

# Grenzen von Mehrwortsequenzen

Carolyn Odebrecht, Institut für deutsche Sprache und Linguistik  
Betreuerin: Prof. Dr. Anke Lüdeling



## 1. Fragestellung

Wie entwickeln sich Konstruktionen diachron?

→ **Methodische Fragestellung:**  
Wie können Grenzen von Mehrwortsequenzen (MWS) definiert werden?

Wie können daraus Konstruktionen ermittelt werden?

Welches theoretische Modell kann diese Phänomene erklären?

**Methode:**

- Korpuslinguistische Untersuchung von Konstruktionen im Sprachwandel
- diachron und quantitativ
- mit einer operationalisierten Definition von Konstruktionen

**Evaluierung von MWS als Datengrundlage**

Datengrundlage der Untersuchung:  
Korpus „Parlamentsreden\_Deutscher\_Bundestag“ (3.134.192 Token) und Korpus „SPD\_Oppositionsmitglieder\_“ (228.282 Token)

**Konstruktionen** werden im Sinne der Konstruktionsgrammatik als Form-Bedeutungs-Paar verstanden (Goldberg 1995).

Der Untersuchungsgegenstand soll bestimmt werden durch

- die Häufigkeit der Verwendung (Frequenz)
- zu definierende Grenzen (u.a. Länge)
- kategoriale Zuordnung (Struktur & Funktion)
- Begrenzung auf eine Art von Konstruktion
- Datengrundlage (Register & zeitliche Einordnung).

**Annahme:** Kookkurrenzen sind nicht immer zufällig. Aus Kookkurrenzen können Konstruktionen abstrahiert werden – bottom-up-Analyse.

## 2. Beispiele MWS

(1) *Außerdem ist es für mich in diesem Zusammenhang und vor allen Dingen auch vor dem Hintergrund der Debatte am heutigen Vormittag völlig unverständlich [...]*  
(17 Vorkommen)

(2) *Vor dem Hintergrund der Anfragen und der Antworten ist mir heute etwas unverständlich, warum Sie diesen Antrag vorlegen [...]*  
(18 Vorkommen)

(3) *Das ist Populismus, wenn auch vor dem Hintergrund der Landtagswahl in Sachsen-Anhalt – das ist schon gesagt worden – ein verständlicher.*  
(59 Vorkommen)

(4) *Insgesamt halte ich es für begrüßenswert, dass sich die Bundeswehr als Berufsfeld vorstellt.*  
(30 Vorkommen)

(5) *Ich halte es für ein Gebot der Höflichkeit und des Respektes voreinander, dass man von Menschen [...] mehr weiß als einen Namen.*  
(60 Vorkommen)

(6) *Ich sage Ihnen eines, Herr Steinmeier – den Zwischenruf mit der Münchner Sicherheitskonferenz habe ich gehört – [...]*  
(9 Vorkommen)

(7) *Ich sage Ihnen eines: Wir Grüne sind für deutlich mehr und längere Lastzüge, aber bitte schön auf der Schiene.*  
(9 Vorkommen)  
(Die Zahlen stellen absolute Vorkommen in beiden Korpora dar.)

**MWS nach Biber et al. (2004) – Lexical Bundles (LB)**

- 4-Wort-Sequenzen
- auf Wortformen basierend
- cut off: Vorkommen in einem Text
- cut off: Vorkommen in einem Korpus
- strukturell blind erhoben

→ Adaption dieser Methode in Odebrecht (2011)

## 3. Problematisierung von LB als MWS

**Flexionsmorphologie und Orthografie (2 & 3)**  
vor dem Hintergrund der (59 Vorkommen)  
Vor dem Hintergrund der (18 Vorkommen)  
vor dem Hintergrund des (13 Vorkommen)

Berücksichtigung von Flexionsendungen bei der Erhebung:  
→ separate Zählung verschiedener Kasusendungen und orthografischer Varianten  
→ Problematisch: Type-Token-Verhältnis

**Wortstellungsvarianz im Deutschen (4 & 5)**

Strikt serielles Vorgehen:  
→ separate Zählung von Wortstellungsvarianten  
→ Problematisch: Type-Token-Verhältnis

**Strukturelle Grenzen (6 & 7)**

Theoretischer Status unklar:  
→ LBs nur innerhalb eines Satzes  
→ ohne Berücksichtigung des Kontextes  
→ (6) LB als ein Teil eines Satzes, (7) als vollständiger Satz

→ keine theoretischen Vorannahmen über LB  
Methode zur Erhebung als Definition von LB

**Zwischenfazit:** Strukturell blinde Erhebung, strikte Serialität, ungenügende, strukturelle Kategorisierungsmöglichkeiten aufgrund der Syntax und Flexionsmorphologie sowie Orthographie des Deutschen zeigen, dass MWS (basierend auf der Wortform-Ebene) allein keinen genügender Ansatz für die Identifikation von Konstruktionen darstellen.

