

Amir Zeldes

Komposition als Konstruktionsnetzwerk im fortgeschrittenen L2-Deutsch*

Abstract: Compounding is one of the most productive and flexible word formation processes in German, yet it represents a challenge for German learners by virtue of its heterogeneous semantics and morphological complexity. This paper examines, on the basis of corpus data, how advanced German learners employ compounding in practice, how learner usage of compounding differs from native usage, and what types of compounds can be distinguished in these respects. The analysis of the data shows that much like their native counterparts, learners acquire very many semantically distinct compounding constructions at all levels of abstraction, which can be described as a hierarchical constructional network. Although advanced learners produce remarkably few formal errors in compound formation it will be shown that compounding occurs significantly more rarely and is used substantially less productively in learner data as compared to native controls, and that learners have added difficulties in the formation of especially deverbal compounds with internal argument structure.

Komposition zählt zu den produktivsten und vielfältigsten Wortbildungsprozessen im Deutschen, stellt jedoch durch ihre heterogene Semantik und morphologische Komplexität eine Herausforderung für Deutschlerner dar. In diesem Artikel wird auf der Basis von Korpusdaten der Frage nachgegangen, wie fortgeschrittene Lerner des Deutschen Komposition einsetzen, wie ihr Gebrauch von Komposita vom muttersprachlichen Gebrauch abweicht und welche Kompositionstypen dabei unterschieden werden können. Die Analyse der Daten belegt, dass Lerner ähnlich wie Muttersprachler sehr viele semantisch verschiedene Kompositionskonstruktionen auf allen Abstraktionsebenen erwerben, die als eine Art hierarchisches Konstruktionsnetzwerk erfasst werden können. Obwohl Lerner insgesamt verhältnismäßig wenige formale Fehler in der Kompositabildung produzieren, wird gezeigt, dass sie Komposition signifikant seltener und weniger produktiv verwenden als Muttersprachler, sowie besondere Schwierigkeiten bei der Bildung von v.a. deverbalen Komposita mit internen Rektionsbeziehungen aufweisen.

0. Einleitung – Komposition als komplexer Erwerbsgegenstand

1. Theoretische Grundlagen zum Begriff des Konstruktionsnetzwerks

2. Datengrundlage und Komposita-Extraktion

3. Fehleranalyse

4. Die oberste Ebene des Konstruktionsnetzwerks – Komposition und konkurrierende Konstruktionen

4.1 Häufigkeit

4.2 Konkurrierende Konstruktionen

4.3 Produktivität

5. Die mittleren Ebenen – Kompositionsklassen

6. Die unteren Ebenen – teillexikalisierte Minikonstruktionen

7. Zusammenfassung und Diskussion

A. Anhang: Drei-Buchstaben-Kennungen nach ISO 639-3

0. Einleitung – Komposition als komplexer Erwerbsgegenstand

Komposition bildet einen besonders prominenten Wortbildungstyp im Deutschen (s. Schlücker 2012 zum „kompositionsfreudigen“ Deutschen), der Deutschlernern oft

* Dieser Artikel ist aus einem Vortrag im Rahmen eines Treffens des wissenschaftlichen Netzwerks Kobalt-DaF entstanden. Für viele Hinweise und Ideen danke ich den Teilnehmern des Arbeitstreffens. Insbesondere möchte ich mich bei Marc Reznicek und Heike Zinsmeister sowie bei zwei anonymen GutachterInnen für zahlreiche wertvolle Kommentare zu früheren Versionen des Manuskripts bedanken. Alle verbleibenden Fehler gehen selbstverständlich zu meinen Lasten.

bereits in der Grundstufe vermittelt wird (bspw. Heidermann 1997:58-60, Fandrych & Tallowitz 2008:202) und nicht zuletzt durch die in vielen Sprachen fehlende Zusammenschreibung Lernern auch auffällig erscheinen mag (vgl. Neef 2009:386). Gleichzeitig bietet diese Wortbildungsstrategie, die durch ihre Häufigkeit dem Lerner ohnehin zugänglich ist, eine verhältnismäßig leichte Möglichkeit sein Vokabular zu erweitern, was gerade in der Übergangsphase vom Anfängersprachstand zu den fortgeschrittenen Niveaus sehr wichtig ist.

Trotz einiger Studien zur Komposition im Rahmen der Fremdsprachenforschung, v.a. im Kontext der Lehrwerkgestaltung und der Lexikographie (bspw. Rak 1990, Fandrych & Thurmair 1994, Barz 1995, Trynkle 2010), ist bisher nicht hinreichend geklärt, was der Erwerbsgegenstand Komposition insgesamt beinhaltet bzw. wie er strukturiert ist. Was lernt man genau, wenn man Komposition als Wortbildungsstrategie im Deutschen erwirbt? Lernt man allgemeine, ggf. rekursive Regeln bzw. Konstruktionen der Art $[N-N]_N$, oder mehrere abstrakte Untermuster wie bspw. $[N-[V]er]_N$ für Rektionskomposita wie *Taxifahrer*, oder sogar einzelne mehr oder weniger lexikalisierte Muster, bspw. mit reihenbildenden Modifikatoren ($[Profi-N]_N$) und Kopfnomina ($[N-Leiter]_N$)? Werden solche unterschiedlichen Klassen von Komposita gleich gut beherrscht? Werden sie von Lernern gleich oft verwendet, wie von Muttersprachlern? Werden Muster nur im Rahmen von vertrauten Lexikalisierungen verwendet oder auch produktiv eingesetzt? Kommen andere Kompositionsbedeutungen vor, und falls ja, sind diese dann immer fehlerhaft oder auffällig? Diese Fragen verlangen eine nähere Betrachtung der Kompositabildung im Gebrauch, so wie sie in L2- und vergleichbaren L1-Daten belegt ist.

In diesem Aufsatz werde ich die These vertreten, dass dem Erwerb von Komposition als produktivem Wortbildungsmechanismus viel mehr zugrunde liegt, als eine kurze Liste abstrakter Regeln, mit denen die allgemeine Struktur deutscher Komposita theoretisch erfasst werden kann. Von formalen Aspekten wie der Wahl von Fugenelementen, über die zulässigen bzw. bevorzugten semantischen Muster in der Beziehung zwischen Kopf und Modifikator, bis hin zur Erweiterbarkeit produktiver Reihen im Gebrauch soll die Produktion von Komposita bei fortgeschrittenen Deutschlernern anhand von Korpusdaten empirisch untersucht werden.¹ Dabei wird ein Netzwerkmodell vorgestellt, in dem unterschiedlich ausspezifizierte Konstruktionen im Sinne der Konstruktionsgrammatik (s. z.B. Goldberg 2006 i.a. und insb. Booij 2010 zum morphologischen Bereich; für einen Überblick über verschiedene Strömungen der Konstruktionsgrammatik s. Fischer & Stefanowitsch 2006) gleichzeitig repräsentiert und aktiviert werden können, die unterschiedliche Aspekte der Verwendung von Komposita beleuchten.

¹ Der Fokus auf fortgeschrittene Lerner in der vorliegenden Untersuchung wird u.a. durch die Erwartung motiviert, dass systematische L2-bedingte Effekte im fortgeschrittenen Sprachstand leichter erkennbar sind, als bei der vergleichsweise heterogenen Sprachproduktion von Anfängern. Die aktuellen Ergebnisse können und sollen jedoch zukünftige Studien zu früheren Stufen im Erwerb des Kompositionsmechanismus ergänzen.

Der nächste Abschnitt gibt einen kurzen Überblick über die theoretischen Grundlagen der gebrauchsbasierten Grammatik, die für die Darstellung eines Konstruktionsnetzwerks „Komposition“ relevant sind. Abschnitt 2 stellt das Lernerkorpus Falko vor, aus dem die Datengrundlage für diese Studie gewonnen wurde. Abschnitt 3 widmet sich der Analyse von Fehlern in der Kompositabildung, wie bspw. fehlerhafter Verwendung von Fugenelementen, Verletzung von Wortartbeschränkungen und ungrammatischen bzw. blockierten Kombinationen von Kopf und Modifikator. Abschnitte 4-6 bieten anschließend einen Überblick über nicht fehlerhafte qualitative und quantitative Unterschiede in der produktiven Verwendung von Kompositionskategorien auf unterschiedlichen Abstraktionsebenen: die allgemeinsten Kompositionskonstruktionen, Untermuster mit spezifischen Beschränkungen und Interpretationen sowie teilweise lexikalisierte Muster bzw. sog. Minikonstruktionen, die für die Ausdehnung von Komposition in Reihen mit gemeinsamem Kopf- bzw. Modifikatorlexem verantwortlich gemacht werden können. Abschnitt 7 fasst die Ergebnisse der Untersuchung zusammen und bespricht einige Konsequenzen.

1. Theoretische Grundlagen zum Begriff des Konstruktionsnetzwerks

Um Unterschiede nicht nur bei der qualitativen Verwendung der allgemeinsten Kompositionskonstruktionen, wie bspw. $[N-N]_N$ (rekursiv) bzw. $[A-N]_N$ (nicht-rekursiv) untersuchen zu können, ist ein gebrauchsbasierter Ansatz unerlässlich. In vielen Fällen kann man erst von einer abweichenden Verwendung in L2-Daten sprechen, wenn man eine Gebrauchsnorm anhand von muttersprachlichen Daten etabliert hat (vgl. Lüdeling & Walter 2010:319), mit der man zeigen kann, inwiefern zunächst scheinbar regelmäßige Bildungen doch auffällig sind, gemessen an den Erwartungen eines Muttersprachlers oder auch eines Fremdsprachendidaktikers. Gebrauchsbasierte Sprachbeschreibungen nehmen an, dass Sprecher nicht nur über das Wissen von kontextfreien Ableitungsregeln verfügen, sondern auch einiges über lexikalisierte Vertreter solcher Regeln wissen, einschließlich angemessener Kontexte für ihre Verwendung, Wissen über relative Frequenzen unterschiedlicher lexikalischer Typen, sowie die Erweiterbarkeit einzelner Untermuster und deren dazugehörige Semantik, die ausgehend von bereits vertrauten Prototypen ausgedehnt wird. Wir werden sehen, dass L2-Gebrauch in jedem dieser Aspekte abweichen kann, was u.a. auf unvollständigem Gebrauchswissen beruht und auch didaktische Konsequenzen haben kann.

Die vorliegende Untersuchung basiert auf dem konstruktionsgrammatischen Ansatz (Construction Grammar, abgekürzt CxG), so wie er in Goldberg (2006) größtenteils für die Syntax und in Booij (2010) für die morphologische Domäne beschrieben wird. Die Konstruktionsgrammatik geht davon aus, dass sprachliches Wissen im mentalen Lexikon als eine vernetzte Menge von Konstruktionen gespeichert wird, in der jede Konstruktion ein Form-Bedeutungspaar darstellt, wobei die Bedeutungsseite der Konstruktion breit gefasst wird, um auch pragmatische Aspekte und Gebrauchswissen über Wahrheitswertsemantik hinaus erfassen zu können. Konstruktionen werden typischerweise im Falle von beobachteten

Bedeutungen postuliert, die nicht vollständig kompositionell von den Bedeutungen ihrer Bestandteile erschließbar sind. Die meisten CxG-Ansätze betonen jedoch, dass mit zunehmender Frequenz auch kompositionelle Fügungen lexikalisiert werden, d.h. im mentalen Lexikon eingetragen werden können.

Für den momentanen Zweck sind folgende Merkmale des Modells wichtig: es wird angenommen, dass jedes Exemplar einer Konstruktion zunächst gespeichert wird und dass die Repräsentation der entsprechenden Konstruktion bei jeder Wiederholung gestärkt wird. Diese Repräsentationsstärke wird oft als „Entrenchment“ bezeichnet, d.h. Verwurzelung oder Verankerung im mentalen Lexikon. Allgemeine, lexikalisch unterspezifizierte Konstruktionen entstehen durch Generalisierungen über bereits verwurzelte Exemplare, unter denen die häufigsten Vertreter der Konstruktion als Prototypen gelten, die auch maßgebend sind für ihre Semantik im Falle einer Erweiterung auf Lexeme, die in der Konstruktion nicht verwurzelt sind. Auf diese Weise verkörpert jedes Vorkommen einer komplexen, abstrakten Konstruktion das Vorhandensein mehrerer Konstruktionen auf unterschiedlichen Ebenen (vgl. Goldberg 2006:10). So wird z.B. bei der Produktion und entsprechend in der Perzeption der Phrase in (1) eine Reihe von Konstruktionen aktiviert:

- (1) *Gib mir das Buch!*
 [V DP DP]_{VP} (Ditransitivkonstruktion)
 [*gib* DP DP]_{VP} (*gib*-VP)
 [*gib mir* DP]_{VP} (*gib-mir*-VP)
 VP_[+imperativ] (Imperativkonstruktion)

Unterschiedliche Abstraktionsebenen anzunehmen, mag zunächst unmotiviert erscheinen, betrachtet man jedoch die hohe Frequenz solcher Phrasentypen sowie den Bedeutungsbeitrag von jeder Ebene, so wird klar, dass die Annahmen des CxG-Paradigmas direkt zu einer solchen Struktur führen. Der Satz in (1) ist gleichzeitig ein Vorkommen der Ditransitivkonstruktion und der Imperativkonstruktion, aber auch ein Vorkommen der *gib*- bzw. *gib-mir*-Konstruktionen, die durch ihre Häufigkeit ggf. auch Bedeutungsbeiträge entwickeln können. So wird z.B. die pragmatische und distributionelle Bedeutung kodiert, die bestimmt, wann die Verwendung eines Imperativs angemessen ist, wann er in Verbindung mit dem Lemma *geben* verwendet wird, was der Einfluss der Tatsache ist, dass die erste Person (der Sprecher) als Rezipient vorkommt (was einen Befehl u.U. mehr oder weniger höflich bzw. gesellschaftlich akzeptabel machen kann), etc.

Im Folgenden werde ich die Strukturierung von Komposita im Gebrauch auf eine ähnliche Art und Weise analysieren. Auch das Kompositum in (2) realisiert mehrere konstruktionelle Ebenen, die jeweils einen Bedeutungsbeitrag und Gebrauchsinformationen mit sich bringen.

