Nominalsystems aufgefasst werden: nämlich der Verlagerung der Kasusflexion weg vom Substantiv hin zum Substantivbegleiter (hierzu auch Abschnitte 4.3.2).

In der vorliegenden Untersuchung beschränken wir uns auf Belege, in denen die syntaktische Bedingung für das Auftreten von Kasusflexion eindeutig erfüllt ist. Wir konzentrieren uns auf Fälle, in denen dem Substantiv der bestimmte oder unbestimmte Artikel und – optional – ein Adjektiv vorangeht. Zum bestimmten Artikel zählen wir auch dessen klitische Realisierung nach vorangehender Präposition (**zum Adjutanten**).⁵

5 Wie die Auszählung in Weber (2021: 21) ergibt, handelt es sich bei den Artikeln um die mit Abstand Token-frequentesten Determinierer.

3 Methodik der Korpusuntersuchung

3.1 Lemmaliste

In einem ersten Schritt wurden anhand des auf DeReKo basierenden Untersuchungskorpus KoGra (siehe Abschnitt 3.2) möglichst viele Lemmata ermittelt, für die die schwache Flexion mindestens einmal belegt ist. Dabei wurden nicht nur Simplexe, sondern auch Wortbildungen berücksichtigt – insbesondere deshalb, weil zahlreiche Maskulina überhaupt erst durch das Vorkommen eines bestimmten Suffixes zur Klasse der schwachen Maskulina gehören (*Doktorand*, *Flötist*, *Dirigent* etc., vgl. z. B. Dudengrammatik 2016: 212). Berücksichtigt wurden auch Lexeme mit (Prä)Konfix (z. B. *Urmensch*, *Expolizist*). Ausgeschlossen wurden jedoch prototypische Komposita, verstanden als Wörter, die mindestens zwei Elemente enthalten, die frei vorkommen könnten (z. B. *Militärpolizist*, *Landesverkehrsreferent*).

Für die Suche nach nicht-nominativischen Vorkommen von Substantiven bietet sich eine Suche nach nicht-nominativischen Determiniererformen an. Problematisch dabei ist jedoch der Synkretismus zwischen Akkusativ Singular und Dativ Plural (vgl. *den/diesen/meinen*/etc. *Bären/Autoren*). Die z. T. umfangreichen Suchergebnisse (allein die Suche nach der Sequenz *den Autoren* z. B. liefert im Gesamtkorpus 8.494 Treffer) müssten manuell daraufhin durchgesehen werden, ob mindestens einer der Belege als Singular aufzufassen ist. Schäfer (2019: 397–398) beschränkt seine Untersuchung aus diesem Grund auf Vorkommen nach dem indefiniten Artikel *ein-*, da dieser ja nur im Singular vorkommt. Wir folgen hier dieser Strategie – allerdings nur zum Zwecke der Erstellung der Lemmaliste. Bei der eigentlichen, Token-basierten Variationsanalyse, bei der wir uns auf manuell nachannotierbare Stichproben beschränken, berücksichtigen wir auch Vorkommen nach dem Definitartikel.

Gesucht wurde nach Substantivwortformen, i) die unmittelbar nach einer Akkusativ- oder Dativform des indefiniten Artikels *ein-* (*einen/einem*) auftreten, ii) die auf *-n* enden, iii) die einem Lemma zugeordnet sind, das nicht auf *-n* endet und iv) die auf oberster Wortbildungsebene nicht als Kompositum klassifiziert sind.⁶ In einem nächsten Schritt wurde die Liste der zugehörigen Lemmata manuell überprüft. Ausgeschlossen wurden nominalisierte Adjektive und Partizipien (*Kleine*, *Kämpfende*) sowie verbleibende Komposita. Für jedes Lemma wurde manuell überprüft, ob das Korpus tatsächlich mindestens einen validen Beleg für die Verwendung als schwach flektiertes Substantiv enthält. In einem letzten Schritt wurden orthographische Varianten (z. B. *Hagiograph* und *Hagiograf*, *Ko-Referent* und *Koreferent*) zusammengefasst. Die resultierende Liste enthielt

6 Die CQL-Suchanfrage lautet: 1:[word= "einen"|word="einem"] 2:[tag="NN" & toprule="^[^c].*" & lemma!="unknown" & word = "^.+n\$"] & 2.lc=2.lemma_lc.

1.151 Lemmata. Sie wurde anschließend ergänzt um die Substantive *Friede(n)*, *Funke(n)*, *Gefalle(n)*, *Hode(n)* und *Schade(n)*, die zuvor nicht erfasst wurden, da das finale *-n* dort bereits als Teil der Lemmaform hinterlegt war. Die finale Liste enthält damit 1.156 Lemmata.

3.2 Stichproben

Für die Token-basierten Analysen zur Flexion im Nicht-Nominativ (vgl. Abschnitte 4.2, 4.4, 5.1 und 5.3), die den Schwerpunkt dieses Kapitels bilden, wurden Daten aus zwei Korpora erhoben, dem DeReKo-basierten KoGra und dem Foren-Teil des DECOW_{16B}-Webkorpus (Schäfer & Bildhauer 2012), einem Teil desjenigen Korpus, auf dem die Untersuchung Schäfers (2019) basiert. Das Untersuchungskorpus KoGra umfasst zurzeit 7 Mrd. Tokens und wurde mit Dokumentmetadaten wie ‚Land‘, ‚Region‘, ‚Register‘, ‚Medium‘, ‚(inhaltliche) Domäne‘ und ‚Entstehungszeit‘ versehen und stratifiziert (vgl. Bubenhofer et al. 2014: 54–76)⁷. Das Subkorpus Forum des DECOW_{16B} umfasst 4,2 Mrd. Tokens. Das DeReKo bzw. KoGra enthält zum Großteil Presstexte. Diese entstehen typischerweise in Situationen mit vergleichsweise hohem Normdruck und werden in der Regel sorgfältig redigiert. Viele Forentexte des DECOW_{16B}-Webkorpus hingegen entstehen in nächsprachlichen Situationen, unter niedrigerem Normdruck, eher spontan und ohne redaktionelles Korrektiv. Der Gegensatz zwischen DeReKo bzw. KoGra und den Webforen erlaubt uns somit, mögliche Unterschiede zwischen informelleren und formelleren Textsorten zu untersuchen.

Für jedes Lemma unserer Liste (vgl. Abschnitt 3.1) wurde nach Belegen gesucht, in denen das Substantiv nach dem definiten oder indefiniten Artikel auftritt, optional mit einem Adjektiv dazwischen. Dabei wurden alle drei obliquen Kasus Akkusativ, Dativ und Genitiv berücksichtigt (d. h. die Artikelformen *den/dem/des*, *einen/einem/eines* sowie Verschmelzungen aus Präposition und Definitartikel wie z. B. *am*, *zum* und *beim*). Für jedes Substantiv wurde nach starken und schwachen Formen gesucht. Starke Akkusativ- und Dativformen wurden operationalisiert als Formen, die mit der Lemmaform übereinstimmen. Starke Genitivformen wurden operationalisiert als Lemmaformen plus *-(e)s*. Schwache Formen wurden operationalisiert als Lemmaformen plus *-(e)n*. Zudem wurden die kausaugmentierenden Genitivformen auf *-(e)ns* erfasst, operationalisiert als Lemmaformen plus *-(e)ns*. Eine ausführlichere Darstellung der Suchanfragen findet sich im Anhang (Abschnitt 7.1).

7 Für genauere Informationen zu KoGra vgl. <https://grammis.ids-mannheim.de/korpusgrammatik/6615> (abgerufen am 15.12.2023).

Aus den Millionen von Treffern wurden mehrere Stichproben gezogen. Zum einen einfache, nicht weiter stratifizierte Stichproben zur Erstellung deskriptiver Statistiken (Abschnitte 4.2 und 5.1). Diese geben u. a. Auskunft darüber, wie hoch die Anteile starker und schwacher Formen im tatsächlichen Sprachgebrauch in den beiden Korpora ausfallen. Wie bereits Schäfer (2019: 401) zeigt und wie sich auch hier bestätigen wird, sind Norm-divergente Formen (starke Formen schwacher Maskulina und schwache Formen starker Maskulina) insgesamt selten. Aus diesem Grund folgen wir Schäfer darin, für die inferenzstatistische Analyse der Verteilung der beiden Varianten (Abschnitte 4.4 und 5.3) mit stratifizierten Stichproben zu arbeiten, in denen wir den Anteil der jeweils selteneren Variante künstlich erhöhen (vgl. Schäfer 2019: 403–404). Damit erlauben diese Stichproben zwar keine Aussage mehr darüber, in welchem quantitativen Verhältnis starke und schwache Formen in der Sprachpraxis auftreten (dafür die separaten, nicht-stratifizierten Stichproben). Sie ermöglichen aber eine Bestimmung der relativen Stärke und der Richtung des Einflusses der untersuchten (potentiell) variationssteuernden Faktoren (zu diesen siehe Abschnitte 4.3 und 5.2). Für Näheres zur inferenzstatistischen Methodik siehe Abschnitte 4.4 und 5.3 sowie im Anhang (Abschnitt 7.3).

4 Schwache Maskulina

Der Begriff „schwache Maskulina“ wird traditionell für diejenigen Maskulina verwendet, die **historisch** schwach flektieren. Im vorliegenden Grammatikkapitel können wir jedoch keine genuin sprachgeschichtliche Perspektive einnehmen. Stattdessen verwenden wir den Begriff hier für diejenigen Maskulina, für die das schwache Kasusallomorph *-en* **kodifiziert** ist. Dabei orientieren wir uns am entsprechenden Eintrag bei Duden Online.⁸ Konkret werten wir als schwache Maskulina all diejenigen, für die eine nicht-nominativische Singularform mit dem Suffix *-(e)n* angegeben wird. Wie sich zeigt, lassen sich dieser Definition gemäß die allermeisten Substantive unserer Lemmaliste als schwache Maskulina klassifizieren (nämlich 1.042 von 1.156, entspricht 90,1 %, z. B. *Aktivist*, *Kollege*). Mit anderen Worten: Die große Mehrheit der schwach belegbaren Substantive sind auch als schwach flektierende Substantive kodifiziert. Für einen kleinen Teil davon (31 von 1.042, entspricht 3 %) wird neben dem schwachen auch das starke Muster angegeben. Zum Teil wird dabei angegeben, dass die Wahl von der Verwendung/Bedeutung abhängt. So wird z. B. für *Typ* in der Bedeutung ‚Art, Kategorie‘ nur das starke Muster angegeben, für *Typ* in der Bedeutung ‚(junge) männliche Person‘ hingegen

8 <https://www.duden.de/woerterbuch>. Bei Präkonfixbildungen (z. B. *Urgedanke*), die keinen eigenen Eintrag haben, wurde auf den Eintrag für das entsprechende präkonfixlose Lexem (*Gedanke*) zurückgegriffen.

auch das schwache. Zum Teil wird regionale Variation berücksichtigt. So z.B. bei *Hirsch*, wo das schwache Muster als süddeutsch und österreichisch bezeichnet wird. 17 Substantive (1,5%) mit finalelem Schwa sind als Maskulina kodifiziert, die die schwache Endung *-n* auch im Nominativ aufweisen (können) (z.B. *Wille*, *Glaube*), einhergehend mit Genitivformen auf *-ns*. Für ein Substantiv (*Buchstabe*) wird genitivisches *-ns* angegeben, ohne dass auch eine Nominativform auf *-n* angegeben wird.

4.1 Nominativ Singular

Bevor wir uns dem Nicht-Nominativ zuwenden, untersuchen wir die Substantive hinsichtlich ihrer Form im Nominativ Singular. Dabei geht es um die Frage, inwiefern sich Formen auf *-(e)n* möglicherweise auch bei Substantiven finden, für die dies nicht kodifiziert ist. Wie in Abschnitt 2.2 beschrieben, würde *-(e)n* im Nominativ dafür sprechen, dass das Substantiv eine Stammvariante auf *-(e)n* kennt. Damit käme der Variation zwischen *-(e)n*-haltigen und *-(e)n*-losen Formen im Akkusativ und Dativ ein anderer Status zu als bei den übrigen Substantiven.

Um entsprechende Formen zu identifizieren, wurde in KoGra nach der *(e)n*-Form des jeweiligen Substantivs unmittelbar nach der Nominativ-Singular-Form des indefiniten Artikels (*ein*) gesucht.⁹ Wie sich zeigt, ergibt diese Suche für 72 der 1.042 Substantive mit kodifizierter schwacher Flexion mindestens einen – möglicherweise validen – Treffer. Darunter sind 14 der 17 Substantive, für die eine solche Nominativform kodifiziert ist; es fehlen lediglich die insgesamt selten belegten Präkonfixbildungen *Elektrofunke*, *Urgedanke* und *Urglaube*. Es bleiben 58 Substantive, für die sich entsprechende Formen finden lassen, ohne dass diese kodifiziert sind. In vielen Fällen liegt jedoch der Verdacht nahe, dass es sich um Tippfehler handelt (z.B. *ein Menschen*, *ein Diplomaten*, *ein Athleten*). Berücksichtigt man nur diejenigen Lemmata, für die man mindestens fünf Treffer erhält, dann bleiben noch 23 Lemmata: 13, für die die Form kodifiziert ist (*Drache*, *Friede*, *Funke*, *Gedanke*, *Gefalle*, *Glaube*, *Haufe*, *Hode*, *Name*, *Same*, *Schade*, *Sparre* und *Wille*), und 10, für die dies nicht gilt (*Bauer*, *Experte*, *Herr*, *Junge*, *Kollege*, *Kunde*, *Mensch*, *Nachbar*, *Polizist* und *Soldat*). Absolute Trefferzahlen sind jedoch mit Vorsicht zu genießen: So ergibt z.B. die Suche nach der Sequenz *ein Menschen* zwar immerhin 131 Treffer (vor Ausschluss von Fehltreffern wie z.B. *ein Menschen fressender*, *einäugiger Riese*). Dem stehen insgesamt aber fast 60.000 Belege für *ein Mensch* gegenüber. Aus diesem Grund wurde für jedes der 23 Lemmata mit mindestens fünf Belegen für nominativisches *-(e)n* eine Zufallsstichprobe von 200 Belegen erhoben (Suche nach Belegen mit oder ohne *-(e)n* nach *ein*), um den Status der *(e)n*-Form im

9 Mit dem definiten Artikel ergibt sich das Problem, dass dieser ambig zwischen Nominativ Singular und Genitiv Plural ist (z.B. *der Funken*).

Verhältnis zur *(e)n*-losen Form zu bestimmen. Wo die Gesamtrefferzahl unter 200 lag, wurden alle Belege ausgewertet. Bei *Drache* wurde je eine Stichprobe für die Lesart ‚Fluggerät‘ und die Lesart ‚Fabelwesen‘ erhoben.

Die Ergebnisse für die Substantive, für die nominativisches *-n* (zumindest als Variante) kodifiziert ist, sind in Tabelle 1 dargestellt. Die Stichproben legen nahe, dass die *-n*-Form bei *Gefalle*, *Schade*, *Haufe* und *Hode* fast ganz fest ist bzw. zumindest klar dominiert. Auch die wenigen Belege für *Sparre* zeigen *-n*. Bei *Drache* (Fluggerät), *Same*, *Funke*, *Friede*, *Drache* (Fabelwesen) und *Glaube* herrscht (mehr) Variation, wobei nur bei *Drache* (Fluggerät) und *Same* die *-n*-Form überwiegt. *Drache* neigt also in der Fluggerät-Lesart stärker zur *n*-Form, in der Fabelwesen-Lesart stärker zur *n*-losen Form. Dies passt dazu, dass die *n*-Ausweitung insgesamt fast ausschließlich Lexeme mit unbelebtem Denotat betrifft (vgl. sämtliche Lexeme in Tabelle 1 jenseits von *Drache*). Bei *Gedanke*, *Name* und *Wille* schließlich kommt die *-n*-Form fast gar nicht vor. Diese Befunde decken sich weitestgehend mit den Angaben bei Thieroff & Vogel (2009: 45–46) – abgesehen davon, dass bei *Hode* nicht beide Formen „gleichermaßen“ (Thieroff & Vogel 2009: 45) vorkommen, sondern die *-n*-Form doch sehr deutlich überwiegt. Bei den 10 Substantiven, für die die *-(e)n*-Form nicht kodifiziert ist, lässt diese sich in den 200er-Stichproben nicht nachweisen. Dies deutet darauf hin, dass sie trotz absoluter Beleganzahlen ≥ 5 im Gesamtkorpus im Verhältnis zur endungslosen Form kaum eine Rolle spielt.

