

Martin Hilpert (Neuchâtel)

Wie viele Konstruktionen stecken in einem Wortbildungsmuster?

Eine Problematisierung des Produktivitätsbegriffs aus konstruktionsgrammatischer Sicht

Abstract: Dieser Beitrag geht der Frage nach, wie sich der Begriff der Produktivität konstruktionsgrammatisch fassen lässt und zu welchen neuen Einsichten ein solches Verständnis von Produktivität führt. Zu diesem Zweck werden einige Aspekte von Produktivität diskutiert und auf das konstruktionsgrammatische Konzept eines Netzwerks von sprachlichen Mustern bezogen. Während Produktivität typischerweise als globale Eigenschaft eines Wortbildungsmusters verstanden wird, erlaubt die Konstruktionsgrammatik eine differenziertere Sicht. Am Beispiel englischer Partizipialkomposita wird illustriert, dass verschiedene Teile eines Konstruktionsnetzwerks unterschiedlich produktiv sein können, und dass Produktivität in niedrigen Ebenen des Netzwerks nicht mit der Produktivität höherer Netzwerkknoten gleichzusetzen ist.

1 Einleitung

Zur morphologischen Produktivität existiert eine weitläufige Literatur, in der dieser Begriff unterschiedlich ausgelegt und ihm auch eine gewisse inhärente Unklarheit bescheinigt wird (z.B. Aronoff 1976; Mayerthaler 1981; Bauer 2001; Baayen 2005). Der vorliegende Beitrag widmet sich der Frage, ob die Konstruktionsgrammatik (Goldberg 1995, 2006; Ziem/Lasch 2013) etwas Nützliches zum Verständnis des Begriffes beitragen kann und ob eine konstruktionsgrammatische Perspektive auf Produktivität (wie sie auch in Barðdal 2008 oder Perek 2016 erarbeitet wird) zu Einsichten führt, die über das bisher gesagte hinausgehen. Wie im Titel dieses Beitrags bereits angedeutet, wird dabei eine Rolle spielen, dass ein einzelner Wortbildungsprozess als eine Vielzahl von mehr oder minder abstrakten, untereinander verbundenen Konstruktionen aufgefasst werden kann. Anhand empirischer Daten soll gezeigt werden, dass diese Konstruktionen in Bezug auf den Grad ihrer jeweiligen Produktivität variieren, und dass mehrere produktive Konstruktionen sich nicht notwendigerweise zu einem einzigen produktiven Wortbildungsprozess aufaddieren. Als Fallstudie wird die historische Entwicklung von

Partizipialkomposita wie *doctor-recommended* oder *chocolate-covered* im Amerikanischen Englisch der letzten 150 Jahre untersucht.

Der vorliegende Beitrag gliedert sich wie folgt. Abschnitt 2 skizziert zunächst einige Grundzüge der Konstruktionsgrammatik und erörtert, wie morphologische Produktivität aus einer konstruktionsgrammatischen Sichtweise heraus verstanden werden kann. Abschnitt 3 stellt die empirische Datengrundlage dieses Beitrags vor und diskutiert die Methoden, die für ihre Untersuchung herangezogen wurden. Im Anschluss werden die Ergebnisse präsentiert und im Kontext der theoretisch motivierten Frage nach einem konstruktionsgrammatischen Verständnis von Produktivität interpretiert. Abschnitt 4 schließt mit einigen weiterführenden Überlegungen.

2 Produktivität aus konstruktionsgrammatischer Sicht

Die gebrauchsbasierte Konstruktionsgrammatik (Goldberg 1995; 2006; Ziem/Lasch 2013) konnte sich in den vergangenen Jahren als neuer theoretischer Ansatz in der Sprachwissenschaft etablieren. Ein Grundgedanke dieser Theorie ist es, dass sich sprachliches Wissen als ein strukturiertes Inventar von symbolischen Einheiten, sogenannten Konstruktionen, verstehen lässt (Hilpert 2014, S. 2). Das Konstruktionsinventar wird von Sprechern aus dem Sprachgebrauch heraus induziert (Tomasello 2003) und unterliegt fortwährender Veränderung durch jedes neue Sprachereignis (Bybee 2010). Die Konstruktionen in diesem Inventar variieren in Komplexität und Schematizität, so dass die vertretene Bandbreite von monomorphemischen Wörtern bis hin zu abstrakten syntaktischen Gefügen reicht. Einheitliches Merkmal von Konstruktionen ist lediglich, dass sie eine Formseite und eine Inhaltsseite haben, was sie zu bedeutungstragenden Strukturen macht. Der theoretische Anspruch der Konstruktionsgrammatik ist es, sprachliches Wissen in seiner Gesamtheit auf diese Strukturen zu reduzieren. Dieser Ansatz, für den Goldberg (2006, S. 18) den Slogan „it’s constructions all the way down“ geprägt hat, wendet sich damit gegen eine modulare Auffassung von sprachlichem Wissen, in der Lexikon und Grammatik strikt getrennt sind.

Ein zentraler Punkt der konstruktionsgrammatischen Theorie ist die Metapher eines Netzwerks, in dem Konstruktionen miteinander verbunden und hierarchisch geordnet sind. Da dieses Netzwerk sprachliches Wissen abbilden soll, ist es als eine kognitive Struktur zu verstehen. Innerhalb dieser Struktur sind Konstruktionen durch verschiedene Typen assoziativer Verbindungen miteinander verknüpft, die sich auf mehrere Prinzipien gründen. Das wichtigste dieser Prinzipien,