- (2) *Abteilungsleiter*
 [N-N]_N (Nominalkompositum)
 [N_[+arg]-N]_N (Rektionskompositum)

[N-Ver_N]_N (Deverbales Agensnomenkompositum)
 [N-Leiter_N]_N (Leiter-Kompositum)

Die Muster unterscheiden sich zwar in ihrem Abstraktheitsgrad, sind aber aus Sicht der CxG vergleichbare Objekte, die alle in der Lage sind, im Zuge des Kompositionsprozesses eigene Bedeutungsbeiträge zu liefern. Ordnet man diese Ebenen graphisch, so bekommt man eine Darstellung wie in Abbildung 1 (vgl. die hierarchisch geordneten Kompositionsschemata bei Tuggy 2005:248-251, die v.a. den feinkörnigen unteren Ebenen des Netzwerk entsprechen; s. Abschnitte 5-6 im Detail). Die Knoten im Graph entsprechen einzelnen Konstruktionen, welche nach unten schrittweise spezifischer werden und über immer weniger Exemplare generalisieren. In der untersten Ebene befinden sich auch lexikalisierte Wortformen, die den gemeinsamen Konstruktionsstatus von Lexemen und teillexikalisierten Konstruktionen zeigen. Gleichzeitig kann man davon ausgehen, dass es zu Beziehungen zwischen solchen Konstruktionsnetzwerken kommen kann, die mit den Mitgliedern des Netzwerks konkurrieren und ähnliche Bedeutungen ausdrücken. Dies wird deutlich beim Vergleich von bspw. [A-N]_N-Komposita mit durch Adjektive modifizierten Nomina (s. Barz 1996, Schlücker & Hüning 2009) bzw. durch das Verhältnis zwischen deverbalen Rektionskomposita und Verbalphrasen (s. Gaeta & Zeldes 2012).

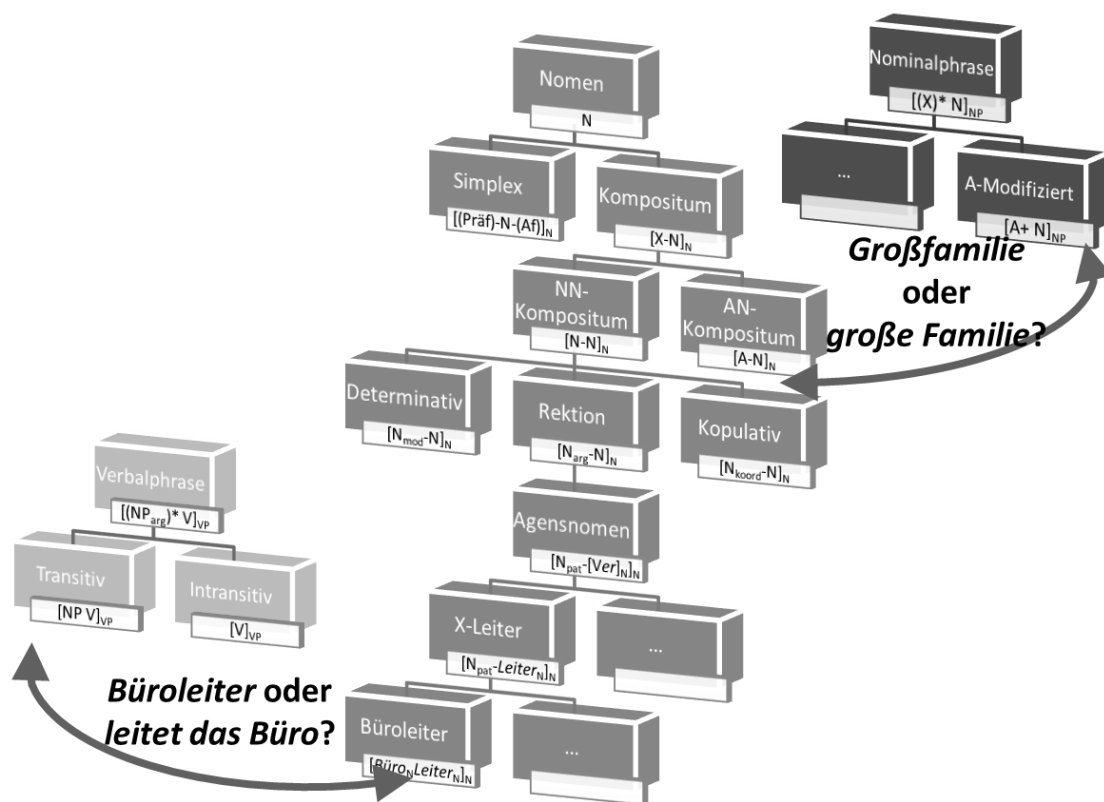


Abbildung 1. Komposition als Teil eines Konstruktionsnetzwerks. Doppelpfeile verbinden konkurrierende Konstruktionen außerhalb des Kompositionsnetzwerks. Die Zeichen ‘*’ und ‘+’ bezeichnen schematisch beliebig viele bzw. mindestens eine Wiederholung des vorangehenden Elements.

Obwohl die Knoten im Diagramm durch die zunehmende Abstraktion nach oben eher taxonomie- bzw. baumartig wirken, sollte man hier dennoch viel mehr von einem Netzwerk sprechen, da die genaue Zuordnung von Eltern- und Kinderknoten unterschiedlich betrachtet werden kann. So könnte man z.B. zunächst Rektionskomposita von Kopulativkomposita unterscheiden, und dann jede Gruppe in $[A-N]_N$ - bzw. $[N-N]_N$ -Komposita unterteilen. In einem Netzwerk-Modell fällt die Klassifikation solcher mehrfachen Gemeinsamkeiten leichter und man kann direkte Verbindungen zwischen allen Knoten zulassen, obgleich viele mögliche Pfade nicht oder kaum zum Ausdruck kommen.

Im Laufe der Untersuchung unten soll klar werden, dass alle Abstraktionsebenen wie im Beispiel (2) Teil des Erwerbsgegenstandes darstellen, den der Lerner bewältigen muss, um Komposition möglichst zielsprachennah zu verwenden, und dass solche mehr oder weniger abstrakten Konstruktionen wie im muttersprachlichen System auch im Sprachsystem des fortgeschrittenen Lernalters angenommen werden müssen (vgl. Gries & Wulff 2005, Haberzettl 2006 zu Konstruktionen bei Lernern). Dazu gehören der Erwerb der Konstruktionsebenen, eine Beherrschung der Abwechslungen zwischen ihnen und anderen relevanten Konkurrenten, sowie lexikalisches Wissen über Prototypen, welche die Erweiterung einer jeden Konstruktion motivieren und semantisch prägen.

2. Datengrundlage und Komposita-Extraktion

Die Untersuchung von Komposita im Gebrauch wird erst anhand einer empirischen Datengrundlage möglich, in der nicht nur Fehler, sondern auch Unterschiede in der korrekten Verwendung von Komposita bei DaF-Lernern zum Ausdruck kommen können. Die wohl größte derzeit verfügbare Datengrundlage, die sich zur Erforschung von Lerneräußerungen eignet, bildet das Lernerkorpus Falko (Reznicek et al. 2010), das geschriebene Sprache fortgeschrittener Deutsch-Lerner² in bestimmten Textsorten zur Verfügung stellt. In dieser Untersuchung soll das größte Falko-Teilkorpus verwendet werden, das sog. Essay-Korpus, das 248 Lerner-Aufsätze zu vier argumentativen Themen enthält,³ mit über 120.000 Wortformen (in der Version 2.2), sowie ein Vergleichskorpus mit Muttersprachlerdaten (95 Texte, über 68.000 Wortformen in der Version 2.2), die unter denselben Bedingungen und zu denselben Themen erhoben wurden. Für einige der Studien unten wurde noch das Ergänzungskorpus Falko-WHIG benutzt,⁴ das weitere L2-Essay-Daten umfasst (92 Texte, über 52.000 Wortformen in der Version 1.0), aber etwas weniger Annotationen enthält (s. unten).

² Der Sprachstand der Lerner wurde vor der Aufnahme in das Korpus anhand von einem c-test kontrolliert, s. Reznicek et al. (2010).

³ Die Aufgabenstellung beinhaltet eine allgemeine Erörterung von Themen wie die fehlende Praxisorientierung von Universitätsabschlüssen, Feminismus, Kriminalität, sowie Arbeit und Belohnung. Diese Themen bilden eine Untermenge der im ICLE-Korpus verwendeten Aufsatzthemen (*International Corpus of Learner English*, Granger et al. 2009).

⁴ Das Korpus wurde im Rahmen des Projekts WHIG (What's Hard in German?) in Zusammenarbeit mit der Universität Bangor erhoben, s. <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/lernschwierigkeiten/WHIG>.

Alle o.g. Korpora sind mit Wortartannotationen nach dem STTS-Tagset (Schiller et al. 1999) automatisch mittels des TreeTaggers (Schmid 1994) annotiert und manuell korrigiert worden, und enthalten zusätzlich zwei Zielhypothesenannotationen, die im Falle einer nicht standardsprachlichen Lerneräußerung eine möglichst minimale Korrektur vorschlagen. In der 1. Zielhypothese (ZH1) wird der Lernertext orthographisch, morphosyntaktisch (z.B. Kongruenz, Argumentstruktur) und in Bezug auf Wortstellung so korrigiert, dass ein minimal grammatischer Text entsteht, der nicht zwangsläufig einer sinnvollen zielsprachlichen Äußerung entspricht. In der 2. Zielhypothese werden auch semantische, pragmatische und stilistische Probleme behoben, sodass der Text für einen Muttersprachler vollständig akzeptabel werden soll. Abbildung 2 veranschaulicht dieses Verfahren und die Verhältnisse zwischen den Annotationsebenen (für eine detaillierte Besprechung der Zielhypothesen s. Reznicek et al., im Druck).

ZH1	in	Berufen	,	in	denen	man		Argumente	machen	und	schützen	muss
ZH1Diff					CHA		MOVS	CHA				MOVT
ZH2	in	Berufen	,	in	denen	man		Argumente	bringen	und	unterstützen	muss
ZH2Diff					CHA		MOVS	CHA	CHA		CHA	MOVT
tok	<i>in</i>	<i>Berufen</i>	<i>,</i>	<i>in</i>	<i>den</i>	<i>man</i>	<i>muss</i>	<i>Argumenten</i>	<i>machen</i>	<i>und</i>	<i>schützen</i>	

Abbildung 2. Zielhypothesen für ein Fragment einer Lerneräußerung. Der ursprüngliche Lernertext steht in der Zeile ‚tok‘ (=Token) unten. In ZH1 werden fehlerhafte Wortformen korrigiert (*Argumenten* > *Argumente*) um eine minimal grammatische Äußerung zu produzieren, während ZH2 auch Wortwahl und Semantik korrigiert (*schützen* > *unterstützen*).

Jede Zielhypothese wird selbst wiederum mit Wortarten versehen, sowie mit sog. Differenz-Tags, welche die Art der Abweichung vom Lernertext klassifizieren: eine orthographische Veränderung des Textes (CHA = Change), Hinzufügung eines Wortes (INS = Insertion), Entfernung eines Wortes (DEL = Deletion) und Bewegung (MOV = Move, bzw. MOVS für die Quelle der Bewegung und MOVT für das Ziel). So lassen sich bestimmte Arten von Korrektur oberflächennah und automatisch erfassen. Diese Annotationsebenen sind jeweils auch für die Lemma- und Wortartebenen vorhanden (im Falle einer Lemmakorrektur etc.) und werden unten in der Fehleranalyse verwendet.

Als letztes bietet das Essay-Korpus noch automatisch erstellte Abhängigkeitsanalysen zu jedem Satz (zum Parser s. Bohnet 2010). Diese wurden auf der Grundlage von ZH1 erstellt, da die Lerneräußerung selbst durch Fehler oft nur schwer parsebar ist. Obwohl die Analyse hier nicht unmittelbar auf den Lernerdaten basiert, wird unten gezeigt, wie sie in der Analyse der Lernerdaten dennoch sehr hilfreich sein kann. Das gesamte Korpus mit allen Subkorpora ist online durchsuchbar unter: <http://korpling.german.hu-berlin.de/falko-suche/>.

Da Komposition selbst im Korpus nicht explizit annotiert wurde, wurden zunächst alle als Substantiv getagte Wortformen (mit dem STTS-Tag „NN“) abgefragt und halbautomatisch weiterverarbeitet. Um Komposita zu erkennen wurden die Nomina mit einer sehr großen Liste von Simplizia (ca. 45.000 Lemmata) abgeglichen, die aus dem Lexikon der SMOR-Morphologie (Schmid et al. 2004)

sowie der Lemma-Liste des morphologisch annotierten Korpus TuePP-D/Z (s. Müller 2004) extrahiert wurde. Dann wurde versucht, jedes Nominallemma möglichst als Simplex in der Liste zu finden, oder andernfalls das längste Simplexnomen zu finden, das rechts im untersuchten Nomen als Kompositionskopf stehen könnte. Falls ein möglicher Kopf gefunden werden konnte, wurde noch überprüft, ob der aus der vorgeschlagenen Zerlegung resultierende Nicht-Kopf ausgeschlossen werden muss, bspw. dann, wenn der Nicht-Kopf ein bekanntes Präfix war (*Un-*, *Ver-* etc.) oder aus einer phonologisch ausgeschlossenen Sequenz bestand (z.B. *Zw-*, *Kr-*). Nach der automatischen Analyse wurden alle noch unanalysierten Nomina in den für die quantitative Studie unten (Abschnitt 4) verwendeten Daten aus dem Essay-Korpus heuristisch und manuell untersucht und korrigiert. Diese bestanden meistens entweder aus Orthographiefehlern oder aus nominalisierten Adjektiven, die im Nomenlexikon nicht eingetragen waren. Auch einige Tagging-Fehler waren in dieser Gruppe enthalten. Die Ergebnisse der korrigierten Analyse der Essay-Daten, die die Datengrundlage für die weiteren Untersuchungen unten darstellt, sind in Tabelle 1 unter den automatisch extrahierten Zahlen zusammengefasst.⁵

		Essays L1	Essays L2	WHIG	Gesamt
automatisch	Simplizia	9451	18239	8624	36314
	Komposita	1393	1981	1158	4532
	unbekannt	593	1283	781	2657
	Gesamt	11437	21503	10563	43503
korrigiert	Simplizia	9789	19067	--	28856
	Komposita	1476	2113	--	3589
	Tagging-Fehler	172	323	--	495

Tabelle 1. Übersicht der Komposita-Extraktion in den verschiedenen Subkorpora.