Tabelle 1: Anteile von Nominativformen auf *-n* (KoGra).

Lexem	n	% <i>-n</i>
<i>Gefalle</i>	83	100
<i>Schade</i>	200	100
<i>(Sparre)</i>	(5)	(100)
<i>Haufe</i>	200	99,5
<i>Hode</i>	62	96,8
<i>Drache</i> (Fluggerät)	200	77,5
<i>Same</i>	97	67,8
<i>Funke</i>	200	42,0
<i>Friede</i>	200	23,0
<i>Drache</i> (Fabelwesen)	200	15,5
<i>Glaube</i>	200	8,5
<i>Gedanke</i>	200	0,5
<i>Name</i>	200	0,5
<i>Wille</i>	200	0,0

Zusammengefasst: Die Ergebnisse deuten darauf hin, dass Nominativ-Singular-Formen (und das heißt: Stammvarianten) auf *-(e)n* zumindest in KoGra im Wesentlichen wohl nur bei denjenigen Substantiven eine Rolle spielen, für die eine solche Form auch kodifiziert ist – bei Substantiven also, die auf Schwa enden (sodass nie das Allomorph *-en*, sondern immer *-n* auftritt) und – mit Ausnahme von *Drache* (Fabelwesen) – ein unbelebtes Denotat haben.

4.2 Anteile schwacher und starker Formen im Nicht-Nominativ

Im nächsten Schritt ermitteln wir die Anteile der Fälle, in denen die schwachen Maskulina tatsächlich schwach flektiert werden, also im Akkusativ, Dativ und Genitiv mit dem Suffix *(e)n* versehen werden, im Verhältnis zu den Fällen, in denen sie – davon abweichend – in starken Formen auftreten, d. h. im Akkusativ und Dativ endungslos bzw. im Genitiv mit dem Suffix *-(e)s*. Darüber hinaus berücksichtigen wir im Genitiv auch die kasusaugmentierende Variante auf *-(e)ns*. Ausgeschlossen werden die im vorangehenden Abschnitt diskutierten Substantive mit kodifizierter Nominativform (und damit Stammform) auf *-n*, da bei ihnen häufig nicht zu entscheiden ist, ob ein Akkusativ/Dativ auf *-n* im konkreten Fall als schwach flektierte Form zu einem *-n*-losen Stamm aufzufassen ist, oder ob das finale *-n* als Teil des Stamms anzusehen ist. Analog dazu sind Genitive auf *-ns* bei den fraglichen Substantiven ambig zwischen kasusaugmentierenden Formen, die das schwache Suffix *-n* mit dem starken Suffix *-s* kombinieren, und solchen, die lediglich das starke *-s* enthalten.¹⁰ Grundlage bilden zwei einfache, nicht-stratifizierte Zufallsstichproben von je 10.000 Belegen (vor dem manuellen Ausschluss von Fehlbelegen) aus KoGra und dem Forenteil des DECOW-Webkorpus. Auf diese Weise sollen Unterschiede zwischen eher formellen/distanzsprachlichen Texten einerseits und eher informellen/nähesprachlichen Texten andererseits erfasst werden.

Die KoGra-Stichprobe enthält nach manuellem Ausschluss von Fehlbelegen noch 7.243 Belege, verteilt auf 440 Lexeme. Das frequenteste Lexem ist *Präsident*; mit 972 Belegen macht es 13,4 % aller Belege aus. Die große Mehrheit von 96,4 % der Belege (6.979) und 97,3 % der Lexeme (428) entfällt auf solche, für die ausschließlich die schwache Flexion (also kein Nebeneinander starker und schwacher Formen) kodifiziert ist.

98,6 % der Belege (7.143) weisen das schwache Suffix *-(e)n* auf. Konzentriert man sich auf diejenigen Lexeme, für die ausschließlich die schwache Flexion kodifiziert ist, erhöht sich der Anteil noch einmal auf 99,1 %. Im Akkusativ tritt

¹⁰ Ebenfalls ausgeschlossen wird der Sonderfall *Buchstabe* (genitivisches *-ns* ohne Nominativform auf *-n*).

in 1,3 % der Belege (31 von 2.347) die endungslose Form auf, verteilt auf 6,8 % der akkusativisch belegten Lexeme (20 von 296). Im Dativ tritt die endungslose Form in 2,6 % der Belege auf (64 von 2.468), verteilt auf 11,3 % der dativisch belegten Lexeme (35 von 309). Im Genitiv tritt die starke Form auf *-(e)s* in nur 0,2 % der Belege auf (4 von 2.428), die kasusaugmentierende Form auf *-(e)ns* in nur einem Fall (entspricht unter 0,1 %). Die Belege für *-(e)s* verteilen sich auf 1,1 % der genitivisch belegten Lexeme (3 von 275), und zwar mit *Hirsch*, *Matador* und *Nachbar* ausschließlich auf solche, für die ein Nebeneinander von schwacher und starker Flexion kodifiziert ist. Die Form auf *-(e)ns* tritt bei *Junge* auf.

Insgesamt deuten die Ergebnisse somit darauf hin, dass starke Formen im überwiegend presssprachlichen KoGra kaum vorkommen. Je nachdem, ob man diejenigen Substantive, für die neben der schwachen auch die starke Flexion kodifiziert ist, einschließt oder ausschließt, kommen Anteile von gerade einmal 1,4 % oder 0,9 % zustande. Diese Werte liegen unter dem, den Schäfer (2019: 401) für das DECOW-Webkorpus ermittelt (2,2 %). Zudem stehen die Befunde im Einklang mit der Beobachtung, dass die Schwankung von schwach zu stark vor allem den Akkusativ und Dativ betrifft (hierzu mehr in Abschnitt 4.3.2). Tatsächlich legen sie sogar nahe, dass zumindest bei den Substantiven, für die ausschließlich die schwache Flexion kodifiziert ist, die Schwankung im Genitiv sehr stark restringiert ist. In der Stichprobe ist die starke Form auf *-(e)s* unter 2.381 Genitivvorkommen dieser Substantive (verteilt auf 266 Lexeme) kein einziges Mal belegt.

Die Webforen-Stichprobe enthält nach manuellem Ausschluss von Fehlbelegen noch 7.720 Belege, verteilt auf 406 Lexeme. Anders als in KoGra ist das häufigste Lexem nicht *Präsident*, sondern *Mensch*; mit 1.521 Belegen macht es 19,7 % aller Belege aus. Wie in KoGra entfallen mit 90,9 % (7.017) die allermeisten Belege und mit 95,3 % (387) die allermeisten Lexeme auf solche, für die ausschließlich die schwache Flexion kodifiziert ist.

Wie in KoGra weist auch in den Webforen mit 94,8 % (7.321) die große Mehrheit der Belege das schwache Suffix *-(e)n* auf. Konzentriert man sich auf diejenigen Lexeme, für die ausschließlich die schwache Flexion kodifiziert ist, erhöht sich der Anteil auf 96,7 % (6.786 von 7.017). Im Akkusativ tritt in 6,1 % der Belege (179 von 2.954) die endungslose Form auf, verteilt auf 18,5 % der akkusativisch belegten Lexeme (53 von 287). Im Dativ tritt die endungslose Form in 6,6 % der Fälle auf (209 von 3.190), verteilt auf 25,0 % der dativisch belegten Lexeme (71 von 284). Im Genitiv tritt die starke Form *-(e)s* in nur 0,6 % der Belege auf (9 von 1.576), die kasusaugmentierende Form *-(e)ns* in nur 0,1 % der Belege (2 von 1.576). Die Belege für *-(e)s* verteilen sich auf 4,2 % der genitivisch belegten Lexeme (5 von 212), und zwar mit *Bauer*, *Elektromagnet*, *Hirsch*, *Papagei* und *Tempomat* wie in KoGra ausschließlich auf solche, für die ein Nebeneinander von schwacher und starker Flexion kodifiziert ist. Die Belege für *-(e)ns* entfallen wie in KoGra auf das Lexem *Junge*.

In den Webforen ist der Anteil starker Formen also etwas höher als in KoGra. Dennoch überwiegt die schwache Form immer noch sehr deutlich. Je nachdem, ob man diejenigen Substantive, für die neben der schwachen auch die starke Flexion kodifiziert ist, einschließt oder ausschließt, ergeben sich für die starke Form Anteile von 5,3 % oder 3,4 %. Diese Werte liegen zugleich auch über jenem, den Schäfer für das DECOW-Webkorpus insgesamt ermittelt (2,2 %, s. o.). Wie in KoGra scheinen auch in den Foren starke Formen in hohem Maße auf endungslose Akkusativ- und Dativformen beschränkt zu sein. Die starke Genitivform auf *-(e)s* lässt sich in der Stichprobe wieder deutlich seltener und nur für solche Substantive belegen, für die das starke Muster neben dem schwachen Muster kodifiziert ist. Für die Substantive, für die nur die schwache Form kodifiziert ist, lässt sie sich in der Stichprobe unter 1.503 Genitivbelegen (verteilt auf 199 Lexeme) kein einziges Mal nachweisen.

4.3 Variationssteuernde Faktoren

Im Folgenden gehen wir einigen Faktoren nach, die im Akkusativ, Dativ und Genitiv die Wahl zwischen den kanonischen schwachen Formen auf *-(e)n* und den selteneren, starken Formen (endungslosen Akkusativen und Dativen sowie Genitiven auf *-(e)s*) steuern. Neben den Faktoren, die auch von Schäfer (2019) untersucht worden sind (Phonotaktik, Semantik, Kasus und Gebrauchsfrequenz des Substantivs), berücksichtigen wir hier zwei weitere grammatische Faktoren (das (Nicht-)Vorkommen eines attributiven Adjektivs vor dem Substantiv und das (Nicht-)Vorkommen einer engen Apposition nach dem Substantiv) sowie den Gegensatz zwischen dem überwiegend formellen und distanzsprachlichen KoGra und den meist informelleren und nächsprachlichen DECOW-Webforen. In den Abschnitten 4.3.1 bis 4.3.6 stellen wir die Einflussfaktoren und die mit ihnen verbundenen Hypothesen zunächst genauer vor. In Abschnitt 4.4 modellieren wir ihren Einfluss mit einem gemischten logistischen Regressionsmodell.

4.3.1 Prototypikalität: Phonotaktik und Semantik

Köpcke (1995: 168–176) demonstriert, dass sich die schwachen Maskulina anhand phonotaktischer und semantischer Merkmale in typische und untypische Vertreter einteilen lassen. Darauf aufbauend zeigt Schäfer (2019) in seiner Korpusstudie, dass die prototypischen Vertreter am stärksten am schwachen Muster festhalten bzw. – komplementär formuliert – Abweichungen hin zum starken Muster am ehesten bei den untypischen Vertretern zu beobachten sind.

Unter den schwachen Maskulina unserer Lemmaliste (1.025 nach Ausschluss jener, für die eine Nominativform auf *-n* kodifiziert ist) dominieren in

phonotaktischer Hinsicht (vgl. Tabelle 2) Mehrsilber mit Endbetonung (Schäfers Kategorie „polyult“, z. B. *Architekt, Idiot, Mandant*) und Mehrsilber mit finalem Schwa (Schäfers Kategorie „polyschwa“, z. B. *Matrose, Erbe, Immunologe*).¹¹ Die beiden anderen der von Schäfer (2019) unterschiedenen Kategorien, d. h. Mehrsilber mit Nicht-Endbetonung („polynult“, zu Beispielen s. u.) und Einsilber („mono“, z. B. *Mensch, Herr*), sind also vergleichsweise selten. Zu den Mehrsilbern mit Nicht-Endbetonung zählen wir auch die Bildungen mit betontem Präkonfix (z. B. *Exrekrut, Neofaschist, Urahn*), die ohne dieses Präkonfix in den meisten Fällen den Mehrsilbern mit Endbetonung zuzuordnen wären (*Rekrut, Faschist*). Schließt man diese von den Mehrsilbern mit Nicht-Endbetonung aus, bleiben nur noch fünf Substantive dieser Kategorie (*Bauer, Bayer, Nachbar, Pommer, Ungar*; dies entspricht 0,5 % der schwachen Maskulina). Zwei Substantive (*Augur, Cäsar*) schließlich weisen ein morphologisch gesteuertes Betonungsmuster auf: In unflektierter Form tragen sie den Hauptakzent auf der ersten Silbe (*Augur, Cäsar*), in flektierter Form auf der letzten Silbe des Stammes (*Auguren, Cäsaren*, vgl. Tabelle 2). Eine größere Rolle spielt dieser Typus bei den als stark kodifizierten Substantiven (s. u.), denn er umfasst auch die große Gruppe der *or*-Substantive. In semantischer Hinsicht (Tabelle 3) handelt es sich bei über 90 % der Substantive um solche mit menschlichem Denotat (Schäfers Kategorie „hum“). Die übrigen verteilen sich in etwa gleichmäßig auf solche mit belebtem aber nicht menschlichem Denotat (Schäfers Kategorie „ani“) und solche mit unbelebtem Denotat (Schäfers Kategorie „ina“)

Im Vergleich mit einer Zufallsstichprobe von 1.000 maskulinen Substantiven, die **nicht** in unserer Liste schwach belegter Substantive vertreten sind (hierzu Anhang 7.2), d. h. sich in der von uns überprüften Umgebung (nach *einen/einem*) **nicht** schwach flektiert belegen lassen, sind in phonotaktischer Hinsicht die Mehrsilber mit finalem Schwa am stärksten überrepräsentiert. Zwar bilden diese unter den schwachen Maskulina lediglich die zweitgrößte Gruppe (vgl. Tabelle 2), dafür scheinen sie unter den Maskulina jenseits der schwachen Flexionsklasse fast gar nicht vorzukommen. So enthält die Vergleichsstichprobe nur zwei Lexeme (0,2 %) dieses Typs: *Käse* und das selten gebrauchte *Cicerone*. Damit ist ihr Anteil an den schwachen Maskulina 169-mal so hoch wie ihr Anteil an der Vergleichsgruppe. Am zweitstärksten überrepräsentiert sind die Mehrsilber mit Endbetonung. Deren Anteil ist unter den schwachen Maskulina 5,7-mal so hoch wie unter den Substantiven der Vergleichsgruppe (Beispiele aus der Vergleichsgruppe: *Aspekt, Genuss, Vampir*). Einsilber und Mehrsilber mit Nicht-Endbetonung hingegen sind unter

11 Anders als Köpcke (1995: 166, Fn. 9) scheint Schäfer das finale Schwa bei der Silbenanzahl mitzuzählen. Damit gehören auch Substantive wie *Kurde*, die erst durch das Schwa zu Mehrsilbern werden, zur Kategorie „polyschwa“. Zur Vergleichbarkeit unserer Ergebnisse mit denen Schäfers verfahren wir hier ebenfalls so.

Tabelle 2: Schwache Maskulina (Lexeme) nach Phonotaktik.

