

Roland Schäfer und Ulrike Sayatz

# Die Kurzformen des Indefinitartikels im Deutschen

**Abstract:** We discuss cliticized indefinite articles in written German, both “regular” reduced forms (*ein* ‘a’ → *n*) and extended forms (*ein* → *nen*). Previous analyses of factors that favor written realizations of both variants are based on too sparse data for inferential statistics. We fill this gap, demonstrating that quantitative graphemics for non-standard writing is possible.

We use a large web corpus for our studies. By analyzing incoherent usage of cliticized and non-cliticized forms in conjuncts, we demonstrate that the two forms partially alternate randomly. We then show that the general loss of the genitive in informal writing does not explain why cliticized genitives do not occur at all. In two further studies, we apply logistic regression, finding that cliticization is disfavored in certain contexts (sentence-initially, after prepositions) and favored in others (when the non-cliticized form is bisyllabic), but that purely phonological factors (preceding vowel) are not significant. Finally, we show that full written cliticization (*num* ‘only a’) only occurs when the shape of the cliticized article by itself is insufficient as a well-formed graphemic word (like *n*).

Lastly, we argue that the phenomenon is at least partly a genuinely graphemic one, going beyond mere phonographical renderings of spoken language, and we discuss the results from a paradigm perspective.

**Keywords:** cliticization, grammaticalization, indefinite articles, corpus-based graphemics, non-standard writing, web corpora

DOI 10.1515/zfs-2014-0008

---

**Roland Schäfer**, Institut für Deutsche und Niederländische Philologie, Freie Universität Berlin, E-Mail: roland.schaefer@fu-berlin.de

**Ulrike Sayatz**, Institut für Deutsche und Niederländische Philologie, Freie Universität Berlin, E-Mail: sayatz@zedat.fu-berlin.de

# 1 Paradigmen und Analysen des deutschen Indefinitartikels

In diesem Artikel untersuchen wir die formalen und distributionellen Faktoren, welche die Klitisierungsschreibungen des deutschen Indefinitartikels auslösen (also Kurzformen wie *n* zu Vollformen wie *ein*), sowie die genauen morphologischen bzw. graphematischen Realisierungen dieser Schreibungen. In mehreren korpuslinguistischen Studien an einem großen Webkorpus des Deutschen zeigen wir damit erstmals innerhalb der Graphematik des Deutschen, dass sehr große Korpora zur Untersuchung von Gebrauchsschreibungen existieren, und dass graphematische Hypothesen an solchen großen Korpora mit inferenzstatistischen Methoden überprüft werden können.

In Abschnitt 1 führen wir zunächst in das Phänomen ein (Abschnitt 1.1), diskutieren dann die sich ausbildenden Paradigmen der Kurzformen sowie ihre systematische und historische Herleitung (Abschnitt 1.2). Abschnitt 1.3 widmet sich dem Status der Kurzformen (ob sie z. B. ein graphematisches Wort konstituieren) und der Integration ihrer Verschriftungen in das graphematische System des Deutschen. In Abschnitt 1.4 fassen wir die distributionellen Faktoren zusammen, die für die Klitisierungsschreibungen als relevant angesehen werden und leiten schließlich in Abschnitt 1.5 Hypothesen ab, die in Abschnitt 2 empirisch anhand von fünf Korpusstudien überprüft werden. In Abschnitt 3 diskutieren wir unsere Ergebnisse im Hinblick auf unsere Ausgangsfrage der distributionellen Bedingungen für die Etablierung eines neuen verschrifteten Kurzformparadigmas.

## 1.1 Der deutsche Indefinitartikel

Der Indefinitartikel weist im Deutschen neben seinen Vollformen gebrauchssprachliche Kurzformen bzw. Klitisierungsformen auf, die sowohl im gesprochenen Standard als auch im schriftlichen Nonstandard als Endstufe von Reduktionsprozessen aus den Vollformen abgeleitet werden können (Eisenberg 2013b: 190; Nübling 1992: 19; Nübling 2005: 112). Die folgenden Beispiele zeigen die erwartbaren Klitisierungsformen im Nominativ Neutrum (1a), im Akkusativ Maskulinum (1b) und im Dativ Femininum (1c).<sup>1</sup> In (2) finden sich Beispiele

---

<sup>1</sup> Alle Beispiele entstammen dem Webkorpus DECOW2012 (vgl. Abschnitt 2.1), innerhalb des sie über die URL identifiziert werden können. Das Belegdatum wird daher nicht angegeben. Die URLs finden sich im Anhang.

für erweiterte Formen (*nen* statt *n*) im Nominativ Neutrum (2a) und Nominativ Maskulinum (2b).

- (1) (a) *is doch nur n Spiel.*  
 (b) *Claudia hat nen Schäferhund ...*  
 (c) *Frag doch mal in ner Werkstatt nach, ob 125 Grad noch ok sind ...*
- (2) (a) *Ansonsten ists nen schönes Spiel :)*  
 (b) *War mal nen guter Tip vom Sattler.*

Im Fokus verschiedener Arbeiten der letzten Jahre stand die Frage, inwiefern der Indefinitartikel ein neues Kurzformparadigma ausbildet. Die Untersuchungen betrafen dabei zum einen spezifische Gebrauchskontexte wie Standard vs. Substandard (Volmert 1995; Schiering 2002; Ziegler 2011), gesprochene vs. geschriebene Sprache wie z. B. Chatsprache oder Presstexte (Topfink 2002; Burri 2003; Vogel 2006; Ziegler 2012), zum anderen funktionale Systemerweiterungen in Bezug auf *son* (Hole & Klumpp 2000; Lerner & Lohnstein 2005; Eroms 2008; von Heusinger 2012), Paradigmenausgleich in Bezug auf *nen* (Vogel 2006) sowie „Formerosion“ auf dem „Grammatikalisierungspfad vom Zahlwort *eins* zum Indefinitartikel“ (Szczepaniak 2009: 79).

Die Faktoren einer graphematischen Etablierung dieser Kurzformen blieben bislang insofern unbeachtet bzw. unbefriedigend evaluiert, als die jeweils erhobenen Daten entweder nur aus Korpora der gesprochenen Sprache gewonnen wurden, oder aber die zugrunde gelegten Chatkorpora aufgrund ihrer geringen Größe für inferenzstatistische Auswertungen zu klein sind, während die etablierten Korpora standardnaher Schriftsprache das Phänomen klitisierter Indefinitartikel nicht abbilden.<sup>2</sup>

Das primäre Ziel der vorliegenden Studie ist es, die formalen und distributionellen Einflüsse auf die Verschriftung der Kurzformen des Indefinitartikels im Verhältnis zu deren Vollformen auf der Basis des 9,1 Milliarden Tokens umfassenden Webkorpus DECOW2012 quantitativ zu untersuchen.<sup>3</sup> Dabei interessieren uns insbesondere die Faktoren, die die Konstituenz eines „einsilbige[n]

<sup>2</sup> Die bislang einzige quantitative Analyse zur Kurzform *nen* in Zeitungstexten von Ziegler (2012) auf der Basis mehrerer Zeitungskorpora des IDS (Kupietz & Keibel 2009) berücksichtigt keine weiteren Voll- und Kurzformen der Indefinitartikel. Die „Daten [würden] zugunsten der Explizitformen völlig verzerren“ (Ziegler 2012: 301). Im „Bereich standardnahen Sprechens“ sieht sie eine Entwicklung zur Dominanz der Kurzform im Vergleich zur Vollform (im Nominativ mit 78 %, im Akkusativ sogar mit 100 %). Belastbare quantitative Daten für verschriftete Kurzformen im Vergleich zu den Vollformen legt Ziegler nicht vor.

<sup>3</sup> Zur Korpusbeschreibung vgl. Abschnitt 2.1.

autonome[n] Wort[es]“ (Vogel 2006: 188) und damit den „Minimalumfang“ eines phonologischen und graphematischen Wortes sichern (Jacobs 2005; Fuhrhop 2008).

## 1.2 Formenbestand und Genese

In diesem Abschnitt diskutieren wir die beobachtbaren Kurzformen des Indefinitartikels (bzw. das sich daraus ergebende Paradigma) sowie Theorien zur Genese dieses Paradigmas. Die strukturell möglichen Kurzformen des Indefinitartikels sind zunächst in Tabelle 1 zusammengestellt.<sup>4</sup>

Die Reduktionen des Indefinitartikels werden als mehrstufige Prozesse im Sinne sogenannter Abschwächungsreihen (vgl. insbesondere Dedenbach 1987: 46 und Prinz 1991: 92 für das Standarddeutsche und Schiering 2002 für das Ruhrdeutsche) beschrieben. Im Ergebnis entstehen einsilbige Kurzformen, bei denen relativ zu ihren Vollformen drei Typen zu unterscheiden sind.

1. Die einsilbige Vollform *ein* [a<sup>h</sup>ɛn] wird in einem dreistufigen Reduktionsprozess über die Monophthongierung zu [an] mit anschließender Abschwächung zu Schwa [ən] und schließlich durch Schwa-Elision zu [n] (vgl. Dedenbach 1987: 47).<sup>5</sup>
2. Bei den zweisilbigen Vollformen der Maskulina und Neutra im Dativ (*einem* [a<sup>h</sup>ɛnəm]) und der Maskulina im Akkusativ (*einen* [a<sup>h</sup>ɛnən]) wird in einer vierten Reduktionsstufe [nən] zu [n̩] bzw. [nəm] zu [n̩m] mit silbischem Nasal im Auslaut. Als „Endstufe der Abschwächung“ sieht Dedenbach (1987: 48) dann nach Elision des ersten Nasals ein unsilbisches [n] bzw. [m], während Prinz (1991: 93) die Reduktionsformen [ŋ] und [ŋm] mit einem

**Tabelle 1:** Strukturell mögliche Kurzformen des deutschen Indefinitartikels

	Femininum	Maskulinum	Neutrum
Nominativ	<i>ne</i>	<i>n / nen</i>	<i>n / nen</i>
Akkusativ	<i>ne</i>	<i>n / nen</i>	<i>n / nen</i>
Dativ	<i>ner</i>	<i>nem</i>	<i>nem</i>
Genitiv	<i>ner</i>	<i>nes</i>	<i>nes</i>

<sup>4</sup> Eine weitere Kurzform des Akkusativ Maskulinums ist *ein*. Da diese sich formal und paradigmatisch anders verhält als die nasal anlautenden Kurzformen, lassen wir sie hier und im weiteren Verlauf außer Acht.

<sup>5</sup> IPA-Transkriptionen werden hier nach Krech et al. (2009) vereinheitlicht und dadurch ggf. abweichend von den zitierten Autoren wiedergegeben.

Sonoranten als Silbenkern damit begründet, dass „Klitika entweder ein Schwa oder einen vokalisiertes Sonoranten als Silbenkern aufweisen“. Welche der beiden Optionen für die jeweilige klitische Form gelte, ist nach Prinz vorhersagbar: Silbenfinale Sonoranten bilden immer den Silbenkern bei Klitika. Diese Formen seien standardisiert. In allen anderen Fällen wie bei den Feminina mit *ne* [nə] steht im Ableitungsschema nach Prinz (1991: 92) ein Schwa.

3. Für Formen wie *nen* im Akkusativ Maskulinum kommen nach Ziegler (2011: 76) „zwei Reduktionspfade“ infrage. Entweder werde die Reduktion nicht vollständig ausgeführt oder aber es komme nach der vollständigen Reduktion zu einer „phonetischen Auffüllung“ im Sinne von „Jespersen’s Cycle“.<sup>6</sup> Während Prinz (1991: 84) *nen*-Kurzformen generell (also auch im Akkusativ Maskulinum) dem Substandard zuordnet, wird bei Vogel (2006), Burri (2003), Tophinke (2002) und Ziegler (2011, 2012) *nen* nur im Nominativ Maskulinum, Nominativ Neutrum und Akkusativ Neutrum sowie im sogenannten Monokasus (*nen paar*) als neue, nicht standardkonforme „erweiterte“ Kurzform diskutiert, die nicht nur als gesprochene, sondern auch verschriftete Variante neben der Vollform sowie der „normalen“ Kurzform besonders in Texten spontansprachlicher Register (Chattexten) und Presstexten zu beobachten sei.

Kurzformen im Genitiv in allen drei Genera werden dagegen in den vorliegenden Untersuchungen weder im Standard noch im Substandard belegt. Burri (2003: 5) interpretiert dies als „insbesondere im Mündlichen fortschreitenden Genitiv-Abbau“, während Vogel (2006: 178) hier in Anlehnung an Prinz (1991: 112) die Genitivform als lexikalisch blockiert betrachtet.<sup>7</sup> Ziegler (2011: 76) führt *nes* im Paradigma der Maskulina und Neutra als „Aussprachevarianten“ auf, ohne dies jedoch mit Daten zu belegen. Auch für klitisiertes *son* wird eine Genitivform strukturell angenommen, aber ebenfalls empirisch nicht belegt (von Heusinger 2012: 30). Zur marginalen Verankerung genitivischer Kurzformen vgl. die Ergebnisse in den Abschnitten 2.2 und 2.4, zu *sones* besonders Fußnote 23.

<sup>6</sup> Vgl. dazu auch Vogel (2006: 189).