Diese Zahlen geben jedoch bisher weder Auskunft darüber, was für Komposita in den beiden Datensätzen belegt sind, noch darüber, ob alle Komposita in den Lernerdaten grammatisch sind, und falls nicht, warum. Bevor wir zur quantitativen Auswertung der Kompositionstypen und ihrer Häufigkeiten in Abschnitt 4 kommen, soll daher im nächsten Abschnitt die Fehleranalyse zur Grammatikalität der L2-Komposita anhand der Zielhypothesenannotation durchgeführt werden.

3. Fehleranalyse

Bevor die Verwendung der Lernerkomposita mit dem muttersprachlichen Gebrauch verglichen werden kann, soll zunächst die Frage beantwortet werden, inwiefern fortgeschrittene Lerner Komposition korrekt einsetzen. Welche Fehler treten bei der Bildung von Komposita auf und wie oft? Bilden Komposita einen besonders

⁵ Die Rohdaten aus der Komposita-Analyse werden online unter dem folgenden Link zur Verfügung gestellt: http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/mitarbeiterinnen/amir/amir_data/.

schwierigen Bereich? Ist L2-Komposition qualitativ hinreichend ähnlich wie muttersprachliche Komposition, um einen sinnvollen Vergleich zu ermöglichen? Auf diese Fragen soll nun anhand der oben beschriebenen Zielhypothesen eingegangen werden.

Auf der Ebene der minimal grammatischen Zielhypothese ZH1 lassen sich in erster Linie orthographische Fehler erkennen, bzw. allgemeiner gesagt Fälle, in denen das geschriebene Wort durch einen Annotator für ungrammatisch und korrekturbedürftig gehalten wurde. Im L2-Essay-Korpus lassen sich auf diese Weise 218 fehlerhafte Wortformen unter den 2113 Komposita im L2-Essay-Korpus identifizieren, d.h. ein Fehleranteil von 10,31%. Bei näherer Betrachtung der Fehlertypen, die in der Tabelle 2 aufgeführt werden, fällt jedoch auf, dass einige Fehlertypen nicht direkt mit Komposition als Wortbildungsprozess zusammenhängen.⁶

Fehlertyp	Frequenz	% Komposita	% Fehler
<i>Derivation</i>	9	0,43%	3,96%
<i>Lexikalisch</i>	12	0,57%	5,29%
<i>Flexion</i>	20	0,95%	8,81%
<i>Orthographie (Modifikator)</i>	30	1,42%	13,22%
<i>Orthographie (Kopf)</i>	53	2,51%	23,35%
<i>Fugenfehler</i>	103	4,87%	45,37%

Tabelle 2. Verteilung der Fehlertypen in den L2-Komposita.

Orthographische Fehler außerhalb der Fugenelemente erscheinen sowohl bei den Köpfen als auch bei den Modifikatoren und bilden insgesamt 36,57% der Fehler. Diese Fehlertypen sind allerdings kein besonderes Merkmal von Komposita, da Orthographiefehler in den entsprechenden Kompositionsgliedern auch als Simplizia erscheinen können.⁷

- (3) *Bacherler-Fakaltät* (L1=jpn), vgl. Simplex *Bachellor* (L1=ell)
- (4) *Fremdsprachenkenntniss* (L1=ces), vgl. Simplex *Kennnissen* (L1=ita)

Gleichermaßen zeichnen sich Derivations- und Flexionsfehler durch keine besonderen mit der Komposition verbundenen Eigenschaften aus, wie folgende Beispiele zeigen:

- (5) *Arbeitssuchung* (L1=pol), vgl. Simplex *Trainung* (L1=eng)
- (6) *Frauen [...], die immer noch für ihre ersten **Rechten** kämpfen* (L1=fra), vgl.: [...] *haben Frauen **Wahlenrechten** [...] bekommen* (L1=eng)

⁶ Die Zahlen in der Tabelle kommen auf insgesamt 227 Fehler, was durch die Möglichkeit des Auftretens eines Kompositums in mehreren Fehlerkategorien gleichzeitig zustande kommt.

⁷ In den Beispielen wird die erste Muttersprache des Schreibers immer mit der Drei-Buchstaben-Kennung nach der Norm ISO 639-3 gekennzeichnet. Eine Übersicht der hier vorkommenden Abkürzungen findet sich im Anhang A.

Von besonderem Interesse sind hier deshalb die lexikalischen Fehler, die durch eine fehlerhafte Kombination von Lexemen zu fehlerhaften Kompositabildungen führen, sowie die Fehler in den Fugenelementen, die wiederum rein formale Fehler im Kompositionsprozess selbst darstellen. Beispiele für die erste Gruppe stehen in (7)-(9).

- (7) [...] oft befindet sich das Studiumort nicht in der ^{ZH1=Heimatstadt} Heimstadt (L1=ukr)
- (8) Die Frage ist leicht zu beantworten, in der ^{ZH1=Finanzwelt} Finanziellewelt muss mann sehr gut ausgebildet sein (L1=eng)
- (9) [...] ein Kampf nach ^{ZH1=Menschenrechte} Menschlicherechte (L1=fra)

Dieser Fehlertyp ist in ZH1 mit nur 12 Vorkommen insgesamt eher selten belegt, was darauf hinweist, dass die Lerner insgesamt nur selten ungrammatische bzw. blockierte Kompositionsprodukte produzieren, die von einem Annotator für unmöglich gehalten wurden.⁸ Wie die Beispiele oben zeigen, kann dies an der Wahl der Lexeme liegen (*Heim* statt *Heimat*) oder an Wortartbeschränkungen, wie die beiden fehlerhaften Komposita in (8) oder (9). Hier ist wichtig anzumerken, dass der Äußerung nicht unbedingt ein Flexions- bzw. Zusammenschreibungsfehler zugrunde liegt. Da die Lerner sonst auf Groß- bzw. Kleinschreibung und Zusammenschreibung achten (vgl. *Frage*, *Kampf* oben), kann man den Beleg *Finanziellewelt* durchaus als [A-N]_N-Kompositum mit einem Fugenelement *-e* analysieren, wobei die Wahl des Elements *-e* nichtsdestotrotz u.a. durch das homonyme Flexionssuffix motiviert sein kann. Eine ähnliche Analyse wäre auch für *Menschlicherechte* naheliegend.

Weitaus häufiger als lexikalische Fehler (ausgeschlossenes Kompositionsglied) sind die Fugenelementer, die in der Abbildung 3 quantitativ verglichen werden und in der Konfusionsmatrix in der Tabelle 3 noch genauer betrachtet werden können.⁹

⁸ Hierzu muss man jedoch sagen, dass ‚merkwürdige‘ Komposita oft erst in der ZH2 korrigiert wurden, wenn sie als Fehler erst auf der stilistischen oder semantischen Ebene angesehen wurden, aber an sich nicht als unmögliche Wortbildung gelten (zur Unterscheidung zwischen ZH1 und ZH2 s. detailliert in Reznicek et al. 2010).

⁹ Da ein Fehler gleichzeitig fehlendes und überflüssiges Material enthalten kann, kommen die Fälle in der Tabelle auf insgesamt mehr als 103, die Anzahl der Wortformen mit Fugenelementern. Um keine Information zu verlieren, folge ich Neef (2009) in der Unterscheidung aller Formen, einschließlich komplementär verteilter Typen (*-n-* bzw. *-en-* als Allomorphe von *-(e)n-* etc., vgl. Becker 1992). Allerdings werden nur die Typen aufgezählt, für die Fehler belegt sind: so finden sich in den Daten keine Fälle mit Umlaut (*Mütter-* bzw. *Häuser*) oder *-ns-*. Der eine Fall einer fehlerhaften *er-*-Fuge betrifft die Form *Geschlechⁿtenrollen*, d.h. *en-*-Fuge statt reiner *er-*-Fuge.

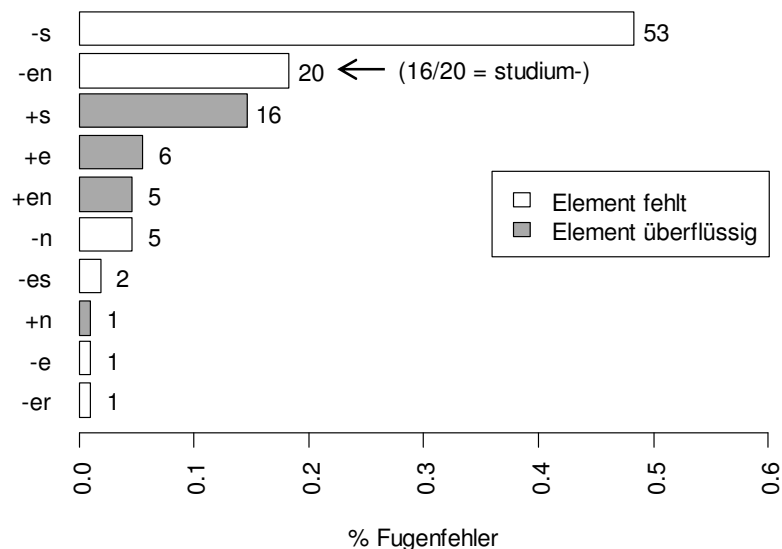


Abbildung 3. Fugenfehler in den Lerner-Komposita. Bei fehlenden *en*-Fugen handelt es sich in 16/20 Fällen um die Form *Studium-* bzw. *Studiums-* anstelle von *Studien-*.

		soll							DEL ¹⁰
		<i>s</i>	<i>es</i>	<i>e</i>	<i>en</i>	<i>n</i>	<i>er</i>	∅	
ist	<i>s</i>	0	0	0	3	0	0	14	0
	<i>es</i>	0	0	0	0	0	0	0	0
	<i>e</i>	2	0	0	0	3	0	3	0
	<i>en</i>	0	2	0	0	0	0	3	0
	<i>n</i>	0	0	0	0	0	1	0	0
	<i>er</i>	0	0	0	0	0	0	0	0
	∅	51	0	1	16	5	0	0	0
	DEL	0	0	1	1	0	0	0	0

Tabelle 3. Konfusionsmatrix der Fugenelemente.

Die Daten zeigen, dass fast die Hälfte der Fehler durch ein fehlendes *-s* verursacht werden, und ein weiteres Drittel der Fehler durch fehlendes *-en* bzw. ein überflüssiges Fugen-*s* zustande kommt. Da die *-en*-Fälle zu über 75% aus dem Vorderglied *Studium-* statt *Studien-* bestehen, was durch das Aufsatzthema über Universitätsabschlüsse unverhältnismäßig häufig wird, kann man sagen, dass Probleme mit dem Fugen-*s* eindeutig die häufigsten sind.¹¹ Die Ursache dieser Fakten

¹⁰ Die Kategorie DEL bezieht sich auf die subtraktive Fugenmarkierung, bspw. *Wollknäuel* (< *Wolle*).

¹¹ Die sonst niedrige Fehlerrate bei den Lernern zeigt, dass sie gerade die konsistenten Regeln der Fugenwahl in vielen Fällen beherrschen (vgl. Hartkamp & Schneider-Wiejowski 2010 zu mehr und weniger konsistenten Regeln der Verfügun). Das Fugen-*s* wurde auch im muttersprachlichen Gebrauch als einerseits derzeit am produktivsten, andererseits in der Alternation mit der Nullfuge am variabelsten bezeichnet (s. Michel 2009, Nübling & Szczepaniak 2011). Die nach den neusten Ergebnissen von Nübling und Szczepaniak für die Variation verantwortlichen Faktoren, wie ein unbetontes Präfix im Vorderglied bzw. eine nicht-trochäische oder einsilbige Struktur, zeichnen die Lernerfehler jedoch nicht aus (vgl. auch Kürschner 2005 zum Einfluss von morphologischer Komplexität bzw. Derivation bei der Fugenwahl). Allerdings könnte die von Nübling und Szczepaniak identifizierte Variabilität für die erhöhte Fehlerrate durchaus mitverantwortlich sein.

kann anhand der vorliegenden Daten nicht eindeutig nachgewiesen werden, es fällt jedoch auf, dass das Vorderglied in den meisten Fällen auf einem alveolaren Plosiv auslautet, d.h. mitunter durch Auslautverhärtung ein phonetisches [t]:

		[t]	[f]/[pf]	[r]	[m]	[l]	[k]/[ŋ]	[n]	[s]/[ts]
<i>s überflüssig</i>	<i>Token</i>	10	1	1	2	1	1	0	0
	<i>Typen</i>	8	1	1	1	1	1	0	0
<i>s fehlt</i>	<i>Token</i>	38	2	0	0	2	5	5	1
	<i>Typen</i>	10	1	0	0	2	3	3	1

Tabelle 4. *s*-Fugenfehler nach Auslautkonsonant des Vorderglieds.

Hier kann man sich vorstellen, dass solche nicht vorhersagbaren Unterschiede wie *Projekt-* in *Projektarbeit* neben *Arbeits-* in *Arbeitsprojekt* (aber *Arbeit-* in *Arbeitgeber*) den Lerner verwirren, während andere Umgebungen konsistenter sind, gerade da, wo Suffixe zu konsequenten Fugenelementen führen (bspw. *-ungs-*). Hinzu kommt, dass ein Sibilant hinter Alveolar perzeptorisch weniger prominent sein dürfte, als es nach artikulatorisch weiter entfernten Konsonanten der Fall ist. Eine mögliche didaktische Konsequenz dieser Analyse wäre, dass es sich lohnen könnte, Lerner explizit auf die Problemfälle des Fugen-*s* aufmerksam zu machen, insbesondere in der Umgebung nach [t]. Hätte man allein die Fehler im Muster [t±s#] und das Problem mit dem Lexem *Studien-* beheben können, so würden 62% der Fugenfehlertoken und 39% der Fugenfehlertypen verschwinden. Wenn man jedoch die Fehlerrate unter den Komposita insgesamt betrachtet und v.a. noch von den für Komposition nicht direkt relevanten Fehlertypen absieht, stellt man ohnehin fest, dass der Fehleranteil bei den Komposita nicht wirklich groß ist, sprich, fortgeschrittene DaF-Lerner scheinen hier wenige Fehler zu machen.