Phonotaktik	n	%
Endbetonung (<i>Architekt</i>)	600	58,5
finale Schwa (<i>Matrose</i>)	346	33,8
Nicht-Endbetonung (<i>Bauer, Ex-Rekrut</i>)	55	5,4
Einsilbigkeit (<i>Mensch</i>)	22	2,1
morphologisch gesteuerte Betonung (<i>Augur</i>)	2	0,2
gesamt	1.025	100

Tabelle 3: Schwache Maskulina (Lexeme) nach Semantik.

Semantik	n	%
menschliches Denotat (<i>Mensch</i>)	940	91,7
unbelebtes Denotat (<i>Automat</i>)	40	3,9
belebtes aber nicht menschliches Denotat (<i>Löwe</i>)	36	3,5
als polysem gewertet (<i>Parasit</i>)	9	0,9
gesamt	1.025	100

den schwachen Maskulina unterrepräsentiert. Der Anteil Ersterer ist in der Vergleichsgruppe 9-mal so hoch wie unter den schwachen Maskulina (Beispiele aus der Vergleichsgruppe: *Hengst, Pakt, Putsch*). Der Anteil Letzterer ist in der Vergleichsgruppe 12,6-mal so hoch wie unter den schwachen Maskulina (Beispiele aus der Vergleichsgruppe: *Ärmel, Motzer, Status*).¹²

In semantischer Hinsicht sind unter den schwachen Maskulina im Vergleich mit der Vergleichsgruppe die Substantive mit menschlichem Denotat überrepräsentiert; ihr Anteil fällt dort 2,6-mal so hoch aus wie unter den Substantiven der Vergleichsgruppe (Beispiele aus der Vergleichsgruppe: *Rüpel, Choleriker, Frisör*). Substantive mit belebtem aber nicht menschlichem Denotat und vor allem Substantive mit unbelebtem Denotat sind unter den schwachen Maskulina hingegen unterrepräsentiert. Der Anteil Ersterer ist in der Vergleichsgruppe 1,4-mal so hoch wie unter den schwachen Maskulina (Beispiele aus der Vergleichsgruppe: *Eber, Ganter, Krebs*). Der Anteil Letzterer ist in der Vergleichsgruppe 15,4-mal so

12 Schließt man unter den schwachen Maskulina die Präkonfixbildungen von den Mehrsilbern mit Nicht-Endbetonung aus (und zählt somit nur die fünf Lexeme *Bauer, Bayer, Nachbar, Pommer* und *Ungar*), ist der Anteil dieser Kategorie an der Vergleichsgruppe sogar 136,4-mal so hoch wie ihr Anteil an der Vergleichsgruppe.

hoch wie unter den schwachen Maskulina (Beispiele aus der Vergleichsgruppe: *Kanton, Lack, Kolonialismus*).

Der Vergleich mit der Vergleichsgruppe stützt damit Köpckes (1995: 170) und Schäfers (2019: 395–396) Annahme, dass der Phonetaktik insgesamt ein höherer Signalwert für die Zugehörigkeit zur Klasse der schwachen Maskulina zukommt als der Semantik: Das mit Abstand am stärksten überrepräsentierte Merkmal ist das phonotaktische Merkmal ‚finales Schwa‘, gefolgt vom ebenfalls phonotaktischen Merkmal ‚Endbetonung‘. Erst dann folgt das semantische Merkmal ‚menschliches Denotat‘.

Schäfer (2019: 405) zeigt in Bezug auf den Faktor Phonetaktik, dass die Substantive mit finalem Schwa am stärksten an der schwachen Flexion festhalten, gefolgt von den Mehrsilbern mit Endbetonung, den Einsilbern und schließlich den Mehrsilbern mit Nicht-Endbetonung. Diese Rangfolge entspricht damit genau der Rangfolge, in der die entsprechenden Kategorien unter den schwachen Maskulina unserer Lemmaliste im Vergleich mit unserer Vergleichsstichprobe über- bzw. unterrepräsentiert sind. In Bezug auf den Faktor Semantik zeigt Schäfer, dass Substantive mit menschlichem Denotat stärker zur schwachen Flexion neigen als Substantive mit belebtem nicht-menschlichem Denotat und Substantive mit unbelebtem Denotat. Auch in semantischer Hinsicht also hält die im Vergleich mit der Vergleichsstichprobe überrepräsentierte Kategorie (menschliches Denotat) stärker an der schwachen Flexion fest als die unterrepräsentierten (nicht-menschliche Denotate). Der Vergleich der schwachen Maskulina unserer Lemmaliste mit der Vergleichsstichprobe stützt damit Schäfers Interpretation seiner Ergebnisse, nämlich, dass Substantive umso eher an der schwachen Flexion festhalten, je prototypischer sie für die Klasse der schwachen Maskulina sind.

In unserer eigenen Variationsanalyse erwarten wir ähnliche Ergebnisse wie bei Schäfer: In phonotaktischer Hinsicht sollten also am stärksten Substantive mit finalem Schwa am schwachen Muster festhalten, gefolgt von Mehrsilbern mit Endbetonung, Einsilbern und schließlich Mehrsilbern mit Nicht-Endbetonung; in semantischer Hinsicht sollten am stärksten Substantive mit menschlichem Denotat am schwachen Muster festhalten.

4.3.2 Kasus

Wie bereits erwähnt, betrifft der bestehenden Literatur zufolge der Wechsel von schwach zu stark den Akkusativ und den Dativ in stärkerem Maße als den Genitiv (Thieroff 2003: 109, 114; Dudenwörterbuch 2021: 907, Schäfer 2019: 405–407). Die Substantive neigen also eher dazu, im Akkusativ und Dativ endungslos aufzutreten, als im Genitiv *-(e)n* durch *-(e)s* zu ersetzen. Die in Abschnitt 4.2 vorgestellten deskriptiven Statistiken stehen im Einklang mit dieser Beobachtung. Tatsächlich deuten sie sogar darauf hin, dass die Schwankung von schwach zu stark im Genitiv

ein sehr stark restringiertes Phänomen darstellt. In der Variationsanalyse erwarten wir daher, dass die Wahrscheinlichkeit für die schwache Flexion im Genitiv höher ist als im Akkusativ und Dativ (ähnlich bei Schäfer 2019: 405, 407).

Thieroff (2003: 14) erklärt den Gegensatz zwischen Genitiv einerseits und Akkusativ und Dativ andererseits strukturalistisch, nämlich damit, dass bei der analogischen Angleichung der schwachen Maskulina an das dominante, starke Paradigma nicht nur die konkreten Formen des starken Paradigmas eine Rolle spielen, sondern auch dessen abstrakte Struktur, nämlich der Gegensatz zwischen einer markierten Genitivform und einer unmarkierten Nicht-Genitiv-Form (*des Tag(e)s* vs. *der/dem/den Tag_*). Damit fällt der analogische Druck im Akkusativ und Dativ stärker aus als im Genitiv. Denn im Akkusativ und Dativ ist der Unterschied zwischen den beiden Mustern vergleichsweise groß: Er besteht im Vorliegen vs. Fehlen einer Markierung (*dem/den Bären* vs. *dem/den Tag_*). Im Genitiv hingegen ist der Unterschied vergleichsweise klein: Er besteht lediglich in der konkreten Form der Markierung (*des Bären* vs. *des Tag(e)s*). In Bezug auf das abstrakte Distinktionssystem ist die Analogie zu den starken Maskulina somit bereits dann hergestellt, wenn lediglich Akkusativ und Dativ angeglichen werden (*des Bären* vs. *der/dem/den Bär*).

Darüber hinaus dürfte der Unterschied an sich eine Rolle spielen, dass der Wechsel von schwach zu stark im Akkusativ und Dativ im **Weglassen** eines Kasussuffixes besteht, im Genitiv dagegen in der **Ersetzung** eines Suffixes durch ein anderes. Denn das Deutsche (bzw. allgemeiner: die germanischen Sprachen, vgl. Harbert 2007: 90) zeigt eine allgemeinere, langfristige Entwicklungstendenz zum Abbau von Kasusmarkierung am Substantiv selbst: „Während die Substantive in den früheren Sprachstufen ihre Kasusmarkierung prototypisch am Wortkörper selber trugen, ist seit dem Althochdeutschen ein Abbauprozess zu beobachten“ (Dal/Eroms 2014: 6).

Zuletzt könnte auch ein soziolinguistischer Grund eine Rolle spielen: Wie z. B. Seiler (2022: 58) feststellt, ist der Genitiv insgesamt „ein stilistischer Marker geworden, der Hoch- oder Schriftsprachlichkeit signalisiert“ (siehe hierzu auch ausführlich Szczepaniak 2014 und Pickl 2020). Wenn nun schon der Gebrauch des Genitivs an sich ein Indikator für eine Orientierung an der schriftlichen Norm sein sollte, dann wäre zu erwarten, dass diese sich auch bei der morphologischen Realisierung dieses Kasus zeigt: Der Genitiv sollte also in stärkerem Maße als der Akkusativ und Dativ zur Norm-konvergenten Form tendieren: Bei den als schwach kodifizierten Substantiven könnte damit auch ein soziolinguistischer Faktor die Bewahrung der schwachen Flexion insbesondere im Genitiv begünstigen.

4.3.3 Apposition

Das Dudenwörterbuch (2021: 907–908) beobachtet, „dass bei einer Reihe von Fremdwörtern [...] die Kasusendung im Dativ und Akkusativ Singular häufig

weggelassen [wird], wenn sie als Rangbezeichnung, Titel oder Berufsbezeichnung und in diesen Fällen oft in Verbindung mit dem Eigennamen gebraucht wird“. Als Beispiele werden u. a. *dem Fabrikant(en) Meier* und *für den Präsident(en) Müller* angegeben. Mit der Dudengrammatik (2016: 998–999, 1002) kann man in solchen Strukturen Konstruktionen mit einer explikativen Apposition sehen. Unter diesen Konstruktionstyp fallen auch Fälle mit anderen Gattungsbezeichnungen als erstem Substantiv wie z. B. bei *dem Planet(en) Erde*. Gemäß Dudengrammatik ist in der fraglichen Konstruktion das erste der beiden Substantive (*Fabrikant, Präsident, Planet*) der – zu flektierende – Hauptkern, das zweite (*Meier, Müller, Erde*) ein „appositiver Nebenkern“. Ein Grund für die Tendenz zur Flexionslosigkeit könnte darin liegen, dass viele Kombinationen von Rangbezeichnung und Eigenname häufig ohne Determinierer auftreten (*mit Fabrikant Meier, für Präsident Müller*). Und in dieser determiniererlosen Konstruktionsvariante gilt, umgekehrt, nicht das erste, sondern das zweite Substantiv als zu flektierender Hauptkern, während das erste Substantiv als (determiniererloser) Nebenkern im Regelfall unflektiert bleibt (vgl. Dudengrammatik 2016: 999). Die Flexionslosigkeit des ersten Substantivs in der determiniererlosen Konstruktionsvariante könnte qua Analogie die Flexionslosigkeit in der determinierhaltigen Variante begünstigen. Wenn diese Erklärung auf dem richtigen Weg ist, dann würde das bedeuten, dass die Tendenz zur Unterlassung der Kasusflexion in Konstruktionen mit Apposition indirekt aus der Unterlassung der Kasusflexion in Kontexten ohne Substantivbegleiter folgt („syntaktisch bedingte Unterlassung der Kasusflexion“, vgl. Abschnitt 2.3).

In unserer Variationsanalyse klassifizieren wir sämtliche Belege in Bezug darauf, ob das schwach flektierbare Substantiv zusammen mit einem unmittelbar nachfolgenden appositiven Nebenkern auftritt. Es steht zu vermuten, dass das Vorkommen eines solchen Nebenkerns im Akkusativ und Dativ die „starken“ (endungslosen) Formen begünstigt.

4.3.4 Attributives Adjektiv

Es ist denkbar, dass auch das (Nicht-)Vorkommen eines attributiven Adjektivs vor dem Substantiv einen Einfluss auf das (Nicht-)Auftreten des schwachen *-(e)n*-Suffixes hat. Einen entsprechenden Befund macht Rohdenburg (1988, 1993) für nordniederdeutsche Dialekte (siehe auch Berg 2013: 244–266). In diesen Dialekten sind die *-(e)n*-Formen schwacher Maskulina in den Funktionsbereich des Nominativs eingedrungen; sie treten bisweilen auch in prädikativer Funktion auf, und zwar insbesondere in NPs mit indefinitem Artikel und attributivem Adjektiv (vgl. *de [...] is ja ok'n halven Buurn*, ‚der ist ja auch ein halber Bauer‘, unsere Hervorhebung). Rohdenburg vermutet, dass es sich dabei um eine Angleichung an die *-(e)n*-Form des Adjektivs handelt. Berg (2013: 226–227) bezeichnet dieses Phänomen als

„Prinzip der phonologischen Korrespondenz“. ¹³ Es stellt sich die Frage, ob sich ein ähnlicher Effekt auch in unseren standardsprachlichen bzw. standardnahen Daten finden lässt. Denkbar wäre, dass die Präsenz eines vorangehenden attributiven Adjektivs auf *-(e)n* (vgl. z. B. *von einem ehrlichen Menschen*) die Wahrscheinlichkeit für die Weglassung der Substantivendung *-(e)n* im Akkusativ und Dativ des nachfolgenden schwachen Maskulinums verringert.

4.3.5 Frequenz des Substantivs

Schäfers (2019: 406) Untersuchung ergibt, dass auch die Frequenz des Substantivs einen Einfluss auf das Flexionsverhalten hat. So neigen bei den schwachen Maskulina die frequenteren Lemmata weniger stark zur starken Flexion. Das wird damit erklärt, dass bei diesen Substantiven das starke Muster das neuere ist, und hochfrequente Formen resistenter gegenüber analogischem Wandel sind, da sie stärker verfestigt (*entrenched*) sind. Wir erwarten hier denselben Effekt. Um die Auswirkung der Frequenz des Substantivs auf die Flexion zu untersuchen, wurden die Häufigkeitsklassen der Lemmata im verwendeten DeReKo-Teilkorpus und für den Webforenteil des DECOW berechnet und annotiert (vgl. Perkuhn et al. 2012: 80). Bei der Berechnung der Häufigkeitsklasse wurde die absolute Frequenz eines Lemmas in Bezug gesetzt zur absoluten Frequenz des häufigsten Lemmas (der bestimmte Artikel *der/die/das*), so dass ein Lemma *x* die Häufigkeitsklasse *k* hat, wenn das häufigste Lemma etwa 2^k -mal häufiger vorkommt als *x*. Das bedeutet: Je niedriger die Häufigkeitsklasse, desto häufiger sind die darin enthaltenen Wörter.¹⁴ Um den Einfluss der Häufigkeitsklasse auf beide Teilkorpora in einem statistischen Modell anwenden zu können, wurde der Mittelwert der Häufigkeitsklassen für jedes Lemma für beide Teilkorpora ermittelt.

4.3.6 Korpus: KoGra vs. Webforen

Wertet man – wie wir es hier gemacht haben – den Eintrag bei Duden Online als Hinweis auf die kodifizierte Norm, dann handelt es sich bei den starken Formen in den allermeisten Fällen um Norm-divergente Formen. (Ausnahmen bilden die wenigen Substantive, für die bei Duden Online schwache und starke Formen nebeneinander angegeben werden.) Vor diesem Hintergrund ist davon auszugehen,

13 In ihrem traditionellen, nicht-nominativischen Funktionsbereich sei die *-(e)n*-Form in diesen Dialekten insgesamt rückläufig, wobei ein vorangehendes Adjektiv sie aber noch stütze. So trete in einer Mundart die *n*-Form überhaupt nur noch dann auf, wenn dem Substantiv ein Adjektiv vorangeht (z. B. *[bi] eenen Buur* ‚bei einem Bauern‘ vs. *von eenen olen Buurn* ‚von einem alten Bauern‘).