<sup>7</sup> „Möglicherweise lässt sich [die Beschränkung des verkürzten Genitivs auf präpositionsregierte Kontexte] besser damit begründen, dass der Genitiv für die Markierung der Attribution

### 1.3 Klitisierungsschreibungen im graphematischen System des Deutschen

Zunächst ist zu klären, was ganz allgemein der kategoriale Status von Klitika ist. Im zweiten Schritt ist die Etablierung ihrer Verschriftung im graphematischen System des Deutschen zu bewerten.

Die kategoriale Bestimmung von Klitika liegt nach Nübling (1992: 11) auf einer „Skala zwischen den Polen Wort [...] und Flexiv [...]“, wobei die Klitika mehrheitlich „ihre syntaktische und phonologische Selbstständigkeit eingebüßt haben“, aber noch nicht als Flexive grammatikalisiert seien.<sup>8</sup> Der kategoriale Status indefiniter Kurzformen interessiert in der vorliegenden Studie vorrangig hinsichtlich der Frage, inwiefern diese sowohl auf phonologischer als auch graphematischer Ebene mit anderen sprachlichen Einheiten interagieren, und zwar unter der allgemeinen Annahme der Präferenz einer Enklise.<sup>9</sup> Dabei ist zunächst davon auszugehen, dass in fokusmarkierten Positionen die klitischen Kurzformen generell auszuschließen sind (Prinz 1991: 140; Topfink 2002: 173). Eine funktionale Begründung für die gegenüber der Proklise präferierte Enklise sieht Nübling (2005) bei Reduktionsformen des definiten wie indefiniten Artikels darin, dass zwischen Artikel und Substantiv in der Nominalphrase noch weitere (attributive) Einheiten stehen können.

Im Vergleich zum definiten Artikel ist der Indefinitartikel nach Nübling (2005: 119) „bezüglich jeder seiner Einzelformen in der gesprochenen Sprache gleichermaßen verschmelzungsfreudig, wobei der Diphthong getilgt wird und der nasal anlautende, silbische Rest an die Präposition agglutiniert, d.h., die Integrität der Präposition bleibt immer gewahrt (*in + eine* → *in'ne*, *auf + einem* → *auf'nem*, *mit + einer* → *mit'ner* usw.)“. Dieses und die geringe und nicht nach Flexionsformen gestaffelte Grammatikalisierung des klitisierten Indefinitartikels unterscheiden ihn vom definiten Artikel. Dafür gebe es dann auch Klitisierungen wie *mit'ner*, die beim definiten Artikel mit Präposition nicht vorkämen. Dedenbach (1987: 95) analysiert dagegen Formen wie *durch'n* als synkretistische Verschmelzungen von Präposition mit bestimmtem oder unbestimmtem Artikel.

Wenn wir uns nun der graphematischen Etablierung der Kurzformen des Indefinitartikels zuwenden, so ist diese (wie bei allen Klitika) grundsätzlich vor

---

so wichtig ist, dass nur die Vollform diese Funktion adäquat erfüllen kann. In der Präpositionalphrase liegt das Hauptgewicht jedoch auf der Präposition selbst, so dass hier prinzipiell auch die Kurzform stehen kann.“ (Vogel 2006: 178).

<sup>8</sup> Zur Diskussion vgl. Nübling (1992: 36) und Prinz (1991: 98).

<sup>9</sup> Proklise tritt bei Artikeln möglicherweise nur als Allegroverbindung auf (Nübling 1992).

dem Hintergrund der Schreibprinzipien des Deutschen zu sehen. Das Deutsche als tief verschriftete Sprache bildet mit seinen die phonographischen Schreibungen überlagernden silbischen, morphologischen und syntaktischen Schreibprinzipien lautliche Reduktionsprozesse im Standard zunächst nicht graphematisch ab (Eisenberg 2013a: 309). Die graphematische Etablierung von Klitika in standardnaher Schriftsprache ist vielmehr an einen hohen Grammatikalisierungsgrad gebunden. Für den definiten Artikel heißt das nach Nübling (1992: 304), dass nur die grammatikalisierten speziellen klitischen Verbindungen wie *im*, *zum*, *zur* in der Standardschrift etabliert sind, Allegroformen und einfache Klitika wie *durchs*, *aufs*, *ums* dagegen „vom Parameter der sog. konzeptionellen Mündlichkeit“ eines Textes abhängen (Nübling 1992: 305). Dabei bilden Schreibungen mit Apostroph wie *auf's*, *mit'm*, *für'n* keine prototypischen graphematischen Wörter (Fuhrhop 2008) und signalisieren nach Nübling (1992: 308) „einen geringen Grammatikalisierungsgrad und damit eine geringere Akzeptanz der betreffenden Verschmelzung“. Der Apostroph kennzeichnet nach Bredel (2008, 2011) einen irreversiblen Defekt, d.h., er zeigt an, dass eine spezifische Wortinformation fehlt, die nicht in der Textumgebung aufgefunden werden kann. Bredel (2008: 103) postuliert eine sogenannte „Monofunktion“ des Apostrophs als „graphisches Reparaturzeichen für den Leseprozess“. Dem Leser wird demnach durch den Apostroph signalisiert, dass „eine für den Dekodierprozess erforderliche Wortheigenschaft nicht durch Buchstaben ausgedrückt ist“ (Bredel 2011: 42). Wichtig ist nach Bredel (2008: 43) dabei, dass die fehlende Wortinformation nicht allein auf ihr „zugrunde liegendes Vorhandensein“ bezogen wird. Als graphematisches Zeichen kann der Apostroph phonologische, morphologische oder auch syntaktische Informationen kodieren. Die traditionellen Subtypen des Elisions- und Stammformapostrophs werden so unter einer einheitlichen Funktion gefasst.<sup>10</sup>

Als grundlegender Befund lässt sich nun feststellen, dass die graphematischen Realisierungsvarianten der Kurzformen des Indefinitartikels von enklitischer Schreibung mit und ohne Apostroph (3) bis zur autonomen Spatienschreibung mit und ohne Apostroph (4) reichen.

- (3) (a) *Michael – der auch gerade darüber nachdenkt seine PWM-Schaltung rauszuschmeissen und so'nen Relais an das jetzt gefundene D+ anzuschließen.*  
 (b) *Dort gibt's 3 Tische undne Theke und mit 20 Leuten ist der Laden voll.*

<sup>10</sup> Dass beide Funktionen nicht immer genau getrennt werden können, zeigt sich gerade bei Klitisierungen. Bei *auf's*, *auf'm* oder *für'n* markiert der Apostroph sowohl fehlendes graphisches Material (möglicherweise unter Rekurs auf fehlendes phonologisches) als auch eine Morphemgrenze. Zur Diskussion vgl. Nübling (1992: 307) und Bredel (2011: 41, 2008: 104).

- (4) (a) *Der Typ ist ein Panzer auf zwei Beinen – mit ’nem Schuß wie ein Stier.*  
 (b) *Ich mein jetzt wenn er zuhause schon meistens nicht richtig in der Abteilung läuft, dann klappts auf nem Tunier ja sowieso nicht!*

Diese Daten widersprechen Prinz (1991), der annimmt, dass Kurzformen „immer mit vorangehendem Apostroph versehen und niemals mit dem vorangehenden oder folgenden Wort zusammengeschrieben“ würden, also *was für’n Ding*, aber *\*was fürn Ding* (Prinz 1991: 114). Ein Fehlen von Spatien wertet Prinz als „typisch für im Lexikon affigierete Wörter“ (Prinz 1991: 101). Die graphematische Realisierung der Form „vom Partner durch ein Apostroph getrennt“ erhebt er zur Regel für den reduzierten Indefinitartikel als ungebundene lexikalische Einheit.

Tophinke (2002: 173) belegt in ihrer Untersuchung bis auf die Ausnahme *fürn* nur autonome Verschriftungen der Kurzformen (*ne, nen, n, en, ner, nem*). Ziegler (2012: 312) interpretiert Apostrophschreibungen bei *nen* als explizite Kennzeichnung eines „Schreiben[s] gegen die Regel“ und postuliert hier ein „Spannungsfeld“ zwischen Mündlichkeit und Schriftlichkeit.<sup>11</sup> Burri (2003: 7) belegt im Gegensatz zu Prinz (1991) in ihrer Untersuchung zu Chattertexten sowohl enklitische Kurzformen an die Partikel *so* und Formen wie *dasn, isn, nurn, nochn* als auch nicht standardkonforme Verschriftungen von klitisierten definiten Artikeln mit Präpositionen wie *aufm, aufn, aufer* (Burri 2003: 10) und vermutet hier eine Analogie zu entsprechenden Formen beim definiten Artikel wie *im, zur*, die sich zum Teil auch in der Standardsprache durchgesetzt haben.

Für die Frage nach dem graphematischen (Wort-)Status der Kurzformen des Indefinitartikels sowie ihrer graphematischen Interaktion mit anderen Einheiten spielen also folgende Aspekte eine Rolle:

1. graphematische Enklise als Zusammenschreibung wie in *son, fürn*.
2. Apostrophschreibung wie *so’n, für’n*
3. autonome Schreibung mit Spatium *so n, für n*.<sup>12</sup>

<sup>11</sup> Unsere Daten (vgl. Abschnitt 2.6) zeigen jedoch, dass gerade bei *nen* die Apostrophschreibung dispräferiert ist. Der Apostroph markiert also keinen (orthographischen) Regelverstoß an sich. Er ist schließlich Bestandteil des (standardsprachlichen) Interpunktionssystems des Deutschen und so in der Standardschrift etabliert. Seine Funktion geht grundsätzlich über die Markierung phonologischer Reduktionen hinaus. Er zeigt auch Silben- oder Morphemgrenzen an und referiert insofern nicht auf Mündlichkeit, wie von Ziegler behauptet, sondern auf Schriftlichkeit. Weiter dazu die Diskussion in Abschnitt 3.

<sup>12</sup> Wir diskutieren den Unterschied zwischen Apostrophschreibungen mit und ohne Spatium nicht, da die Aufbereitung des von uns verwendeten Korpus diesen Unterschied ohnehin nicht abbildet. In der Korpusaufbereitung werden Spatien grundsätzlich als (mögliche) Wort- bzw. Tokengrenze interpretiert und anschließend gelöscht. Vgl. z. B. Schäfer & Bildhauer (2013: Kap. 4).

Nach Jacobs (2005: 99) ist Zusammenschreibung von Ausdrücken immer dann gefordert, wenn sie nicht „den Minimalumfang“ eines phonologischen Wortes erreichen.<sup>13</sup> Jacobs sieht darin die Erklärung, dass „Klitika, auch wenn sie keine morphologisch gebildeten Ausdrücke“ seien, „in vielen Schriftsystemen generell nicht durch Spatien (sondern allenfalls durch Apostrophen oder Ähnliches) abgetrennt werden“ (Jacobs 2005: 99). Danach müssten Schreibungen mit Apostroph einer Schreibung mit Spatium vorgezogen werden. Die erweiterte Kurzform *nen* könnte dann im Sinne einer „lautliche[n] Auffüllung“ nach Vogel (2006: 188) sowohl auf phonologischer als auch graphematischer Ebene das Erreichen eines Minimalumfangs im Sinne von Jacobs (2005) sichern. Dieser graphematische Minimalumfang wird auch durch Zusammenschreibungen wie bei *son* erfüllt. Zu beachten ist, dass über die Form *n* hinaus auch einsilbige graphische Formen wie *ne* nach Fuhrhop (2008: 198) nicht den Umfang eines graphematischen Worts erreichen. Dieses besteht minimal aus einer graphischen Vollsilbe (Konsonantgraphem plus Vollvokalgraphem). Das Graphem <e> bedürfte eines Längenzeichens, um hier nicht als Schwa-Graphem interpretiert zu werden (vgl. *Reh*). Die einzigen deutschen Wörter von ähnlicher graphischer Form wie *ne* sind *je* und *Re*, in denen aber irregulär <e> als Vollvokalgraphem steht (vgl. weiter vor allem Studie 5 in Abschnitt 2.7).

Angesichts der allgemein angenommenen Standardferne des Phänomens liegt nun die Hypothese auf der Hand, dass eine Verschriftung von Kurzformen der Indefinitartikel nur in Texten solcher Register auftritt, die zumindest partielle Abweichung vom Standard tolerieren. Die Formen werden dementsprechend im Allgemeinen (und im Besonderen mit Bezug auf die erweiterte Kurzform *nen* im Neutrum und Nominativ Maskulinum) als „Mündlichkeitsmarker“ (Ziegler 2011, 2012) bzw. „orate Charakterisierung“ (Topfink 2002: 171) eines „neuen schriftsprachlichen Registers“ (Weingarten 1997, zitiert nach Burri 2003: 3) quasi-spontaner Schriftlichkeit, aber auch als spezielle Form der Redewiedergabe in Presstexten (Ziegler 2012) bewertet, wobei sich die Annahme einer varietätenspezifischen Auszeichnung insbesondere von *nen* nicht belegen lässt (vgl. Vogel 2006: 183). Dabei wird die Nähe zur gesprochenen Sprache ebenso wie die „Manifestation von Umgangssprache“ (Burri 2003: 4) neben anderem auch an den Besonderheiten der verschrifteten klitisierten Indefinitartikel festgemacht. Diese Verschriftungen stehen damit neben anderen phonographischen Schreibungen, welche Reduktionen und Assimilationen der gesprochenen Sprache graphisch manifestieren und damit nicht standardkonform

---

<sup>13</sup> Die Definition des phonologischen Wortes wird in Jacobs (2005: 22) gegeben und bezieht sich im Wesentlichen auf Ausdrücke, deren phonologische Repräsentation eine hierarchisch wohlgeformte Domänenstruktur aufweist, die einen Wortknoten enthält.

abbilden (Burri 2003; Tophinke 2002; Schuster & Tophinke 2012). Dazu gehören z. B. Apokopierung des <t> nach Frikativ wie *is*, Verschriftung von Schwa-Elisionen und Assimilationen der Reduktionssilbe wie *holm*, Enklise oder Ausfall nachgestellter Personalpronomen der zweiten und dritten Person wie *kannste* usw.