4. Die oberste Ebene des Konstruktionsnetzwerks – Komposition und konkurrierende Konstruktionen

4.1 Häufigkeit

Nachdem wir festgestellt haben, dass fortgeschrittene Lerner Komposita im Allgemeinen korrekt verwenden, können wir nun der Frage nachgehen, ob Lerner Komposition so oft einsetzen wie Muttersprachler, sowie ob es Auffälligkeiten in der Verteilung bzw. in den Umgebungen gibt, die bei der Komposition in den Lernaufsätzen vorkommen. Ein solcher Vergleich kann nur im Rahmen der sog. „Contrastive Interlanguage Analysis“ (CIA, s. Granger 2002) stattfinden, in dem davon ausgegangen wird, dass dem Sprachgebrauch der Lerner ein mindestens teilweise konsistentes System (die sog. „Interlanguage“, s. Selinker 1972) zugrunde liegt, wodurch Abweichungen vom muttersprachlichen Gebrauch eine entsprechende Systematik aufweisen. Hierbei könnte man evtl. erwarten, dass Komposition als komplexer Prozess von den meisten Lernern mindergebraucht wird, entweder weil seine Komplexität zu mehr oder weniger bewusster Vermeidung führt oder weil Komposita schwieriger zu erwerben sind und entsprechend seltener produziert werden, als in dem zu erlernenden Sprachgebrauch (vgl. Zeldes et al. 2008 zu

Mindergebrauch als Hinweis auf Schwierigkeitsgrad). Wie die ersten Daten zum Extraktionsverfahren in der Tabelle 1 angedeutet haben, kommen Komposita unter den Lernernomina tatsächlich signifikant seltener vor. Die Normalisierung für den Vergleich mit den Muttersprachlern bezieht sich hier auf den Anteil der Komposita unter allen Nomina.¹²

Im L1-Korpus sind 13,1% der Nomina Komposita, im L2-Korpus nur 9,97%. D.h., die Lerner bilden im Schnitt etwa drei Komposita weniger für alle 100 Nomina, bzw. sie lassen ca. 23% (3/13) der im muttersprachlichen Gebrauch belegten Komposita weg. Wenn man alle Lerner mit allen Muttersprachlern als zwei homogene Gruppen vergleicht, ist der Unterschied hoch signifikant ($\chi^2=72.73$, $p<2.2e-16$). Ob dieser Vergleich berechtigt ist, hängt jedoch von der Frage ab, ob die Lerner sich einheitlich genug verhalten, um diesen Vergleich zu ziehen. Eine Varianzanalyse (ANOVA) belegt, dass die Unterschiede auch bezogen auf die einzelnen Sprecher immer noch hoch signifikant sind ($p<9.038e-05$). Allerdings zeigt Abbildung 4, dass die Streuung im Anteil der Komposita unter den Lernern relativ hoch ist: es gibt Lerner, die auffällig mehr oder auch weniger Komposita produzieren, während der Schwankungsbereich in den L1-Daten kleiner ist.

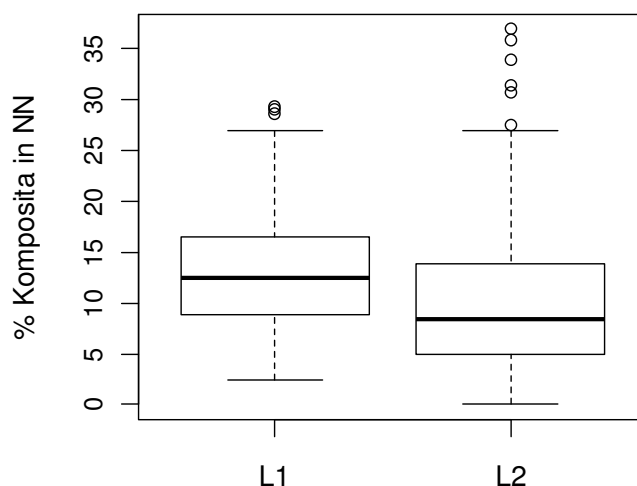


Abbildung 4. Boxplot des Kompositaanteils in L1- und L2-Texten. Die fetten Linien geben den Medianwert für den Kompositaanteil an, die unteren und oberen Kanten der Rechtecke jeweils das 1. bzw. 3. Quartil. Die gestrichelten Linien decken Ergebnisse im Bereich 1,5*Interquartil-Abstand, einzelne Ergebnisse jenseits dieser Linien werden mit Kreisen als Ausreißerwerte gesondert markiert.

Eine naheliegende Hypothese wäre, dass Kompositionshäufigkeit evtl. durch die Eigenschaften der Muttersprache bzw. der Sprachfamilie des jeweiligen DaF-Lerners beeinflusst wird, d.h., dass der Kompositionsseltenheit eine Art Interferenzphänomen zugrunde gelegt werden könnte. Die Unterteilung der Häufigkeitsdaten in die unterschiedlichen Muttersprachen im Korpus wird in der

¹² Andere Normalisierungen wären auch denkbar, wie bspw. Komposita pro Satz, pro Text, pro tausend Wortformen, etc. In diesem Fall bietet sich aber der Vergleich mit allen Nomina an, da die L2-Texte bzw. -sätze tendenziell etwas kürzer sind als die L1-Äquivalenten, und auch die wortformbasierte Normalisierung durch unterschiedliche Nominaldichte beeinflusst werden kann. Diese Überlegungen bedingten die Entscheidung, Komposition im Vergleich zu allen Nomenbildungen im L1- bzw. L2-Korpus zu untersuchen.

Abbildung 5 für die 15 am besten belegten Muttersprachen dargestellt (weitere Muttersprachen haben jeweils nur ca. 400 Nomina bzw. 5 Texte oder weniger und liefern deshalb weniger zuverlässige Ergebnisse).

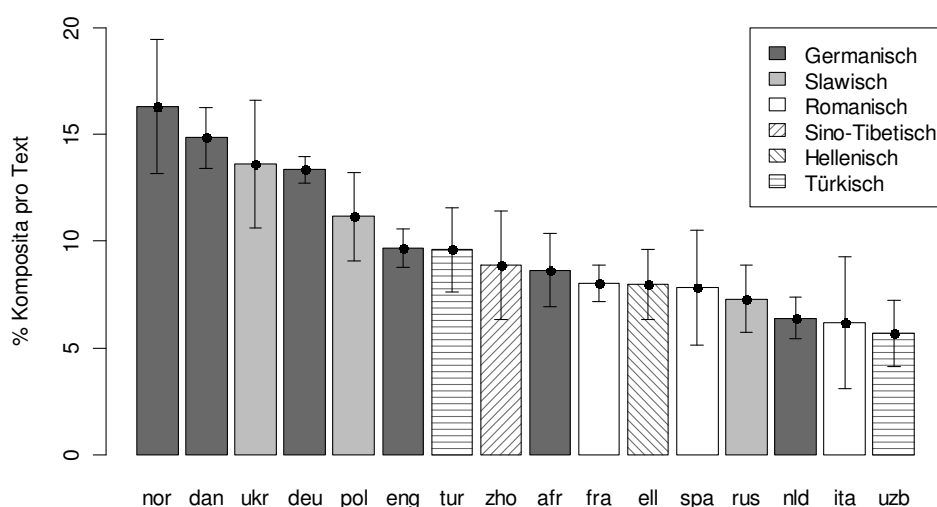


Abbildung 5. Durchschnittlicher Anteil der Komposita nach Muttersprache des Schreibers. Der Standardfehler wird durch Fehlerbalken über und unter jedem Balken angegeben. Der Farbton bzw. die Textur in jedem Balken kodiert die Sprachfamilie der jeweiligen Muttersprache.

Wie die Abbildung zeigt, lässt sich kein offensichtlicher Zusammenhang zwischen der Sprachfamilie der jeweiligen Muttersprache und der Kompositionshäufigkeit feststellen. Zwar findet man verhältnismäßig kompositionsfreudige germanische Sprachen wie Norwegisch und Dänisch auf der linken Seite, mit etwas mehr Komposita (1-2%, $p=0.03662$), als bei den Muttersprachlern (4. Balken), aber andere germanische Sprachen wie Afrikaans und insbesondere das Niederländische haben weitaus weniger Komposita. Ukrainisch und Polnisch unterscheiden sich nicht signifikant von den deutschen Muttersprachlern, während man ab dem Englischen nach rechts wieder sehr signifikant weniger Komposita in allen L1 findet ($p<0.0001$), einschließlich der weiteren slawischen Sprache Russisch. Aus Sicht der Sprachfamilien verhalten sich nur die nicht sehr kompositionsfreudigen romanischen Sprachen auf der rechten Seite vergleichbar (kein signifikanter Unterschied), da Türkisch und Usbekisch trotz ihrer Verwandtschaft wieder auch signifikante Unterschiede aufweisen ($p=0.008506$). Darüber hinaus sieht man, dass größere Fehlerschwankungen bei den Lerner Mittelwerten erwartet werden, im Unterschied zum vergleichsweise sicheren Muttersprachler Mittelwert (kleinerer Fehlerbereich).

Angesichts dieser Ergebnisse stellt sich die Frage, ob die individuelle Sprachkompetenz des Lerners vielleicht einen größeren Einfluss auf den Anteil der Komposita haben könnte. Diese Kompetenz lässt sich anhand der Daten objektiv schwer abschätzen, zumal viele leicht messbare Faktoren (durchschnittliche Satzlänge, Textlänge und andere Komplexitätsmaße, s. Lu 2010) auch selbst mit der Verwendung von Komposita zusammenhängen (Komposita können z.B. die Satzlänge in Wörtern bei gleichem semantischen Inhalt reduzieren). Die einzige textunabhängige Möglichkeit der Kompetenzeinschätzung liefern die C-Test-Werte,

die für jeden Lerner als Metadaten erhoben wurden, um den fortgeschrittenen Sprachstand zu überprüfen.¹³

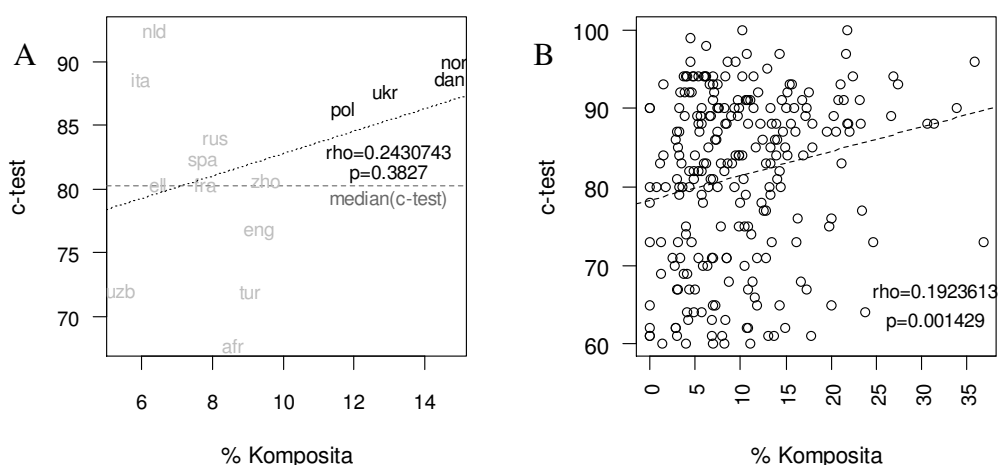


Abbildung 6. Keine signifikante Korrelation zwischen Kompositionsmittelwert und C-Test-Mittelwert pro Sprache (A) und eine schwache Korrelation zwischen Kompositionshäufigkeit und C-Test-Wert pro Lerner (B).

Eine signifikante Korrelation zwischen den C-Test-Mittelwerten für jede L1-Gruppe und der Kompositionsrate besteht nicht (Abbildung 6A). Zwar findet man die Sprachen mit den meisten Komposita oben rechts im höheren C-Test-Werte-Bereich, die Streuung der restlichen Sprachen führt aber insgesamt zu keinem signifikanten Ergebnis. Eine bessere Korrelation ($p=0.001429$) findet man zwischen C-Test-Werten und Kompositionsraten für einzelne Sprecher, unabhängig von der Muttersprache, obgleich diese relativ schwach ist ($\rho=0.1923613$),¹⁴ wie auch die erhebliche Streuung der Datenpunkte in Abbildung 6B erkennen lässt.

Insgesamt ist also ein signifikanter Mindergebrauch der Komposita bei den fortgeschrittenen Lernern nachweisbar, obwohl gleichzeitig große Schwankungen von Lerner zu Lerner feststellbar sind. Es gibt keine überzeugende Evidenz, die einen Einfluss der typologischen Eigenschaften der Muttersprache belegt (eine mögliche Ausnahme bilden die Daten aus den komposita-armen romanischen Muttersprachen). Die Kompetenz des Lerners, so wie sie mit dem C-Test gemessen werden konnte, hat ggf. einen schwachen Einfluss auf die Kompositionsrate, wobei die Korrelationsstärke verhältnismäßig gering ist.

4.2 Konkurrierende Konstruktionen

Wenn Komposition in den Lernerdaten insgesamt deutlich seltener ist, stellt sich als nächstes die Frage, welche sprachlichen Mittel statt der Komposition verwendet werden, um vergleichbare kommunikative Inhalte auszudrücken. Kompensieren die Lerner die Abwesenheit von Komposita durch andere Konstruktionen bzw. treffen sie andere Entscheidungen, die auf die Konkurrenzbeziehungen zwischen Konstruktionen

¹³ Gleichzeitig muss man sagen, dass die c-test-Werte im Falco-Korpus nicht unbedingt mit subjektiv eingeschätzter Textqualität korrelieren und für die Auswertung von Sprachstand in Bezug auf Wortbildungsstrategien nicht zwangsläufig aussagekräftig sind.

¹⁴ Ich verwende hier die Spearman-Korrelation mit entsprechenden rho-Werten statt der üblicheren Pearson-Korrelation mit r^2 -Werten, da die c-test-Werte nicht normalverteilt sind.

in ihrer Interlanguage schließen lassen? Im Idealfall würde man überprüfen wollen, ob spezifische Kompositionsköpfe und -modifikatoren bei den Lernern in anderen syntaktischen Konstruktionen häufiger vorkommen, als bei den Muttersprachlern, z.B. ob man neben *Holztisch* auch *Tisch aus Holz* häufiger findet als erwartet (vgl. Gaeta & Zeldes 2012 zur Häufigkeit solcher Fügungen in einem großen Web-Korpus bezogen auf Kompositionstypen). Leider lässt die Datenlage dieses Vorgehen nicht zu: Es gibt viel zu wenige Belege von jedem Lexempar und die meisten sind in diesem kleinen Korpus außerhalb von Komposita nicht belegt. Als Alternative können wir jedoch grundsätzlich abstraktere Konstruktionen aus den oberen Strukturebenen des Konstruktionsnetzwerks betrachten, welche naturgemäß auch häufiger sind.