das für sogenannte Vererbungsrelationen eine zentrale Rolle spielt, ist Kategorialität: Eine Kollokation wie *You drive me crazy* instanziiert die übergeordnete Kategorie eines idiomatischen Musters, in dem das Verb *drive* mit einem Objekt und einem Element wie *mad*, *nuts*, oder eben *crazy* kombinierbar ist. Dieses idiomatische Muster wiederum instanziiert eine abstraktere Kategorie, nämlich die englische Resultativkonstruktion (z.B. *John hammered the metal flat*). Diese Konstruktion kombiniert ein Verb mit einem Objekt und einem resultativen Element. Die Resultativkonstruktion ihrerseits instanziiert die noch abstraktere Konstruktion einer Verbalphrase, die sämtliche möglichen Kombinationen eines Verbs mit seinen Begleitern abbildet. Verbindungen zwischen abstrakten Konstruktionen und ihren konkreteren Instanzierungen werden unter dem Begriff der Vererbung (*inheritance*, Goldberg 1995, S. 72) diskutiert. Die theoretische Annahme hinter dieser Begriffswahl ist, dass abstrakte Konstruktionen Informationen an ihre konkreten Instanzierungen „vererben“, so dass grammatische Informationen, die auf der höchsten Hierarchieebene repräsentiert sind, von dort aus von sämtlichen konkreteren Instanzierungen abgerufen werden können. Innerhalb der Konstruktionsgrammatik konkurrieren zu der Idee der Vererbung zwei Sichtweisen. Die Theorie der vollständigen Vererbung (vgl. Fillmore et al. 1988) vertritt den Standpunkt, dass vererbte Informationen ausschließlich auf der höchstmöglichen Ebene repräsentiert sind. Verfechter einer gebrauchsbasierten Konstruktionsgrammatik (Goldberg 2006) nehmen dahingegen an, dass grammatische Informationen im Netzwerk sprachlichen Wissens redundant repräsentiert sind, sich also auf mehreren Ebenen des Netzwerks in identischer Form wiederfinden.

Neben Kategorialität ist Ähnlichkeit ein weiteres Prinzip, nach dem Konstruktionen im Netzwerk geordnet sind. Wenn Konstruktionen in Bezug auf Form oder Bedeutung Überlappungen aufweisen, dann führen diese Überlappungen dazu, dass die Konstruktionen miteinander assoziiert werden. Ein Beispiel für solche Assoziationen wäre etwa die Verbindung zwischen der ditransitiven Konstruktion und der transitiven Konstruktion. Ein Satz wie *John wrote Mary a letter* ähnelt *John wrote a letter* in Form und Bedeutung, so dass Sprecher die beiden miteinander assoziieren. Ähnlich lässt sich argumentieren, dass die deutschen Subjunktionen *dass*, *ob* und *weil* im Sprecherwissen miteinander verbunden sind, da sie alle mit Verbletzstellung auftreten.

Neben Kategorialität und Ähnlichkeit spielt auch das Prinzip der Kontiguität eine Rolle für die Verknüpfungen zwischen Konstruktionen. Zwei Konstruktionen sind dann assoziativ miteinander verbunden, wenn der Gebrauch der einen Konstruktion im Hörer die Erwartung der anderen weckt, wie es etwa bei Kollokationen wie *Bescheid geben* oder *eine Genehmigung einholen* der Fall ist. Das jeweils erste Element evoziert hier das zweite Element des Wortpaares. Kontiguität besteht auch zwischen morpho-syntaktischen Mustern und lexikalischen Elemen-

ten. Assoziationen dieser Art sind vielfach durch die sogenannte Kollostruktionsanalyse nach Stefanowitsch/Gries (2003) untersucht worden.

Wie passt jetzt das Konzept der Produktivität in dieses Verständnis von sprachlichem Wissen? Davon ausgehend, dass die Konstruktionsgrammatik eine Theorie sprachlichen Wissens ist, versteht sich Produktivität in diesem Beitrag als eine kognitive Größe. Angelehnt an eine Definition in van Marle (1985) soll der Begriff hier beschreiben, mit welchem Maß an kognitivem Aufwand komplexe Formen prozessiert werden. Diese Definition impliziert, dass Produktivität kein diskretes Phänomen ist, sondern ein graduelles. Sie beinhaltet weiterhin, dass die Produktivität einer Konstruktion umso stärker ist, je leichter es Sprechern fällt, neue Instanzen der Konstruktion zu produzieren oder zu prozessieren. Der Umgang mit neuen Instanzen einer Konstruktion soll im Folgenden anhand von Wortbildungsmustern näher diskutiert werden.

Wortbildungsmuster, wie beispielsweise das englische Nominalisierungsschema *Adjektiv + ness*, sind ebenfalls als Konstruktionen zu verstehen (Booij 2010). Sie stellen eine schematische Verbindung dar, in der ein Affix mit einem abstrakten Platzhalter kombiniert wird und diese Kombination eine bestimmte Bedeutung trägt. Eine Auffassung von Wortbildungsprozessen als Konstruktionen ist mehr als nur eine neue Benennung morphologischer Phänomene. Die Idee eines hierarchisch geordneten Netzwerks impliziert, dass die abstrakte Form eines Wortbildungsmusters (wie etwa im Englischen *Adjektiv + ness*, *Verb + er*, etc.) lediglich den obersten Knotenpunkt in einem Netzwerk mit vielen weiteren Verästelungen darstellt. Die untersten dieser Verästelungen sind konkrete Worttypen wie etwa *darkness* oder *sweetness*. Zusätzlich zu der schematischen Form eines Wortbildungsmusters und dessen Worttypen erlaubt die Konstruktionsgrammatik aber die Annahme weiterer Knotenpunkte im Netzwerk, die sich auf einer mittleren Abstraktionsebene zwischen diesen beiden Extremen befinden. Diese Knotenpunkte lassen sich als untergeordnete Muster oder Subschemata bezeichnen. Traugott (2008, S. 236) schlägt hierfür den Begriff der Meso-Konstruktion vor und steckt mit den Termini der Makro-, Meso- und Mikro-Konstruktion das gesamte Kontinuum der unterschiedlichen Abstraktionsgrade ab. Als konkretes Beispiel für Meso-Konstruktionen lassen sich für die englische Konstruktion *Verb + er* mehrere untergeordnete Muster annehmen, die in ihrer Semantik jeweils spezifischer sind als das abstrakte, übergeordnete Muster. Formen wie *worker*, *singer* oder *teacher* bringen agentive Rollen zum Ausdruck, Formen wie *opener*, *grinder* oder *stapler* bezeichnen Werkzeuge. Weiterhin finden sich idiosynkratische Formen wie *five*, die sich in die genannten semantischen Kategorien nicht ohne weiteres einfügen und auch morphologisch von dem übergeordneten Schema abweichen, da mit dem Zahlwort *five* kein verbaler Stamm vorliegt. Eine konstruktionsgrammatische Analyse von Wortbildungsprozessen erklärt diese Be-

funde durch die Annahme eines Netzwerks von unterschiedlich abstrakten Subschemata, die Aspekte ihrer Form und Bedeutung teilen, aber jeweils auch eigenständige Züge aufweisen. Wie weiter unten dargelegt werden soll, können sich diese Subschemata in ihrer jeweiligen Produktivität stark unterscheiden.