Abschließend sei darauf verwiesen, dass das Verhältnis von Nichtstandard- und Standardverschriftungen auch hinsichtlich der Frage diskutiert wird, inwieweit sich diese systemisch entfalten und damit tendenziell Sprachwandel bewirken (Haase et al. 1997; Naumann 1998; Burri 2003; Schuster & Tophinke 2012). So vermutet Nübling (1992: 307) eine „Brückenfunktion“ von konzeptionell eher gesprochenen Textsorten derart, dass die Verschriftung gesprochener Klitika ein Weg in die Hochsprache wäre. Aussagen dazu bedürfen (wie von Nübling gefordert) einer genaueren statistisch fundierten Untersuchung, die bislang jedoch für keines der beschriebenen Phänomene nicht standardkonformer Verschriftungen vorliegt. Es ist allerdings davon auszugehen, dass keine (vor allem diachron) ausreichend differenzierten Korpora existieren, die entsprechende Untersuchungen erlauben würden.

#### 1.4 Distributionelle Faktoren: Kurzform vs. Vollform sowie *n* vs. *nen*

Es wurde bereits gesagt, dass die Verschriftung der Kurzformen des Indefinitartikels auf bestimmte Positionen und Register beschränkt ist. Zusätzlich kommen als distributionelle Faktoren für die Verschriftung von Kurzformen sowie das Auftreten der „erweiterten“ Kurzform (*nen* als Nominativ Maskulinum und Nominativ bzw. Akkusativ Neutrum) folgende Faktoren infrage:

1. Affinität der Kurzform oder der Erweiterung zu bestimmten Wortarten im Kontext,
2. lautliche bzw. graphische Umgebungen (z. B. Hiatusvermeidung),
3. topologische Aspekte, z. B. satzinitiale Positionen, Position in der Präpositionalphrase,
4. spezifische Kasus-Genus-Kombinationen,

*Ad 1:* Die Verteilung der Kurzformen und Vollformen in den vorliegenden qualitativen Analysen lässt nicht auf eine Kontextsensitivität hinsichtlich bestimmter Wortarten schließen. Nach Vogel (2006: 182) ist bei den von ihr untersuchten 75 Substantiv-Lemmata „keine eindeutige lautliche, morphologische oder semantische Spezifizierung“ erkennbar.

*Ad 2:* Hinsichtlich der lautlichen Umgebungen wird für den Gegensatz von *nen* und *n* die Frage diskutiert, ob Hiatusvermeidung oder andere kombinatorische Beschränkungen eine Rolle spielen. Burri (2003: 7) sieht *nen* als Variante von *n* nach vorausgehendem vokalischem oder nasal auslautendem Wort (*dann nen jahr*). Vogel (2006: 185) kann hingegen die Hypothese einer Hiatusvermeidung durch *nen* nicht bestätigen. Gerade auch bei graphematisch markiertem vokalischem Auslaut (und nicht nur bei phonetischem wie bei *aber* [a:bə]) kommt es eher zu einer Enklise (wie bei *son*) als zu einer Erweiterung auf *nen*.

*Ad 3:* Unter topologischem Aspekt sprechen die wenigen Daten von Vogel (nur sechs Fälle im Korpus) für eine Präferenz von *nen* am Satzanfang. In Bezug auf die Alternation von *nen* und *n* im Nominativ Maskulinum diskutiert Vogel auch eine Verteilung auf die syntaktischen Funktionen und sieht dort eine Präferenz für *nen* im thematischen Subjekt. Dies spräche gegen die Annahme einer Übergeneralisierung der *nen*-Form des Akkusativ Maskulinum auf den Nominativ. Vogel nimmt in diesem Fall eher untypische syntaktische Kontexte wie rhematisches Subjekt und Prädikativ an (Vogel 2006: 184). In Bezug auf verschriebene Vorkommen von *nen* vertritt Ziegler (2012: 301) die These einer textpragmatisch begründeten Verwendung im Fließtext „in der Funktion eines Autorenkommentars“, während die Verwendung in Schlagzeilen als „Zitat wörtlicher Rede“ (Ziegler 2012: 301) nachrangig sei.

*Ad 4:* In den vorliegenden Studien zeigt sich eine distributionelle Präferenz der Kurzformen gegenüber den Vollformen zunächst bei den Feminina in allen Kasus bis auf den Genitiv. Vogel (2006: 185) ermittelt ein Verhältnis von 1:2 für *eine* vs. *ne* sowie von 2:3 für *einer* (nur Dativ) vs. *ner*. Bei den Maskulina und Neutra ist eine Präferenz der Kurzform gegenüber der Vollform deutlich an die Zweisilbigkeit der Vollform gebunden.<sup>14</sup> Diese Korrelation wird nach Vogel vor allem daraus ersichtlich, dass im Nominativ Maskulinum trotz belegter Varianten der Kurzformen *n* und *nen* die Vollform *ein* in einem Verhältnis von 3:1 gegenüber den Kurzformen dominiere. Vogel schlussfolgert daraus eine generelle Dominanz einsilbiger Formen im Paradigma des Indefinitartikels. Die „falschen“ Kurzformen *nen* im Nominativ Maskulinum sowie bei den Neutra im Nominativ und Akkusativ seien dann als Sicherung eines autonomen funktionalen Wortes zu interpretieren, in dessen Paradigma Einsilbigkeit dominiere. Die „gegenwartssprachliche Epenthese von *n*“ (Vogel 2006: 189) verhindere da-

---

<sup>14</sup> Das wird auch an den Daten von Tophinke (2002: 173) deutlich, die bei *eine* und *einen* mehrheitlich Reduktionsschreibungen feststellt, allerdings ohne auf die Zweisilbigkeit Bezug zu nehmen. Tophinke interpretiert die Reduktionsschreibungen vielmehr als Realisierungen lautlicher Klitika im Kontrast zu „orthographischen Schreibungen“ solcher Fälle, „in denen in der gesprochenen Sprache keine Klitisierung vorliegt“.

mit eine Enklise und vereinheitliche das Kurzformparadigma auch silbenstrukturell durch nasalen Anlaut. Da diese Ergebnisse anhand einer relativ kleinen Datenmenge gewonnen wurden, stellt sich die Frage, ob sie an großen Datenmengen repliziert bzw. verfeinert werden können. In Abschnitt 2 beheben wir dieses Desideratum.

## 1.5 Eingrenzung der Fragestellung und Hypothesen

Keiner der in den existierenden Untersuchungen diskutierten morphophonologischen und distributionellen Faktoren scheint allein stehend eine hinreichende Bedingung zur Erklärung der Verteilung von Kurzformen und Vollformen darzustellen. Dasselbe gilt für die Faktoren, die gegebenenfalls die Verwendung von *n* vs. *nen* steuern. Unsere Korpusstudien haben daher ein multifaktorielles Modell zum Ziel, das verschiedene formale und distributionelle Faktoren auf phonologischer, morphologischer und graphematischer Ebene berücksichtigt, die lediglich als Tendenzen interpretiert werden. Im Rest dieses Abschnitts soll der per Hypothese angenommene Einfluss dieser Variablen prägnant formuliert und theoretisch motiviert werden.

1. Für die Verteilung von Kurzform vs. Vollform ist eine Präferenz einsilbiger gegenüber zweisilbiger Formen zu vermuten, so dass im ganzen Paradigma der Tendenz zur Einsilbigkeit nachgegeben wird. Hypothese: Die Kurzform wird in den Genus-Kasus-Positionen des Paradigmas präferiert, in denen die Vollform zweisilbig ist (also in allen außer Nominativ Maskulinum sowie Nominativ und Akkusativ Neutrum).
2. Hinsichtlich der Genus-Kasus-Kombinationen wurden in den vorliegenden Untersuchungen Kurzformen des Indefinitartikels in genitivischen Nominalphrasen nur marginal oder gar nicht belegt. Im Sinne der Grammatik kommen als Erklärung allgemeiner Genitivabbau im betreffenden Register oder die geringe Salienz bzw. starke Obliquität des Genitivs infrage. Hypothese: Die Kurzformen des Genitivs sind zwar strukturell möglich, aber faktisch blockiert.
3. Sowohl für die prinzipielle Wahl der Kurzform gegenüber der Vollform als auch für die Alternation von *n* vs. *nen* wurde Hiatusvermeidung als möglicher Auslöser diskutiert. Hypothese: Vorgehender vokalischer Auslaut und/oder nachfolgender vokalischer Anlaut korrelieren mit der Wahl der Kurzform im Allgemeinen und/oder mit der Wahl von *n* vs. *nen*.
4. Außerdem wird untersucht, ob die satzinitiale Position einen Einfluss auf das Auftreten der Kurzform hat, da in satzinitialer Position Enklise blockiert ist und wir übereinstimmend mit der älteren Literatur bei der Kurzformen-

- bildung von einem präferierten Prozess der Enklise ausgehen. Hypothese: In satzinitialer Position ist die Kurzform dispräferiert.
5. Unabhängig von der Vokalizität des vorangehenden Auslauts und der Satzinitialität könnten diverse mehr oder wenig explizit in der Literatur genannte morphosyntaktische Faktoren die Wahl der Kurzform gegenüber der Vollform beeinflussen, wobei wir uns hier auf zwei spezifische Kontexte (vs. alle anderen) beschränken. Hypothese: Bestimmte syntaktische Wörter links vom Indefinitartikel (Präpositionen, *so* oder alle anderen) beeinflussen die Präferenz/Dispräferenz für die Kurzformen des Indefinitartikels.
  6. Weiterhin geht es neben der Verteilung von Vollform und Kurzform auch um die Variation von *n* vs. *nen*. Wir gehen davon aus, dass die *nen*-Kurzform graphematisch im Unterschied zur *n*-Kurzform das autonome Wort sichert, da sie die Minimalanforderungen an ein graphematisches Wort erfüllt, eine wohlgeformte graphematische Silbe mit einem Vokalgraphem zu enthalten (vgl. Fuhrhop 2008: 197).<sup>15</sup> Es ist also zu erwarten, dass die Tendenz zur Auffüllung der regulären Kurzform *n* zu *nen* rein formal gesteuert wird, nicht etwa funktional (also mit einer Präferenz in bestimmten Numeri oder Kasus). Hypothese: Unabhängig von der zu erwartenden höheren Auftretenshäufigkeit der Form *nen* im Akkusativ Maskulinum ist die Auftretenshäufigkeit von *nen* im Nominativ Maskulinum sowie im Nominativ und Akkusativ Neutrum prinzipiell gleich.
  7. Als weitere Determinante für die autonome Wortschreibung wird schließlich der links stehende Apostroph untersucht. Es wird angenommen, dass Apostrophschreibung und graphematischer Wortstatus korrelieren. Hypothese: Das Auftreten der erweiterten Form *nen* (statt *n*) korreliert positiv mit der Schreibung ohne Apostroph.

In Abschnitt 2.5 und 2.6 werden diese Hypothesen empirisch in zwei Korpusstudien überprüft. Zur Einordnung des Phänomens führen wir in den Abschnitten 2.3 und 2.4 zunächst zwei Auxiliarstudien zur Registerspezifik und zum Fehlen der Genitiv-Formen durch. Abschnitt 2.7 widmet sich schließlich in einer kleineren Studie den ansonsten nicht berücksichtigten Zusammenschreibungen des Indefinitartikels mit dem links stehenden Wort.

## 2 Korpusstudien

In diesem Abschnitt wird zunächst die Wahl des Korpus für die Studien begründet (Abschnitt 2.1), bevor in Abschnitt 2.2 eine allgemeine deskriptive Quantifi-

---

<sup>15</sup> Zur graphematischen Silbe vgl. weiterhin Fuhrhop & Buchmann (2009).

zierung der Verteilung der Kurzformen erfolgt. In den Abschnitten 2.3–2.7 werden fünf Korpusstudien zur Überprüfung der in Abschnitt 1.5 aufgestellten Hypothesen vorgestellt.