Einige naheliegende Möglichkeiten, kompositionsähnliche semantische Relationen auf phrasaler Ebene auszudrücken, sind attributive Adjektive (*große Familie* statt *Großfamilie*), Präpositionalphrasen (*Holztisch* : *Tisch aus Holz*), Genitivattribute (*Hauswand* : *Wand des Hauses* bzw. als *von-PP* *Wand von dem Haus*) und Relativsätze (*Wohnort* : *Ort, an dem man wohnt*).¹⁵ Die Häufigkeit der ersten zwei Kategorien lässt sich mit der Wortartannotation im Korpus relativ leicht ermitteln, während die letzten zwei Konstruktionen anhand der Dependenzbäume im Korpus untersucht werden können. Die Tabelle 5 gibt einen Überblick über die absoluten und normalisierten Frequenzen der Kategorien in den L1- bzw. L2-Daten im Vergleich.

	f(L1)	norm(L1)	f(L2)	norm(L2)	Übergebrauch?	p-val
<i>pos=ADJA</i>	2759	0,04029	5068	0,0412801	2,45%	ns
<i>deprel=AG</i>	771	0,011259	1325	0,0107925	-4,20%	ns
<i>deprel=RC</i>	843	0,012311	1292	0,0105237	-14,52%	0,00039
<i>pos=APPR</i>	4173	0,060939	7014	0,0571308	-6,25%	0,00069
<i>von</i>	408	0,005958	836	0,0068094	14,29%	0,02848
<i>¬von</i>	3765	0,054981	6178	0,0503213	-8,48%	0,00001

Tabelle 5. Frequenzen für einige mit der Komposition konkurrierende Konstruktionen: Attributive Adjektive (ADJA), Präpositionen (APPR), Genitivmodifikatoren (AG) und Relativsätze (RC).

Die Wortart ADJA (attributives Adjektiv) und die Dependenzrelation AG (Attribut, Genitiv) zeigen keine signifikante Abweichung in den Daten. Relativsätze (RC steht für ‚Relative Clause‘) und Präpositionen (APPR=Adposition, präpositional) hingegen sind beide mindergebraucht, d.h. sie kompensieren die fehlenden Komposita nicht, sondern stellen ein weiteres Problem in der Vermittlung von Modifikationsrelationen in den Lernertexten dar. Wenn man allerdings die Präposition *von* getrennt untersucht, die der allgemeinen Semantik der Komposition auf der syntaktischen Ebene am nächsten kommt, findet man einen signifikanten Übergebrauch von 14,29%. Auch wenn dieser noch bescheidene Übergebrauch die gesamte Seltenheit von Komposita nicht erklären kann, kann man versuchen, Fälle zu finden, in denen eine *von*-Phrase evtl. anstelle der Komposition verwendet wurde. Anhand der 2. Zielhypothese (ZH2), in der auch stilistische Korrekturen sonst

¹⁵ Zur Konkurrenz zwischen Komposition, Genitivattributen und Präpositionalphrasen als Modifikatoren s. auch Lauterbach (1993:106ff)

grammatischer Sätze vorgenommen wurden, kann man solche Beispiele finden (eine Suche nach *von* auf der Text-Ebene, wo ZH2 ein Kompositum hat):

- (10) *Oft in Zeit von Kriegen [...] > ZH2: Kriegszeiten (L1=eng)*
- (11) *[...] schlechte Benützung von Zeit und Geld >
ZH2: Zeit- und Geldverschwendung (L1=afr)*
- (12) *[...] eine Position von Autorität > ZH2: Führungsposition (L1=eng)*

Diese vereinzelt Belege zeigen, dass Konkurrenzphänomene durchaus für fehlende Komposita verantwortlich gemacht werden können. Eine substantielle Ersetzungsstrategie für Komposition bieten sie jedoch quantitativ nicht an, denn die Lernerdaten enthalten 23% weniger Komposita als die Muttersprachlerdaten, was nur durch mehrere hunderte solcher Belege verursacht werden könnte, wenn die Konkurrenz durch *von*-Phrasen die Verantwortung für den Mindergebrauch getragen hätte. Daraus können wir deshalb schließen, dass die Lerner entweder andere, bisher unerkannte Strukturen verwenden, um die Erweiterungssemantik der Komposita zu ersetzen, oder dass die Lerner grundsätzlich weniger Modifikationen vornehmen. Letztere Vermutung wird auch durch Studien zum Mindergebrauch syntaktischer Modifikation (etwa von adverbialen Modifikatoren) in den Falko-Daten unterstützt, s. Hirschmann et al. (eingereicht).

4.3 Produktivität

Die bisherigen Untersuchungen haben uns nur gezeigt, wie sich Lerner und Muttersprachler im Einsatz von Komposition unterscheiden, ohne jedoch auf die Frage einzugehen, ob die erhobenen Komposita produktiv gebildet wurden oder lediglich vertraute Kombinationen darstellen, die u.U. auch als ganze Wörter im mentalen Lexikon gespeichert werden. Bei stark lexikalisierten Einheiten könnte man annehmen, dass auf das Kompositum zugegriffen werden kann, ohne allgemeine Regeln bzw. Muster der Art $[N-N]_N$ anzuwenden.¹⁶ Eine direkte Zählung der produktiven Bildungen ist deshalb nicht möglich, weil hinreichende und notwendige Bedingungen für die eindeutige Erkennung von Neubildungen im Einzelfall nicht vorhanden sind (vgl. Bauer 2001:34ff). Als Diagnostiken für produktives Verhalten hat man daher in der morphologischen Produktivitätsforschung operationalisierbare Maße entwickelt, die u.a. auf Faktoren wie Wortschatzgröße und dem Anteil seltener lexikalischer Typen basieren (s. Baayen & Lieber 1991, Baayen 2001, 2009). Letztere betreffen v.a. die Anzahl sog. *Hapax Legomena*, d.h. Typen, die im Korpus nur einmal belegt sind. Diese entsprechen nicht zwangsläufig Neologismen, werden aber mit zunehmender Korpusgröße immer weniger wahrscheinlich vertraute Lexeme sein, sondern eher nicht-lexikalisierte Spontanbildungen, bspw. die Hapax Legomena:

¹⁶ Trotzdem würde man im CxG-Framework mindestens bei transparenten Komposita annehmen wollen, dass solche lexikalisierten Vertreter die Kompositionsstruktur aktivieren bzw. stärken und als Prototypen des Prozesses seine Semantik beeinflussen.

- (13) *Topjobs* (L1=eng),
Arbeitsbildungssystem (L1=fra),
Rentnerleben (L1=dan),
 ...

Das Verhalten des Wortschatzes unter Lernern und Muttersprachlern lässt sich am einfachsten anhand von Wortschatzwachstumskurven (s. Baayen 2001, Evert & Baroni 2007) vergleichen, wie in Abbildung 7. Die Abbildung zeigt die Entwicklung der Anzahl der Lemmatypen V (für Vokabular) mit wachsender Stichprobengröße N . Wie man sieht, übertrifft die Muttersprachlerkurve den Wert der Lernerkurve an jeder Stelle. Die Signifikanz des Unterschieds und die Effektstärke dürfen nur bei einer gleich großen Stichprobe erhoben werden, da es in größeren Stichproben immer schwieriger wird, neue Vokabeln zu finden (s. Gaeta & Ricca 2006). Diese Überprüfung darf hier bei $N=1393$ stattfinden, der Größe des kleineren L1-Datensatzes, und ergibt einen sehr signifikanten Unterschied bei $p=4.809e-14$ mit einer Effektstärke von $\phi=0.2019401$.¹⁷ Das Verhältnis von Typen zu Token (Type-Token-Ratio, TTR) beträgt 56,1% bei den Lernern gegenüber 69,9% bei den Muttersprachlern, bzw. einen Unterschied von $V=782$ gegenüber $V=975$. Die Muttersprachler haben somit einen 24,6% größeren Wortschatz in ihrem Kompositabestand für die hier erhobene Stichprobengröße.

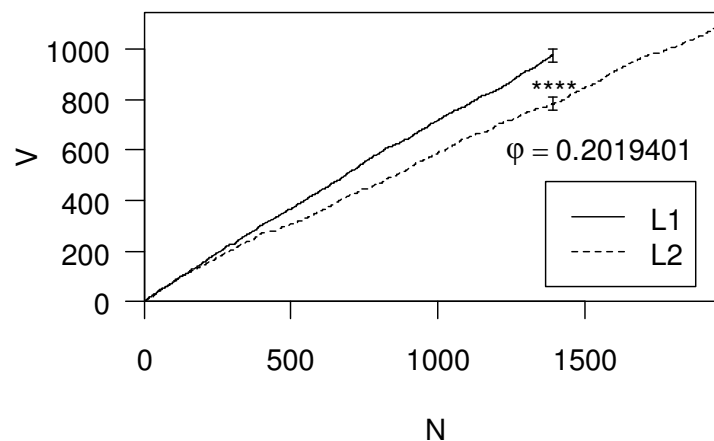


Abbildung 7. Wortschatzwachstumskurven für Vokabulargröße V in L1- und L2-Daten. Der Vergleich der Kurven geschieht am Ende der L1-Daten bei $N=1393$. Die Fehlerbalken an dieser Stelle geben Konfidenzintervalle für den tatsächlichen Wert von V mit 95% Wahrscheinlichkeit an.

Diese Beobachtung könnte man zumindest teilweise mit der bereits bekannten Tatsache erklären, dass Lerner tendenziell den häufigsten Vertreter einer Kategorie übergebrauchen und seltenere Varianten noch seltener verwenden als Muttersprachler (vgl. Ringbom 1998, der einen Übergebrauch der häufigsten Typen in einigen englischen Kategorien feststellen konnte, wie bspw. *think* als häufigstes bzw. bei

¹⁷ Hier wurden nur die 1393 korrekt lemmatisierbaren L1-Komposita mit den 1981 lemmatisierbaren L2-Komposita verglichen, da die übrigen 83 bzw. 132 ggf. durch Orthographiefehler die Typenanzahl aufblähen (bspw. *Universitätsabschluss*, *Universitätsabschlüs* neben *Universitätsabschluss* usw.). Allerdings würde deren Berücksichtigung nur wenig an den Ergebnissen ändern ($\phi \sim .19$), zumal auch die Muttersprachler-Daten nicht fehlerfrei sind (z.B. *Universitätsabschluss*).

Lernern zu häufiges Verb des Denkens). Die gleiche Beobachtung wurde auch systematisch von Ellis & Ferreira-Junior (2009) als Eigenschaft der typischen Verteilung sowohl von Vokabeln als auch von abstrakten grammatischen Konstruktionen bei Lernern erfasst, eine Verteilung, die die Daten auf eine kleine Auswahl von Typen mit übermäßig vielen Token für die „Lieblingsvokabeln“ verteilt, und auffallend wenige seltene Typen beinhaltet.

Ein solches Verhalten würde auch einen ähnlichen Unterschied bei der Anzahl der Hapax Legomena VI vorhersagen (VI steht für Vokabular mit Frequenz 1). Auch dieser Unterschied lässt sich tatsächlich beobachten: Hier erreichen die Lerner $VI=612$ gegenüber $VI=795$ bei den Muttersprachlern, ein signifikanter Unterschied ($p=5.328e-12$, $\phi=0.184781$) mit 29,9% mehr Hapax Legomena für die Muttersprachler (s. Abbildung 8). Darüber hinaus kann der Anteil der Hapax Legomena VI/N als Schätzer für die Wahrscheinlichkeit einer bisher ungesesehenen Neubildung verwendet werden (s. Baayen & Lieber 1991): damit besteht eine Wahrscheinlichkeit von 57% für einen neuen Typ nach 1393 Token bei den Muttersprachlern, aber nur 43,9% bei den Lernern.

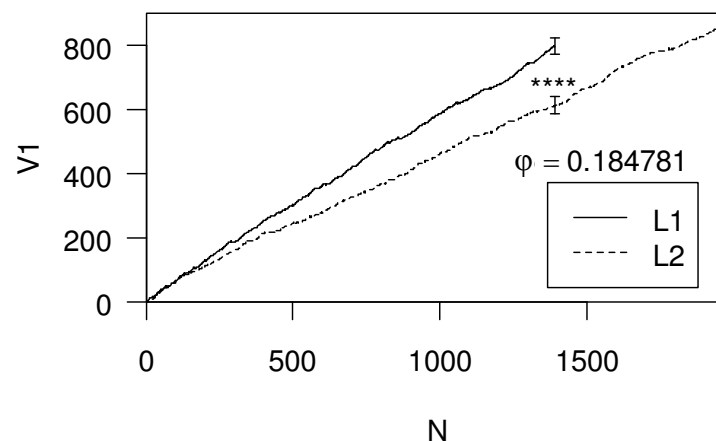


Abbildung 8. Wortschatzwachstumskurven für die Anzahl der Hapax Legomena VI in L1- und L2-Daten. Der Vergleich geschieht wieder bei $N=1393$ mit Fehlerbalken für den tatsächlichen Wert von VI (95% Konfidenzintervall).

Insgesamt zeigen diese Ergebnisse, dass die fortgeschrittenen Lerner Komposita nicht nur seltener, sondern auch deutlich weniger produktiv verwenden, was mit bisherigen Studien zu Typenverteilungen und Frequenzeffekten bei Lernern übereinstimmt (Ellis & Ferreira-Junior 2009, Berth 2009, Madlener 2011).¹⁸ Dies betrifft sowohl die Größe ihres aktiven Wortschatzes (einschließlich stark lexikalisierten Fällen), als auch die Vorkommenswahrscheinlichkeit spontan gebildeter neuer Zusammensetzungen, sofern diese mit dem Anteil der Hapax Legomena geschätzt werden kann.

¹⁸ Diese Aussage soll im Kontext des Baayen'schen Produktivitätsparadigmas rein quantitativ verstanden werden, d.h. der Wortschatz und die Menge an Hapax Legomena sind bei Lernern tendenziell kleiner. Gleichzeitig können sich Lerner qualitativ ‚übermäßig‘ produktiv verhalten, indem sie regelmäßige Formen produzieren können, die für Muttersprachler blockiert sind, bspw. *Treffung* oder *Sinkung* (s. Berth 2009).