## 4. Lösungsansatz

Welche anderen verwandten Phänomene können hinzugezogen werden, um Konstruktionen zu ermitteln?

- Kookkurrenzen (MWS) basierend auf Lemmata – Type-Token-Verhältnis
- Kollokationen nicht strikt seriell – Wortstellungsvarianz
- kategoriale Zuordnung durch Part of Speech (PoS)-Annotationen und Dependenzrelationen
- keine Festlegung auf eine bestimmte Anzahl an Wörtern einer MWS

→ Anwendung auch auf historischen Korpora möglich, wenn Normalisierung und Lemmatisierung und PoS-Annotationen im Korpus vorhanden sind.

**Existierende Ansätze**

→ bereits Unterscheidung zwischen lexikalischen und grammatischen Kollokationen und Mehrebenen-MWS

„Collostructions“ (Stefanowitsch und Gries, 2003)

„Syntax-based Collocation Extraction“ (Seretan, 2011)

„Hybrid n-grams“ (Wible und Tsao, 2010)

→ kaum Anwendung von Diagnostiken für Grenzen von sprachlichen Einheiten wie Konstituens, Kongruenz oder Phrasenzugehörigkeit möglich:

- Grenzen durch die gleiche Frequenz von Kookkurrenzen bestimmt
- offene wie geschlossene Strukturen nach generativer Sicht
- Kongruenz von „fehlenden“ Elemente nicht bestimmbar.

## 5. Beispielanalyse von kombinierten frequenzbasierten Phänomenen

(8)		MWS	MWS	WWS	MWS						gram. Koll
tok		dass		es	sich	um	einen	unsäglichen	Skandal		handelt
PoS		KOUS		PPER	PRF	APPR	ART	ADJA	NN		VVFIN
lemma		dass		es	er sie es	um	ein	unsäglich	Skandal		handeln
(MWS 34 Vorkommen, Bsp. aus Odebrecht 2011)											

- Ermittlung auf Grundlage der Lemma-Ebene
- strukturelle Parameter erzeugen Frame der Konstruktion
- mittels Frame-Semantik Einordnung der Konstruktion (vgl. dazu Boas und Sag 2010) Abstrahierung von konkreten Daten

**Annahmen:**

Kollokationen als empirisch fundierte Aussage über die Wahrscheinlichkeit benachbarter Elemente (Evert 2009, 1218).

**Annahme:** Verschiedene Abstraktionsgrade von Konstruktionen.

(9) [...] bei denen **es sich nicht um Staatsbeihilfen der Mitgliedstaaten handelt** [...] (1 Vorkommen, Bsp. aus Odebrecht 2011)

- Prinzip: Kollokationen beinhalten Kollokationen (s. Stefanowitsch und Gries 2003).
- (8) & (9) bildet das Verb *handeln* dem semantischen Kern,
- Herstellung eines Bezuges zwischen (8) & (9)
- Ermittlung Grenzen von Konstruktionen demnach über Gemeinsamkeiten möglich

Berücksichtigung von nicht nur Oberflächenmerkmalen sondern auch von syntaktischen und semantischen Eigenschaften

## 6. Eigener Untersuchungsgegenstand

→ **Nominale oder pronominale Kookkurrenzen**

(10) *Das sage ich auch vor dem Hintergrund, dass wir [...]* (47 Vorkommen, Bsp. Odebrecht 2011)

(11) *Gibt es für den Fall, dass die italienische Regierung das wahrmacht [...]* (21 Vorkommen, Bsp. Odebrecht 2011)

- (10) & (11) unterschiedliche Einbettung der MWS in den Kontext bzw. Satz im Vergleich zu (8) & (9)
- unterschiedlicher semantischer Kern, aber Einleitung einer gleichen syntaktischen Struktur

→ Belege (10) & (11) weisen linguistisch interessante Strukturen auf!

**Offene Fragen:**

Haben die solche Belege gemeinsame Strukturen oder semantische Eigenschaften? Gibt es hier eine gemeinsame Konstruktion? Wie kann diese theoretisch eingebettet werden? Wie lassen sich solche Phänomene theoretisch z.B. innerhalb der „Formulaic Language“ (Wray 2008) diskutieren?

**Ausblick:**

Diese Belege näher zu betrachten wird ein weiterer Schritt sein. Wenn eine operationalisierbare Definition und theoretische Motivation einer nominalen Konstruktion gelingt, soll diese diachron untersucht werden.