## 2.1 Wahl des Korpus

Ein für die gegebene Fragestellung geeignetes Korpus muss einerseits Texte solcher Register enthalten, in denen überhaupt Kurzformen der Indefinitartikel als Spontanschreibungen (und nicht als Wiedergabe der wörtlichen Rede) verschriftlicht sind. Andererseits muss das Korpus groß genug sein, um ein niederfrequentes Phänomen zumindest so zahlreich zu repräsentieren, dass inferenzstatistische Methoden zur Anwendung kommen können. Die etablierten standardnahen Korpora des Deutschen erfüllen diese Anforderungen nicht. Sowohl das DeReKo (z. B. Kupietz & Keibel 2009) als auch das DWDS-Kernkorpus (Geyken 2007) enthalten ausschließlich standardnahe Schriftsprache. Das DWDS-Kernkorpus ist zudem mit 122,8 Millionen Tokens vergleichsweise klein. Da das Ziel unserer Untersuchungen auch bzw. besonders ein graphematisches ist, kommen (ohnein sehr kleine) Korpora gesprochener Sprache nicht infrage, da es sich bei ihnen notwendigerweise um Transkripte handelt. Verschriftlichung von nicht standardkonformer gesprochener Sprache findet sich allerdings in Texten, die primär quasi-spontane Kommunikation aus Internet-Foren, Blogs usw. enthalten. Diese Texte sind in großer Zahl frei im Internet – und damit in Webkorpora – verfügbar.<sup>16,17</sup>

Das deutsche Korpus DECOW2012 enthält 9,1 Milliarden Tokens in 7,8 Millionen Dokumenten, die im Jahr 2011 von Seiten aus der nationalen Domäne *.de* heruntergeladen wurden (Schäfer & Bildhauer 2012 und ähnlich auch Baroni et al. 2009). Die Genre- und Registerstratifizierung wurde in Schäfer & Bildhauer (2012) stichprobenhaft nach einer Taxonomie erfasst, die grundlegend in Sharoff (2006) beschrieben wurde. Für den gegebenen Anlass ist insbesondere die Klassifikation nach *Mode* (umgangssprachlich für „Register“) interessant, innerhalb derer die Kategorie *Quasi-Spontan* als der bereits erwähnte Typ von vergleichsweise spontansprachlichen und weitgehend nicht redigierten Texten aus Foren, Blog-Diskussionen usw. definiert ist. Es ist davon auszugehen, dass

---

**16** Über die Stratifizierung von Webkorpora vgl. weiter Schäfer & Bildhauer (2013: Kap. 5) und Biemann et al. (2013: 46).

**17** Ergebnisse von Suchmaschinen sind hingegen aus konzeptuellen und technischen Gründen keine Alternative zu gezielt für linguistische Zwecke konstruierten Webkorpora. Für eine Diskussion dazu vgl. Kilgarriff (2006) und Schäfer & Bildhauer (2013: Kap. 1).

quasi-spontane Texte Merkmale der gesprochenen Sprache enthalten, aber nicht mit dieser gleichzusetzen sind. Die Kategorie *Blogmix* ist definiert als eine Mischung aus nicht spontanen und quasi-spontanen Anteilen (z. B. Blog-Einträge mit Kommentaren). Schäfer & Bildhauer (2012: 492) ermitteln für das gesamte Korpus DECOW2012 einen Anteil von 22,5 % an Quasi-Spontan-Dokumenten und einen Anteil von 4,5 % an Blogmix-Dokumenten. Insgesamt besteht das Korpus also zu geschätzt 27 % aus Dokumenten, die potentiell grammatische und graphematische Variation jenseits des Standards enthalten (ca. 1,95 Millionen Dokumente). Für unsere Überzeugung, dass nur ein Korpus von der Art des DECOW2012 für die gegebene Untersuchung infrage kommt, sprechen die in den folgenden Abschnitten vorgelegten Ergebnisse.

## 2.2 Allgemeine Verteilung der Kurzformen und Vollformen im Korpus

Zunächst geben wir in diesem Abschnitt eine Schätzung der absoluten Häufigkeiten der einzelnen Formen. Dies ist zielführend, um die allgemeine Produktivität der Klitisierung des Indefinitartikels im Korpus abschätzen zu können.

Wir haben nach Vorkommen von konkreten Kurz- und Vollformen des Indefinitartikels gesucht, optional gefolgt von Adjektiven und obligatorisch gefolgt von einem Substantiv. Insgesamt finden sich so 130.112.190 Belege, davon 126.575.593 mit Vollformen und 3.536.597 mit Kurzformen. Es ergibt sich die geschätzte bereinigte Verteilung in Tabelle 2. Der Schätzcharakter der Zahlen rührt daher, dass die Konkordanzen neben erwünschten Treffern auch Fehlbelege enthalten. Der Anteil der Fehlbelege wurde mit hoher Konfidenz an je einer Stichprobe von 1.000 Belegen (pro Zeile in Tabelle 2) geschätzt. Es ist zu bemerken, dass trotz der absolut gesehen hohen Zahl von 3.142.482 Kurzformen ihr Anteil an allen Vorkommen des Indefinitartikels lediglich 2,51 % ausmacht.

Augenfällig sind bereits deskriptiv einige Ungleichverteilungen. Z.B. fallen die in gleicher Weise in der Vollform und der Kurzform für den Nominativ und Akkusativ synkretistischen femininen Formen *eine* und *ne* auf. Bei den Vollformen macht *eine* nur 17,6 % aus, aber bei den Kurzformen fallen auf *ne* 35,6 % der Fälle, also ziemlich genau doppelt so viele. Sehr stark scheint die Tendenz zur Vermeidung der Kurzform *nes* im Genitiv Singular des Maskulinums zu sein. Die Belege, in denen diese Form vorkommt, sind zwar ansonsten einwandfrei (5), und die Form muss damit als eine zumindest marginal im System verankerte angesehen werden, die absoluten Frequenzen sind aber auch im DECOW2012 so gering, dass der Genitiv aus der inferenzstatistischen Untersuchung ausgeschlossen werden muss (s. Abschnitt 2.5).

**Tabelle 2:** Anzahlen der Belege für Artikel-Substantiv und Artikel-Adjektiv-Substantiv Kombinationen im gesamten DECOW2012 (Adjektive mit der angegebenen Endung). Die Quote der Fehlbelege wurde an Stichproben mit  $n = 1.000$  geschätzt, und die Zähl­daten wurden gemäß dieser Schätzung korrigiert. Die Konkordanzen für Kombinationen mit *nes* sind exhaustiv korrigiert.

Vollformen			Kurzformen		
NP-Form	Geschätzte Anzahl in DECOW2012	%-Anteil an Vollformen in DECOW2012	NP-Form	Geschätzte Anzahl in DECOW2012	%-Anteil an Kurzformen in DECOW2012
<i>ein</i> N	22.727.202	18,6	<i>n</i> N	106.805	3,4
<i>ein</i> ADJ-er N	5.091.204	4,2	<i>n</i> ADJ-en N	11.110	0,4
<i>ein</i> ADJ-es N	7.573.257	6,2	<i>n</i> ADJ-er N	12.338	0,4
<i>eine</i> N	21.527.079	17,6	<i>n</i> ADJ-es N	18.397	0,6
<i>eine</i> ADJ-e N	13.836.512	11,3	<i>ne</i> N	1.118.628	35,6
<i>einem</i> N	9.267.382	7,6	<i>ne</i> ADJ-e N	477.270	15,2
<i>einem</i> ADJ-en N	4.656.023	3,8	<i>nem</i> N	259.872	8,3
<i>einen</i> N	11.703.929	9,6	<i>nem</i> ADJ-en N	79.097	2,5
<i>einen</i> ADJ-en N	5.955.271	4,9	<i>nen</i> N	608.040	19,3
<i>einer</i> N	9.935.744	8,1	<i>nen</i> ADJ-en N	177.098	5,6
<i>einer</i> ADJ-en N	5.270.574	4,3	<i>nen</i> ADJ-er N	16.240	0,5
<i>eines</i> N	3.416.945	2,8	<i>nen</i> ADJ-es N	28.099	0,9
<i>eines</i> ADJ-en N	1.276.629	1,0	<i>ner</i> N	175.046	5,6
			<i>ner</i> ADJ-en N	54.418	1,7
			<i>nes</i> N	13	< 0,001
			<i>nes</i> ADJ-en N	11	< 0,001
Gesamt	122.237.751	100	Gesamt	3.142.482	100

- (5) (a) *Wen du dich während nes Schubs ans Bett fesselst bleiben die offenen Stellen mit sicherheit gröstenteils aus*  
 (b) *Das Zweite erinnert mich in punkto Fixieren daran, wie ich in der Schule einmal den Rat nes Lehrers annahm und mein Bild fixieren wollte ... und zum falschen Spray gegriffen hab ...*

Die Form *ner* im Femininum existiert zwar, aber im Wesentlichen auch nicht als Genitiv. Von den 1.000 Belegen für *ner* ohne Adjektiv sind zwei eindeutige attributive Genitive, drei weitere werden von *während* und *bezüglich* regiert und können daher auch Dative sein. In der Vergleichskonkordanz für *einer* finden sich dagegen 251 Genitive. Bei den 1.000 Belegen mit *ner* und Adjektiv sind es immerhin neun eindeutige attributive und verbregierte Genitive (6) und ein

Beleg mit der Präposition *innerhalb*. In der Vergleichskonkordanz für *einer* mit Adjektiv sind es mit 235 wieder wesentlich mehr Genitive. Eine Beschränkung der Kurzform auf präpositionsregierte Genitive (Vogel 2006: 178) kann jedoch nicht festgestellt werden, vgl. (5) und (6).

(6) *Naja, ich denke mal er ist ner ähnlichen Meinung wie ich ;-)*

Schließlich ist zu klären, ob es sich bei den Kurzformen der Gestalt *n* im Akkusativ Maskulinum tatsächlich um Indefinita (zu *einen*) handelt, da prinzipiell auch eine Interpretation als Definitum (zu *den*) denkbar ist. Dazu wurden die 227 einschlägigen Belege aus der Eintausender-Stichprobe für *n* ohne Adjektiv individuell durchgesehen. Für 217 Belege besteht kein Zweifel, dass es sich um Indefinita handelt, vgl. (7a). Zwei Beispiele sind eindeutig Definita, vgl. (7b) und (7c). Für acht weitere bestand seitens der Autoren Zweifel, wie sie zu lesen seien. Insgesamt wurden daher neun Linguistinnen und Linguisten nach ihrer Einschätzung der Lesart gefragt, wobei die gesamte Webseite als Kontext gegeben wurde.<sup>18</sup> Zu den Beispielen in (8) geben wir die Ergebnisse der Befragung in eckigen Klammern, jeweils die Anzahl der Antworten in der Form „[definit/indefinit/ambig]“.

- (7) (a) *kannst du mir nochmal n tipp geben?*  
 (b) *Ja, gut, aber die Kurse die dort angeboten werden, sind teilweise auch für 'n Hintern.*  
 (c) *Mittlerweile muss ich schon fast die Hälfte der Songs skippen, weil sie mir dermaßen auf n Sack gehen.*
- (8) (a) *ich bin dann mit dem auto von meiner mama auf n feldweg gefahren [...]* [1/4/4]  
 (b) *Wenn Monkey Island 5 genauso werden würde wie Teil 4 (was ich nach dem Erfolg von Sam & Max nicht hoffe), würde ich mir n Kauf zweimal überlegen.* [5/0/4]  
 (c) *und hat n durchmesser eines ess tellers* [8/0/1]  
 (d) *Und wenn gesagt wird Win 2003 denke ich immer zuerst an Windows SERVER 2003 und das is fuer n Desktop nur nervig/gefährlich.* [4/3/2]  
 (e) *Auf n kutter darf man die nicht anziehen weil man in einer warthose ertrinkt könnte wenn man über board geht ...* [2/5/2]  
 (f) *hab grad voll n Blackout ...* [5/3/1]

<sup>18</sup> Weil es sich im Kontext dieses Artikels um einen untergeordneten Punkt handelt, wurde keine Studie im eigentlichen Sinn durchgeführt.

- (g) *ich wollte zwar schon immer mal auf n prüfstand, aber hab's bisher aus geldgründen gelassen [6/0/3]*
- (h) *batze wird bejubelt wenn er rey auf n stuhl rammt xD? [4/2/3]*

Es herrscht weitgehende Uneinigkeit. Wir gehen davon aus, dass es sich bei der Form *n* fast immer um einen Indefinitartikel handelt, und dass sich nur wenige ambige Fälle und noch weniger eindeutig definite Fälle unter den Belegen befinden. Für die Stichproben zu Studie 3 und 4 (Abschnitte 2.5 und 2.6) wurde sichergestellt, dass nur eindeutige Indefinita enthalten sind.

### 2.3 Studie 1: Kurzformen als Alternation in quasi-spontaner Schriftsprache

Wir überprüfen nun zunächst unsere Annahme, dass es sich bei den verschrifteten Kurzformen um eine partiell unsystematische Alternation in quasi-spontaner Schriftlichkeit handelt. Für die in Studie 3 (Abschnitt 2.5) verwendete Stichprobe von 2.499 Kurzformen haben wir ermittelt, aus welchen Typen von Dokumenten sie stammen. Insgesamt stammen 2.263 aus Foren, 123 aus Blogs mit Kommentaren, 47 aus Blogs ohne Kommentare, 15 aus Amateurdichtung, neun aus Web-Gästebüchern, fünf aus transkribierten Interviews und zwei aus Webshops mit Kommentarfunktion.<sup>19</sup> Dass die Verschriftung der Kurzformen an ein bestimmtes Register gebunden ist, kann also sicher angenommen werden.

Wichtiger ist allerdings die Frage, wie konsequent im einschlägigen Register die Kurzformen verwendet werden. Besonders aufschlussreich ist hierzu die Betrachtung von Koordinationsstrukturen. Sollten merkmalsgleiche NPs in einer Koordinationsstruktur als Mischung von Vollformen und Kurzformen realisiert werden, wäre dies ein guter Indikator dafür, dass eine Alternation im eigentlichen Sinn des Wortes vorliegt, also eine nicht vollständig durch kontextuelle Merkmale motivierbare parallele Verfügbarkeit mehrerer grammatischer Optionen. Genau dies ist der Fall, wie (9) und (10) illustrieren. Alternationen mit Vollform gefolgt von Kurzform (9a), (10a) und umgekehrt (9b), (10b) treten in zusammengehörigen Konjunktionen (9) sowie Disjunktionen (10) auf, sowohl verbdependent (9) als auch präpositionsdependent (10).