5. Die mittleren Ebenen – Kompositionsklassen

Neben der Untersuchung von Komposition als einheitlicher Prozess soll als nächstes untersucht werden, ob die in Abschnitt 1 skizzierten Kompositaklassen von den Lernern gleich behandelt werden, bzw. ob es Teilbereiche in der Komposition gibt, bei denen die Komposition unter den Lernern beliebter oder weniger beliebt zu sein scheint. Erwerben die DaF-Lerner solche Konstruktionen wie (14)-(15)?

(14) Zweckkompositum: $[N-N]_N \leftrightarrow |N \text{ für } N|$, z.B. *Arbeitstisch*

(15) Agens-Kompositum: $[N-[V]er]_N \leftrightarrow |jemand, \text{ der } N \text{ Vt}|$, z.B. *Büroleiter*

Hieraus folgt zunächst die Frage, ob man untergeordnete Kompositionsmuster prinzipiell als Erwerbsgegenstand annehmen soll bzw. muss (vgl. Booij 2009, 2010, der im Rahmen des Construction Morphology-Ansatzes solche Konstruktionen im L1-Lexikon annimmt). Als *advocatus diaboli* könnte man die Nullhypothese formulieren, Konstruktionen der Art $[X-N]_N$ (mit den dazugehörigen morphologischen Angaben zur Fugewahl usw.) reichen aus, und die Menge der tatsächlich realisierten Kompositionsbeziehungen bzw. -semantiken würde sich aus kommunikativen Bedürfnissen ergeben.

Um zu zeigen, dass dies nicht ausreicht, kann man Beispiele von denkbaren, kommunikativ nützlichen Kompositionsemantiken finden, die weder grammatisch sind noch von den Lernern gebildet werden. Ein Beispiel hierfür sind die unangemessenen deutschen Komposita für Staatsoberhäupter in (16)-(17), die z.B. in semitischen Sprachen durchaus grammatisch sind und zu einer produktiven Konstruktion gehören (18)-(19):

(16) * *Die Dänemarkkönigin*

(17) * *Der Deutschlandpräsident*

(18) Arab. *malikat-adDanmārik*, wörtl. ‚*Die Dänemarkkönigin*‘

(19) Heb. *nesi-Germanja*, wörtl. ‚*Der Deutschlandpräsident*‘

Solche Bildungen kommen in den fortgeschrittenen Lernerdaten nicht vor, obwohl sie sonst denkbar wären. Die Vermeidung solcher Komposita bzw. die allgemeine Seltenheit von kombinatorisch fehlerhaften oder unprototypischen Bildungen muss daher gelernt werden (s. Abschnitt 3 sowie die denkbaren aber ungrammatischen teillexikalisierten Minikonstruktionen in Abschnitt 6). Auch für das Verstehen selbst von motivierten Komposita reichen die Bedeutungen der Kompositionsglieder oft nicht aus, wenn man eine konventionalisierte Kompositionsbedeutung nicht kennt oder nicht erkennen kann (vgl. z.B. *Blitzschlag* : *Blitzstart* oder das ambige Kompositum *Goethebild* in Barz 1995:14; s. auch Booij 2009 und weitere Referenzen dort).

Wenn man den Sinn einer Untersuchung von Kompositionsuntertypen wie in (14)-(15) akzeptiert, bleibt die Frage immer noch offen, welche Typen vorhanden sind bzw. gelernt werden müssen. Eine ausführliche Klassifikation der Kompositionsemantiken im hier untersuchten Korpus geht über die Grenzen dieser Studie hinaus. Eine etwas grobe aber trotzdem einleuchtende Möglichkeit ist, die

Komposita mit Phrasen zu paraphrasieren (vgl. Gaeta & Zeldes 2012 für ein ähnliches Vorgehen mit Komposita aus Web-Daten). Hierbei entstehen Paraphrasen mit Präpositionalphrasen, Nominalphrasen und Verbalphrasen, wie die folgenden Beispiele zeigen:¹⁹

- (20) *Holztisch* → *Tisch aus Holz* (PP aus)
 (21) *Großfamilie* → *große Familie* (ADJ)
 (22) *Fußballspieler* → *spielt Fußball* (RK Akk.)
 (23) *Mitspracherecht* → *Recht auf Mitsprache* (RK auf)

Es versteht sich, dass solche Paraphrasen letztendlich immer z.T. subjektiv und in vielen Fällen nur schwer entscheidbar sind, doch das hier vorgeschlagene Verfahren bietet zumindest eine realistische Möglichkeit, Komposita ohne sehr spezifische theoretische Annahmen auf sinnvolle Weise in Klassen zu unterteilen.²⁰

Um die Verteilung der Klassen in den L2-Aufsätzen mit dem L1-Korpus zu vergleichen, wurden jeweils 200 zufällig ausgewählte Hapax Legomena unter den Komposita in jedem Korpus untersucht. Der Fokus auf den Hapax Legomena wird wieder dadurch motiviert, dass diese wahrscheinlicher unlexikalisierte oder nur wenig vertraute Komposita darstellen und daher einen besseren Einblick in die Verwendung von Komposition als produktiven Wortbildungsprozess ermöglichen. Produktivere Kompositionsmuster sind auch wahrscheinlicher transparent bzw. motiviert (vgl. Grimm 2001:220), was eine Kategorisierung anhand von Paraphrasen praktikabler macht: Vollkommen opake Fälle wie etwa *Löwenzahn* können einer solchen Analyse natürlich kaum unterliegen. Die Frequenzen der Klassen als Determinativ- bzw. Rektionskomposita werden in der Tabelle 6 geschildert.²¹

L1				L2			
Klasse	Det	Rekt	gesamt	Klasse	Det	Rekt	gesamt
<i>ADJ</i>	28	1	29	<i>ADJ</i>	31	0	31
<i>Akk.</i>	0	*28	*28	<i>Akk.</i>	0	*13	*13
<i>PP an</i>	*2	1	*3	<i>PP an</i>	*12	1	*13
<i>PP auf</i>	1	3	4	<i>PP auf</i>	2	5	7
<i>PP aus</i>	5	0	5	<i>PP aus</i>	2	0	2
<i>PP bei</i>	1	1	2	<i>PP bei</i>	2	0	2
<i>PP für</i>	33	0	33	<i>PP für</i>	25	3	28

¹⁹ Die Paraphrasen sollen keineswegs einen transformationellen Ansatz zur Generierung der Komposita aus einer syntaktischen Tiefenstruktur suggerieren (s. Kürschner 1974, Rivet 1999, Siebert 1999).

²⁰ Der paraphrasierende Ansatz geht z.T. auf frühere Versuche zur Systematisierung der Kompositionsemantiken zurück. So schlägt z.B. Levi (1978) die neun Prädikatstypen CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM und ABOUT für primäre Nominalkomposita und ACT, PRODUCT, AGENT, PATIENT für deverbale Rektionskomposita vor (s. auch Fanselow 1981 für eine formalsemantische Analyse zum Deutschen sowie Motsch 2004:376-420 für eine große Liste von Klassen für das Deutsche). Ein wesentlicher Unterschied im aktuellen Ansatz ist, dass aufgrund der unbegrenzten Menge an möglichen Kompositionsemantiken (vgl. Jespersen 1942:142-147) keine vollständige Liste vorgeschlagen wird, und dass jede PP-Paraphrase entweder einem Determinativ- oder einem Rektionskompositum entsprechen kann, in Abhängigkeit von der Valenz des Kopfnomens.

²¹ Die wenigen Kopulativkomposita werden weder als Determinativ- noch als Rektionskomposita zugeordnet. Possessivkomposita waren in der Stichprobe nicht enthalten.

<i>PP in</i>	13	3	16	<i>PP in</i>	12	0	12
<i>PP mit</i>	2	1	3	<i>PP mit</i>	4	0	4
<i>PP nach</i>	1	0	1	<i>PP nach</i>	0	2	2
<i>PP über</i>	0	1	1	<i>PP über</i>	0	2	2
<i>PP von</i>	**34	*28	62	<i>PP von</i>	**58	*14	72
<i>PP wegen</i>	1	0	1	<i>PP wegen</i>	4	0	4
<i>PP zu</i>	1	2	3	<i>PP zu</i>	0	1	1
<i>Kopulativ</i>		0	0	<i>Kopulativ</i>		2	2
<i>(sonstige)</i>	7	2	9	<i>(sonstige)</i>	5	0	5
gesamt	129	***71	200	gesamt	159	***41	200

Tabelle 6. Frequenzen der Kompositionsklassen in den L1- und in den L2-Daten. Signifikante Unterschiede zwischen den Gruppen werden markiert (* p<.05; ** p<.01; *** p<.005).

Der auffälligste, signifikanteste Unterschied (p<.005) liegt wohl im Gesamtanteil der Rektionskomposita in den beiden Stichproben: 71/200 (35,5%) bei den Muttersprachlern und nur 41/200 (20,5%) bei den Lernern. Dieser Unterschied stammt größtenteils nicht aus den insgesamt seltenen relationalen Nomina (wie etwa *Mitspracherecht* in (23) oben), sondern aus dem Kernbereich der Rektionskomposita, den deverbale Ableitungen mit akkusativischer Lesart (28 : 13, p<.05) bzw. mit *von*-Phrasen paraphrasierbaren Komposita mit agentivischem Modifikator (28 : 14, p<.05), wie in den folgenden Beispielen:

- (24) *Kartenverkäufer* (L1=deu) → *verkauft Karten*
- (25) *Strassensauberer* (L1=eng) → *macht die Straßen sauber*
- (26) *Verdienstausschlag* (L1=deu) → *Ausschlag von Verdienst*
- (27) *Gesellschaftsentwicklung* (L1=fra) → *Entwicklung von der Gesellschaft*

Umgekehrt stellt sich heraus, dass die mit *von*-Phrasen paraphrasierbaren Komposita, die keine offensichtliche Rektionsrelation aufweisen, bei den Lernern häufiger vorkommen als bei den Muttersprachlern (34 : 58, p<.01). Die Komposita in dieser Kategorie erscheinen in vielen Fällen trotz ihrer niedrigen Frequenz eher lexikalisiert zu sein als offensichtliche Spontanbildungen, und haben auch oft Entsprechungen in den Muttersprachen der Lerner:

- (28) *Arbeitwelt* [sic] (L1=fra, vgl. Fra. *monde [du] travail*)
- (29) *Lebensstil* (L1=eng, vgl. Eng. *lifestyle*)
- (30) *Hautfarbe* (L1=deu)
- (31) *Jugendkultur* (L1=deu)

Warum solche weniger argumentalen Relationen (*die Welt von der Arbeit, der Stil von dem Leben*) in den Lernerdaten häufiger belegt sind, kann man an dieser Stelle nur vermuten. Wenn Rektionskomposita jedoch öfter auf weniger lexikalisierte Kombinationen von Kopf und Argument zurückgehen, könnte diese Beobachtung auch mit der niedrigen Produktivitätsrate der Lerner zusammenhängen.²² Eine

²² Interessanterweise werden gerade deverbale Rektionskomposita in der DaF-Literatur manchmal für leichter lernbar gehalten (s. Fandrych & Thurmair 1994:38), weil die intendierte semantische Relation zwischen Kopf und Modifikator durch die Argumentstruktur des entsprechenden Verbs eingeschränkt

interessante Frage, die sich wahrscheinlich erst unter experimentellen Bedingungen aussagekräftig beantworten lässt, ist, ob Lerner Komposita öfter verwenden und sich leichter merken, die eine Entsprechung in ihrer eigenen Muttersprache finden. Eine weitere Hypothese, die wir von den Daten ableiten können, wäre, dass die Argumentstruktur deverbaler Nomina an sich eine Schwierigkeit darstellt, z.B. weil Erfahrung mit der Argumentwahl in VPen auf der syntaktischen Ebene weniger gut auf die nominale Umgebung übertragen wird. Auch diese Möglichkeit müsste in Online-Studien genauer untersucht werden.

6. Die unteren Ebenen – teillexikalisierte Minikonstruktionen

Den Abschluss der Untersuchung des Kompositionsnetzwerks bilden die untersten Ebenen, die (teil-)lexikalisierte Muster bzw. sog. „Minikonstruktionen“²³ mit konkreten Kopf- bzw. Modifikatorlexemen enthalten. Dass man solche Minikonstruktionen zur Erfassung der Korpusdaten benötigt, kann man anhand von Beispielen erkennen wie (32)-(36):

- (32) *Profi-akademiker* [sic] (L1=eng)
- (33) *Profi-Basketballer* (L1=eng)
- (34) *Profi-Sportler* (L1=eng)
- (35) * *Expertenakademiker*
- (36) * *Expertensportler*

Wenn Lerner nicht auf spezifische reihenbildende Modifikator-Lexeme achten würden, könnte man erwarten, dass man Komposita wie in (35)-(36) findet und nicht nur solche wie (32)-(34), denn die Semantik von *Experte* und *Profi* ist sehr ähnlich. Dies ist aber nicht der Fall. Umgekehrt kann man sogar zeigen, dass die Wahl des Modifikators die Erkennung der Konstruktion beeinflussen kann. Wenn das Kopfnomen ein deverbales Agensnomen ist, würde man trotz der vergleichbaren Semantik der Lexeme die Konstruktion in (37) wahrscheinlich eher als Exemplar des Musters [N-[V]er]_N deuten und nur (38) als ein adjektivisches Kompositum nach dem Muster [*Profi*-N]_N ↔ |professionelles Nl.²⁴

- (37) #*Experten-Programmierer* ↔ |jemand, der Experten programmiert|
- (38) *Profi-Programmierer* ↔ |professioneller Programmierer|

wird. Die hier vorgestellten Daten belegen jedoch, dass Lerner solche Konstruktionen nicht unbedingt zielsprachlich verwenden, selbst wenn sie diese passiv leichter verstehen können.

²³ Der Begriff der Minikonstruktion wurde von Boas (2003, 2011) in der syntaktischen Domäne geprägt und dient der Unterscheidung zwischen dem Bedeutungsbeitrag abstrakter Konstruktionen wie bspw. der Resultativkonstruktion und dem spezielleren Bedeutungsbeitrag einer teillexikalisierten Ausführung der Konstruktion, z.B. mit einem bestimmten Verb. Diesen Begriff können wir hier auf den morphologischen Bereich ausdehnen, um teillexikalisierte Kompositionsmuster mit zusätzlichen Bedeutungskomponenten zu erfassen.