Die Netzwerkmetapher der Konstruktionsgrammatik erlaubt eine neue Perspektive auf bestehende korpuslinguistische Ansätze zur Produktivität. Über verschiedene Aspekte von Produktivität herrscht trotz der Vielfalt der Definitionen weitgehende Einigkeit. So ist etwa unstrittig, dass die Typenfrequenz eines Wortbildungsmusters einen wesentlichen Aspekt von Produktivität ausmacht. Je höher die Typenfrequenz eines Musters, desto wahrscheinlicher ist es, dass das Muster stark produktiv ist (Bybee 2010, S. 67; Barðdal/Gildea 2015, S. 33). Ausnahmen von dieser Generalisierung wie etwa das frequente, aber unproduktive englische Derivationssuffix *-ment* (Hilpert 2013, S. 113) zeigen allerdings, dass weitere Aspekte eine Rolle spielen. Korpuslinguistische Messungen von Produktivität (Baayen 2005) beziehen daher üblicherweise die Frequenz von niederfrequenten Typen in die Berechnungen mit ein. Die Überlegung hier ist, dass das häufige Auftreten von seltenen Formen ein Indiz dafür ist, dass Sprecher einen Wortbildungsprozess dazu nutzen, Neologismen zu produzieren. Hapax Legomena aus einem Korpus sind zwar nicht notwendigerweise Neologismen, andererseits haben echte Neologismen eine überproportional große Wahrscheinlichkeit, mit niedriger Frequenz im Korpus aufzutreten. Aus den Kennzahlen von Tokenfrequenz und Anteil der Hapax Legomena lassen sich Produktivitätswerte berechnen, die sich mit Ergebnissen aus psycholinguistischen Experimenten gut in Einklang bringen lassen (ebd., S. 245), so dass hiermit ein nützliches Werkzeug vorliegt. Als weitere Komplikation lassen sich allerdings Beschränkungen diskutieren, die selbst hochproduktive Muster betreffen. Das Nominalisierungssuffix *-er* ist nach allen verfügbaren Maßen stark produktiv, aber trotzdem ist es nicht grundsätzlich mit allen bestehenden Verben kombinierbar. Verbstämme wie *have*, *die* oder *seem* führen bei einer Kombination mit *-er* zu Nominalisierungen, die von Muttersprachlern entweder abgelehnt oder als unidiomatisch eingestuft werden. Goldberg (2016, S. 371) führt als weitere Beispiele Kombinationsbeschränkungen von Verben in Argumentstrukturkonstruktionen an. So kann etwa die hochproduktive ditransitive Konstruktion nicht mit dem Verb *explain* verwendet werden (**Can you explain me the problem?*), und das Verb *vanish* passt nicht zur ebenfalls hochproduktiven transitiven Konstruktion (**The magician vanished the rabbit*). Wie erklären sich solche Beschränkungen?

Goldberg (ebd., S. 375) nähert sich dieser Frage mithilfe des Konzeptes eines semantischen Raums, in dem sich die etablierten Typen eines produktiven Musters befinden. Nach Goldberg hängt die Akzeptabilität eines Neologismus von zwei Punkten ab. Zum einen spielt die Verteilung der etablierten Typen im seman-

tischen Raum eine Rolle. Goldberg unterscheidet hier zwischen einer flächendeckenden Streuung (*high coverage*), bei der ein semantischer Raum gleichmäßig ausgefüllt ist, und einer spärlichen Streuung (*low coverage*), bei der weite Teile des semantischen Raums unausgefüllt bleiben oder aber starke Ungleichmäßigkeit in ihrer Füllung aufweisen. Zum anderen ist die Position eines Neologismus im semantischen Raum. Neologismen in flächendeckend gefüllten semantischen Räumen sollten dabei einen hohen Grad von Akzeptabilität aufweisen, während Neologismen in spärlich gefüllten semantischen Räumen eine geringere Akzeptabilität erfahren. Suttle/Goldberg (2011) weisen in einem psycholinguistischen Experiment mit einer künstlich erstellten Sprache nach, dass dieselben Stimuli, abhängig von der Streuung etablierter Typen in einem semantischen Raum, tatsächlich als verschieden akzeptabel eingestuft werden.

Ein fiktives Beispiel hilft, das Phänomen zu veranschaulichen: Zwei Lerner des Englischen haben kein Vorwissen zum Wortbildungsmuster mit dem Suffix *-ness*. Der erste der beiden erhält Lernmaterialien mit den Typen *sickness*, *illness*, *dizziness* und *queasiness*, die sich semantisch stark ähneln. Der zweite erlernt *sickness*, *sweetness*, *greatness* und *softness*, die semantisch heterogener sind. Nach dieser Vorbereitung wird beiden ein neuer Typ präsentiert, nämlich *carefulness*. Wie reagieren die beiden Lerner? Für den ersten Lerner war mit den ersten vier Typen ein relativ enger semantischer Rahmen gesteckt, der mit der Akzeptanz der neuen Form nun erheblich erweitert werden müsste. Durch die Akzeptanz ergäbe sich zudem eine spärliche, ungleichmäßige Streuung im semantischen Raum. Der zweite Lerner konnte bereits in den Lernmaterialien eine flächendeckende semantische Streuung beobachten, so dass der neue Typ keine allzu große Überraschung darstellt. Durch das jeweilige Vorwissen wurden also verschiedene Erwartungshaltungen geprägt, die sich auf die relative Akzeptabilität eines neuen Typen auswirken: Neologismen in flächendeckend gefüllten semantischen Räumen sind akzeptabel, Neologismen in spärlich gefüllten semantischen Räumen sind es nur dann, wenn sie eine semantische Nähe zu bereits etablierten Typen aufweisen.