---

<sup>19</sup> Für 35 Belege ließ sich der Dokumenttyp nicht mehr eindeutig bestimmen, da der Text alleine (wie im Korpus gespeichert) oft nicht eindeutig ist und die Webseite mit dem vollständigen Layout in diesen Fällen nicht mehr erreichbar waren. Die als Blog klassifizierten Dokumente schließen Vereinsseiten u. ä. ein, auf denen von nichtprofessionellen Schreiberinnen und Schreibern über Feste, Versammlungen usw. berichtet wird. Regionale oder überregionale Pressedokumente befanden sich nicht in der Stichprobe.

**Tabelle 3:** Frequenzen der Koordinationsstrukturen mit und oder oder mit Indefinitartikel im Dativ Singular des Maskulinums/Neutrums, verschiedene Kombinationen von Vollformen und Kurzformen

Koordinations-Konstruktion	absolute Frequenz	Anteil an Koordinations-Konstruktionen mit mindestens einer Kurzform
<i>einem N und/oder einem N</i>	33.119	
<i>einem N und/oder nem N</i>	110	11,71 %
<i>nem N und/oder einem N</i>	115	12,25 %
<i>nem N und/oder nem N</i>	714	76,04 %

**Tabelle 4:** Parallele Ergebnisse zu Tabelle 3 für den Akkusativ (marginal auch Nominativ) Singular des Maskulinums und des erweiterten *nen*-Neutrums.

Koordinations-Konstruktion	absolute Frequenz	Anteil an Koordinations-Konstruktionen mit mindestens einer Kurzform
<i>einen N und/oder einen N</i>	26.102	
<i>einen N und/oder nen N</i>	202	9,36 %
<i>nen N und/oder einen N</i>	189	8,75 %
<i>nen N und/oder nen N</i>	1.768	81,89 %

- (9) (a) *Phyliss Jo, hatte ich ja auch, bei mir wars eher [ein Rauschen und n Vibrieren] oder so ...*  
 (b) *auf der hinfahrt gibts [n quiz und ein tippspiel] wo man geile sachen gewinnen kann ...*
- (10) (a) *ich suche bei euch nach [einem Newsletter oder nem RSS-Fedd] für die Releas-List, gibt es so was?*  
 (b) *sollte ich mir welche zulegen dann nur von [nem Züchter oder einem Händler] mit Züchternachweis.*

Wir haben beispielhaft an separaten Stichproben für die Koordinationsstrukturen mit *und* oder *oder* im Akkusativ Singular des Maskulinums (und ggf. erweiterten *nen*-Neutrums) und im Dativ Singular des Maskulinums und Neutrums die Frequenzen im gesamten DECOW2012 für alle Permutationen von Voll- bzw. Kurzformen ermittelt. Beide Formen bieten sich wegen der weitgehenden Abwesenheit von Synkretismen besonders für eine automatische Erhebung an. Die Tabellen 3 und 4 zeigen die Ergebnisse.

Wenn auch eine konsistente Verwendung von Vollformen und Kurzformen in Koordinationsstrukturen gemäß Tabellen 3 und 4 bevorzugt wird, ist der

**Tabelle 5:** Ergebnisse zum allgemeinen Genitivschwund in quasi-spontanen Dokumenten.

	Stichprobengröße (Tokens)	Anzahl Genitiv- Vollformen	Genitiv-Vollformen pro 1.000 Tokens
Dokumente ohne Kurzformen	4.762.910	3.370	0,708
Dokumente mit Kurzformen	18.504.080	4.573	0,247

Anteil gemischter Konstruktionen doch beträchtlich. Eine Präferenz für eine bestimmte Abfolge von Vollform und Kurzform in Koordinationsstrukturen gibt es dabei nicht. Der Gebrauch der Kurzformen ist also als inkonsequent zu bezeichnen.

## 2.4 Studie 2: Genitivabbau

In den multifaktoriellen Studien in den Abschnitten 2.5 und 2.6 kann der Genitiv wegen seiner geringen Frequenz bei den Kurzformen nicht berücksichtigt werden. In diesem Abschnitt wird daher zunächst untersucht, ob der Genitivabbau in quasi-spontaner Schriftsprache eine hinreichende Erklärung für das gänzliche Fehlen der Kurzformen des Genitivs ist, wie es z. B. von Burri (2003: 5) vertreten wurde, oder ob tiefere strukturelle Gründe vorliegen, bezüglich derer bisher keine Hypothesen aufgestellt wurden.

Wir betrachten zunächst in einer Auxiliarstudie zwei weitere Stichproben mit je 5.000 Dokumenten aus dem gesamten DECOW2012, um anhand eines Vergleichs mit den Vollformen festzustellen, ob der Genitivabbau in Dokumenten, die Kurzformen des Indefinitartikels enthalten, tatsächlich weiter vorangeschritten ist. Die erste Stichprobe besteht aus Dokumenten, in denen mindestens drei Kurzformen des Indefinitartikels vorkommen, die zweite besteht aus Dokumenten, in denen gar keine Kurzform vorkommt. In beiden wurde die Vorkommenshäufigkeit von Vollformen des maskulinen/neutralen Genitivs in den Konstellationen *eines* N und *eines* ADJ-en N als Schätzung der Vorkommenshäufigkeit entsprechender Nominalphrasen ermittelt. Wir betrachten die Anzahl der Genitiv-NPs mit Vollformen des Indefinitartikels über die gesamten Stichproben relativ zur Anzahl der in ihnen enthaltenen Tokens. Es zeigt sich in Tabelle 5, dass in den Dokumenten mit Kurzformen tatsächlich der Genitivabbau stärker vorangeschritten ist.

Um nun zu ermitteln, ob die Kurzformen im Genitiv tatsächlich eine Frequenz signifikant unterhalb des im Rahmen des Genitivabbaus Erwartbaren haben, haben wir alle 478.551 Dokumente des DECOW2012, in denen mindes-

**Tabelle 6:** Genitivabbau im Maskulinum/Neutrum als nicht hinreichende Erklärung für das Fehlen der Genitive der Kurzformen in DECOW2012-Dokumenten, in denen insgesamt mindestens drei Kurzformen des Indefinitartikels vorkommen. Es wurde gesucht: die konkrete Form des Indefinitartikels, gefolgt von einem optionalen Adjektiv (mit einer im Paradigma erwartbaren Flexionsform), gefolgt von einem Substantiv. Für den Genitiv wurde zusätzlich gefordert, dass das Substantiv auf <s> endet. Fisher-Exakt:  $p < 0,001$  \*\*\*, Chancenverhältnis = 347,86.

	beobachtet		erwartet	
	Genitiv	nicht Genitiv	Genitiv	nicht Genitiv
Vollform	1.306.264	12.660.544	1.273.467	12.693.341
Kurzform	<b>107</b>	360.773	<b>32.904</b>	327.976

tens drei Kurzformen des Indefinitartikels vorkommen, als Subkorpus exportiert. Innerhalb dieser Dokumente haben wir die Frequenz der maskulinen und neutralen Vollformen und Kurzformen in der Kombination von Artikel mit oder ohne Adjektiv über (nicht händisch bereinigte) heuristische Suchen geschätzt. Wir wissen aus der in Abschnitt 2.2 beschriebenen Bereinigung des vollständigen DECOW2012-Datensatzes, dass solche Suchanfragen den Anteil der Genitiv-Kurzformen sogar eher überschätzen, weil der typischerweise enthaltene Anteil von verrauschten Ergebnissen angesichts der sehr geringen tatsächlichen Treffermengen (echte Genitive) eher ins Gewicht fällt als bei reich belegten Formen.<sup>20</sup> Tabelle 6 zeigt die Gegenüberstellung der Frequenzen für Vollform und Kurzform, jeweils im Genitiv und in allen anderen Kasus. Über die Randsummen kann für eine solche Tabelle die erwartete Häufigkeit für die einzelnen Felder ausgerechnet werden (vgl. z. B. Gries 2008: 170). Es ergibt sich für die Genitiv-Kurzformen eine Erwartung von 32.904 Belegen gegenüber 107 beobachteten. Ein Fisher-Exakt-Test zeigt, dass das Ergebnis mit  $p < 0,001$  signifikant ist. Das Chancenverhältnis ist mit 347,86 sehr hoch.<sup>21</sup> Der allgemeine Geni-

<sup>20</sup> Mit „verrauschten Belegen“ sind hier solche gemeint, die zwar durch die Suchanfrage gefunden werden, die aber nicht der gesuchten Struktur entsprechen. Solche Fehlbelege können z. B. durch Fehlschreibungen oder eine ungenaue bzw. fehlerhafte linguistische Annotation des Korpus zustande kommen. Bei den Vollformen waren bei der Suche nach *eines* plus Substantiv (ohne Adjektiv) 91 von 1.000 Belegen Fehlbelege, von den insgesamt 662 Treffern für *nes* waren 649 Fehlbelege, überwiegend in Form von Akronymen (z. B. das „Nintendo Entertainment System“). Das sehr eindeutige statistische Ergebnis wird dadurch nicht merklich verfälscht (s. auch Fußnote 22).

<sup>21</sup> Das gilt auch, wenn man mit fünftausend *nes*-Genitiven rechnet und die restlichen Felder konstant hält (Fisher-Exakt:  $p < 0,001$  \*\*\*, Chancenverhältnis = 7,44). Trotz der verrauschten Daten ist das Ergebnis (das überproportionale Fehlen der Genitiv-Kurzformen) so stark, dass es ohne händische Bereinigung der Stichproben als robust gelten kann.

**Tabelle 7:** Kodierschema für die händisch kodierte Stichprobe zu Studie 3. Der Doppelstrich trennt unabhängige von der abhängigen Variablen.

Variable	Ausprägungen	Erläuterung
Gen(us)	Mask(ulinum), Neut(rum), Fem(ininum)	Genus des NP-Kopfs
Kas(us)	Nom(inativ), Akk(usativ), Dat(iv), Gen(itiv)	Kasus der NP
Adj(ektiv)	1, 0	Die NP enthält ein Adjektiv (= 1).
Linksv(okal)	1, 0	Das Wort-Token links vom Artikel hat vokalischen Anlaut (= 1).
Rechtstv(okal)	1, 0	Das Wort-Token rechts vom Artikel hat vokalischen Auslaut (= 1).
Linksk(ontext)	Praep(osition), So, Int(er- punktion), And(eres)	Art des dem Artikelwort unmittelbar vorausgehenden Tokens.
Kurz(form)	1, 0	Der Artikel ist eine Kurzform (= 1).

tivabbau in den einschlägigen Dokumenten erklärt somit das Fehlen der genitivischen Kurzformen nicht.<sup>22</sup>

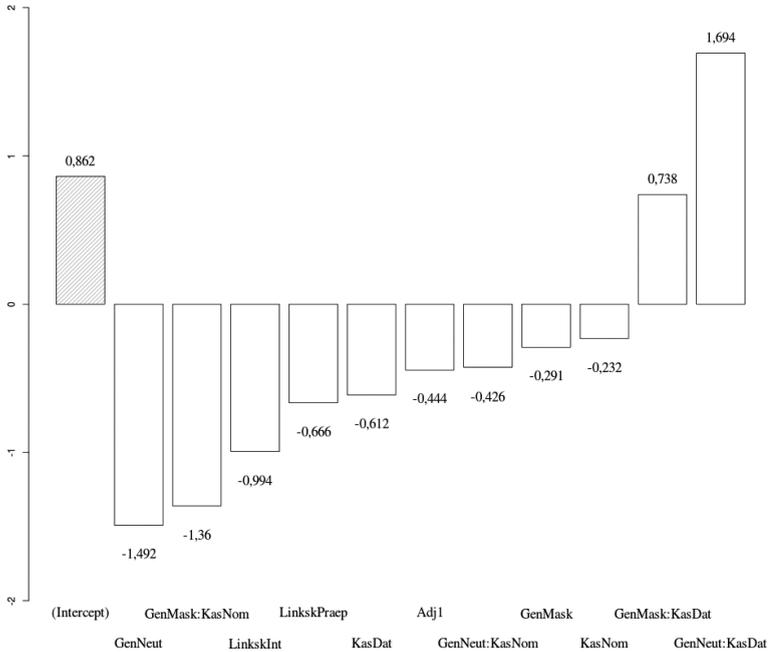
## 2.5 Studie 3: Vollform und Kurzform

Wollen wir ermitteln, was die Wahl der Kurzform (zunächst ohne Berücksichtigung der Alternation zwischen *n* und *nen*) positiv oder negativ beeinflusst, bietet sich ein multifaktorielles Modell an, da aus theoretischer Sicht keine Hypothese besteht, dass ein einzelner Faktor, wie z. B. die Einbettung in eine PP oder ein vokalischer Auslaut des vorangehenden Worts, alleine darüber entscheidet, ob die Kurzform geschrieben wird oder nicht. Vielmehr ist anzunehmen, dass viele Faktoren zusammenwirken, die jeweils eine Präferenz für oder gegen die Kurzform beisteuern (vgl. Abschnitt 1.5).