14 Zur Anwendung und Berechnung von Häufigkeitsklassen vgl. ausführlich Perkuhn et al. (2012: 80–83).

dass sie eher dort auftreten, wo ein vergleichsweise geringer Normdruck herrscht. Dies gilt zum einen für die gesprochene im Vergleich mit der geschriebenen Sprache (vgl. Dudengrammatik 2016: 1191).¹⁵ Darüber hinaus ist auch innerhalb der Schriftlichkeit mit Unterschieden zwischen Textsorten zu rechnen, die sich hinsichtlich ihres Normierungsgrades unterscheiden. Wir testen dies anhand des Gegensatzes zwischen KoGra und dem Forenteil des DECOW-Webkorpus. Wie bereits in Abschnitt 3.2 beschrieben, besteht KoGra überwiegend aus Preetexten, die typischerweise unter hohem Normdruck stehen und sorgfältig redigiert werden, während viele Forentexte unter geringerem Normdruck und ohne redaktionelles Korrektiv entstehen. Wir erwarten daher, dass die Webforen in stärkerem Maße zu starken Formen tendieren als KoGra. Die in Abschnitt 4.2 präsentierten deskriptiven Statistiken stehen im Einklang mit dieser Annahme.

Wie in Abschnitt 4.3.2 beschrieben, ist der Wechsel im Akkusativ und Dativ, wo er von einer endungshaltigen zu einer endungslosen Form führt, auch vor dem Hintergrund einer allgemeineren Entwicklungstendenz des Deutschen zu sehen, Kasusmarkierungen am Substantiv abzubauen. In vielen Nicht-Standard-Varietäten des Deutschen ist diese Entwicklung weiter fortgeschritten als in der Standardsprache. So z. B. im Pfälzischen, wo Substantive (nach Post 1992: 121) keinerlei Kasusendungen mehr kennen. Wenn die Foren in stärkerem Maße als KoGra zu endungslosen Formen neigen, dann würde das dafür sprechen, dass diese allgemeinere Entwicklungstendenz des Deutschen auch in diesen informelleren Texten stärker vorangeschritten ist als in der Pressesprache.

4.4 Variationsanalyse

4.4.1 Methodik der Variationsanalyse

Wie bereits Schäfer (2019) zeigt und wir bestätigen können (vgl. Abschnitt 4.2), sind die beiden Ausprägungen unserer abhängigen Variable (schwache vs. starke Formen) sehr ungleich verteilt. So sind starke Formen deutlich seltener als schwache. Aus diesem Grund untersuchen wir den Einfluss der in den Abschnitten 4.3.1 bis 4.3.6 vorgestellten Einflussfaktoren anhand einer stratifizierten Stichprobe, in der wir den Anteil der selteneren, starken Form erhöhen (so auch Schäfer 2019). Aus der kombinierten Gesamtextraktion der Daten (KoGra und Webforen) wurden

15 So scheint z. B. das Zweifelsfälle-Wörterbuch (2021: 907) das Weglassen der Endung *-(e)n* im Akkusativ und Dativ implizit als Merkmal insbesondere der gesprochenen Sprache anzusehen, wenn es heißt, es sei „**auch** im geschriebenen Standarddeutsch so weit verbreitet, dass es nicht einfach als inkorrekt bezeichnet werden kann“ (unsere Hervorhebung).

10.000 Belegkandidaten für schwache Formen und 10.000 Belegkandidaten für starke Formen bzw. Genitivformen auf *-(e)ns* zufällig ausgewählt. Anschließend wurden manuell Fehltreffer und unklare Belege ausgeschlossen. Ebenfalls ausgeschlossen wurden die Lexeme *Matador* und *Cäsar*, die beiden einzigen Lexeme mit morphologisch gesteuerter Betonung. Es verblieben 10.044 Belege. 7.933 davon sind Belege für schwache Formen, 2.106 sind Belege für starke Formen, und fünf sind Belege für Genitivformen auf *-(e)ns*. Letztere können auf Grund ihrer geringen Anzahl im Folgenden nicht weiter berücksichtigt werden. Die Daten wurden mit einem gemischten logistischen Regressionsmodell analysiert (vgl. Bates et al. 2015). Als Prädiktorvariablen (= unabhängige Variablen) wurden Phonotaktik, Semantik, Kasus, Apposition, attributives Adjektiv, Häufigkeitsklasse des Substantivs und Korpus aufgenommen; das konkrete Substantivlemma wurde als Zufallsfaktor berücksichtigt. In Abhängigkeit der verschiedenen Prädiktoren wurde die Wahrscheinlichkeit modelliert, mit der in einem gegebenen Fall starke oder schwache Flexion auftritt (Kriteriumsvariable = abhängige Variable). Zur besseren Vergleichbarkeit unserer Ergebnisse mit jenen Schäfers (2019) haben wir zusätzlich ein Regressionsmodell ohne Lemma als Zufallsfaktor gerechnet. Die Informationen zu beiden Modellen befinden sich im Anhang (Abschnitt 7.3).

4.4.2 Ergebnisse

Abbildung 1 zeigt die Koeffizientenschätzungen und 95 %-Konfidenzintervalle für das gemischte logistische Regressionsmodell. Als Referenzausprägung (intercept) setzen wir (Schäfer 2019: 405 folgend) für jeden Prädiktor diejenige Ausprägung an, die gemäß Hypothese am stärksten die schwache Form begünstigt. Die abgetragenen Ausprägungen geben an, ob und ggf. wie stark sie im Vergleich mit der zugehörigen Referenzausprägung die Wahrscheinlichkeit für die starke Flexion erhöhen (oder ggf. wider Erwarten verringern).

Das Modell zeigt einen Einfluss für die Prädiktorvariablen Phonotaktik, Semantik, Kasus, Apposition und Korpus. Zum Prädiktor Phonotaktik: Im Vergleich zu Mehrsilbern mit finalem Schwa (Referenzausprägung) neigen vor allem Einsilber und Mehrsilber mit Nicht-Endbetonung zur starken Flexion, in geringerem Maße auch Mehrsilber mit Endbetonung. Anschaulich ist hier ein Blick in die einfachen, nicht-stratifizierten Stichproben: In der KoGra-Stichprobe kommen bei den Mehrsilbern mit finalem Schwa starke Formen in nur einem von 1.821 Belegen vor (0,1 %). Starke Formen lassen sich also fast nur bei den anderen phonotaktischen Kategorien beobachten: bei den Mehrsilbern mit Endbetonung (1,3 %, 50 von 3.785), den Einsilbern (2,5 %, 35 von 1.392) und den Mehrsilbern mit Nicht-Endbetonung (5,3 %, 13 von 245). In den Webforen macht die starke Form bei den Mehrsilbern mit finalem Schwa immerhin 1 % aus (23 von 2.358). Aber auch in den Foren fallen die Anteile in den anderen Kategorien höher aus (Mehrsilber

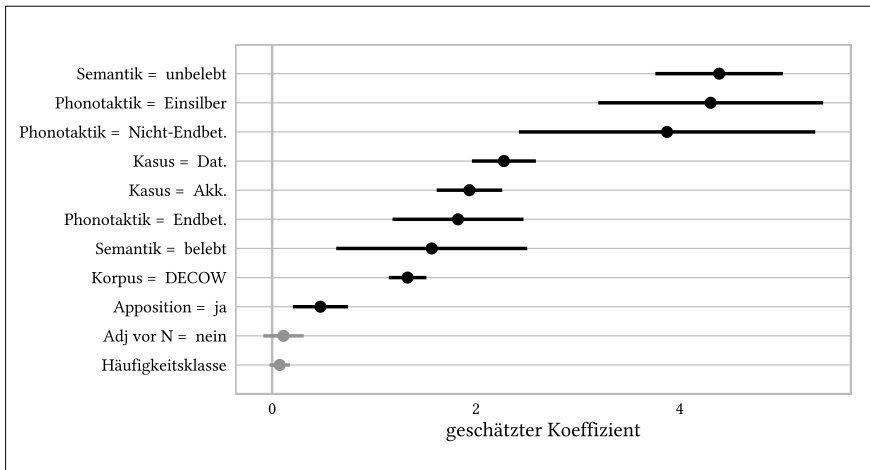


Abbildung 1: Koeffizientenschätzungen für das gemischte logistische Regressionsmodell für die Substantive mit kodifizierter schwacher Flexion. Die horizontalen Linien zeigen 95 %-Konfidenzintervalle. Positive Koeffizienten erhöhen die Wahrscheinlichkeit für die starke Flexion, negative Koeffizienten verringern sie. Marginales R^2 (nur feste Effekte) = 0,53, konditionales R^2 (gesamtes Modell) = 0,78. „Nicht-Endbet.“ = Mehrsilber mit Nicht-Endbetonung, „Dat.“ = Dativ, „Akk.“ = Akkusativ, „Endbet.“ = Mehrsilber mit Endbetonung, „Adj vor N“ = Adjektiv vor Nomen.

mit Endbetonung: 6,5 %, 154 von 2.370; Einsilber: 6,5 %, 174 von 2.667; Mehrsilber mit Nicht-Endbetonung: 14,2 %, 46 von 325). Das bedeutet: Bei den historisch schwachen Maskulina auf Schwa scheint sich der Flexionsklassenwechsel im Wesentlichen auf jene zu beschränken, bei denen er sich dergestalt vollzieht, dass das Flexiv *-n* als Teil des Stammes reanalysiert wird. Wie in Abschnitt 4.1 gezeigt, handelt es sich bei diesen um eine Teilmenge jener Substantive mit finalem Schwa, die das für schwache Maskulina untypische Merkmal aufweisen, ein unbelebtes Denotat zu haben.

Zum Prädiktor Semantik: Im Vergleich zu Substantivverwendungen mit menschlichem Referenten (Referenzausprägung) neigen vor allem Verwendungen mit unbelebtem Referenten zur starken Flexion, in geringerem Maße auch Verwendungen mit nicht-menschlichem aber belebtem Referenten. Die Ergebnisse für Phonotaktik und Semantik ähneln denen Schäfers (2019). Sie stützen die Annahme, dass Abweichungen vom kanonischen, schwachen Muster vor allem bei Substantiven (bzw. Substantivverwendungen) auftreten, für die in phonotaktischer und/oder semantischer Hinsicht dieses Muster eigentlich untypisch ist.

Zum Prädiktor Kasus: Im Vergleich zum Genitiv (Referenzausprägung) neigen Dativ und Akkusativ stärker zur starken Flexion. Mit Thieroff (2003: 109, 114) lässt dieser Befund sich zum einen strukturalistisch erklären: Um das abstrakte Distinktionsmuster der starken Flexion zu erreichen, die als analogisches Vorbild

fungiert (nämlich markierte Genitivform vs. unmarkierte Nicht-Genitivform), ist es ausreichend, lediglich Akkusativ und Dativ anzugleichen. Darüber hinaus steht der Wechsel im Akkusativ und Dativ, wo er zu Endungslosigkeit führt, im Einklang mit der allgemeineren Tendenz des Deutschen hin zu einem Abbau substantivischer Kasusflexion. Schließlich lässt sich der Gegensatz zwischen Genitiv einerseits und Akkusativ und Dativ andererseits auch soziolinguistisch motivieren: Wenn der Genitiv ein Marker für Hoch- und Schriftsprachlichkeit ist (Seiler 2022: 58), dann ist insbesondere für diesen Kasus zu erwarten, dass er auch normgerecht – hier also schwach – realisiert wird.

Für den Prädiktor Apposition finden wir einen schwachen Effekt, der in die zu erwartende Richtung zeigt: Wenn das Substantiv mit einem appositiven Nebenkern auftritt, erhöht sich die Wahrscheinlichkeit für die starke Flexion dadurch, dass endungslose Akkusativ- und Dativformen begünstigt werden. Im Vergleich mit den Effekten für Phonetik und Semantik scheint dieser Faktor aber nur eine geringe Rolle zu spielen.

Für den Prädiktor Adjektiv vor Nomen wird durch das gemischte Modell kein signifikanter Einfluss ermittelt. Dort lässt sich also nicht beobachten, dass in NPs ohne attributives Adjektiv eher stark flektiert wird als in NPs mit Adjektiv, der Hypothese gemäß dadurch, dass im Akkusativ und Dativ die Endung am ehesten in NPs ohne Adjektiv wegfällt – wenngleich immerhin die Tendenz in diese Richtung deutet. Im einfachen Regressionsmodell ohne Lemma als Zufallsfaktor finden wir einen schwachen Effekt in diese Richtung. Eine mögliche Interpretation dieses Unterschieds zwischen den beiden Modellen könnte sein, dass der Adjektiv-Effekt an einzelne Substantiv-Lexeme gebunden ist (was aber wiederum selbst einer Erklärung bedürfte).

Auch für die Häufigkeitsklasse des Substantivs finden wir im gemischten Modell keinen signifikanten Effekt, wohl aber im einfachen Modell ohne Lemma als Zufallsfaktor (vgl. Schäfer 2019: 406), auch wenn dieser Effekt im Vergleich zu den anderen relativ schwach ausgeprägt ist. Die Wahrscheinlichkeit für Abweichungen hin zum starken Muster erhöht sich bei eher weniger frequenten Substantiven (= mit steigender Häufigkeitsklasse). Unsere Vermutung ist, dass der Effekt im gemischten Modell mit Lemma als Zufallsfaktor verschwindet, weil die Häufigkeitsklasse unmittelbar mit dem Substantiv verbunden ist und der Effekt deshalb nicht mehr zum Tragen kommt, wenn man das Lemma als Zufallsfaktor modelliert.

Für den außersprachlichen Prädiktor Korpus ergibt sich, dass die Webforen in stärkerem Maße zur starken Flexion neigen als KoGra. Dieser Befund war erwartet: Er stützt die Annahme, dass das Norm-divergente – und hier zugleich Flexiv-ärmere – Flexionsmuster eher in informellen Registern auftritt.

5 Starke Maskulina

Analog zum Vorgehen bei den schwachen Maskulina werten wir als „starke Maskulina“ diejenigen, für die bei Duden Online nur das starke Muster angegeben wird (Genitiv auf *-(e)s*). Anhand dieses Kriteriums lassen sich 69 der insgesamt 1.156 schwach belegbaren Substantive unserer Lemmaliste (entspricht 5,9 %) als starke Maskulina klassifizieren.¹⁶ 39 davon (entspricht 57 %) sind *or*-Substantive wie *Autor* (vgl. 4–6); 30 haben eine andere Struktur (z. B. *Muslim*, *Pensionär*, *Pimpf*), vgl. z. B. (9).