Zusammenfassend lässt sich festhalten, dass mit Typenfrequenz, dem Anteil niederfrequenter Typen sowie der Verteilung etablierter Typen im semantischen Raum Kriterien vorliegen, die für die korpusbasierte Vermessung von Produktivität nützlich sind und die sich durch Triangulierung mit den Ergebnissen psycholinguistischer Forschung auch durch Verhaltensdaten bestätigt sehen. Im Folgenden sollen diese Konzepte daher nicht an sich kritisiert werden, sondern es soll lediglich argumentiert werden, dass ein weiterer Aspekt von Bedeutung ist, der in der bisherigen Diskussion zu Produktivität zu wenig Beachtung gefunden hat. Dieser Aspekt betrifft die Tatsache, dass zwei Konstruktionen mit gleichen Typenfrequenzen, gleichen Anteilen von Hapax Legomena und identischer semanti-

scher Streuung ihrer Typen dennoch unterschiedlich in der Struktur ihres Konstruktionsnetzwerks sein können, was mit Auswirkungen auf ihre Produktivität verbunden ist.

Die bis jetzt diskutierten Operationalisierungen von Produktivität teilen das zentrale Merkmal, dass sie einem sprachlichen Muster einen globalen Produktivitätswert zuweisen. Sie beziehen sich also immer auf die schematische Form eines Musters, das den höchsten Knotenpunkt in einem hierarchischen Netzwerk von Konstruktionen darstellt. Bei Derivationsmorphemen wie *-ness* oder *-er* wären es also die Schemata *Adjektiv + ness* bzw. *Verb + er*, denen ein Produktivitätswert zugewiesen wird. Ist es aber gerechtfertigt, Produktivität als Eigenschaft eines einzigen Knotenpunkts in einem solchen Netzwerk anzunehmen? Im Folgenden soll gezeigt werden, dass dies eine gewisse Vereinfachung darstellt. Mit der Konstruktionsgrammatik lässt sich fragen, welche Subschemata im Netzwerk eines bestimmten abstrakten Musters mehr oder weniger produktiv sind. Im Hinblick auf das bereits erwähnte Beispiel von *-er* wird deutlich, dass die Produktivität eines Subschemas mit agentiven Verbstämmen nahezu unbeschränkt ist (*talker, builder, runner, etc.*), während stative Verbstämme stärkeren Beschränkungen unterliegen (*believer, lover, owner, ?preferer, ?knower*) und Zahlwörter als Stämme nur sehr bedingt verwendet werden (*fiver, sixer, tenner, ?twoer, ?hundreder, ?thousander*). Ein konstruktionsgrammatisches Verständnis von Produktivität sollte daher über etablierte Operationalisierungen hinausgehen, indem es die Netzwerkstruktur zwischen dem übergeordneten, abstrakten Muster, davon ausgehenden Subschemata und konkreten Instanzierungen dieser Subschemata mit einbezieht. Wie dies konkret bewerkstelligt werden soll, wird in den nun folgenden Abschnitten diskutiert.

3 Daten, Methoden und Ergebnisse

Der empirische Teil dieses Beitrags wendet sich englischen Partizipialkomposita zu. Formen wie *chocolate-covered* oder *road-tested* exemplifizieren ein hochproduktives Wortbildungsmuster, das in Standardwerken zur Wortbildung im Englischen (Fabb 2001, S. 68; Plag 2003, S. 153; Bauer 2006, S. 490; Bauer/Lieber/Plag 2012, S. 470) sowie in Referenzgrammatiken (Biber et al. 1999, S. 534; Huddleston/Pullum 2002, S. 1659; Quirk et al. 1985, S. 1577) Erwähnung findet. Die Analyse in diesem Beitrag stützt sich auf Daten aus dem COHA (Corpus of Historical American English; Davies 2010). Das Korpus, das insgesamt 400 Millionen Tokens enthält, repräsentiert vier schriftsprachliche Textsorten (Belletristik, Zeitschriften, Zeitungstexte, und Fachliteratur), die in ihrer Verwendung von 1810 bis 2009 ab-

gebildet sind. Die vorliegende Untersuchung verwendet einen bereits bestehenden Datensatz, den Hilpert (2015) daraufhin untersucht, ob die Entwicklung englischer Partizipialkomposita einen Fall von Grammatikalisierung darstellt. Für diesen Datensatz wurden adjektivisch getaggte Komposita wie *government-sponsored*, die sich aus einem Substantiv und einem Partizip in der Vergangenheitsform zusammensetzen, aus dem COHA extrahiert. Der Datensatz enthält 127.000 Tokens, die sich auf etwa 32.000 Typen aufteilen. Bei einzelner Betrachtung der Substantive und Partizipien ergeben sich Typenfrequenzen von 8.300 für die Substantive und 3.200 für die Partizipien. In den Daten zeigt sich ein substanzieller diachroner Anstieg von Token- und Typenfrequenzen (Hilpert 2015, S. 2). Der relative Anteil von Hapax Legomena dagegen bleibt über die gesamte Zeit des COHA hinweg auf hohem Niveau stabil (ebd., S. 13). Eine Untersuchung der Daten im Hinblick auf die Partizipien zeigt, dass die ansteigenden Token- und Typenfrequenzen primär durch Entwicklungen in einer relativ geringen Anzahl von Partizipien bedingt sind. Abbildung 1 (ebd., S. 15) gibt einen Einblick in diesen Prozess. Das Schaubild zeigt die zwanzig Partizipien, die in dem Datensatz mit den größten Wortfamilien auftreten. Auf der x-Achse ist der Verlauf der Zeit abgebildet, die y-Achse zeigt normalisierte Typenfrequenzen (Typen per Million Wörter in der jeweiligen Dekade des COHA). Im Vergleich wird deutlich, dass insbesondere das Partizip *based* sich in der zweiten Hälfte des zwanzigsten Jahrhunderts überdurchschnittlich stark entwickelt. Andere Partizipien, wie *sized*, *related*, und *driven*, zeigen ein ähnliches Profil, wachsen aber weniger stark. Wieder andere, wie *colored*, *shaped*, und *covered*, weisen ein langsames, aber stetiges Wachstum auf.