Für die Stichprobe haben wir genau 2.499 Kurzformen und 2.500 Vollformen zufällig extrahiert, und es wurden die Faktoren aus Tabelle 7 händisch für diese Belege kodiert.<sup>23</sup> Die Gleichverteilung bei der abhängigen Variable

<sup>22</sup> Interessanterweise ist die zusammengeschiedene Form *sones* ebenfalls nicht verfügbar. Von 137 Belegen für *Sones* bzw. *sones* im gesamten DECOW2012 bleibt nach Abzug der Eigennamen und Verschreibungen (überwiegend für *Sohnes*) keiner übrig, der einen Genitiv des emergenten Pronomens *son* darstellt.

<sup>23</sup> Die Merkmale entsprechen den Hypothesen aus Abschnitt 1.5. Lediglich die An- bzw. Abwesenheit eines Adjektivs in der NP wurde heuristisch als Variable hinzugefügt.



**Abbildung 1:** Graphische Darstellung der Koeffizienten des Modells gemäß Tabelle 8. Positive Koeffizienten zeigen eine Präferenz für die Kurzform.

(„Kurzform“) wurde nötig, da, wie in Abschnitt 2.2 dargelegt, eine Zufallsstichprobe aus allen Indefinitartikeln lediglich 2,51% Kurzformen enthalten hätte, was für jedes inferenzstatistische Verfahren eine ungeeignete Datenlage wäre.<sup>24</sup> Auch die Belege für die Vollformen wurden ausschließlich aus solchen Dokumenten gewählt, die mindestens drei Kurzformen enthalten, da eine Evaluation gegen Belege aus einem Register, in dem die Kurzform gar nicht verschriftet wird, nicht zielführend ist. Schließlich haben wir im Sinne von Abschnitt 2.2 sichergestellt, dass die 33 Belege für *n* im Akkusativ Maskulinum ohne Zweifel Indefinita waren.

In einer binären logistischen Regression (einem binomialen generalisierten linearen Modell) betrachten wir die abhängige Variable „Kurzform“ (1 oder 0).<sup>25</sup>

<sup>24</sup> Es handelt sich um eine Situation mit sogenannten seltenen Ereignissen. Zur Regression mit seltenen Ereignissen vgl. z. B. King & Zeng (2001).

<sup>25</sup> Für mehr oder weniger technische Einführungen vgl. z. B. Backhaus et al. (2006), Fahrmeir et al. (2009) oder Zuur et al. (2009). Linguistisch orientiert, aber weniger ausführlich als die genannten ist Johnson (2008). Ein wegweisender Artikel mit einer logistischen Regression in der Korpuslinguistik war Bresnan et al. (2007).

**Tabelle 8:** Koeffizienten und Chancenverhältnisse für das Modell zu Studie 3.

Unabhängige	Signifikanz	Koeffizient	Chancenverhältnis
Intercept (GenFem, KasAkk, Adj0, LinkskAnd)	***	0,862	2,369
GenMask	***	-0,291	0,747
GenNeut	***	-1,492	0,225
KasNom	*	-0,232	0,793
KasDat	***	-0,612	0,542
Adj1	***	-0,444	0,641
LinkskPraep	***	-0,666	0,514
LinkskInt	***	-0,994	0,370
LinkskSo		0,216	1,241
GenMask:KasNom	***	-1,360	0,257
GenNeut:KasNom	.	-0,426	0,653
GenMask:KasDat	***	0,738	2,092
GenNeut:KasDat	***	1,694	5,439

Abhängige Variable: Kurz(form). Positive Koeffizienten zeigen eine Präferenz für Kurz1, negative für Kurz0.  $n = 4.999$ . Wald-Test für das Modell:  $p < 0,001$ . Dispersion:  $\varphi = 1,000$ . Signifikanzkodierung:  $p < 0,001$  \*\*\*;  $p < 0,01$  \*\*;  $p < 0,05$  \*;  $p < 0,1$  .

Das Modell beantwortet uns die Frage, ob (und wenn ja wie stark) das Auftreten der Ausprägungen der unabhängigen Variablen mit dem Auftreten der Kurzform korreliert. Der Genitiv muss wegen der zu geringen Vorkommenshäufigkeit der Kurzform (in Zufallsstichproben der hier verwendeten Größen ist diese i. d. R. null) ausgeschlossen werden (Abschnitte 2.2 und 2.4). Von den unabhängigen Variablen erweisen sich zunächst „Linksvokal“ (vokalischer Auslaut des dem Artikelwort direkt vorangehenden Wortes) und „Rechtsvokal“ (vokalischer Anlaut des dem Artikelwort direkt folgenden Wortes) als nicht signifikant und werden wie üblich aus dem Modell genommen. Damit bleiben die signifikanten Faktoren in Tabelle 8 übrig.<sup>26</sup> So als Linkskontext hat keinen signifikanten Einfluss, verbleibt aber als Ausprägung eines ansonsten signifikanten Faktors im Modell.

Die signifikanten Koeffizienten gemäß Tabelle 8 werden in Abbildung 1 graphisch dargestellt. Positive Werte zeigen eine Präferenz für die Kurzform an, negative eine Präferenz für die Vollform. Bei der Interpretation muss nun zunächst beachtet werden, dass die Ausprägungen GenFem, KasAkk, Adj0 und

<sup>26</sup> Wir verwenden die im Statistikpaket R (R Core Team 2014) übliche Notation für Faktor und Ausprägung mit Binnenmajuskel, so dass „GenMask“ als „Genus = Maskulinum“ und „Adj1“ als „Adjektiv = ja (vorhanden)“ zu lesen sind. Der Doppelpunkt markiert Interaktionen (Zusammenwirkungen) zwischen Ausprägungen zweier Faktoren.

**Tabelle 9:** Summierte Koeffizienten des Modells aus Tabelle 8 für Genus und Kasus. Negative Koeffizienten sind grau hinterlegt und zeigen eine Präferenz für Vollformen. Die schwach signifikante Interaktion GenNeut:KasNom wurde eingerechnet.

	Femininum	Maskulinum	Neutrum
Nominativ	0,630	-1,021	-1,288
Akkusativ	0,862	0,571	-0,630
Dativ	0,250	0,697	0,452

LinksAnd vom Intercept abgebildet werden, der mit einem Koeffizienten von 0,862 eine Präferenz für die Kurzform anzeigt. Dies ist vor allem bezüglich des Femininums konsistent mit den Beobachtungen zu Tabelle 2 in Abschnitt 2.2, wo bereits deskriptiv festgestellt wurde, dass die Kurzformen des Femininums eine überproportional hohe Auftrittshäufigkeit gegenüber ihren Vollformen haben. Ebenso ist aber die Abwesenheit eines Adjektivs und der Akkusativ bei den Kurzformen häufiger, als bei einer Gleichverteilung zu erwarten wäre.

Die Einzelausprägung, die besonders stark die Vollform bevorzugt, ist links vom Artikelwort stehende Satzinterpunktion (interpretiert als NP-Position am Satzanfang). Signifikant, aber weniger stark fallen die Einbettung in eine PP (LinksPraep) und die Anwesenheit eines attributiven Adjektivs (Adj1) ins Gewicht. Die Interpretation für Kasus und Genus ist vor allem wegen der bestehenden Interaktion komplexer. Die Tendenzen sind einfacher zu verstehen, wenn wir die Koeffizienten für konkrete Kombinationen von Genus und Kasus aufaddieren. Damit können (*ceteris paribus*) die tatsächlichen Einflüsse von Genus und Kasus in den Feldern des Paradigmas direkt abgelesen werden, vgl. Tabelle 9. Es existiert eine Tendenz zur Vollform in den Positionen des Paradigmas, in denen die Vollform einsilbig ist.

## 2.6 Studie 4: Normale und erweiterte Kurzform

Wie in Abschnitt 1 (besonders 1.5) erläutert, gibt es ein Feld von vier Genus-Kasus-Kombinationen, in denen *n* und *nen* als Kurzformen konkurrieren, nämlich Maskulinum und Neutrum im Nominativ und Akkusativ. Über eine Stichprobe, die 1.172 Belege mit *n* und 1.151 Belege mit *nen* enthielt, wurden für Studie 4 die Variablen ermittelt, die mit der Wahl der jeweiligen Kurzform korrelieren. Deskriptiv ist zunächst festzuhalten, dass bei den Neutra insgesamt 22,69% der Belege mit *nen* formuliert sind. Bei den Maskulina sind es zwar mit 60,94% deutlich mehr, aber die erweiterte Kurzform kann beim Neutrum

**Tabelle 10:** Absolute Verteilung der Formen auf die Genus- und Kasuskombinationen in Studie 4.

	Maskulinum		Neutrum	
	<n>	<nen>	<n>	<nen>
Nominativ	323	90	142	40
Akkusativ	314	904	393	117

keinesfalls als marginal gelten. Zur Verdeutlichung zeigt Tabelle 10 die absolute Verteilung innerhalb der Stichprobe.

Wie in Studie 3 (Abschnitt 2.5) wurde eine binäre logistische Regression mit den unabhängigen Variablen aus Tabelle 7 gerechnet. Es kam die Variable „Apostroph“ hinzu, die genau dann die Ausprägung 1 hat, wenn das Artikelwort mit Apostroph geschrieben ist (sonst 0). Die Abhängige hat die Ausprägung 1, wenn die Form *nen* lautet und 0, wenn sie *n* lautet. Die meisten Faktoren mussten als nicht signifikant eliminiert werden. Es bleiben die signifikanten Faktoren in Tabelle 11.

Wie erwartet bevorzugen das Neutrum und der Nominativ die Form *n*. Zu beachten ist, dass die starke Interaktion GenNeut:KasNom die Präferenz der Einzelausprägungen GenNeut und KasNom für *n* im Nominativ Neutrum deutlich abschwächt. Konkret bedeutet dies, dass sich die unabhängigen Präferenzen für *n* gegenüber *nen*, die vom Neutrum und vom Nominativ ausgehen, nicht im Nominativ Neutrum akkumulieren. Das Auftreten der Form *nen* ist damit sehr wahrscheinlich durch morphophonologische und graphematische Merkmale (Dispräferenz für *n*) und nicht durch morphosyntaktische bzw. funktionale (Genus und Kasus) motiviert. Die Form *nen* ist also im nicht-femininen Paradigma kein allgemein prototypischer Akkusativ. Tabelle 12 zeigt wie Tabelle 9 die aufsummierten Koeffizienten für Kasus und Numerus. Die Unterschiede zwischen dem Nominativ in beiden Genera (-1,192 und -1,135) sowie dem Akkusativ Neutrum (-1,108) sind äußerst gering bzw. zu vernachlässigen.

Abschließend bleibt festzustellen, dass die Apostrophschreibung bei der langen Kurzform *nen* signifikant seltener ist als bei der kurzen Kurzform *n*. Die phonologisch motivierten Variablen Linksvokal (vor allem zur Hiatusvermeidung) und Rechtsvokal sind wie schon im ersten Modell nicht signifikant. Die Vermutung von Vogel (2006), dass Satzinitialität *nen* bevorzuge, kann ebenfalls nicht bestätigt werden.

## 2.7 Studie 5: Vollständige graphematische Enklise

In einer letzten Studie wird jetzt die Verteilung der Zusammenschreibungen – also vollständige graphematische Enklise – des links stehenden Wortes mit

**Tabelle 11:** Koeffizienten und Chancenverhältnisse für das Modell zu Studie 4.

Unabhängige	Signifikanz	Koeffizient	Chancenverhältnis
Intercept (GenMask, KasAkk, Apostroph0)	***	1,192	3,294
GenNeut	***	-2,300	0,100
KasNom	***	-2,384	0,092
Apostroph1	***	-1,828	0,161
GenNeut:KasNom	***	2,357	10,554

Abhängige Variable: Variante (*n* oder *nen*). Negative Koeffizienten favorisieren *n*, positive *nen*.

$n = 2.323$ . Wald-Test für das Modell:  $p < 0,001$ . Dispersion:  $\varphi = 0,985$ .

Signifikanzkodierung:  $p < 0,001$  \*\*\*;  $p < 0,01$  \*\*;  $p < 0,05$  \*;  $p < 0,1$  ·.

**Tabelle 12:** Summierte Koeffizienten des Modells aus Tabelle 11 für Genus und Kasus.

Positive Koeffizienten zeigen eine Präferenz für *nen*.

	Maskulinum	Neutrum
Nominativ	-1,192	-1,135
Akkusativ	1,192	-1,108

dem verkürzten Indefinitartikel exemplarisch untersucht. In allen bisherigen Studien waren diese Zusammenschreibungen aus mehr oder weniger technischen Gründen nicht enthalten, denn die allgemeine Suche nach zusammengeschiedenen Formen im Korpus ist vor allem deshalb nicht möglich, da z. B. Formen wie *undne* nicht von anderen Wörtern zu unterscheiden sind, die auf die Zeichenkette *ne* enden.