- (9) Kernthese ist eine Beteiligung von Geheimdiensten, die verhindern wollten, dass Diana, deren Sohn William einmal König werden soll, **einen Muslimen** heiraten würde. (Süddeutsche Zeitung, 4.9.1999, S. 14)

In phonotaktischer Hinsicht (Tabelle 4) handelt es sich bei den Substantiven mehrheitlich um solche mit morphologisch gesteuerter Betonung. Den größten Anteil an dieser Gruppe haben die *or*-Substantive. Diese tragen in unflektierter Form den Hauptakzent auf der Silbe **vor** dem stammfinalen *-or* (*Autor*, *Senator*, *Zensor*); nach der Suffigierung mittels *-en* fällt der Hauptakzent **auf** die stammfinale Silbe (*Autoren*, *Senatoren*, *Zensoren*).¹⁷ Hinzu kommen die Substantive *Dämon*, *Embryo* und *Pharao* (*Dämon* vs. *Dämonen*, *Embryo* vs. *Embryonen*, *Pharao* vs. *Pharaonen*). In unflektierter Form könnte man die Substantive mit morphologisch gesteuerter Betonung somit den Mehrsilbern mit Nicht-Endbetonung zuzuordnen, in ihrer flektierten *-en*-Form hingegen den Mehrsilbern mit Endbetonung. Die zweitgrößte Gruppe bilden Substantive, die bereits in unflektierter Form endbetont werden (z. B. *Muslim*, *Pensionär*, *Ysop*). Bei den übrigen Substantiven handelt es sich um Einsilber (*Elch*, *Mast*, *Pimpf*, *Schelm*, *Zeh* und *Zwerg*) und um Mehrsilber mit Nicht-Endbetonung (*Elektromast*, *Enkel*, *Phänotyp*, *Prototyp*, *Urtyp*, *Vetter*). Wie schon bei den schwachen Maskulina besteht die letztgenannte Gruppe überwiegend aus Präkonfixbildungen (*Elektromast*, *Prototyp*, *Phänotyp* und *Urtyp*). Ein Substantiv, nämlich *Kontinent*, klassifizieren wir gemäß Duden Online als Substantiv mit variabler Betonung (*Kontinent* vs. *Kontinent*). In semantischer Hinsicht

16 Neben diesen 69 starken Maskulina und den 1.042 schwachen Maskulina (vgl. Abschnitt 4) enthält unsere Lemmaliste 40 Maskulina ohne Eintrag bei Duden Online (die sich anhand ihrer Wortausgänge überwiegend den schwachen Maskulina zuordnen lassen, vgl. z. B. *Säkuralist*, *Trilobit*, *Mythomane*), das Neutrum *Herz* sowie die vier Sonderfälle in den Beispielen (10–13).

17 Abgesehen von Präkonfixbildungen mit dem Primärakzent auf dem Konfix (z. B. *Ko-Administrator*).

Tabelle 4: Starke Maskulina mit belegter schwacher Flexion (Lexeme) nach Phonotaktik.

Phonotaktik	n	%
morphologisch gesteuerte Betonung (<i>Autor</i>)	42	60,9
Endbetonung (<i>Pensionär</i>)	14	20,3
Einsilbigkeit (<i>Zwerg</i>)	6	8,7
Nicht-Endbetonung (<i>Vetter</i>)	6	8,7
variable Betonung (<i>Kontinent</i>)	1	1,4
gesamt	69	100

Tabelle 5: Starke Maskulina mit belegter schwacher Flexion (Lexeme) nach Semantik.

Semantik	n	%
menschliches Denotat (<i>Autor</i>)	43	62,3
unbelebtes Denotat (<i>Smaragd</i>)	18	26,1
belebtes aber nicht menschliches Denotat (<i>Alligator</i>)	5	7,2
als polysem gewertet (<i>Zwerg</i>)	3	4,3
gesamt	69	99,9

(vgl. Tabelle 5) handelt es sich bei über 60% der Substantive um solche mit menschlichem Denotat.

Im Vergleich mit der bereits in Abschnitt 4.3.1 herangezogenen Zufallsstichprobe von 1.000 maskulinen Substantiven, die sich **nicht** schwach flektiert belegen lassen (siehe Anhang 7.2), sind unter den 69 starken Maskulina, die sich ihrer Kodifizierung entgegen sehr wohl mindestens einmal schwach flektiert belegen lassen, die Substantive mit morphologisch gesteuerter Betonung mit Abstand am stärksten überrepräsentiert. So ist ihr Anteil dort 25,4-mal so hoch wie unter den Substantiven der Vergleichsgruppe (Beispiele aus der Vergleichsgruppe: *Aggressor*, *Tumor*). Dieser Befund steht im Einklang mit dem Eindruck, dass insbesondere *or*-Substantive dazu neigen, sich entgegen der Norm der schwachen Flexion anzuschließen. Ebenfalls überrepräsentiert sind die Mehrsilber mit Endbetonung. Deren Anteil an den starken Substantiven mit belegter schwacher Flexion ist doppelt so hoch wie ihr Anteil an der Vergleichsgruppe. Einsilber und vor allem Mehrsilber mit Nicht-Endbetonung hingegen sind unterrepräsentiert. Der Anteil der Einsilber fällt in der Vergleichsgruppe 2,2-mal so hoch aus wie unter den schwach belegbaren starken Maskulina; der Anteil der Mehrsilber mit Nicht-Endbetonung ist in der Vergleichsstichprobe sogar 7,8-mal so hoch.

In semantischer Hinsicht sind unter den schwach belegbaren starken Maskulina die Substantive mit menschlichem Denotat und die Substantive mit belebtem

aber nicht menschlichem Denotat überrepräsentiert. Der Anteil ersterer fällt unter den schwach belegbaren starken Maskulina 1,8-mal so hoch aus wie in der Vergleichsgruppe, der Anteil letzterer 1,5-mal so hoch. Die Substantive mit unbelebtem Denotat hingegen sind unter den schwach belegbaren starken Maskulina unterrepräsentiert. Deren Anteil ist in der Vergleichsgruppe 2,2-mal so hoch.

Die Überrepräsentation der Mehrsilber mit Endbetonung und der Substantive mit menschlichem Denotat sowie die Unterrepräsentation der Einsilber, der Mehrsilber mit Nicht-Endbetonung und der Substantive mit unbelebtem Denotat teilen die schwach belegbaren starken Maskulina mit den schwachen Maskulina (vgl. Abschnitt 4.3.1). Das lässt sich so interpretieren, dass sich insbesondere solche starken Maskulina entgegen der Norm schwach flektiert belegen lassen, die Eigenschaften aufweisen, die typisch für die schwachen Maskulina sind. Die Stärke der Überrepräsentation der Substantive mit morphologisch gesteuerter Betonung lässt sich hingegen nicht damit erklären, dass genau diese Art von Substantiv prototypisch für die schwachen Maskulina wäre. (Wie in Tabelle 2 gezeigt, macht sie dort nur einen verschwindend geringen Anteil aus, und im Vergleich mit der Vergleichsgruppe sind sie dort deutlich unterrepräsentiert: So ist ihr Anteil an der Vergleichsgruppe 12-mal so hoch wie ihr Anteil unter den schwachen Maskulina.) Eher lässt sich die Neigung der Substantive mit morphologisch gesteuerter Betonung, sich dem schwachen Muster anzuschließen, wohl damit erklären, dass sie in ihrer *-(e)n*-Form das – tatsächlich prototypische – Merkmal der Endbetonung aufweisen. Der Prototypen-Ansatz liefert damit nicht nur eine Erklärung dafür, welche schwachen Maskulina dazu neigen, stark flektiert zu werden, sondern auch dafür, welche starken Maskulina sich den schwachen Maskulina anschließen.

Darüber hinaus dürfte der Prototypenansatz auch zum Verständnis für ein paar auffällige Sonderfälle unter den Substantiven unserer Lemmaliste beitragen, darunter das als Neutrum kodifizierte *Patriarchat* (10), die als Feminina kodifizierten *Brosame* und *Waise* (11–12) sowie das als Pluraletantum kodifizierte *Nibelungen* (13), die in den Belegen anscheinend als schwache Maskulina verwendet werden. Bei *Patriarchat* in (10) liegt vermutlich eine Verwechslung mit dem Lexem *Patriarch* vor, bei *Brosame* in (11) eine volksetymologische Anlehnung an *Same(n)* und bei *Waise* in (12) ein Sexus-bedingter Genuswechsel (es geht um eine männliche Waise). *Waise*, *Nibelunge* und *Brosame* weisen in phonotaktischer Hinsicht das für schwache Maskulina prototypische finale Schwa auf, *Patriarchat* die prototypische Endbetonung. Zudem haben *Waise* und *Nibelunge* ein menschliches Denotat; und auch *Patriarchat* wird in (10) mit menschlicher Referenz verwendet. Es mag auffällig bleiben, dass die Lexeme überhaupt als Maskulina verwendet werden. Ihre Zuordnung speziell zur Untergruppe der schwachen Maskulina hingegen ist angesichts ihrer Phonotaktik (und z. T. auch Semantik) durchaus nachvollziehbar.

- (10) Nachforschungen über **einen Patriarchaten** (die tageszeitung, 17.1.1997, Beilage Le Monde diplomatique, S. 21–22)
- (11) Gebt den Kommunen **einen Brosamen** und lasst die das vor Ort machen! (Protokoll der Sitzung des Parlaments Landtag Nordrhein-Westfalen am 20.9.2007)
- (12) Die Brüder Dardenne haben einen bezaubernden Film über **einen Waisen** und seine Ziehmutter gedreht. (Braunschweiger Zeitung, 10.2.2012, o. S.)
- (13) Bereits als Kind hatte Dietrich diese besondere Luft geschnuppert, da spielte er im „Rheingold“ an der Nürnberger Oper **einen Nibelungen**. (Nürnberger Nachrichten, 5.9.1998, S. 3.)

5.1 Anteile starker und schwacher Formen

Im nächsten Schritt ermitteln wir die Anteile der Fälle, in denen die 69 mindestens einmal schwach belegbaren starken Maskulina ihrer Kodifizierung entsprechend stark verwendet werden, d. h. im Akkusativ und Dativ endungslos bzw. im Genitiv mit dem Suffix *-(e)s*, und die Anteile der Fälle, in denen sie davon abweichend schwach flektiert werden, d. h. im Akkusativ, Dativ oder Genitiv mit dem Suffix *-(e)n* auftreten. Grundlage bilden (wie bei den schwachen Maskulina, vgl. Abschnitt 4.2) zwei einfache, nicht-stratifizierte Zufallsstichproben von je 10.000 Belegen (vor dem manuellen Ausschluss von Fehlbelegen) aus KoGra und dem Webforenteil des DECOW-Webkorpus. Auf diese Weise sollen wieder Unterschiede zwischen eher formellen/distanzsprachlichen Texten einerseits und eher informellen/nähesprachlichen Texten andererseits erfasst werden.¹⁸

Die KoGra-Stichprobe enthält nach manuellem Ausschluss von Fehlbelegen noch 8.603 Belege, verteilt auf 60 Lexeme. Das frequenteste Lexem ist *Autor*, mit 1.848 Belegen macht es 21,5 % aller Belege aus. Die große Mehrheit von 97,9 % der Belege tritt in starken Formen auf (8.423 von 8.603). Im Akkusativ tritt in 3,4 % der Belege die schwache Form auf (93 von 2.702), verteilt auf 23,6 % der akkusativisch belegten Lexeme (13 von 55), im Dativ in 1,7 % der Belege (53 von 3.164), verteilt auf 25,9 % der dativisch belegten Lexeme (15 von 58), und im Genitiv in

18 Nicht berücksichtigt werden konnten hier wie auch in der Variationsanalyse (vgl. Abschnitt 5.3) die Lexeme *Embryo* und *Pharao*, da unsere Suchanfragen (vgl. Anhang 7.1) die erwartbaren schwachen Formen mit *n*-Epenthese nicht erfassen (vgl. *Embryonen*, *Pharaonen*).

1,2 % der Belege (34 von 2.737), verteilt auf 24,0 % der genitivisch belegten Lexeme (12 von 50). Kasusaugmentierende Formen auf *-(e)ns* sind in der Stichprobe nicht belegt. Insgesamt deuten die Ergebnisse damit darauf hin, dass auch bei denjenigen starken Substantiven, die grundsätzlich schwach auftreten – nur solche wurden hier ja überhaupt berücksichtigt – die schwache Form mit insgesamt 2,1 % in KoGra nur eine seltene Nebenform darstellt.

Die Webforen-Stichprobe enthält nach manuellem Ausschluss von Fehlbelegen noch 8.919 Belege, verteilt auf 59 Lexeme. Das frequenteste Lexem ist wie in der KoGra-Stichprobe *Autor*; mit 1.726 Belegen macht es 19,4 % aller Belege aus. Die große Mehrheit von 94,8 % der Belege tritt in starken Formen auf (8.457 von 8.919). Im Akkusativ tritt in 5,9 % der Belege die schwache Form auf (206 von 3.492), verteilt auf 39,3 % der akkusativisch belegten Lexeme (22 von 56), im Dativ in 4,4 % der Belege (163 von 3.689), verteilt auf 35,2 % der dativisch belegten Lexeme (19 von 54), und im Genitiv in 5,4 % der Belege (93 von 1.738), verteilt auf 36,4 % der genitivisch belegten Lexeme (16 von 44). Kasusaugmentierende Formen auf *-(e)ns* sind wie in der KoGra-Stichprobe nicht belegt.

Insgesamt ist der Anteil schwacher Formen starker Maskulina in der Foren-Stichprobe mit 5,2 % etwas höher als in der KoGra-Stichprobe (2,1 % s. o.). Trotzdem ist die schwache Form auch hier nur eine seltene Nebenform. Anders als bei der Schwankung von schwach zu stark deutet sich zudem kein deutlicher Gegensatz zwischen Akkusativ und Dativ einerseits und Genitiv andererseits an. Dieser Befund scheint sich auf den ersten Blick nicht mit jenem Schäfers (2019: 407) für eine Auswahl von *or*-Substantiven zu decken. Zu beachten ist dabei jedoch, dass hier (anders als bei Schäfer) nur solche Lexeme berücksichtigt werden, die sich überhaupt schon mindestens einmal im Akkusativ und/oder Dativ schwach flektiert belegen lassen (vgl. Abschnitt 3.1 zur Erhebung der Lemmaliste). Zudem wurde die Analyse hier nicht auf *or*-Substantive beschränkt.

Als bekanntestes Beispiel für die Schwankung von stark zu schwach kann wohl das Lexem *Autor* gelten. In der KoGra-Stichprobe ist absolut betrachtet die Form *Autoren* die am zweithäufigsten belegte schwache Form eines starken Substantivs; an erster Stelle steht die Form *Prototypen*. In Bezug auf den **Anteil** starker Realisierungen scheint das Lexem *Autor* jedoch eine weniger prominente Rolle zu spielen. Betrachtet man diejenigen 39 Lexeme der Stichprobe, die mindestens 10-mal belegt sind, dann ergeben sich die höchsten Anteile für *Prototyp* (37,4 %, 58 von 155), *Mast*¹⁹ (17,3 %, 22 von 127) und *Muslim* (8,3 %, 3 von 33). *Autor* kommt

19 *Mast* ist das einzige **starke** Maskulinum unserer Lemmaliste, für das sich die *-en*-Form auch im Nominativ (unmittelbar nach *ein*) belegen lässt (vgl. z. B. *Ungefähr 400 Meter hinter dem Pfarrhof soll ein Masten aufgestellt werden*, Burgenländische Volkszeitung, 6.1.2010, S. 22). Wie eine Stichprobe von 200 zufällig ausgewählten Belegen für *Mast/Masten* unmittelbar nach *ein* zeigt, macht die *-en*-Form ca. 5 % der Fälle aus.

erst an achter Stelle (2,6 %, 48 von 1.848). In der Foren-Stichprobe ergeben sich die höchsten Anteile für *Prototyp* (61,2 %, 153 von 250), *Archetyp* (50 %, 5 von 10) und *Dämon* (20,7 %, 82 von 397). *Autor* kommt an elfter Stelle (5,6 %, 97 von 1.726).

5.2 Variationssteuernde Faktoren

Im Folgenden untersuchen wir, inwiefern dieselben Faktoren, die die Wahl zwischen schwachen und starken Formen bei den schwachen Maskulina steuern (bzw. für die sich zumindest entsprechende Hypothesen formulieren ließen), diese Wahl auch bei den starken Maskulina beeinflussen. Wie sich bereits gezeigt hat, weisen diese Faktoren bei den starken Maskulina z. T. andere Ausprägungen auf. So gehören hinsichtlich ihrer Phonotaktik ja die meisten der schwach belegbaren starken Maskulina dem Typus ‚morphologisch gesteuerte Betonung‘ an, der bei den schwachen Maskulina kaum vorkommt. Dafür ist, umgekehrt, bei den starken Maskulina das Merkmal ‚finales Schwa‘ nicht vertreten.