Die Abbildung legt nahe, dass sich Muster wie *Substantiv + based* und *Substantiv + related* als produktive Subschemata etablieren. Wenn also im Jahr 1997 eine neue Form wie *silicone-based* im COHA aufkommt, ist dies nicht auf die Produktivität des abstrakten Musters *Substantiv + Partizip* zurückzuführen, sondern auf die Produktivität des konkreteren Subschemas *Substantiv + based*.

Die Frequenzverläufe in Abbildung 1 liefern einige Anhaltspunkte dazu, welche konkreten Partizipien sich zu produktiven Subschemata herausbilden, allerdings erlauben sie keine Rückschlüsse auf Subschemata, an denen gleichzeitig mehrere, semantisch verwandte Partizipien beteiligt sind. Mit Goldberg (2016) kann die Frage gestellt werden, wie sich die Partizipien aus dem Datensatz im semantischen Raum verteilen und ob die englischen Partizipialkomposita einen Fall flächendeckender Streuung oder eher spärlicher Streuung darstellen. Um sich dieser Frage zu nähern, kann das Instrument einer semantischen Vektorraumanalyse (Turney/Pantel 2010; Levshina 2015) benutzt werden, die es erlaubt, die im Datensatz vertretenen Partizipien im Hinblick auf ihre semantische Ähnlichkeit hin zu vergleichen. Aus einem solchen Vergleich lässt sich ableiten, ob es im

semantischen Raum dieser Partizipien Bereiche mit ungewöhnlich hoher oder ungewöhnlich niedriger Typendichte gibt. Die nächsten Absätze skizzieren die Funktionsweise dieser Methode und wie sie im vorliegenden Fall angewendet wurde.

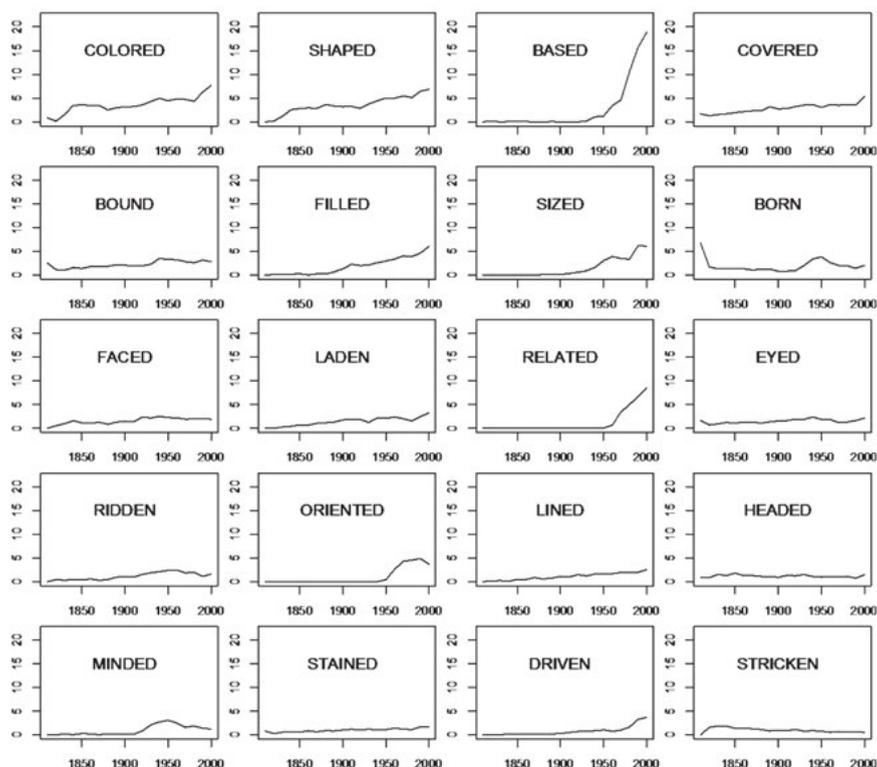


Abb. 1: Normalisierte Typenfrequenzen der 20 Partizipien mit den größten Wortfamilien im COHA

Im ersten Arbeitsschritt wurden Frequenzvektoren für sämtliche Partizipialtypen erstellt. Hierzu wurde für jedes einzelne Partizip eine Konkordanz aus dem COCA-Korpus (Davies 2008) abgerufen. Kollokatfrequenzen wurden ermittelt, indem alle Wörter innerhalb eines Kontextfensters von vier Wörtern links und rechts vom Suchbegriff ausgezählt wurden. Funktionswörter und Satzzeichen wurden von diesen Zählungen ausgenommen. Der zweite Arbeitsschritt ist die Erstellung einer Datentabelle aus den Frequenzvektoren. Jede Spalte in dieser Tabelle repräsentiert ein Partizip, jede Zeile repräsentiert ein Kollokat, das mit unterschiedlichen Frequenzen (oder auch gar nicht) mit jedem einzelnen Partizip auf-

treten kann. In einem dritten Arbeitsschritt wurde für sämtliche Frequenzwerte in der Tabelle das statistische Maß der *Pointwise Mutual Information* (PMI; Levshina 2015, S. 234) errechnet. Die Ähnlichkeiten zwischen Kollokatvektoren, die im Folgenden diskutiert werden, stützen sich also nicht auf rohe Kookkurrenzfrequenzen, sondern auf Vektoren von PMI-Werten. Als vierter Arbeitsschritt wird aus der Datentabelle eine Distanzmatrix gewonnen, indem für sämtliche paarweisen Kombinationen von Kollokatvektoren ein Ähnlichkeitsmaß mithilfe des Kosinus berechnet wird. Wenn zwei Partizipien große Überlappungen in ihren jeweiligen Kollokaten aufweisen und diese Kollokate mit ähnlichen Frequenzen auftreten, so ergibt sich ein hoher Wert, der diese starke Ähnlichkeit reflektiert. Eine Distanzmatrix dieser Art kann in einem fünften Arbeitsschritt durch ein Verfahren der Dimensionsreduktion auf ein zweidimensionales Schaubild projiziert werden, das dann als semantische Karte interpretierbar ist. In der vorliegenden Analyse wurde für diesen Arbeitsschritt das sogenannte t-SNE Verfahren (van der Maaten/Hinton 2008) zum Einsatz gebracht.