Für diese Studie haben wir daher als Heuristik die zehn häufigsten Wörter, die im in Abschnitt 2.4 beschriebenen quasi-spontansprachlichen Subkorpus links vom (absolut gesehen häufigsten getrennt geschriebenen) Artikel *ne* auftreten, in der entsprechenden Zusammenschreibung gesucht und die Häufigkeiten der Zusammen- und der Getrennschreibungen im gesamten DECOW2012 verglichen. Im Besonderen geht es hier darum, den graphematisch stark defektiven Artikel *n* mit anderen weniger defektiven Formen wie *ne* zu kontrastieren. Da es sich um eine erste Auxiliarstudie handelt, wurden Formen mit Apostroph aufgrund des hohen zusätzlichen Rechercheaufwands nicht berücksichtigt und nur *n*, *ne* und *nen* untersucht. Einige Fälle mit Formen wie *maln* (überwiegend verkürztes oder verschriebenes *malen*) oder *wien* (ausschließlich Referenzen auf die Stadt Wien, ggf. in Kleinschreibung) sowie Kombinationen mit Präpositionen wurden der automatischen Auswertung halber nicht betrachtet. An einer Stichprobe von 200 Treffern mit Präpositionen und zusammengeschiedenem *n*

**Tabelle 13:** Verteilung der Zusammen- und Getrennschreibung in DECOW2012 für die zehn Wörter, die am häufigsten links des getrennt geschriebenen klitisierten Artikels *ne* auftreten (außer Präpositionen und Fälle, in denen Ergebnisse zu erwarten sind, die nicht der Zielstruktur entsprechen, z. B. *wien*). Verbleibende Fehlbelege wurden nicht aussortiert.

Linkskontext	n			ne			nen		
	mit Spatium	ohne Spatium	% ohne	mit Spatium	ohne Spatium	% ohne	mit Spatium	ohne Spatium	% ohne
noch	19.097	8.694	0,31	122.728	313	< 0,01	62.970	64	< 0,01
so	11.123	110.846	0,91	103.996	21.923	0,17	42.847	4.927	0,10
auch	10.242	2.084	0,17	80.215	89	< 0,01	38.040	31	< 0,01
ist	4.496	1.295	0,22	58.406	25	< 0,01	9.161	5	< 0,01
und	10.500	1.188	0,10	51.310	13	< 0,01	40.338	8	< 0,01
nur	7.077	2.766	0,28	46.348	36	< 0,01	18.750	10	< 0,01
ich	6.574	1.697	0,21	44.388	30	< 0,01	34.417	28	< 0,01
mir	3.669	637	0,15	27.485	13	< 0,01	20.187	8	< 0,01
wieder	3.905	359	0,08	24.199	9	< 0,01	13.517	2	< 0,01
hat	2.234	1.054	0,32	23.698	22	< 0,01	18.073	8	< 0,01

zeigte sich in der Durchsicht, dass 120 Fälle eindeutig Definita enthielten, 64 eindeutig Indefinita. Zwölf Belege waren ambig, und vier waren Fehlbelege. Wir gehen davon aus, dass in der Zusammenschreibung mit Präposition typischerweise definite Artikel vorkommen und haben Präpositionen als Linkskontexte des Artikels daher hier unberücksichtigt gelassen. Tabelle 13 zeigt die Ergebnisse.

Das Bild ist ohne inferenzstatistische Auswertung eindeutig. Mit Ausnahme der Kombination mit *so* findet die Zusammenschreibung nur mit dem klitischen Artikel *n* in beachtenswerter Menge statt. Genau diese Form *n* konstituiert für sich allein weder ein minimales graphematisches Wort noch eine minimale graphematische Silbe des Deutschen (s. Jacobs 2005; Fuhrhop 2008; vgl. auch Abschnitt 1.3). Formen mit <e> wie *ne* enthalten zwar keine graphematische Vollsilbe, immerhin aber eine graphematische Silbe, und sie entziehen sich daher weitgehend der Zusammenschreibung. Dieses Ergebnis ist von grundlegender Bedeutung. Nicht-graphematische Erklärungen sind unserer Ansicht nach für diesen Befund nahezu auszuschließen, da z. B. morphophonologisch keine Hypothese besteht, dass Enklise bei *nur n* gegenüber der bei *nur ne* präferiert ist. Damit liegt eindeutige Evidenz für autonome Prinzipien der Gebrauchsschreibung unabhängig sowohl von Abbildungsrelationen zwischen Grammatik und Graphematik als auch von orthographischen Regeln vor.

### 3 Schlussfolgerungen

Die Ergebnisse unserer Korpusstudien sollen nun abschließend auf die in der Literatur diskutierten und in dieser Studie in Abschnitt 1.5 aufgestellten Hypothesen zu den distributionellen Faktoren und der Registerspezifität der Kurzformen des Indefinitartikels bezogen und gewertet werden. Außerdem erfolgt die Interpretation der Ergebnisse unter einem paradigmatischen Gesichtspunkt. Wir beziehen uns dabei primär auf die nasal anlautenden Kurzformen.

Wir haben in Abschnitt 2.3 gezeigt, dass die Kurzformen in unserem Korpus nur in Texten eines bestimmten Registers auftreten, vor allem in Forendiskussionen und Blogs. Das Register des jeweiligen Textes ist der wichtigste Faktor für das Auftreten der Verschriftung der Kurzformen. Nicht zuletzt aufgrund der Größe des Korpus konnten wir außerdem zeigen, dass der Gebrauch der Kurzformen nicht konsistent erfolgt, dass es sich im gegebenen Register also partiell um eine echte Alternation handelt. Die Kurzformen des Indefinitartikels stehen selbst bei Schreiberinnen und Schreibern, die sie verwenden, also nicht überall dort, wo sie in der gesprochenen Sprache möglich und erwartbar sind. Insofern

greift eine Bewertung der Kurzformen als rein phonographische Verschriftung zu kurz. Ebenso ist die lange Kurzform *nen* nicht in allen vorkommenden Positionen schlüssig über Phonem-Graphem-Korrespondenzen erklärbar. Eine phonographische Verschriftung der Kurzform des Indefinitartikels im Nominativ Maskulinum sowie Neutrum und im Akkusativ Neutrum müsste *n* lauten, wenn man eine „reguläre“ Reduktion zugrunde legt. Hörbelege wie *Wir machen nen Foto* deuten auf einen sich möglicherweise auch auf phonologischer Ebene vollziehenden Auffüllungsprozess hin (Vogel 2006), der phonologisch ein autonomes Wort sichert. Die Wechselwirkungen von graphematischer und phonologischer Ebene sind im Rahmen korpuslinguistischer Studien aber nicht erfassbar. Unser Blick richtete sich deshalb auf die innergraphematischen Bedingungen der Verschriftungen von Kurzformen, die wir an der Form des graphematischen Wortes und seiner Umgebung festgemacht haben, konkret: Spatium und graphematische Enklise, Apostrophschreibung sowie der Prototyp des graphematischen Wortes aus mindestens einer wohlgeformten graphematischen Silbe.

Nun sind vor allem die in Abschnitt 1.5 konkret benannten und der empirischen Überprüfung direkt zugänglichen Hypothesen zu evaluieren (Nummerierung parallel zu Abschnitt 1.5).

1. Wir haben gezeigt, dass eine Tendenz zur Einsilbigkeit im Gesamtparadigma des Indefinitartikels besteht. Die Vollform ist in den Positionen des Paradigmas, in denen sie selber einsilbig ist, gegenüber der Kurzform präferiert (Abschnitt 2.5).
2. Vom letzten Punkt auszunehmen ist der Genitiv, da wir die faktische Nicht-Existenz der verkürzten Genitivformen (*nes* und *ner*) zeigen konnten (Abschnitt 2.2), obschon der Genitiv in der Vollform zweisilbig ist. Dass das Fehlen der Formen aber nicht auf den allgemeinen Genitivabbau reduzierbar ist, haben wir anhand hochsignifikanter Ergebnisse erwiesen (Studie 2 in Abschnitt 2.4).
3. Wir haben gezeigt, dass Hiatusvermeidung als Auslöser für die Kurzformen und als Determinante der Alternation zwischen *n* und *nen* in unseren Daten nicht signifikant ist (Abschnitt 2.5 und 2.6). Generell zeigt sich also, dass phonologische Faktoren eine untergeordnete Rolle bei der Verschriftung der Klitisierungsformen spielen.
4. Für die Präferenz der Enklise sprach eine signifikante Korrelation zwischen Satzinitialität und dem Unterbleiben der Kurzform (Abschnitt 2.5). Satzinitial ist Enklise i. d. R. auszuschließen. Eine Präferenz für die aufgefällten Formen *nen* in satzinitialer Position ließ sich nicht feststellen (Abschnitt 2.6).
5. Während wir in Abschnitt 2.5 zeigen konnten, dass Präpositionen als linker Kontext signifikant die Vollformen bevorzugen, erweist sich die Partikel so

**Tabelle 14:** Nasal anlautendes Kurzformenparadigma des Indefinitartikels. Grau hinterlegt sind Felder, in denen die Vollformen einsilbig sind. Die jeweils gleich starke Tendenz zur Auffüllung wird durch > („ist präferierter als“) angezeigt. Eingeklammert sind faktisch nur marginal im Gebrauch befindliche Formen.

	Femininum	Maskulinum	Neutrum
Nominativ	<i>ne</i>	<i>n &gt; nen</i>	<i>n &gt; nen</i>
Akkusativ	<i>ne</i>	<i>nen &gt; n</i>	<i>n &gt; nen</i>
Dativ	<i>ner</i>	<i>nem</i>	<i>nem</i>
Genitiv	( <i>ner</i> )	( <i>nes</i> )	( <i>nes</i> )

als Linkskontext der Kurzformen als nicht signifikant bevorzugt. Allerdings wurden in Studie 3 aus technischen Gründen keine Daten für vollständige Zusammenschreibungen berücksichtigt. Die Daten aus Studie 5 (Abschnitt 2.7) legen die Vermutung nahe, dass anstelle von *so* mit der Kurzform des Artikel das emergente Pronomen mit dem Stamm *son* (in Zusammenschreibung) verwendet wird.

- Wir haben gezeigt, dass die Auffüllung von *n* zu *nen* in allen Positionen des Paradigmas, in denen *n* die erwartbare Kurzform ist, in gleichem Ausmaß stattfindet (Abschnitt 2.6). Die Tendenz zur Auffüllung ist dabei nicht marginalisierbar, wie schon die absoluten Zählzahlen in Tabelle 10 gezeigt haben.
- Wie aus graphematischer Perspektive erwartet, zeigt unsere Korpusstudie, dass in den Fällen, in denen eine Alternation der Kurzformen *n* und *nen* vorliegt, die Apostrophschreibung bei der Form *n* signifikant häufiger ist, also graphisch *'n*, aber seltener *'nen*; umgekehrt häufiger *nen*, aber seltener *n*.

Abschließend stellen wir nun Überlegungen dazu an, wie sich die Kurzformen und die zusätzliche Alternation im Gesamtparadigma positionieren. Es ergibt sich Tabelle 14 in Anlehnung an Tabelle 1. In Übereinstimmung mit der existierenden Literatur und unseren Ergebnissen gehen wir davon aus, dass die Kurzformenbildung einer Tendenz zur Einsilbigkeit und zum nasalen Anlaut im Paradigma des Indefinitartikels entspricht. Die Präferenz zur Verkürzung ist allerdings herabgesetzt in den Formen, in denen die Vollform ohnehin einsilbig ist, die Tendenz zur Einsilbigkeit ist also stärker als die zum nasalen Anlaut.

Die Genitive des Femininums und des Nicht-Femininums weisen keine morphophonologisch oder graphematisch dispräferierten Merkmale auf, und der Genitivabbau im einschlägigen Register ist nicht vollständig abgeschlossen. Zur Klärung der Frage, was die Nichtexistenz des Genitivs motiviert, wäre unserer

Ansicht nach zunächst eine umfassende Untersuchung dieses Phänomens in gesprochener Sprache erforderlich.

Wir haben gezeigt, dass die Auffüllung von *n* zu *nen* in allen drei betreffenden Positionen des Paradigmas (Nominativ Maskulinum sowie Nominativ und Akkusativ Neutrum) gleichermaßen präferiert ist. Damit kann als erwiesen gelten, dass es sich um einen primär an die morphologische und vor allem graphematische Form gebundenen Prozess handelt, weitgehend unabhängig von phonologischen Regularitäten.

Unter einer paradigmatischen Perspektive ergibt sich mit den aufgefüllten *nen*-Formen und ohne die Genitive die besonders präferierte Verteilung von Synkretismen aus Tabelle 15. Sowohl die Gruppierung in Femininum und Nicht-Femininum als auch der Synkretismus von Nominativ und Akkusativ innerhalb aller Genera ist im nominalen Flexionssystem des Deutschen allgegenwärtig. Da man nicht davon sprechen kann, dass das System in Tabelle 15 im gegenwärtigen Deutschen (in irgendeinem Register) als Endpunkt einer Entwicklung erreicht wäre, ist diese Überlegung freilich spekulativ. Es liegt aber nahe, dass auch dieser Systemausgleich die Auffüllung zu *nen* befördert.

**Tabelle 15:** Synkretismusfelder im aufgefüllten Kurzformparadigma ohne den faktisch nicht existenten Genitiv.

	Femininum	Nicht-Femininum
Nominativ	<i>ne</i>	<i>nen</i>
Akkusativ		
Dativ	<i>ner</i>	<i>nem</i>

Aus korpuslinguistischer Perspektive ergeben sich aus der hier vorgelegten Studie einige klare Desiderata. Zunächst wäre es wünschenswert, die von uns weitgehend eingenommene agnostische Position bezüglich der Korrelation von graphischen und phonologischen Realisierungen durch die bereits erwähnte empirisch fundierte Untersuchung zu den Kurzformen in gesprochener Sprache aufzulösen. Diese Aufgabe ist momentan nicht innerhalb der Korpuslinguistik zu leisten, weil ausreichend große Korpora der gesprochenen Sprache nicht existieren, und weil eine eindeutige Transkription der Kurzformen ohne eine speziell auf die Fragestellung zugeschnittene apparative phonetische Analyse in verfügbaren Korpora der gesprochenen Sprache nicht zu erwarten ist. Eine solche Analyse wäre nötig, um z. B. den Unterschied zwischen [nən], [nŋ] und [ŋ] genau zu ermitteln.