Darüber hinaus fallen die mit den Faktoren verbundenen Hypothesen z. T. anders aus. Das liegt daran, dass nun die starken Formen die kanonischen und die schwachen Formen die divergenten sind.

5.2.1 Prototypikalität: Phonotaktik und Semantik

Gegenstand unserer Variationsanalyse sind von vornherein nur diejenigen starken Maskulina, die sich mindestens einmal im Akkusativ und/oder Dativ schwach flektiert belegen lassen (vgl. Abschnitt 3.1 zur Erstellung der Lemmaliste). Wie zu Beginn von Abschnitt 5 gezeigt, sind unter diesen Substantiven solche überrepräsentiert, die Eigenschaften aufweisen, die typisch für schwache Maskulina sind: In phonotaktischer Hinsicht solche, die (mindestens in flektierter *-(e)n*-Form) die für schwache Maskulina typische Endbetonung aufweisen. Und in semantischer Hinsicht solche, die ein menschliches Denotat haben. Im Rahmen unserer Variationsanalyse überprüfen wir nun, inwiefern auch noch einmal innerhalb der Substantive, die sich mindestens einmal schwach flektiert belegen lassen, diejenigen mit typischen Eigenschaften in stärkerem Maße zu schwachen Formen tendieren als diejenigen mit weniger typischen Eigenschaften. Wenn das der Fall ist, dann sollten in phonotaktischer Hinsicht die Mehrsilber mit Endbetonung in stärkerem Maße zu schwachen Formen neigen als die Einsilber und die Mehrsilber mit Nicht-Endbetonung. Für die große Gruppe der Substantive mit morphologisch gesteuerter Betonung ließe sich vermuten, dass sie sich ähnlich wie die

Akkusativ- und Dativformen mit *-en* ließen sich demnach auch als unflektierte Formen zu einer Stammvariante auf *-en* auffassen.

Mehrsilber mit Endbetonung verhalten, mit denen sie in flektierter *-(e)n*-Form zusammenfallen. In semantischer Hinsicht wäre zu erwarten, dass Substantivverwendungen mit menschlichem Referenten stärker zu schwachen Formen neigen als Substantivverwendungen mit belebtem nicht-menschlichem Referenten oder unbelebtem Referenten.

5.2.2 Kasus

Für die starken Maskulina erscheint uns weniger klar als für die schwachen, ob und ggf. wie sich der Faktor Kasus auf die Wahl zwischen starker und schwacher Flexion auswirkt. Das strukturalistische Argument Thieroffs (vgl. Abschnitt 4.3.2) lässt auch bei den starken Maskulina die schwache Form am ehesten im Genitiv erwarten. Denn im Genitiv, und nur im Genitiv, bliebe trotz der Angleichung an die schwache Flexion die abstrakte Distinktionsstruktur des starken Paradigmas, d. h. markierter Genitiv vs. nicht-markierter Nicht-Genitiv, erhalten (vgl. z. B. *des Autoren* vs. *der/dem/den Autor*). Aus soziolinguistischer Perspektive wäre wieder zu erwarten, dass insbesondere im Genitiv die Norm-konvergente Form gewählt wird. Das ist hier nun aber die starke Form. Das bedeutet: Während bei den schwachen Maskulina der strukturalistische und der soziolinguistische Faktor in dieselbe Richtung wirken (beide begünstigen die schwache Form eher im Genitiv als im Akkusativ und Dativ), wirken sie bei den starken Maskulina einander entgegen. Vor diesem Hintergrund ließe sich die Vermutung aufstellen, dass der Gegensatz zwischen Genitiv einerseits und Akkusativ und Dativ andererseits bei den starken Maskulina eine geringere Rolle spielt als bei den schwachen.

5.2.3 Apposition und attributives Adjektiv

Für die Faktoren Apposition und attributives Adjektiv ist unklar, ob sie einen Einfluss haben. Für den Fall, dass sie sich auswirken, ließe sich annehmen, dass sie jeweils in dieselbe Richtung wirken, wie es für die schwachen Maskulina angenommen wurde (und zumindest für den Faktor Apposition hier auch nachgewiesen werden konnte). Die Präsenz einer Apposition könnte im Akkusativ und Dativ also die starke, endungslose Form begünstigen, die Präsenz eines attributiven Adjektivs die schwache, endungshaltige Form.

5.2.4 Frequenz des Substantivs

Nimmt man wie für die schwachen Maskulina an, dass frequentere Substantive resistenter gegenüber analogischem Wandel sind als weniger frequente, dann könnten bei den starken Maskulina die frequenteren in stärkerem Maße am starken Muster festhalten als die weniger frequenten.

5.2.5 Korpus: KoGra vs. Webforen

Bei der Verwendung schwacher Formen für starke Maskulina handelt es sich genau wie bei der Verwendung starker Formen für schwache Maskulina um eine Abweichung von der kodifizierten Norm. Vor diesem Hintergrund könnte man erwarten, dass beide Phänomene in den – informelleren – Webforen eher zu beobachten sind als im pressesprachlichen KoGra. Zu erwarten wäre demnach, dass bei den starken Maskulina (anders als bei den schwachen Maskulina) die Foren in stärkerem Maße zum schwachen Muster tendieren als KoGra. Zu berücksichtigen ist jedoch, dass das schwache Paradigma in allen drei nicht-nominativischen Kasus mit der *-(e)n*-Form eine flektierte Form verlangt, während es sich bei starken Akkusativ- und Dativformen um die endungslose Form handelt. Das bedeutet: Während die Normabweichung bei den schwachen Maskulina zu einem Abbau von substantivischer Kasusflexion führt, wie er typisch für nicht bzw. weniger normierte Formen des Deutschen ist, geht die Schwankung von stark zu schwach, umgekehrt, mit einem Aufbau von Kasusflexion einher. Vor diesem Hintergrund erscheint uns nicht offensichtlich, inwiefern für die Schwankung von stark zu schwach genau wie für die Schwankung von schwach zu stark zu erwarten ist, dass sie eher in den Webforen als in KoGra zu beobachten ist. Die in Abschnitt 5.1 präsentierten deskriptiven Statistiken deuten allerdings tatsächlich an, dass auch bei den starken Maskulina die divergenten Formen in den Foren häufiger vorkommen als in KoGra.

5.3 Variationsanalyse

5.3.1 Methodik der Variationsanalyse

Wie bei den schwachen Maskulina sind auch bei den starken Maskulina die beiden Ausprägungen unserer abhängigen Variable (starke vs. schwache Formen) quantitativ sehr ungleich verteilt. Die Norm-divergenten Formen (hier: die schwachen) sind auch hier deutlich seltener als die kanonischen (die starken) (vgl. Abschnitt 5.1). Aus diesem Grund arbeiten wir auch hier mit einer stratifizierten Stichprobe, in der wir den Anteil der selteneren Form erhöhen. Aus der kombinierten Gesamtextraktion der Daten (KoGra und Webforen) wurden wieder je 10.000 Belegkandidaten für schwache Formen und 10.000 Belegkandidaten für starke Formen zufällig ausgewählt und manuell durchgesehen. Anschließend wurden manuell Fehltreffer und unklare Belege ausgeschlossen. Ebenfalls ausgeschlossen wurde das Lexem *Kontinent*, das einzige Lexem mit variabler Betonung. Es verbleiben 9.444 Belege. 8.041 davon sind Belege für starke Formen, 1.403 davon sind Belege für schwache Formen. Die Daten wurden wie bei den schwachen

Maskulina mit einem gemischten logistischen Regressionsmodell analysiert. Die Informationen zum Modell befinden sich in Anhang 7.3.

5.3.2 Ergebnisse

Abbildung 2 zeigt die Koeffizientenschätzungen und 95 %-Konfidenzintervalle für das gemischte logistische Regressionsmodell. Als Referenzausprägung (intercept) setzen wir wie bei den schwachen Maskulina für jeden Prädiktor diejenige Ausprägung an, die gemäß Hypothese am stärksten die schwache Form begünstigt. Für die Prädiktoren Kasus und Korpus ist unserer Auffassung nach bei den starken Substantiven weniger klar, in welche Richtung sie ggf. wirken (vgl. Abschnitte 5.2.2 und 5.2.5). Analog zum Modell für die schwachen Substantive (Abbildung 1) legen wir auch hier die Ausprägungen Genitiv bzw. KoGra als Referenzausprägungen fest.

Durch das statistische Modell wird nur für die Prädiktorvariablen Phonotaktik, Semantik und Korpus ein signifikanter Einfluss auf die Wahrscheinlichkeit der Wahl der Flexion ermittelt. Für den Prädiktor Phonotaktik ergibt sich, dass die Mehrsilber mit Endbetonung (Referenzausprägung) stärker zur schwachen Flexion neigen als die große Gruppe der Substantive mit morphologisch konditionierter

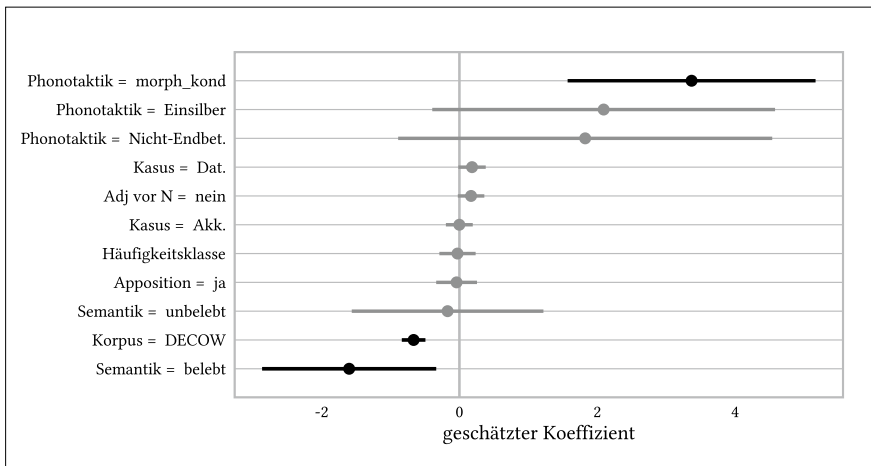


Abbildung 2: Grafische Darstellung der Koeffizientenschätzungen für das gemischte logistische Regressionsmodell für die Substantive mit kodifizierter starker Flexion. Die horizontalen Linien zeigen 95 %-Konfidenzintervalle. Positive Koeffizienten erhöhen die Wahrscheinlichkeit für die starke Flexion, negative Koeffizienten verringern sie. Marginales R^2 (nur feste Effekte) = 0,17, konditionales R^2 (gesamtes Modell) = 0,66. „morph_kond“ = Substantive mit morphologisch konditionierter Betonung, „Nicht-Endbet.“ = Mehrsilber mit Nicht-Endbetonung, „Dat.“ = Dativ, „Adj vor N“ = Adjektiv vor Nomen, „Akk.“ = Akkusativ.

Betonung („morph_kond“: *or*-Substantive und *Dämon*). Für die übrigen phonotaktischen Typen, d. h. die Einsilber und die Mehrsilber mit Nicht-Endbetonung, wird kein signifikanter Unterschied im Vergleich zur Referenzausprägung festgestellt. Für diese Typen geht die Koeffizientenschätzung allerdings auch mit einer größeren Unsicherheit einher, ersichtlich aus den größeren Konfidenzintervallen. Zwischen den Substantiven mit Endbetonung und den Substantiven mit morphologisch gesteuerter Betonung wurde nicht zwingend ein Unterschied erwartet. Er ließe sich aber damit erklären, dass sich die Mehrsilber mit Endbetonung bereits in unflektierter Form wie prototypische schwache Maskulina verhalten, die Substantive mit morphologisch gesteuerter Betonung hingegen erst in ihrer flektierten *-(e)n*-Form.

Zum Prädiktor Semantik: In dem von uns ausgewerteten Datensatz neigen nicht, wie erwartbar, die Substantive bzw. Substantivverwendungen mit menschlichen Referenten (Referenzausprägung) stärker als andere zur schwachen Flexion, sondern die Substantive bzw. Substantivverwendungen mit belebten nicht-menschlichen Referenten. Als solche klassifiziert haben wir die Belege für die Substantive *Alligator*, *Avatar*, *Dämon*, *Elch* und *Fasan* sowie die Mehrheit der Belege für *Zwerg* (‘kleinwüchsiges Fabelwesen’, im Gegensatz z. B. zu Verwendungen zur Bezeichnung eines Kindes).

Für den Prädiktor Kasus ergibt die Analyse keinen signifikanten Effekt. Bezogen auf die hier berücksichtigten Substantive, d. h. solche, die zumindest schon einmal im Akkusativ und/oder Dativ schwach flektiert vorkommen, kann hier also nicht beobachtet werden, dass sie im Genitiv stärker als im Akkusativ und Dativ zur schwachen Flexion neigen. Eine mögliche Erklärung dafür könnte sein, dass der strukturalistische und der soziolinguistische Faktor sich bei den starken Substantiven gegenseitig aufheben (vgl. Abschnitt 5.2.2): Aus soziolinguistischer Perspektive ist insbesondere im Genitiv mit der jeweils Norm-konformen Variante zu rechnen. Bei den schwachen Maskulina ist es mit der schwachen Flexion dieselbe Variante, die auch aus strukturalistischer Sicht begünstigt wird. Bei den starken Substantiven hingegen ist es mit der starken Flexion die andere Variante.

Für den Prädiktor Korpus zeigt sich, dass die Webforen in stärkerem Maße zur schwachen Flexion neigen als KoGra. Das ist dadurch erklärbar, dass hier ja nun das schwache Muster das Norm-divergente ist. Allerdings wäre auch das umgekehrte Ergebnis durchaus erklärbar: Eine stärkere Tendenz der Webforen in Richtung des Norm-konvergenten, starken Musters wäre insofern nicht undenkbar, als dieses insgesamt Flexiv-ärmer ist, und viele weniger streng normierte Formen des Deutschen zu einem Abbau substantivischer Kasusmorphologie neigen. Die Webforen verhalten sich nun aber in beide Richtungen Norm-divergenter – d. h. nicht nur bei der Schwankung von schwach zu stark, sondern auch bei der von stark zu schwach. Und damit auch dort, wo dies zu einem Mehr an Kasusflexion führt.

Ein auffälliger Unterschied zwischen dem Modell für die starken Substantive im Vergleich zum Modell für die schwachen liegt im Anteil der Varianz, der jeweils durch die festen Effekte bzw. den Zufallsfaktor (Lemma) erklärt wird, approximiert durch Pseudo- R^2 von Nakagawa et al. (2017) für gemischte Modelle (vgl. Abschnitt 7.3). Bei den als stark kodifizierten Substantiven (marginales $R^2 = 0,17$) fällt der Anteil der Varianzaufklärung der festen Effekte deutlich kleiner aus als bei den als schwach kodifizierten (marginales $R^2 = 0,53$). Bei den starken Maskulina wird demnach ein Großteil der Varianz erst durch den Zufallsfaktor Lemma erklärt. Dieses Ergebnis lässt sich so interpretieren, dass die Schwankung von stark zu schwach in stärkerem Maße ein Merkmal einzelner Lexeme ist.