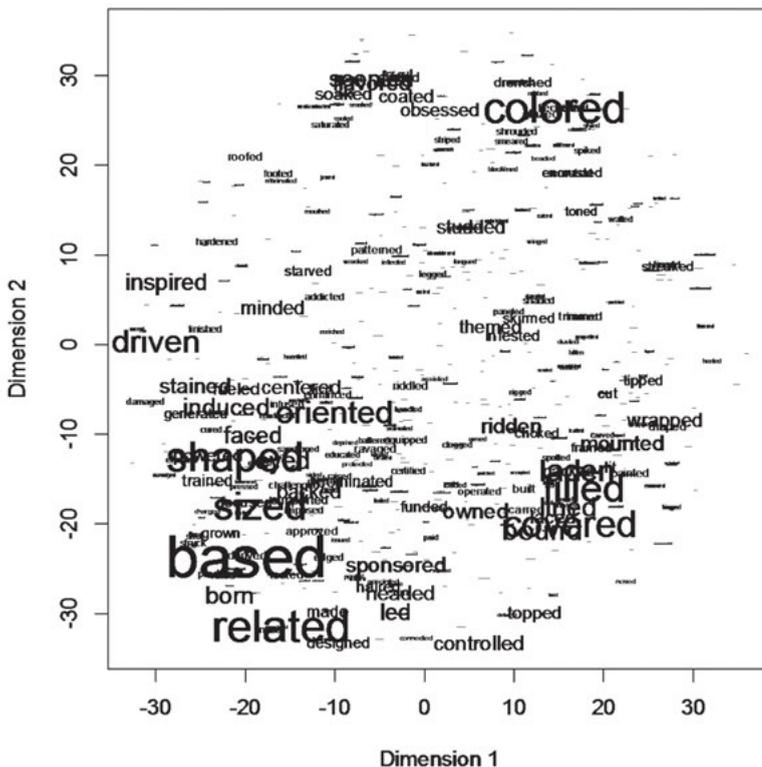


Abb. 2: Semantische Distribution einer Auswahl hochfrequenter Partizipien

Abbildung 2 zeigt als Resultat der fünf Arbeitsschritte die Verteilung der Partizipien, die in der letzten Dekade des COHA Korpus am häufigsten auftreten. Distanzen zwischen den einzelnen Formen zeigen an, ob zwei oder mehrere Partizipien mit ähnlichen Kollokaten auftreten; Textgröße bildet normalisierte Tokenfrequenzen in der letzten Dekade des COHA ab.

Die in der Abbildung zu erkennenden Partizipialtypen geben einen ersten Anhaltspunkt dazu, wie der semantische Raum von englischen Partizipialkomposita beschaffen ist. In der linken unteren Ecke des Schaubilds sind die hochfrequenten Partizipien *based*, *sized*, und *shaped* erkennbar, die zwar nah beieinander abgebildet sind, aber trotzdem klar eigene Kollokationsprofile aufweisen. In der rechten unteren Ecke lässt sich eine erste semantische Gruppierung ausmachen. Die Partizipien *covered*, *filled* und *laden* verdecken weitere Formen wie *packed*, *crammed*, oder *littered*. Gemeinsam ist diesen Partizipien, dass sie einen Sachverhalt zum Ausdruck bringen, der hier als applikativ bezeichnet werden soll. Komposita wie *chocolate-covered* oder *trash-littered* bezeichnen einen Zustand, bei dem ein Objekt mit einer Substanz bedeckt ist. Weiter oben im Schaubild ist das Partizip *colored* zu sehen, das hinter sich eine Reihe von semantisch verwandten Formen verdeckt, wie etwa *tinted*, *dyed*, *lacquered*, *glazed*, und *sprayed*, die ebenfalls auf Farben referieren. Ebenfalls in dieser Region des Schaubilds finden sich Partizipien wie *crowned*, *styled* und *engraved*, die visuelle Qualitäten verbalisieren. Etwas links von der Gruppe visueller Partizipien befindet sich eine dicht gestreute Ansammlung von Partizipien, deren frequenteste Mitglieder *flavored* und *scented* sind, die fast übereinanderliegen. In ihrer unmittelbaren Nähe findet sich eine große Gruppe niederfrequenter Partizipien mit übereinstimmenden semantischen Merkmalen, nämlich *cooked*, *chopped*, *sliced*, *sweetened*, *charred*, *fried*, *baked*, *roasted*, *heated*, *dried*, *grilled* neben vielen weiteren. Diese Gruppe lässt sich auf den gemeinsamen Nenner der kulinarischen Bedeutung bringen.

Was kann nun aus diesem Rundgang durch den semantischen Raum der englischen Partizipialkomposita gefolgert werden? Der erste Eindruck legt nahe, dass verschiedene Regionen innerhalb des semantischen Raumes unterschiedliche Streuungsdichten aufweisen, sowohl was die relative Dichte von Typen angeht als auch die Dichte von Tokens, die sich aus der Frequenz der Typen ergibt. Weiterhin zeigt sich in dem Schaubild ein Unterschied zwischen den dicht bevölkerten Regionen des semantischen Raums. Bei der applikativen Gruppe und der visuellen Gruppe lassen sich die Partizipien *filled* und *covered* beziehungsweise *colored* als hochfrequente Prototypen identifizieren, die das Zentrum einer semantischen Kategorie mit mehreren Mitgliedern bilden. Die kulinarische Gruppe ist anders aufgebaut, nämlich als Verbund niederfrequenter, semantisch verwandter Partizipien, die ohne ein prototypisches, hochfrequentes Mitglied auskommt. Der entgegengesetzte Fall betrifft Partizipien wie *based* oder *related*, die als einzelne

Partizipien zu einer Gruppe Komposita mit hoher Typen- und Tokenfrequenz gehören. Diese Heterogenität in der Füllung des semantischen Raumes reflektiert die Struktur des Netzwerks, in dem die englischen Partizipialkomposita organisiert sind. Das abstrakte Wortbildungsmuster, nach dem Substantive und Partizipien zu einem Kompositum zusammengesetzt werden können, gliedert sich in ein Netzwerk von Subschemata, zu denen „Solitäre“ wie Substantiv-*based* gehören, aber auch „Kollektive“ wie die Verbindung eines Substantivs mit einem kulinarischen Partizip. Ein möglicher Einwand gegen diese Interpretation ist, dass Abbildung 2 die tatsächliche Dichte der semantischen Streuung nur unvollkommen wiedergibt, da die kleine Schriftgröße niederfrequenter Typen eine spärliche Streuung suggeriert. Um diesem Einwand zu begegnen, zeigt Abbildung 3 die semantische Streuung auf eine andere Weise. Das Schaubild stützt sich auf dieselben Urdaten, die auch Abbildung 2 zugrunde liegen. Anstatt jedoch die einzelnen Partizipien typographisch im Raum zu zeigen, werden die Tokenfrequenzen aller Typen für jedes Planquadrat im semantischen Raum aufaddiert, so dass sich eine Karte der semantischen Streuungsdichte ergibt, die sämtliche Partizipien mit in Betracht zieht. Das Schaubild gliedert sich in vier Graphen, die Veränderungen in der Streuungsdichte über die letzten vier Dekaden des COHA abbilden.