Wünschenswert und möglich wäre hingegen eine genauere Untersuchung von vollständigen Zusammenschreibungen klitisierter Artikel. Diese wurde hier nur oberflächlich durchgeführt, vor allem durch die mit ihr verbundenen erheblichen technischen Probleme. Der zusammengeschriebene Indefinitartikel ist nicht als Token im Korpus indexiert und suchbar, und es müsste in einer größeren Studie mit einer komplizierten Such-Heuristik und einer aufwändigen Bereinigung der Ergebnisse gerechnet werden. Dass dies aber fruchtbringend wäre, zeigt die in Studie 5 (Abschnitt 2.7) vorgebrachte Evidenz für unabhängige Prinzipien der Gebrauchsschreibung, die jenseits orthographischer Regeln existieren.

**Danksagung:** Wir danken den folgenden Personen (in alphabetischer Reihenfolge) für Diskussionen, Hinweise und Korrekturen: Felix Bildhauer, Thea Ditrach, Peter Eisenberg, Karsten Schmidt, Niklas Schreiber und Bjarne Ørnes. Sämtliche verbleibenden Fehler und Schwächen haben wir selber zu verantworten.

## Literatur

- Backhaus, Klaus, Bernd Erichson, Wulff Plinke & Rolf Weiber. 2006. *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 11. Aufl. Heidelberg: Springer.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226.
- Biemann, Chris, Felix Bildhauer, Stefan Evert, Dirk Goldhahn, Uwe Quasthoff, Roland Schäfer, Johannes Simon, Leonard Swiezinski & Torsten Zesch. 2013. Scalable construction of high-quality web corpora. *Journal for Language Technology and Computational Linguistics* 28(2). 23–60.
- Bredel, Ursula. 2008. *Die Interpunktion des Deutschen. Ein kompositionelles System zur Online-Steuerung des Lesens*. Tübingen: Niemeyer.
- Bredel, Ursula. 2011. *Interpunktion*. Heidelberg: Winter.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Kraemer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Burri, Gabriela. 2003. Spontanschreibung im Chat. *Linguistik online* 15(3). 3–31.
- Dedenbach, Beate. 1987. *Reduktions- und Verschmelzungsformen im Deutschen. Schwache Formen bei Artikeln und Pronomina*. Frankfurt/M.: Lang.
- Eisenberg, Peter. 2013a. *Grundriss der deutschen Grammatik. Das Wort*. 4., aktualisierte und überarbeitete Auflage unter Mitarbeit von Nanna Fuhrhop. Stuttgart: Metzler.

- Eisenberg, Peter. 2013b. *Grundriss der deutschen Grammatik. Der Satz. 4.*, aktualisierte und überarbeitete Auflage unter Mitarbeit von Rolf Thieroff. Stuttgart: Metzler.
- Eroms, Hans-Werner. 2008. Erweiterung des Systems der Artikel im Deutschen. In Daniel Baudot & Maurice Kauffer (eds.), *Wort und Text. Lexikologische und textsyntaktische Studien im Deutschen und Französischen. Festschrift für René Métrich zum 60. Geburtstag*, 57–67. Tübingen: Stauffenburg.
- Fahrmeir, Ludwig, Thomas Kneib & Stefan Lang. 2009. *Regression – Modelle, Methoden und Anwendungen*. Heidelberg: Springer.
- Fuhrhop, Nanna. 2008. Das graphematische Wort (im Deutschen). Eine erste Annäherung. *Zeitschrift für Sprachwissenschaft* 27(2). 189–228.
- Fuhrhop, Nanna & Franziska Buchmann. 2009. Die Längenhierarchie: Zum Bau der graphematischen Silbe. *Linguistische Berichte* 218. 127–155.
- Geyken, Alexander. 2007. The DWDS corpus: A reference corpus for the German language of the 20<sup>th</sup> century. In Christiane Fellbaum (ed.), *Collocations and idioms: Linguistic, lexicographic, and computational aspects*, 23–41. London: Continuum.
- Gries, Stefan Thomas. 2008. *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck & Ruprecht.
- Haase, Martin, Michael Huber, Alexander Krumeich & Georg Rehm. 1997. Internetkommunikation und Sprachwandel. In Rüdiger Weingarten (ed.), *Sprachwandel durch Computer*, 51–85. Opladen: VS Verlag für Sozialwissenschaften.
- von Heusinger, Klaus. 2012. Referentialität, Spezifität und Diskursprominenz im Sprachvergleich am Beispiel von indefiniten Demonstrativpronomen. In Lutz Gunkel & Gisela Zifonun (eds.), *Deutsch im Sprachvergleich – Grammatische Kontraste und Konvergenzen. Jahrbuch des Instituts für Deutsche Sprache*, 1–40. Berlin: de Gruyter.
- Hole, Daniel & Gerson Klumpp. 2000. Definite type and indefinite token: The article “son” in colloquial German. *Linguistische Berichte* 182. 231–244.
- Jacobs, Joachim. 2005. *Spatien. Zum System der Getrennt- und Zusammenschreibung im heutigen Deutsch*. Berlin: de Gruyter.
- Johnson, Keith. 2008. *Quantitative methods in linguistics*. Hoboken, NJ: Wiley-Blackwell.
- Kilgarriff, Adam. 2006. Googleology is bad science. *Computational Linguistics* 33(1). 147–151.
- King, Gary & Langche Zeng. 2001. Logistic regression in rare events data. *Political Analysis* 9(2). 137–163.
- Krech, Eva-Maria, Eberhard Stock, Ursula Hirschfeld & Lutz Christian Anders (Hrsg.). 2009. *Deutsches Aussprachewörterbuch*. Berlin: de Gruyter.
- Kupietz, Marc & Holger Keibel. 2009. The Mannheim German Reference Corpus (DeReKo) as a basis for empirical linguistic research. *Working Papers in Corpus-based Linguistics and Language Education* 3. 53–59.
- Lenerz, Jürgen & Horst Lohnstein. 2005. „Nur so.“ Strukturaspekte der Vergleichskonstruktion. In Franz Josef d’Avis (ed.), *Deutsche Syntax: Empirie und Theorie*, 81–103. Göteborg: Acta Universitatis Gothoburgensis.
- Naumann, Bernd. 1998. Stirbt die deutsche Sprache? Überlegungen zum Sprachwandel durch IRC (Internet Relay Chat). In Světlá Čmejrková, Jana Hoffmannová, Olga Müllerová & Jindra Světlá (eds.), *Dialoganalyse VI. Referate der 6. Arbeitstagung Prag 1996*, 249–262. Tübingen: Niemeyer.
- Nübling, Damaris. 1992. *Klitika im Deutschen. Schriftsprache. Umgangssprache, alemannische Dialekte*. Tübingen: Narr.

- Nübling, Damaris. 2005. Von „in die“ über „in‘ne“ und „ins“ bis „im.“ Die Klitisierung von Präposition und Artikel als „Grammatikalisierungsbaustelle“. In Thorsten Leuschner, Tanja Mortelmans & Sarah de Groot (eds.), *Grammatikalisierung im Deutschen*, 105–131. Berlin: de Gruyter.
- Prinz, Michael. 1991. *Klitisierung im Deutschen und Neugriechischen*. Tübingen: Niemeyer.
- R Core Team. 2014. *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.
- 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, Jan Odijk & Stelios Piperidis (eds.), *Eighth International Conference on Language Resources and Evaluation*, 486–493. Istanbul: ELRA.
- Schäfer, Roland & Felix Bildhauer. 2013. *Web corpus construction*. San Francisco, CA: Morgan & Claypool.
- Szczepaniak, Renata. 2009. *Grammatikalisierung im Deutschen. Eine Einführung*. Tübingen: Narr.
- Schiering, René. 2002. *Klitisierung von Pronomina und Artikelformen. Eine empirische Untersuchung am Beispiel des Ruhrdeutschen*. Köln: Institut für Sprachwissenschaft, Universität zu Köln.
- Schuster, Britt-Marie & Doris Tophinke (eds.). 2012. *Andersschreiben. Formen, Funktionen, Traditionen*. Berlin: Erich Schmidt.
- Sharoff, Serge. 2006. Creating general-purpose corpora using automated search engine queries. In Marco Baroni & Silvia Bernardini (eds.), *WaCky! Working papers on the web as corpus*, 63–98. Bologna: GEDIT.
- Tophinke, Doris. 2002. Schreiben gegen die Regel – Formen und Funktionen orthografischer Abweichungen im Internet Relay Chat (IRC). In Michael Bommers, Christina Noack & Doris Tophinke (eds.), *Sprache als Form. Festschrift für Utz Maas zum 60. Geburtstag*, 170–182. Wiesbaden: Westdeutscher Verlag.
- Vogel, Petra. 2006. „Ich hab da nen kleines Problem!“ Zur neuen Kurzform des indefiniten Artikels im Deutschen. *Zeitschrift für Dialektologie und Linguistik* 73(2). 176–193.
- Volmert, Johannes. 1995. Jugend und Ruhrgebietssprache: Die regionale Varietät in der Freizeit – und als Unterrichtsgegenstand? In Konrad Ehlich, Wilhelm Elmer, Rainer Noltenius & Stephan Schlickau (eds.), *Sprache und Literatur an der Ruhr*, 57–97. Essen: Klartext.
- Weingarten, Rüdiger. 1997. Einleitung. In Rüdiger Weingarten (ed.), *Sprachwandel durch Computer*, 7–20. Wiesbaden: Westdeutscher Verlag.
- Ziegler, Evelyn. 2011. Subsidente Normen und Sprachkompetenz: ihre Bedeutung für den Deutschunterricht. *Bulletin suisse de linguistique appliquée* 94. 69–85.
- Ziegler, Evelyn. 2012. „Merkel hat ’nen Neuen!“ Die Kurzform „nen“ – ein Mündlichkeitsmarker par excellence. In Britt-Marie Schuster & Doris Tophinke (eds.), *Andersschreiben. Formen, Funktionen, Traditionen*, 295–316. Berlin: Erich Schmidt.
- Zuur, Alain F., Elena N. Ieno, Neil Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. Heidelberg: Springer.

## Anhang: URLs der Korpusbelege

- (1a) <http://www.fifaplanet.de/archive/index.php/t-6644.html>
- (1b) <http://www.techno.de/forum/archive/index.php/t-12646.html>
- (1c) <http://www.diva-di-bologna.de/lofiversion/index.php?t4131.html>
- (2a) <http://www.xblsenioren.de/forum/archive/index.php/t-9832-p-2.html>
- (2b) <http://www.r129-forum.de/archive/index.php/t-1564.html>
- (3a) <http://www.a2-freun.de/forum/archive/index.php/t-15784-p-2.html>
- (3b) <http://www.dvdboard.de/forum/archive/index.php/t-117404-p-14.html>
- (4a) <http://forum.express.de/archive/index.php/t-10045-p-18.html>
- (4b) <http://www.reitsportforum.de/pferde-allgemein/17930-hoffnungsloser-fall.html>
- (5a) [http://www.piercing-magazin.de/tattoo/f\\_tattoo-37424.html](http://www.piercing-magazin.de/tattoo/f_tattoo-37424.html)
- (5b) [http://www.conceptarts.de/forum/Mellorys-Test-Area%2851%29\\_714.html](http://www.conceptarts.de/forum/Mellorys-Test-Area%2851%29_714.html)
- (6) <http://www.900r.de/gpz-forum/archive/index.php/t-15164.html>
- (7a) <http://www.chemieonline.de/forum/archive/index.php/t-68108.html>
- (7b) <http://www.ruhrpottforum.de/rund-um-pott/politik-wirtschaft/32012-erhoehung-hartz-iv-regelsatzes-420-a-64/drucken.html>
- (7c) <http://meinrap.de/forum/archive/index.php/t-506.html>
- (8a) [http://www.fahrschulforum.de/fuehrerschein/Angst\\_beim\\_Auto\\_fahren,212218\\_msg.html](http://www.fahrschulforum.de/fuehrerschein/Angst_beim_Auto_fahren,212218_msg.html)
- (8b) <http://forum.gaming-universe.de/lofiversion/index.php/t9976.html>
- (8c) <http://www.roboternetz.de/community/archive/index.php/t-1383.html>
- (8d) <http://www.niewiederalleine.de/hobbys-co/computerecke/9322-betriebssystem/>
- (8e) <http://www.wirsindfisch.de/Anglerboard/archive/index.php/t-1617.html>
- (8f) <http://www.ioff.de/archive/index.php/t-313919-p-33.html>
- (8g) <http://www.citforum.de/archive/index.php/t-60451.html>
- (8h) <http://www.wrestling-infos.de/7455.html>
- (9a) <http://www.nachtwelten.de/vB/history/topic/8028-2.html>
- (9b) <http://www.hsv-forum.de/archive/index.php/t-56855.html>
- (10a) <http://forum.gaming-universe.de/lofiversion/index.php/t11339.html>
- (10b) <http://www.dghtserver.de/foren/archive/index.php?t-10503.html>