6 Fazit und Ausblick

In diesem Kapitel haben wir Schwankungen zwischen schwacher und starker Maskulinflexion untersucht. Wir haben dazu den Prototypen-Ansatz von Köpcke (1995) und Schäfer (2019) übernommen und um weitere Sprach-interne Faktoren sowie um den Einfluss von Textsortenunterschieden (KoGra vs. Webforen) erweitert. Außerdem haben wir nicht nur die als schwach kodifizierten, sondern auch die als stark kodifizierten Substantive mit einem gemischten logistischen Regressionsmodell untersucht, das die einzelnen Faktoren in ihrem Zusammenspiel analysiert.

Die Auswertung der einfachen, nicht-stratifizierten Zufallsstichproben stützt den Befund Schäfers (2019), dass starke Formen schwacher Maskulina insgesamt selten sind. Und zwar nicht nur im überwiegend pressesprachlichen KoGra, sondern auch im Forenteil des DECOW-Webkorpus. Selbst dort liegt der Anteil endungsloser Akkusativ- und Dativformen bei unter 7%. Schwache Formen starker Substantive sind ähnlich selten, und zwar selbst dann, wenn man – wie hier – nur diejenigen Lexeme berücksichtigt, die überhaupt variieren.

Die Ergebnisse bestätigen zudem den Einfluss der Prototypikalität des Substantivs: Für die schwachen Maskulina bestätigt unsere Variationsanalyse, dass bei Substantiven mit prototypischen Eigenschaften im Sprachgebrauch stärker am schwachen Muster festgehalten wird als bei weniger prototypischen Substantiven. Das gilt insbesondere für die Mehrsilber mit finalem Schwa, für die sich zumindest in KoGra kaum divergente Formen beobachten lassen.

Zudem leistet der Prototypenansatz auch eine Erklärung dafür, welche starken Maskulina sich der schwachen Flexion anschließen. In unserer Untersuchung zeigt sich dies insofern, als unter den starken Maskulina, die sich überhaupt schwach belegen lassen, solche mit Eigenschaften, die prototypisch für schwache Maskulina sind, deutlich überrepräsentiert sind. In phonotaktischer Hinsicht sind dies insbesondere die *or*-Substantive sowie eine Handvoll anderer Substantive mit morphologisch gesteuerter Betonung, die zumindest in ihrer *-(e)n*-Form die für schwache

Maskulina typische Endbetonung aufweisen. Daneben sind es Substantive, die bereits in unflektierter Form endbetont werden. In semantischer Hinsicht sind es vor allem Substantive mit menschlichem Denotat. **Innerhalb** dieser starken Maskulina, die sich überhaupt mindestens einmal schwach belegen lassen, neigen in unserem Datensatz die Substantive mit fester Endbetonung stärker zur schwachen Flexion als die Substantive mit morphologisch gesteuerter Betonung. In semantischer Hinsicht neigen in unserem Datensatz insbesondere diejenigen Substantive bzw. Substantivverwendungen zur schwachen Flexion, deren Referenten wir als belebt aber nicht-menschlich kategorisiert haben.

Für die Schwankung von schwach zu stark konnten wir bestätigen, dass sie vor allem im Akkusativ und Dativ auftritt: Endungslose Akkusativ- und Dativformen sind deutlich wahrscheinlicher als Genitivformen auf *-(e)s*. In den nicht-stratifizierten Zufallsstichproben ließen sich letztere nur marginal belegen – und zwar nur für Substantive mit kodifiziertem Nebeneinander von schwacher und starker Flexion. Für die Schwankung von stark zu schwach hingegen finden wir für die hier berücksichtigten Substantive keinen Effekt. Als Erklärung dafür haben wir vorgeschlagen, dass der Genitiv als hoch- und schriftsprachlicher Kasus grundsätzlich in stärkerem Maße als Akkusativ und Dativ zu einer Norm-konformen Realisierung neigt. Bei den schwachen Maskulina verstärkt dieser soziolinguistische Faktor damit die Tendenz des Genitivs zur schwachen Flexion, wie sie auch aus strukturalistischer Sicht zu erwarten ist. Bei den starken Substantiven hingegen wirkt der soziolinguistische Faktor der aus strukturalistischer Sicht erwartbaren Tendenz entgegen.

Für den Faktor Apposition konnte nur ein schwacher Effekt nachgewiesen werden, und das auch nur bei den schwachen Maskulina. Für den Faktor Adjektiv konnten wir zumindest in den gemischten logistischen Regressionsanalysen keinen Effekt feststellen.

Für die – informelleren – Webforen hat sich gezeigt, dass sie nicht nur bei den als schwach kodifizierten Substantiven stärker zur Norm-divergenten Variante tendieren als KoGra, sondern auch bei den als stark kodifizierten. Die Tendenz der Webforen hin zur jeweils Norm-divergenten Variante gilt damit nicht nur dort, wo dies zur Wahl des Flexiv-ärmeren Musters führt (schwache Maskulina), sondern auch dort, wo es dadurch zu einem Aufbau von Kasusflexion kommt (starke Maskulina).

Für den Genitiv der schwachen Maskulina bleibt abzusehen, inwiefern langfristig mit einer bedeutenden Zunahme der starken Form zu rechnen ist. Es steht zu vermuten, dass der Flexionsklassenwechsel hier von einer anderen Entwicklungstendenz des deutschen Kasussystems überholt wird (oder bereits überholt worden ist): nämlich dem Rückgang des Genitivs als Kasus-kategorie insgesamt und seiner Refunktionalisierung als Prestigemarker. Wie z. B. bei Seiler (2022: 43–46) referiert, ist die Geschichte des Genitivs im Deutschen geprägt durch den Rückzug auf den adnominalen Bereich. In den allermeisten Dialekten ist der Genitiv auch aus dieser Domäne und damit gänzlich verschwunden (Fleischer & Schallert 2011: 84–87).

Statt des Genitivs (*der Hut des Lehrers*) begegnet dort die *von*-Periphrase (*der Hut vom Lehrer*) und/oder der possessive Dativ/Obliquus (*dem Lehrer sein Hut*). Die *von*-Periphrase konkurriert mit dem adnominalen Genitiv auch standardsprachlich, und wie Kopf/Bildhauer (angenommen) zeigen, neigen informellere Register (vertreten durch dieselben Webforen, wie sie auch hier untersucht wurden) unter ansonsten vergleichbaren Bedingungen in stärkerem Maße zu dieser Variante als formellere Register (wie auch hier vertreten durch KoGra). Zu diesem Befund passt die Beobachtung, dass in unseren einfachen, nicht-stratifizierten Stichproben für die schwachen Maskulina genitivische Belege in den Webforen seltener vorkommen als in KoGra: Während in der KoGra-Stichprobe 33,5 % der Belege (2.428 von 7.243) auf den Genitiv entfallen, sind es in der Webforen-Stichprobe nur 20,4 % (1.576 von 7.720). Dass der Genitiv im Standarddeutschen – im Gegensatz zur großen Mehrheit der Nicht-Standard-Varietäten – überhaupt erhalten ist, erklärt Seiler (2022: 58–59) mit der Refunktionalisierung dieses Kasus als Prestigeform (siehe Abschnitt 4.3.2). Diese Konstellation scheint wenig Raum zu lassen für den Flexionsklassenwechsel im Genitiv: Wo Sprecherinnen und Sprecher diesen Kasus überhaupt wählen, haben sie sich damit für eine schriftsprachliche Prestigeform entschieden, für die dann konsequenterweise auch eine normgerechte, also schwache, Realisierung zu erwarten ist. Wo, umgekehrt, die schriftsprachliche Norm eine geringere Rolle spielt, haben Sprecherinnen und Sprecher die Möglichkeit, den Genitiv gar nicht zu verwenden, statt ihn Norm-divergent zu realisieren.

In Bezug auf die Opposition zwischen einer Klasse schwacher Maskulina und einer Klasse starker Maskulina hat der Rückgang des Genitivs allerdings einen ähnlichen Effekt wie die Angleichung der schwachen Genitivmarkierung an die starke. Durch die Nicht-Verwendung ausgerechnet desjenigen Kasus, in dem die stark/schwach-Unterscheidung noch am stabilsten ist, wird im Singular eine Neutralisierung der Flexionsklassenunterscheidung möglich. Das Ergebnis ist ein einziges „Paradigma“ aus einer einzigen – der unflektierten – Form, die in allen drei verbliebenen Kasus (Nominativ, Akkusativ, Dativ) verwendet werden kann (*der/den/dem Tag/Mensch*).

7 Anhang

7.1 Suchanfragen für die Token-basierten Untersuchungen

Die Token-Stichproben, auf denen die Analysen in den Abschnitten 4.2, 4.4, 5.1 und 5.3 basieren, wurden erhoben, indem für jedes Lemma der Lemmaliste (vgl. Abschnitt 3.1) 28 Suchanfragen formuliert wurden. Diese ergeben sich aus der Kombination der Merkmale Artikel, Kasus, Adjektiv und Form des Substantivs. Die genauen Suchanfragen werden in Tabelle 6 am Beispiel des Lemmas

Tabelle 6: Suchanfragen.

1	[word="einen"&tag="ART"][word="MENSCH"]
2	[word="einen"&tag="ART"][word="MENSCHen" word="MENSCHn"]
3	[word="einen"&tag="ART"][tag="ADJA"][word="MENSCH"]
4	[word="einen"&tag="ART"][tag="ADJA"][word="MENSCHen" word=" MENSCHn"]
5	[word="einem"&tag="ART"][word="MENSCH"]
6	[word="einem"&tag="ART"][word="MENSCHen" word="MENSCHn"]
7	[word="einem"&tag="ART"][tag="ADJA"][word="MENSCH"]
8	[word="einem"&tag="ART"][tag="ADJA"][word="MENSCHen" word="MENSCHn"]
9	[word="eines"&tag="ART"][word="MENSCHes" word="MENSCHs"]
10	[word="eines"&tag="ART"][word="MENSCHens" word="MENSCHns"]
11	[word="eines"&tag="ART"][word="MENSCHen" word="MENSCHn"]
12	[word="eines"&tag="ART"][tag="ADJA"][word="MENSCHes" word="MENSCHs"]
13	[word="eines"&tag="ART"][tag="ADJA"][word="MENSCHens" word="MENSCHns"]
14	[word="eines"&tag="ART"][tag="ADJA"][word="MENSCHen" word="MENSCHn"]
15	[word="den"&tag="ART"][word="MENSCH"]
16	[word!="ab"&word!="aus"&word!="außer"&word!="bei"&word!="dank"&word!="entgegen"&word!="entsprechend"&word!="gemäß"&word!="gegenüber"&word!="laut"&word!="mit"&word!="mitsamt"&word!="nach"&word!="nächst"&word!="nahe"&word!="nebst"&word!="per"&word!="pro"&word!="samt"&word!="seit"&word!="von"&word!="zu"] [word="den"&tag="ART"][word="MENSCHen" word="MENSCHn"]
17	[word="den"&tag="ART"][tag="ADJA"][word="MENSCH"]
18	[word!="ab"&word!="aus"&word!="außer"&word!="bei"&word!="dank"&word!="entgegen"&word!="entsprechend"&word!="gemäß"&word!="gegenüber"&word!="laut"&word!="mit"&word!="mitsamt"&word!="nach"&word!="nächst"&word!="nahe"&word!="nebst"&word!="per"&word!="pro"&word!="samt"&word!="seit"&word!="von"&word!="zu"] [word="den"&tag="ART"][tag="ADJA"][word="MENSCHen" word="MENSCHn"]
19	[(word="dem"&tag="ART") (tag="APPRART")][word="MENSCH"]
20	[(word="dem"&tag="ART") (tag="APPRART")][word="MENSCHen" word="MENSCHn"]
21	[(word="dem"&tag="ART") (tag="APPRART")][tag="ADJA"][word="MENSCH"]
22	[(word="dem"&tag="ART") (tag="APPRART")][tag="ADJA"][word="MENSCHen" word="MENSCHn"]
23	[word="des"&tag="ART"][word="MENSCHes" word="MENSCHs"]
24	[word="des"&tag="ART"][word="MENSCHens" word="MENSCHns"]
25	[word="des"&tag="ART"][word="MENSCHen" word="MENSCHn"]
26	[word="des"&tag="ART"][tag="ADJA"][word="MENSCHes" word="MENSCHs"]
27	[word="des"&tag="ART"][tag="ADJA"][word="MENSCHens" word="MENSCHns"]
28	[word="des"&tag="ART"][tag="ADJA"][word="MENSCHen" word="MENSCHn"]

Mensch aufgeführt. In den Suchanfragen, die auf schwach flektierte Formen nach dem Definitartikel im Akkusativ abzielen (z. B. *den (netten) Menschen*, vgl. Anfragen 16 und 18), wurden Treffer unmittelbar nach einer Reihe Dativ-regierender Präpositionen ausgeschlossen (z. B. *mit den (netten) Menschen*). Auf diese Weise sollte die Anzahl von Fehltreffern (nämlich Dativ-Plural-Treffern) verringert werden.

7.2 Vergleichsstichprobe

Um bestimmen zu können, in welchem Maße die in den Abschnitten 4 und 5 diskutierten phonotaktischen und semantischen Merkmale typisch oder untypisch für die schwachen Maskulina bzw. die schwach belegbaren starken Maskulina unserer Lemmaliste sind, haben wir eine Vergleichsstichprobe von 1.000 maskulinen Substantiven erhoben, die sich unmittelbar nach *einen/einem* **nicht** schwach flektiert belegen lassen. Gesucht wurde im 1. Achtel des KoGra-Untersuchungskorpus. Die Suchanfrage lautete: 1:[word="einen"|word="einem"] 2:[tag="NN" & toprule="^[^c].*" & lemma!="unknown" & morph = "masc"]. Die Anfrage ergab 1.037.539 Treffer, verteilt auf 15.081 Lemmata. Aus diesen Lemmata wurden dann zufällig 1.000 valide Lemmata ausgewählt, die **nicht** in unserer Lemmaliste schwach belegbarer Substantive enthalten sind. Ausgeschlossen wurden (analog zum Vorgehen bei der Ermittlung der schwach belegbaren Substantive) Komposita (einschließlich Bildungen mit betonter Partikel, z. B. *Ausschluss*).

Die Tabellen 7 und 8 zeigen die Verteilung der Substantive der Vergleichsstichprobe auf die untersuchten phonotaktischen und semantischen Merkmale.

Tabelle 7: Substantive der Vergleichsstichprobe nach Phonotaktik.

Phonotaktik	n	%
Nicht-Endbetonung (<i>Ärmel, Motzer</i>)	682	68,2
Einsilbigkeit (<i>Hengst, Pakt</i>)	188	18,8
Endbetonung (<i>Aspekt, Genuss</i>)	103	10,3
morphologisch gesteuerte Betonung (<i>Aggressor, Tumor</i>)	24	2,4
finale Schwa (<i>Käse, Cicerone</i>)	2	0,2
Variable Betonung (<i>Kaffee</i>)	1	0,1
gesamt	1.000	100

Tabelle 8: Substantive der Vergleichsstichprobe nach Semantik.