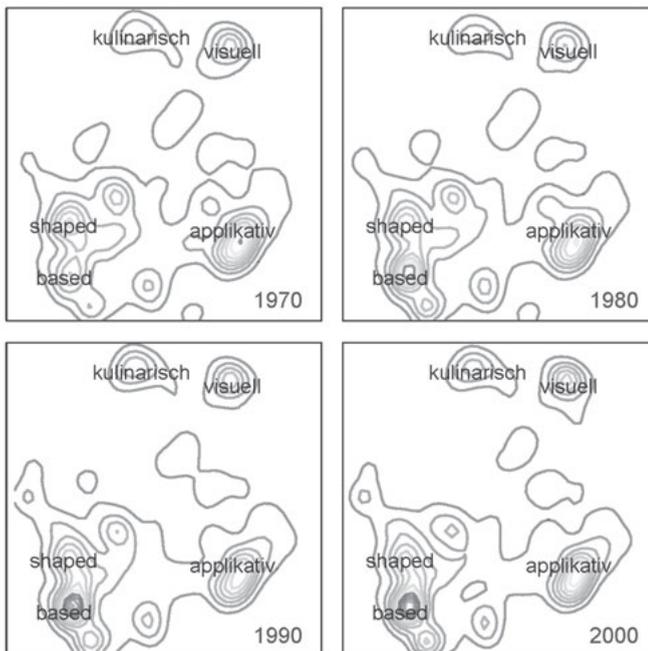


Abb. 3: Veränderungen im semantischen Raum 1970–2000

Am deutlichsten sichtbar ist in Abbildung 3 der Frequenzanstieg des Partizips *based*. Die weiter oben besprochenen semantisch motivierten Gruppierungen sind als relativ konstante Verdichtungen im semantischen Raum zu erkennen. Darüber hinaus zeigen sich kleinere Schwankungen in verschiedenen Bereichen. Das Schaubild untermauert die These, dass sich das Netzwerk der englischen Partizipialkomposita in einige Subschemata aufteilt, von denen manche aus einzelnen Partizipien bestehen und andere größere Gruppen von Partizipien beinhalten.

4 Schlussfolgerungen

In der Einleitung dieses Beitrags wurde die Frage aufgeworfen, wie sich der Begriff der Produktivität konstruktionsgrammatisch fassen lässt und ob ein solches Bemühen zu neuen theoretischen Einsichten führt. Ein Teil der Antwort liegt in der Netzwerkmetapher der Konstruktionsgrammatik begründet. Während sich bestehende Ansätze zur Produktivität generell auf die höchste Abstraktionsebene von Wortbildungsprozessen oder syntaktischen Konstruktionen beziehen, eröffnet die Konstruktionsgrammatik einen Blickwinkel, der auch konkretere Subschemata fokussieren kann. Im vorliegenden Fall lässt sich die zweifellos vorhandene Produktivität der englischen Partizipialkomposita feiner zergliedern, und zwar in die Produktivität einzelner Konstruktionen wie beispielsweise „Substantiv-*based*“ oder die Kombination eines Substantivs mit einem Partizip mit visueller Bedeutung. Diese Sichtweise liefert mehr als nur eine etwas detailliertere Betrachtung der Datenlage. Sie erlaubt auch differenzierte Vorhersagen zum kognitiven Aufwand, den Sprecher für das Verständnis neu geformter Komposita aufbringen müssen. Die Partizipien in Formen wie *ginger-steamed* oder *chipotle-spiced* haben zwar an sich keine hohen Typenfrequenzen, die das Prozessieren erleichtern würden, dafür aber befinden sie sich in einem flächendeckend gefüllten Bereich des semantischen Raums und können daher mit wenig Aufwand prozessiert werden. Zum Vergleich dazu weist ein Kompositum wie *case-hardened* ein Partizip auf, das sich im semantischen Raum in einem spärlich besiedelten Bereich befindet (vgl. Abb. 2, $x=-22$, $y=12$). Das abstrakte Wortbildungsmuster erlaubt prinzipiell neue Formationen in allen Regionen des semantischen Raums, aber wie aus Abbildung 3 deutlich wird, orientieren sich Sprecher an bereits bestehenden Formen. Diese Beobachtung deckt sich mit dem in Perek (2016) diskutierten Ergebnis, dass Dichte im semantischen Raum eine Vorhersage darüber erlaubt, was für neue Typen im weiteren Verlauf der Zeit hinzukommen. Dieses Phänomen lässt sich als Matthäus-Effekt in der Wortbildung bezeichnen: Wo bereits eine hohe Konzentration von Typen vorhanden ist, werden wahrscheinlich auch noch weitere Typen hinzukommen.

Die Regionen im semantischen Raum einer Konstruktion repräsentieren daher verschiedene Subschemata, die sich in ihrer jeweiligen Produktivität unterscheiden. Was in der weiterführenden Forschung noch zu klären sein wird, ist der genaue Aufbau des Konstruktionsnetzwerks. In der vorliegenden Analyse wurden Typen- und Tokenfrequenzen sowie ein qualitativ untersuchter semantischer Vektorraum dazu genutzt, Subschemata eines Wortbildungsprozesses zu identifizieren. Es wäre wünschenswert, die Unterteilung eines Netzwerks in verschiedene Abstraktionsebenen und Subschemata auf eine solidere Basis zu stellen, die intersubjektiv reproduzierbare Ergebnisse liefert. Gries (2011) liefert dazu bereits einige Denkanstöße, die sich auf aktuelle Ansätze wie Goldberg (2016), Perek (2016) und den vorliegenden Beitrag anwenden lassen. Von einer weiteren konstruktionsgrammatischen Auseinandersetzung mit dem Begriff der Produktivität sind also interessante Ergebnisse zu erwarten.