²⁴ Die genaue Bedeutung der Minikonstruktion [*Profi*-N]_N geht hier wahrscheinlich über die Paraphrase |professionelles Nl| hinaus. Dennoch ist eine adjektivische Lesart m.E. passender als bspw. eine kopulative (|ein Profi, der Programmierer ist|), da das Denotat einer subsektiven Bedeutung genügen muss: ein *Profi-Koch*, der schlecht programmiert, ist kein *Profi-Programmierer*.

Wenn auch solche speziellen Minikonstruktionen als Erwerbsgegenstand aufgefasst werden, wird die Aufgabe des Lernalters einerseits sehr kompliziert (es sind sehr viele Muster zu erwerben), andererseits ist Komposition als Mechanismus im Allgemeinen robuster und weniger fehleranfällig, da man trotz Fehler bei einem Muster eine zielsprachennahe Verwendung anderer Muster erreichen kann. Diese Ansicht suggeriert sogar auch, dass Frequenzunterschiede, d.h. Über- bzw. Mindergebrauch, in bestimmten Kompositionstypen unabhängig auftreten können. So finden sich z.B. einige Modifikatorenreihen mit adjektivischen Lesarten, die bei den Lernern nicht auffällig selten erscheinen:

- (39) [*Top-N*]_N: *Topjobs* (L1=eng)
- (40) [*Gesamt-N*]_N: *Gesamtnote* (L1=eng)
- (41) [*Grund-N*]_N: *Grundprinzip* (L1=swe)
- (42) [*Privat-N*]_N: *Privatfirma* (L1=eng)
- (43) [*Haupt-N*]_N: *Hauptziel* (L1=dan)
- (44) [*Riesen-N*]_N: *Riesenschulden* (L1=eng)

Diese Minikonstruktionen sind zwar allgemein relativ selten und es lassen sich daher im begrenzten Umfang des Falko-Korpus keine sicheren quantitativen Schlüsse ziehen, sie weisen jedoch den Mindergebrauch von Komposition und die Unproduktivität, die wir in den letzten Abschnitten gesehen haben, nicht auf. Solche reihenbildenden Modifikatoren mit großen sog. „positional families“²⁵ dürften für den Erwerb deswegen leichter sein, weil ihre Bedeutung oft transparent und konsistent ist, weil sie in vielen Fällen mit Vordergliedern ohne Fugenelement verbunden sind, und nicht zuletzt, weil sie keine Rektionsbeziehungen beinhalten. Diese sind v.a. im letzten Abschnitt als scheinbar problematisch aufgefallen, was bei den Komposita in (39)-(44) entfällt. In anderen Worten könnten die attributiven Modifikator-Minikonstruktionen einfacher sein, weil sie mit sehr vielen Köpfen funktionieren und keinen besonderen Bedingungen unterliegen, wie der Berücksichtigung von Argumentstruktur. Dass sie nicht über eine allgemeine Regel [*A-N*]_N bzw. [*N-N*]_N gelernt werden können ist in vielen Fällen semantisch und syntaktisch klar, vgl. syntaktisch bei Deutung als Adjektive: **ein topper Job*, **ein grundes Prinzip*, etc., bzw. semantisch, wenn man die Lesarten einiger Modifikatoren als Nomina betrachtet: *Riesenschulden* ≠ |Schulden eines Riesen|, *Hauptziel* ≠ |Ziel eines Hauptes|. Diese Bedeutungen entstehen nicht einfach durch die Lexikalisierung des ganzen Kompositums (*Hauptziel*) sondern durch die Lexikalisierung der speziellen Bedeutung von *Haupt-* als Kompositionsmodifikator, sprich als Zeichen einer idiosynkratischen teillexikalisierten Kompositionsstruktur [*Haupt-N*]_N, die über häufige Vertreter wie *Hauptproblem*, *Hauptziel* oder *Hauptargument* auch gelernt werden muss, um produktiv einsatzfähig zu sein.

²⁵ Dieser Begriff wurde von de Jong et al. (2002) verwendet, um mehr oder weniger große Familien von Komposita mit demselben Kompositionsglied zu Bezeichnen. Vgl. auch „Wortbildungsnester“ bei Fleischer & Barz (2007:71-73), die auch verwandte Derivationsprodukte mit einschließen.

Gleichzeitig sollte man nicht annehmen, dass alle Komposita ausschließlich über solche teillexikalisierten Konstruktionen mit sehr spezifischen konventionalisierten Bedeutungen gebildet werden. Wir finden in den L2-Daten wie in den L1-Daten Komposita, die einer freien Komposition mit unvorhersagbarer Bedeutung entstammen, z.T. auch mit fehlerhaften Ergebnissen. So liegen (45)-(49) keine offensichtlichen Minikonstruktionen zugrunde:

Lernerbeispiele:

- (45) [N-N]_N *Lebensblatt*: (L1=ron)
*Eine sehr große Bedeutung hat natürlich aber seine Arbeit, d. h. die Kenntnisse und die Fähigkeiten, die er der Gesellschaft übergibt oder sie für das Wohlbefinden und für die Weiterentwicklung der Gesellschaft verwendet. Der Beruf, den er ausübt lässt mehr oder weniger Spuren auf dem **Lebensblatt** der Gesellschaft.*
- (46) [N-N]_N *Raubplan*: (L1=eng)²⁶
*Größere Geldmengen sind normalerweise sehr gut geschützt und ein Fehler im **Raubplan** kann zur Verhaftung führen.*
- (47) [N_{patiens}-[V]_N]_N *Frauenbedruck*: (L1=eng, ZH1=Unterdrückung von Frauen)
*In manche Ländern, spielt Religion auch ein große Rolle der **Frauenbedruck** (opression [sic] of women).*

Muttersprachlerbeispiele:

- (48) [N-N]_N *MatchosprÜche*: [sic] (L1=deu, ZH1=Matchosprüche)
*[...] wenn sie an die, immernoch, so weit verbreiteten **MatchosprÜche** ihrer männlichen Kollegen denkt [...]*²⁷
- (49) [N-N]_N *Systemsidee*: [sic] (L1=deu, ZH1=Systemidee)
*Jeder könnte kommen und behaupten er arbeite nun härter und hätte somit auch mehr Geld verdient [...] Man kann es nicht überprüfen, da es für so eine moderne Art von **Systemsidee** noch keine Regelungen gibt*

Interessant ist, dass die Lexemwahl in zwei der Lernerbeispiele mit Kombinationen in der jeweiligen L1 zusammenzuhängen scheint. So glossiert die Lernerin in (47) ihr eigenes Kompositum gleich mit der englischen Quelle „opression [sic] of women“, während die rumänische Verfasserin von (45) höchstwahrscheinlich eine rumänische Kollokation übersetzt, Rum. *pagina vieții*, wörtl. ‚Blatt des Lebens‘. In beiden Fällen handelt es sich jedoch nicht um ein Kompositum in der L1, sondern um Modifikation mit einer *of*-PP bzw. einem Genitivattribut. Somit stellen auch diese Beispiele von der Formseite her produktive Anwendungen einer morphologischen Konstruktion zum

²⁶ Dieser Beleg stammt aus der nächsten Version des WHIG-Korpus, die in der derzeit öffentlich verfügbaren Version 1.0 noch nicht enthalten sind.

²⁷ Das Kompositum *Matchosprüche* in (48) kann man auch als deverbales Rektionskompositum lesen, wenn man *Spruch* aktiv von *sprechen* ableitet. Allerdings muss ein *Matchospruch* nicht unbedingt von einem *Matcho* geäußert werden. In beiden Fällen bietet sich keine klare lexikalisierte Minikonstruktion an.

Ausdruck von Lexemen bzw. Lexemkombinationen dar, auch wenn diese von der Zielsprache nicht lizenziert werden. Eine gefühlte Merkwürdigkeit bei solchen Komposita, die zu keinem lexikalischen Muster gut passen, kann aber auch in muttersprachlichen Belegen erscheinen, wie Beispiel (49) zeigt. Ungeachtet der möglichen normativen Einstellung gegenüber solchen Belegen, scheinen die Beispiele allgemeine, abstrakte Konstruktionen zu instanzieren, die wahrscheinlich durch Verallgemeinerung über sehr viele verschiedene Exemplare und entsprechende Untermuster entstehen. Der Status von Komposition im Allgemeinen als eigenständiger Teil des Erwerbsgegenstandes sollte daher m.E. von den bisherigen Befunden unberührt bleiben.

7. Zusammenfassung und Diskussion

In diesem Artikel wurde Komposition als ein komplexes Netzwerk überlappender Konstruktionen dargestellt, das auf unterschiedlichen Abstraktionsebenen von Lernern stückweise erworben werden muss. Mit der Erfahrung, die bis zum fortgeschrittenen Niveau gesammelt wird, kann der Lerner neben der formalen Sprachdidaktik zu Komposition im Allgemeinen auch durch konkrete Exemplare, häufige Prototypen und Beispiele mit konsistenter Form-Bedeutungs-Kopplungen ein solches Netzwerk samt Unterklassen und Minikonstruktionen rekonstruieren und selbst erweitern.²⁸ Es wurde gezeigt, dass das Verhalten der Komponenten dieses Netzwerks auf jeder Ebene nicht durch rein abstrakte Regeln und ein Lexikon der Kompositionsglieder abgeleitet werden kann, sondern anhand von Gebrauchserfahrung mit spezifischen Lexemen erschlossen werden muss.

Die Daten aus dem Lernerkorpus Falko belegen, dass die Komposition von fortgeschrittenen Lernern einerseits ziemlich oft, produktiv und größtenteils fehlerfrei eingesetzt wird. Qualitative Unterschiede, die mit der Komposition als Prozess zusammenhängen, finden sich v.a. in den insgesamt eher seltenen morphologischen Fehlern bei der Wahl des Fugenelements, insbesondere bei der Entscheidung zwischen dem Fugen-*s* und der Nullfuge. Andererseits machen sich einige quantitative Unterschiede bemerkbar: Lerner verwenden Komposition doch signifikant seltener (im Schnitt rund 11% der Nomina gegenüber 15% bei den Muttersprachlern), weisen einen weniger vielfältigen Wortschatz auf und verhalten sich weniger produktiv bei der Bildung neuer Komposita. Besonders auffällig ist der geringe Anteil der Rektionskomposita im Vergleich zu den muttersprachlichen Daten.

Auch wenn der Fokus des vorliegenden Beitrags deskriptiv ist, können und sollten diese Befunde in der korpusgestützten Sprachvermittlung Verwendung finden (vgl. Lüdeling & Walter 2010). Zunächst liefern die Korpusdaten Informationen über muttersprachliche Frequenzen nicht nur von konkreten Komposita, sondern auch von Kompositionsmustern und -semantiken, wie sie im Abschnitt 5 ermittelt wurden.²⁹

²⁸ Einige Lehrwerke unterstützen das Erlernen der relevanten Semantiken sogar durch explizite Hinweise auf bestimmte (allerdings wenige) Klassen, s. Rak (1990:13-14).

²⁹ Solche korpusbasierten Frequenzen können jedoch aus didaktischen Gründen natürlich geändert bzw. systematisch angepasst werden, damit eine optimal lernbare Typenverteilung erreicht wird, s. u.a. Handwerker & Madlener (2009:47-49).

Die Vermittlung dieser häufigeren Strukturen verdient wahrscheinlich auch mehr Ressourcen in der Fremdsprachendidaktik (vgl. Leech 2001 zum Englischen, Tschirner 2005 zum Deutschen) und könnte gleichzeitig von einer Veranschaulichung durch authentische muttersprachliche Korpusbelege profitieren. Insbesondere benötigen aber die Kategorien darunter, die einen Mindergebrauch aufweisen, eine didaktische Unterstützung in der Sprachvermittlung, wie bspw. die Rektionskomposita, sowie die formalen Klassen, denen die meisten morphologischen Fehler geschuldet sind, in erster Linie Komposita mit dem Fugen-s bzw. der Nullfuge. Diese Unterscheidung erscheint gerade im Kontext von Vordergliedern mit einem [t]-Auslaut ggf. auch perzeptorisch schwierig zu sein (s. Abschnitt 3). Darüber hinaus könnte man Lerner auf die produktiveren Kompositionsmuster aufmerksam machen, die im unteren Bereich des Frequenzspektrums erscheinen, d.h. die Muster, die Hapax Legomena hervorbringen. Diese stellen eine weitere Wortbildungsstrategie zum Ausdruck von Modifikationsrelationen dar, die offenbar bei den Lernern auch in anderen Kontexten oft fehlen (Abschnitt 4.1, vgl. auch Hirschmann et al., eingereicht).

Die Befunde aus der Untersuchung der Falko-Daten geben einen ersten quantitativen Einblick in die unterschiedliche Verwendung von Komposita bei fortgeschrittenen Nicht-Muttersprachlern, der erst im Rahmen von einem verhältnismäßig natürlichen und freien Kontext (Verfassung eines Aufsatzes) und unter Beachtung von streng kontrollierten vergleichbaren Erhebungsbedingungen für Lerner und Muttersprachler möglich wird. Diese Datengrundlage sowie die Daten zu den hier extrahierten Komposita sind offen zugänglich. Weitere Studien, die anhand dieser Ressourcen unseren Kenntnisstand über das lexikalisch-konstruktionelle Wissen in der Interlanguage fortgeschrittener Deutschlerner zu erweitern versuchen, könnten deshalb von einer vielversprechenden Quelle für neue Hypothesen über den Sprachgebrauch solcher Lerner profitieren, die neue Erkenntnisse im Bereich der Interlanguage-Grammatik erwarten lässt.

A. Anhang: Drei-Buchstaben-Kennungen nach ISO 639-3

afr - Afrikaans	ita - Italienisch	spa - Spanisch
ces - Tschechisch	jpn - Japanisch	swe - Schwedisch
dan - Dänisch	nld - Niederländisch	tur - Türkisch
deu - Deutsch	nor - Norwegisch	ukr - Ukrainisch
ell - Griechisch	pol - Polnisch	uzb - Usbekisch
eng - Englisch	ron - Rumänisch	zho - Chinesisch
fra - Französisch	rus - Russland	

Literatur

- Baayen, R. Harald (2001). *Word Frequency Distributions*. (Text, Speech and Language Technologies 18). Dordrecht, Boston und London: Kluwer.
- Baayen, R. Harald (2009). Corpus Linguistics in Morphology: Morphological Productivity. In: Anke Lüdeling und Merja Kytö (Hrsg.): *Corpus Linguistics. An International Handbook*. Bd. 2. Berlin: Mouton de Gruyter, 899–919.