Semantik	n	%
unbelebtes Denotat (<i>Kanton, Lack</i>)	584	58,4
menschliches Denotat (<i>Rüpel, Choliker</i>)	352	35,2
belebtes aber nicht menschliches Denotat (<i>Eber, Ganter</i>)	49	4,9
als polysem gewertet (<i>Spender</i>)	15	1,5
gesamt	1.000	100

7.3 Statistik

Die Parameter der statistischen Modelle wurden in R (R Core Team 2022) mit dem lme4-Paket (Bates et al. 2015) geschätzt. Für die beiden gemischten Regressionsmodelle (Substantive mit kodifizierter schwacher Flexion und Substantive mit kodifizierter starker Flexion) ergibt sich jeweils folgende Modellspezifikation²⁰:

$$P(y_i = \text{stark}) = \text{logit}^{-1}(\alpha_{j[i]} + \beta_0 + \beta_1 \text{Semantik}_i + \beta_2 \text{Korpus}_i + \beta_3 \text{Kasus}_i \\ + \beta_4 \text{Phonotaktik}_i + \beta_5 \text{Häufigkeitsklasse}_i + \beta_6 \text{Adj_vor_N}_i \\ + \beta_7 \text{Apposition}_i)$$

mit

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

Durch die verwendete Link-Funktion ist die Wahrscheinlichkeit, im Fall i starke Flexion vorzufinden, nicht linear von den Prädiktorvariablen abhängig. Eine lineare Abhängigkeit besteht auf der Ebene der Logits, d. h. nur auf dieser Ebene lässt sich der Einfluss eines Prädiktors unabhängig von den Werten aller anderen Prädiktoren bemessen. Auch das geschätzte Chancenverhältnis (*odds ratio*, e^β) lässt sich unabhängig von den Werten der übrigen Variablen im Modell interpretieren. Über den Einfluss eines Prädiktors auf die Wahrscheinlichkeit lassen sich genaue Aussagen nur machen, wenn die Ausprägungen aller anderen Prädiktoren berücksichtigt werden. Das Vorzeichen und der Betrag der Koeffizientenschätzer können jedoch wie folgt interpretiert werden: Werden alle anderen Variablenwerte konstant gehalten, dann steigt bei einem Koeffizienten > 0 die vorhergesagte Wahrscheinlichkeit dafür, dass das Substantiv stark flektiert wird, und bei einem

20 Die Formel ist eine Kurzschreibweise und gilt für beide Modelle. Die Notation ist hier abgekürzt, für einen Faktor mit n Ausprägungen werden $n-1$ Koeffizienten geschätzt. Die aufgeführten Faktoren weisen in den beiden gemischten Modellen unterschiedliche Ausprägungen auf. Die Zusammenfassungen in Tabelle 9 und Tabelle 11 machen dies deutlich.

Tabelle 9: Parameterschätzungen für das gemischte Regressionsmodell für Substantive mit kodifizierter schwacher Flexion.

Variable	(Ausprägung)	$\hat{\beta}$	Std.-Fehler	z	p
Intercept		-8,5	0,82	-10,32	< 0,001
Semantik	ani	1,57	0,48	3,27	0,001
	ina	4,39	0,32	13,73	< 0,001
Korpus	DECOW	1,33	0,09	14,12	< 0,001
Kasus	Akkusativ	1,94	0,16	11,81	< 0,001
	Dativ	2,28	0,16	14,2	< 0,001
Phonotaktik	mono	4,3	0,56	7,64	< 0,001
	polynult	3,88	0,74	5,22	< 0,001
	polyult	1,82	0,33	5,56	< 0,001
Häufigkeitsklasse		0,07	0,05	1,46	0,15
Adj_vor_N	nein	0,11	0,1	1,1	0,27
Apposition	ja	0,47	0,14	3,44	< 0,001

Marginales R^2 (nur feste Effekte) = 0,53; konditionales R^2 (gesamtes Modell) = 0,78

negativen Koeffizienten nimmt sie ab. Tabelle 9 (für Substantive mit kodifizierter schwacher Flexion) und 11 (für Substantive mit kodifizierter starker Flexion) zeigen die Koeffizientenschätzungen für die gemischten Modelle. Die Abkürzungen für die Werte der Variablen Phonotaktik und Semantik übernehmen wir größtenteils von Schäfer (2019), siehe Abschnitt 3.4.1 für eine Erläuterung. „morph_kond“ (Tab. 11) steht für Substantive mit morphologisch gesteuerter Betonung.

Wir berichten die vom Modell erklärte Varianz, approximiert durch das im MuMIn-Paket (Bartoń 2019) implementierte Pseudo- R^2 von Nakagawa et al. (2017) für gemischte Modelle. Dieses gibt zum einen die Varianz an, die durch die festen Effekte erklärt wird (marginales $R^2 = R^2_m$), und zum anderen die Varianz, die durch das gesamte Modell einschließlich der zufälligen Effekte erklärt wird (konditionales $R^2 = R^2_c$).

Um unsere Ergebnisse besser mit den Befunden von Schäfer (2019) vergleichen zu können, haben wir für die Substantive mit kodifizierter schwacher Flexion ein einfaches logistisches Regressionsmodell (vgl. Jaeger 2008) ohne Zufallsfaktor mit der folgenden Modellspezifikation berechnet²¹:

$$\begin{aligned}
 P(y_i = \text{stark}) = & \text{logit}^{-1}(\beta_0 + \beta_1 \text{Semantik}_i + \beta_2 \text{Korpus}_i + \beta_3 \text{Kasus}_i \\
 & + \beta_4 \text{Phonotaktik}_i + \beta_5 \text{Häufigkeitsklasse}_i + \beta_6 \text{Adj_vor_N}_i \\
 & + \beta_7 \text{Apposition}_i)
 \end{aligned}$$

21 Das Modell wurde in R (R Core Team 2022) mit dem stats-Paket geschätzt. Die Notation ist auch hier abgekürzt, für einen Faktor mit n Ausprägungen werden n-1 Koeffizienten geschätzt. Die Zusammenfassung in Tabelle 10 macht dies deutlich.

Tabelle 10 zeigt die Koeffizientenschätzungen für dieses Modell.

Tabelle 10: Parameterschätzungen für das einfache logistische Regressionsmodell für Substantive mit kodifizierter schwacher Flexion.

Variable	(Ausprägung)	$\hat{\beta}$	Std.-Fehler	z	p
Intercept		-0,99	0,28	-36	< 0,001
Semantik	ani	1,3	0,12	10,52	< 0,001
	ina	2,24	0,09	23,84	< 0,001
Korpus	DECOW	1,08	0,07	14,78	< 0,001
Kasus	Akkusativ	1,33	0,11	11,97	< 0,001
	Dativ	1,69	0,12	16	< 0,001
Phonotaktik	mono	3,92	0,15	26,34	< 0,001
	polynult	3,81	0,19	21,38	< 0,001
	polyult	1,59	0,14	11,13	< 0,001
Häufigkeitsklasse		0,29	0,01	20,56	< 0,001
Adj_vor_N	nein	0,28	0,09	3,08	0,002
Apposition	ja	0,64	0,09	6,96	< 0,001

Tabelle 11: Parameterschätzungen für das gemischte Regressionsmodell für Substantive mit kodifizierter starker Flexion.

Variable	(Ausprägung)	$\hat{\beta}$	Std.-Fehler	z	p
Intercept		1,23	2,39	0,52	0,61
Korpus	DECOW	-0,67	0,09	-7,6	< 0,001
Semantik	ani	-1,6	0,64	-2,48	0,01
	ina	-0,17	0,71	-0,24	0,81
Kasus	Akkusativ	-0,001	0,1	-0,01	0,09
	Dativ	0,18	0,1	1,8	0,07
Phonotaktik	mono	2,09	1,27	1,65	0,1
	polynult	1,83	1,38	1,32	0,19
	morph_kond	3,37	0,92	3,67	< 0,001
Häufigkeitsklasse		-0,03	0,13	-0,21	0,83
Adj_vor_N	nein	0,17	0,1	1,72	0,09
Apposition	ja	-0,04	0,15	-0,27	0,78

Marginales R^2 (nur feste Effekte) = 0,17; konditionales R^2 (gesamtes Modell) = 0,66

ORCID®

Thilo Weber  <https://orcid.org/0000-0002-0024-9733>

Sandra Hansen  <https://orcid.org/0000-0003-4243-0220>

Literatur

- Augustin, Hagen. 2012. Autor, Doktor, Friede, Funke. Problemfälle der Flexion. In Marek Konopka & Roman Schneider (Hgg.), *Grammatische Stolpersteine digital – Festschrift für Bruno Strecker zum 65. Geburtstag*, 71–82. Mannheim: Institut für deutsche Sprache.
- Bartoń, Kamil. 2019. MuMIn: Multi-Model Inference. R package version 1.43.6.
- Bates, Douglas, Martin Mächler, Ben Bolker & Stev Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Berg, Kristian. 2013. *Morphosyntax nominaler Einheiten im Niederdeutschen*. Heidelberg: Winter.
- Bubenhofer, Noah, Marek Konopka & Roman Schneider. 2014. *Präliminarien einer Korpusgrammatik*. Tübingen: Narr.
- Bubenhofer, Noah, Sandra Hansen-Morath & Marek Konopka. 2014. Korpusbasierte Exploration der Variation der nominalen Genitivmarkierung. *Zeitschrift für germanistische Linguistik* 42(3). 379–419.
- Dal, Ingerid & Hans-Werner Eroms. 2014. *Kurze deutsche Syntax auf historischer Grundlage*. Berlin, Boston: De Gruyter.
- Dudengrammatik 2016 = Wöllstein, Angelika & Dudenredaktion (Hgg.). *Die Grammatik: Unentbehrlich für richtiges Deutsch* (Duden 4), 9. Aufl. Berlin: Dudenverlag.
- Fleischer, Jürg & Oliver Schallert. 2011. *Historische Syntax des Deutschen. Eine Einführung*. Tübingen: Narr.
- Harbert, Wayne. 2007. *The Germanic Languages*. Cambridge: Cambridge University Press.
- Jaeger, Florian T. 2008. Categorical Data analysis: Away From ANOVAs (Transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language* 59(4). 434–446.
- Köpcke, Klaus-Michael. 1995. Die Klassifikation der schwachen Maskulina in der deutschen Gegenwartssprache – ein Beispiel für die Leistungsfähigkeit der Prototypentheorie. *Zeitschrift für Sprachwissenschaft* 14(2), 159–180.
- Kopf, Kristin & Felix Bildhauer. Angenommen. The genitive alternation in German. *Corpus Linguistics and Linguistic Theory*.
- Konopka, Marek. 2012. *Dem Manne kann geholfen werden – Wann kommt das Dativ-e zum Einsatz?* In Marek Konopka & Roman Schneider (Hgg.), *Grammatische Stolpersteine digital – Festschrift für Bruno Strecker zum 65. Geburtstag*, 115–124. Mannheim: Institut für deutsche Sprache.
- Konopka, Marek & Eric Fuß. 2016. Genitiv im Korpus. Untersuchungen zur starken Flexion des Nomens im Deutschen. Tübingen: Narr.
- Krischke, Wolfgang. 2012. Des Menschens Genitive. Normabweichende Genitiv-Varianten bei schwachen Maskulina. *Linguistik Online* 53(3). 55–84.

- Kupietz, Marc, Harald Längen, Paweł Kamocki & Andreas Witt. 2018. The German Reference Corpus DeReKo: New Developments – New Opportunities. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis & Takenobu Tokunaga (Hgg.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 4353–4360. Miyazaki: European Language Resources Association (ELRA). <https://www.aclweb.org/anthology/L18-1689.pdf> (abgerufen am 15.12.2023).
- Kusová, Jana. 2013. Zur morphologischen Variation und Schwankung der Substantivgruppe auf *-or*. In Anja Edith Ference & Libuše Spáčilová (Hgg.), *Deutsch als Sprache der (Geistes)Wissenschaften. Linguistik*. Brno: Tribun EU, 49–59.
- Kusová, Jana. 2014. *Morphologische Varianz der peripheren Substantivparadigmen im geschriebenen Gegenwartsdeutsch. Schwache Maskulina, starke Feminina und gemischte Substantive*. Wien: Praesens.
- Müller, Gereon. 2002. Syntaktisch determinierter Kasuswegfall in der deutschen NP. *Linguistische Berichte* 189. 89–114.
- Nakagawa, Shinichi, Paul C. D. Johnson & Holger Schielzeth. 2017. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface* 134(14). 1–11.
- Nübling, Damaris. 2008. Was tun mit Flexionsklassen? Deklinationsklassen und ihr Wandel im Deutschen und seinen Dialekten. *Zeitschrift für Dialektologie und Linguistik* 75(3). 282–330.
- Paulfranz, Alexandra. 2013. *Kasusmarkierungen der Gegenwartssprache in deutschen Lokal- und Regionaltageszeitungen*. Bamberg: University of Bamberg Press.
- Perkuhn, Rainer, Holger Keibel & Marc Kupietz. 2012. *Korpuslinguistik*. Paderborn: Fink.
- Pickl, Simon. 2020. Polarization and the emergence of a written marker. A diachronic corpus study of the adnominal genitive in German. *Journal of Germanic Linguistics* 32. 145–182.
- Post, Rudolf. 1992. *Pfälzisch. Einführung in eine Sprachlandschaft*, 2. aktualisierte und erweiterte Aufl. Landau: Pfälzische Verlagsanstalt.
- R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> (abgerufen am 15.12.2023).
- Rohdenburg, Günter. 1988. Flexionsangleichung von Substantiven an attributive Adjektive und verwandte Erscheinungen im Nordniedersächsischen. In Heinrich Weber & Ryszard Zuber (Hgg.), *Linguistik Parisette. Akten des 22. linguistischen Kolloquiums, Paris 1987*. Tübingen: Niemeyer. 277–288.

- Rohdenburg, Günter. 1993. Aspekte der Auflösung des Kasussystems im niederdeutschen. In Werner Abraham & Josef Bayer (Hgg.), *Dialektsyntax*. Opladen: Westdeutscher Verlag. 213–229.
- Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson & Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8. 382–439.
- Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (Hgg.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 12)*, 486–493. Istanbul: European Language Resources Association (ELRA).
- Schäfer, Roland. 2019. Prototype-driven alternations: The case of German weak nouns. *Corpus Linguistics and Linguistic Theory* 15(2). 383–417.
- Schmitz, Ulrich. 1999. Ausfahrt waschen. Über den progressiven Untergang der Flexionsfähigkeit. *Sprache an der Jahrtausendwende* (Osnabrücker Beiträge zur Sprachtheorie 60), 135–182.
- Seiler, Guido. 2022. Wie viele Kasus hat das Deutsche? In Sarah Brommer, Kersten Sven Roth & Jürgen Spitzmüller (Hrsg.), *Brückenschläge. Linguistik an den Schnittstellen* (Tübinger Beiträge zur Linguistik), 39–64. Tübingen: Gunter Narr Verlag.
- Szczepaniak, Renata. 2014. Sprachwandel und sprachliche Unsicherheit: Der formale und funktionale Wandel des Genitivs seit dem Frühneuhochdeutschen. In Albrecht Plewnia & Andreas Witt (Hgg.), *Sprachverfall? Dynamik – Wandel – Variation*, 33–49. Berlin, Boston: De Gruyter.
- Thieroff, Rolf. 2000. *Morphosyntax nominaler Einheiten im Deutschen*. Habilitationsschrift. Bonn.
- Thieroff, Rolf. 2003. Die Bedienung des Automaten durch den Mensch. Deklination der schwachen Maskulina als Zweifelsfall. *Linguistik Online* 16(4). 105–117.
- Thieroff, Rolf & Petra Vogel. 2009. *Flexion*. Heidelberg: Winter.
- Weber, Thilo. 2021. Determination in der Nominalphrase – ein Überblick. In Marek Konopka, Angelika Wöllstein & Ekkehard Felder (Hgg.), *Bausteine einer Korpusgrammatik des Deutschen*, Bd. 2, 13–62. Heidelberg: Heidelberg University Publishing.
- Wiese, Bernd. 2008. *Kasusdifferenzierung in der neuhochdeutschen Nominalgruppe*. Arbeitspapier Grammatik des Deutschen im europäischen Vergleich. Mannheim: Institut für Deutsche Sprache.
- Zweifelsfälle-Duden= Hennig, Mathilde (Hg.). 2021. *Dudenband 9 – Sprachliche Zweifelsfälle. Das Wörterbuch für richtiges und gutes Deutsch*, 9. Aufl. Berlin: Dudenverlag.