Literatur

- Aronoff, Mark (1976): *Word formation in generative grammar*. 2. Aufl. Cambridge.
- Baayen, R. Harald (2005): Morphological productivity. In: Köhler, Reinhard/Altmann, Gabriel/Piotrowski, Rajmund G. (Hg.): *Quantitative linguistics. An international handbook*. (= Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science 27). Berlin, S. 243–256.
- Barðdal, Jóhanna (2008): Productivity. Evidence from case and argument structure in Icelandic. (= *Constructional Approaches to Language* 8). Amsterdam.
- Barðdal, Jóhanna/Gildea, Spike (2015): Diachronic construction grammar. Epistemological context, basic assumptions and historical implications. In: Barðdal, Jóhanna et al. (Hg.): *Diachronic construction grammar*. (= *Constructional Approaches to Language* 18). Amsterdam, S. 1–50.
- Bauer, Laurie (2001): *Morphological productivity*. Cambridge.
- Bauer, Laurie (2006): Compounds and minor word-formation types. In: Aarts, Bas/McMahon, April (Hg.): *The handbook of English linguistics*. Malden, S. 483–506.
- Bauer, Laurie/Lieber, Rochelle/Plag, Ingo (2012): *The Oxford reference guide to English morphology*. Oxford.
- Biber, Douglas et al. (1999): *The Longman grammar of spoken and written English*. Harlow.
- Booij, Geert (2010): Construction morphology. In: *Language and Linguistics Compass* 3, 1, S. 1–13.
- Bybee, Joan L. (2010): *Language, usage, and cognition*. Cambridge.
- Davies, Mark (2008): *The Corpus of Contemporary American English (COCA): 450+ million words, 1990–present*. Internet: <http://corpus.byu.edu/coca> (Stand: 31.5.2017).
- Davies, Mark (2010): *The Corpus of Historical American English (COHA): 400+ million words, 1810–2009*. Internet: <http://corpus.byu.edu/coha> (Stand: 31.5.2017).
- Fabb, Nigel (2001): Compounding. In: Zwicky, Arnold /Spencer, Andrew (Hg.): *The handbook of morphology*. Oxford, S. 66–83.

- Fillmore, Charles J./Kay, Paul/O'Connor Mary Catherine (1988): Regularity and idiomaticity in grammatical constructions: The case of *Let alone*. In: *Language* 64, 3, S. 501–538.
- Goldberg, Adele E. (1995): *Constructions. A construction grammar approach to argument structure.* (= *Cognitive Theory of Language and Culture Series*). Chicago.
- Goldberg, Adele E. (2006): *Constructions at work: The nature of generalization in language.* (= *Oxford Linguistics*). Oxford.
- Goldberg, Adele E. (2016): Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption. In: *Language and Cognition* 8, 3, S. 369–390.
- Gries, Stefan Th. (2011): Corpus data in usage-based linguistics. What's the right degree of granularity for the analysis of argument structure constructions? In: Brdar, Mario/Gries, Stefan Th./Žic Fuchs, Milena (Hg.): *Cognitive linguistics. Convergence and expansion.* (= *Human Cognitive Processing* 32). Amsterdam, S. 237–256.
- Hilpert, Martin (2013): *Constructional change in English. Developments in allomorphy, word formation, and syntax.* Cambridge.
- Hilpert, Martin (2014): *Construction grammar and its application to English.* Edinburgh.
- Hilpert, Martin (2015): From hand-carved to computer-based: Noun-participle compounding and the upward-strengthening hypothesis. In: *Cognitive Linguistics* 26, 1, S. 1–36.
- Huddleston, Rodney/Pullum, Geoffrey K. (2002): *The Cambridge grammar of the English language.* Cambridge.
- Levshina, Natalia (2015): *How to do linguistics with R. Data exploration and statistical analysis.* Amsterdam.
- Mayerthaler, Willi (1981): *Morphologische Natürlichkeit.* (= *Linguistische Forschungen* 28). Wiesbaden.
- Perek, Florent (2016): Using distributional semantics to study syntactic productivity in diachrony: A case study. In: *Linguistics* 54, 1, S. 149–188.
- Plag, Ingo (2003): *Word-formation in English.* (= *Cambridge Textbooks in Linguistics*). Cambridge.
- Quirk, Randolph et al. (1985): *A comprehensive grammar of the English language.* New York u.a.
- Stefanowitsch, Anatol/Gries, Stefan Th. (2003): Collocations: Investigating the interaction of words and constructions. In: *International Journal of Corpus Linguistics* 8, 2, S. 209–243.
- Suttle, Laura/Goldberg, Adele E. (2011): The partial productivity of constructions as induction. In: *Linguistics* 49, 6, S. 1237–1269.
- Tomasello, Michael (2003): *Constructing a language. A usage-based theory of first language acquisition.* Cambridge.
- Traugott, Elizabeth C. (2008): Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In: Eckardt, Regine/Jäger, Gerhard/Veenstra, Tonjes (Hg.): *Variation, selection, development: Probing the evolutionary model of language change.* (= *Trends in Linguistics. Studies and Monographs (TiLSM)* 197). Berlin, S. 219–250.
- Turney, Peter D./Pantel, Patrick (2010): From frequency to meaning: Vector space models of semantics. In: *Journal of Artificial Intelligence Research* 37, S. 141–188.
- van der Maaten, Laurens/Hinton, Geoffrey (2008): Visualizing data using t-SNE. In: *Journal of Machine Learning Research* 9, S. 2579–2605.
- van Marle, Jaap (1985): *On the paradigmatic dimension of morphological creativity.* (= *Publications in Language Sciences* 18). Dordrecht.
- Ziem, Alexander/Lasch, Alexander (2013): *Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierte Ansätze.* (= *Germanistische Arbeitshefte* 44). Berlin.