- Baayen, R. Harald und Lieber, Rochelle (1991). Productivity and English Derivation: A Corpus-Based Study. *Linguistics* 29, 801–843.
- Barz, Irmhild (1995). Komposita im Großwörterbuch Deutsch als Fremdsprache. In: Inge Pohl und Horst Ehrhardt (Hrsg.): *Wort und Wortschatz. Beiträge zur Lexikographie*. Tübingen: Niemeyer, 13–24.
- Barz, Irmhild (1996). Komposition und Kollokation. In: Clemens Knobloch und Burkhard Schaefer (Hrsg.): *Nomination - fachsprachlich und gemeinsprachlich*. Opladen: Westdeutscher Verlag, 127–146.
- Bauer, Laurie (2001). *Morphological Productivity*. (Cambridge Studies in Linguistics 95). Cambridge: Cambridge University Press.
- Becker, Thomas (1992). Compounding in German. *Italian Journal of Linguistics / Rivista di Linguistica* 4(1), 5–36.
- Berth, Michael (2009). Treffungen, Sinkung und Benützung. *Korpuslinguistische Untersuchung des Erwerbs von derivationsmorphologischen Wortbildungsregularitäten bei fortgeschrittenen Lernern des Deutschen als Fremdsprache*. Magisterarbeit, Humboldt-Universität zu Berlin. URL: <http://edoc.hu-berlin.de/master/berth-michael-2009-09-17/PDF/berth.pdf>.
- Boas, Hans C. (2003). *Resultative Constructions in English and German*. Stanford: CSLI Publications.
- Boas, Hans C. (2011). Zum Abstraktionsgrad von Resultativkonstruktionen. In: Stefan Engelberg; Anke Holler und Kristel Proost (Hrsg.): *Sprachliches Wissen zwischen Lexikon und Grammatik*. (Institut für Deutsche Sprache, Jahrbuch 2010). Berlin und New York: De Gruyter, 37–70.
- Bohnet, Bernd (2010). Very High Accuracy and Fast Dependency Parsing is not a Contradiction. In: *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. Beijing, 89–97.
- Booij, Geert E. (2009). Compounding and Construction Morphology. In: Rochelle Lieber und Pavol Štekauer (Hrsg.): *The Oxford Handbook of Compounding*. (Oxford Handbooks in Linguistics). Oxford: Oxford University Press, 201–216.
- Booij, Geert E. (2010). *Construction Morphology*. Oxford: Oxford University Press.
- Ellis, Nick C. und Ferreira-Junior, Fernando (2009). Construction Learning as a Function of Frequency, Frequency Distribution, and Function. *The Modern Language Journal* 93(3), 370–385.
- Evert, Stefan und Baroni, Marco (2007). zipfR: Word frequency distributions in R. In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Session*. Prag, 29–32.
- Fandrych, Christian und Tallowitz, Ulrike (2008). *Klipp und Klar. Übungsgrammatik Grundstufe Deutsch*. Stuttgart: Klett.
- Fandrych, Christian und Thurmair, Maria (1994). Ein Interpretationsmodell für Nominalkomposita: linguistische und didaktische Überlegungen. *Deutsch als Fremdsprache* 31, 34–45.
- Fanselow, Gisbert (1981). *Zur Syntax und Semantik der Nominalkomposition. Ein Versuch praktischer Anwendung der Montague-Grammatik auf die Wortbildung im Deutschen*. (Linguistische Arbeiten 107). Tübingen: Niemeyer.

- Fischer, Kerstin und Stefanowitsch, Anatol (2006). Konstruktionsgrammatik: Ein Überblick. In: *Konstruktionsgrammatik: Von der Anwendung zur Theorie*. Tübingen: Stauffenburg, 3–17.
- Fleischer, Wolfgang und Barz, Irmhild (2007). *Wortbildung der deutschen Gegenwartssprache*. 3. Auflage. Tübingen: Niemeyer.
- Gaeta, Livio und Ricca, Davide (2006). Productivity in Italian Word Formation: A Variable-Corpus Approach. *Linguistics* 44(1), 57–89.
- Gaeta, Livio und Zeldes, Amir (2012). Deutsche Komposita zwischen Syntax und Morphologie: Ein korpusbasierter Ansatz. In: Livio Gaeta und Barbara Schlücker (Hrsg.): *Das Deutsche als kompositionsfreudige Sprache: Strukturelle Eigenschaften und systembezogene Aspekte*. (Linguistik - Impulse & Tendenzen 46). Berlin: De Gruyter, 197–217.
- Goldberg, Adele E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Granger, Sylviane (2002). A Bird's-Eye View of Learner Corpus Research. In: Sylviane Granger; Joseph Hung und Stephanie Petch-Tyson (Hrsg.): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam und Philadelphia: John Benjamins, 3–33.
- Granger, Sylviane; Dagneaux, Estelle; Meunier, Fanny und Paquot, Magali (2009). *The International Corpus of Learner English. Version 2. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Gries, Stefan Th. und Wulff, Stefanie (2005). Do Foreign Language Learners also have Constructions? *Annual Review of Cognitive Linguistics* 3, 182–200.
- Grimm, Hans-Jürgen (2001). Kontrastivität in der Lexik. In: Gerhard Helbig; Lutz Götze; Gert Henrici und Hans-Jürgen Krumm (Hrsg.): *Deutsch als Fremdsprache. Ein internationales Handbuch*. (Handbücher zur Sprach- und Kommunikationswissenschaft 19.1). Berlin und New York: Walter de Gruyter, 214–224.
- Haberzettl, Stefanie (2006). Konstruktionen im Zweitspracherwerb. In: Kerstin Fischer und Anatol Stefanowitsch (Hrsg.): *Konstruktionsgrammatik. Von der Anwendung zur Theorie*. Tübingen: Stauffenburg, 55–77.
- Handwerker, Brigitte und Madlener, Karin (2009). *Chunks für DaF. Theoretischer Hintergrund und Prototyp einer multimedialen Lernumgebung*. (Perspektiven Deutsch als Fremdsprache 23). Baltmannsweiler: Schneider Verlag Hohengehren.
- Hartkamp, Sandra und Schneider-Wiejowski, Karina (2010). Die Regelkonformität in der Distribution von Fugenelementen. Eine korpuslinguistische Untersuchung neologistischer Substantivkomposita mit nominalem Erstglied. *Muttersprache* 3/2010, 198–213.
- Heidermann, Werner (1997). *Grammatiktraining Grundstufe*. Ismaning: Verlag für Deutsch.
- Hirschmann, Hagen; Lüdeling, Anke; Rehbein, Ines; Reznicek, Marc und Zeldes, Amir (erscheint). *Underuse of Syntactic Categories in Falko - A Case Study on*

- Modification*. In: 20 Years of Learner Corpus Research. Looking Back, Moving Ahead.
- Jespersen, Otto (1942). *A Modern English Grammar on Historical Principles*. Bd. VI: Morphology. London: George Allen & Unwin.
- Jong, Nivja H. de; Feldman, Laurie B.; Schreuder, Robert; Pastizzo, Matthew und Baayen, R. Harald (2002). The Processing and Representation of Dutch and English Compounds: Peripheral Morphological and Central Orthographic Effects. *Brain and Language* 81(1-3), 555–567.
- Kürschner, Sebastian (2005). Verfügun-g-s-nutzung kontrastiv: Zur Funktion der Fugenelemente im Deutschen und Dänischen. *Tijdschrift voor Skandinavistiek* 26(2), 101–125.
- Kürschner, Wilfried (1974). *Zur syntaktischen Beschreibung deutscher Nominalkomposita. Auf der Grundlage generativer Transformationsgrammatiken*. (Linguistische Arbeiten 18). Tübingen: Niemeyer.
- Leech, Geoffrey (2001). The Role of Frequency in ELT: New Corpus Evidence Brings a Re-appraisal. In: Hu Wenzhong (Hrsg.): *ELT in China 2001: Papers presented at the 3rd International Symposium on ELT in China*. Beijing: Foreign Language Teaching and Research Press, 1–23.
- Levi, Judith N. (1978). *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.
- Lu, Xiaofei (2010). Automatic analysis of Syntactic Complexity. *International Journal of Corpus Linguistics* 15(4), 474–496.
- Lüdeling, Anke und Walter, Maik (2010). Korpuslinguistik. In: Hans-Jürgen Krumm; Christian Fandrych; Britta Hufeisen und Claudia Riemer (Hrsg.): *Handbuch Deutsch als Fremd- und Zweitsprache*. (Handbücher zur Sprach- und Kommunikationswissenschaft 35.) Berlin und New York: Walter de Gruyter, 315–322.
- Madlener, Karin (2011). Developing Productivity with a New Construction: Are there Frequency Effects in Instructed Second Language Acquisition (SLA)? In: Amir Zeldes und Anke Lüdeling (Hrsg.): *Proceedings of Quantitative Investigations in Theoretical Linguistics 4 (QITL-4)*. Berlin, 56–58.
- Michel, Sascha (2009). *Schaden-0-ersatz* vs. *Schaden-s-ersatz*. Ein Erklärungsansatz synchroner Schwankungsfälle bei der Fugenelementbildung von N+N-Komposita. *Deutsche Sprache* 37, 334–351.
- Motsch, Wolfgang (2004). *Deutsche Wortbildung in Grundzügen*. 2. überarbeitete Auflage. (Schriften des Instituts für Deutsche Sprache 8). Berlin und New York: Walter de Gruyter.
- Müller, Frank Henrik (2004). *Stylebook for the Tübingen Partially Parsed Corpus of Written German (TüPP-D/Z)*. Technischer Bericht, Universität Tübingen, Seminar für Sprachwissenschaft.
- Neef, Martin (2009). IE, Germanic: German. In: Rochelle Lieber und Pavol Štekauer (Hrsg.): *The Oxford Handbook of Compounding*. (Oxford Handbooks in Linguistics). Oxford: Oxford University Press, 386-399.

- Nübling, Damaris und Szczepaniak, Renata (2011). *Markmal(s?)analyse, Seminar(s?)arbeit und Essen(s?)ausgabe: Zweifelsfälle der Verfügung als Indikatoren für Sprachwandel. Zeitschrift für Sprachwissenschaft* 30(1), 45–73.
- Rak, Zofia (1990). *Substantivische Komposita im Deutschen - semantische Analyse und Beschreibung für die fachbezogene Fremdsprachenausbildung an polnischen Hochschulen*. Dissertation, Technische Universität Dresden.
- Reznicek, Marc; Lüdeling, Anke und Hirschmann, Hagen (erscheint). Competing Target Hypotheses in the Falko Corpus: A Flexible Multi-Layer Corpus Architecture. In: Ana Díaz-Negrillo (Hrsg.): *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins.
- Reznicek, Marc; Walter, Maik; Schmid, Karin; Lüdeling, Anke; Hirschmann, Hagen; Krummes, Cedric und Andreas, Thorsten (2010). *Das Falko-Handbuch. Korpusaufbau und Annotationen. Version 1.0.1*. Technischer Bericht, Humboldt-Universität zu Berlin.
- Ringbom, Håkan (1998). Vocabulary Frequencies in Advanced Learner English: A Cross-linguistic Approach. In: Sylviane Granger (Hrsg.): *Learner English on Computer*. London und New York: Addison Wesley Longman, 41–52.
- Rivet, Anne (1999). Rektionskomposita und Inkorporationstheorie. *Linguistische Berichte* 179, 307–342.
- Schiller, Anne; Teufel, Simone; Stöckert, Christine und Thielen, Christine (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS*. Technischer Bericht, Universität Stuttgart & Institut für maschinelle Sprachverarbeitung, Universität Tübingen, Seminar für Sprachwissenschaft.
- Schlücker, Barbara (2012). Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung. In: Livio Gaeta und Barbara Schlücker (Hrsg.): *Das deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte*. (Linguistik - Impulse & Tendenzen 46). Berlin: De Gruyter, 1–25.
- Schlücker, Barbara und Hüning, Matthias (2009). Compounds and Phrases. A Functional Comparison between German A + N Compounds and Corresponding Phrases. *Italian Journal of Linguistics / Rivista di Linguistica* 21(1), 209–234.
- Schmid, Helmut (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In: *Proceedings of the Conference on New Methods in Language Processing*. Manchester, UK, 44–49.
- Schmid, Helmut; Fitschen, Arne und Heid, Ulrich (2004). SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. In: *Proceedings of the Fourth Conference on Language Resources and Evaluation (LREC)*. Lisbon, Portugal, 1263–1266.
- Selinker, Larry (1972). Interlanguage. *International Review of Applied Linguistics* 10(3), 209-231.
- Siebert, Susann (1999). *Wortbildung und Grammatik. Syntaktische Restriktionen in der Struktur komplexer Wörter*. (Linguistische Arbeiten 408). Tübingen: Niemeyer.
- Trynkler, Anna (2010). Deutsche Komposita und ihre Erlernungsmöglichkeiten in der Oberschule. *Studia niemcoznawcze. Studien zur Deutschkunde* 44, 467–475.

- Tschirner, Erwin (2005). Korpora, Häufigkeitslisten, Wortschatzerwerb. In: Antje Heine; Mathilde Hennig und Erwin Tschirner (Hrsg.): *Deutsch als Fremdsprache - Konturen und Perspektiven eines Fachs*. München: iudicium, 133–149.
- Tuggy, David (2005). Cognitive Approach To Word-Formation. In: Pavol Štekauer und Rochelle Lieber (Hrsg.): *Handbook of Word-Formation*. (Studies in Natural Language & Linguistic Theory). Dordrecht: Springer, 233–266.
- Zeldes, Amir; Lüdeling, Anke und Hirschmann, Hagen (2008). What's Hard? Quantitative Evidence for Difficult Constructions in German Learner Data. In: Antti Arppe; Kaius Sinnemäki und Urpo Nikanne (Hrsg.): *Proceedings of Quantitative Investigations in Theoretical Linguistics 3 (QITL-3)*. Helsinki, 74–77.

Adresse des Verfassers:

Dr. Amir Zeldes

Institut für deutsche Sprache und Linguistik

Humboldt-Universität zu Berlin

Unter den Linden 6

D-10099 Berlin

E-Mail: amir.zeldes@rz.hu-berlin.de