

KoMuX – Kompositamuster-Explorer

Eine einmalige Datengrundlage für die deutsche Komposita-Forschung

www.owid.de/plus/komux

Annelen Brunner & Katrin Hein

Komposita mit lexikalischem Anker, z. B. X-Mann (*Lederjacken-Mann*, *Museumsmann*)

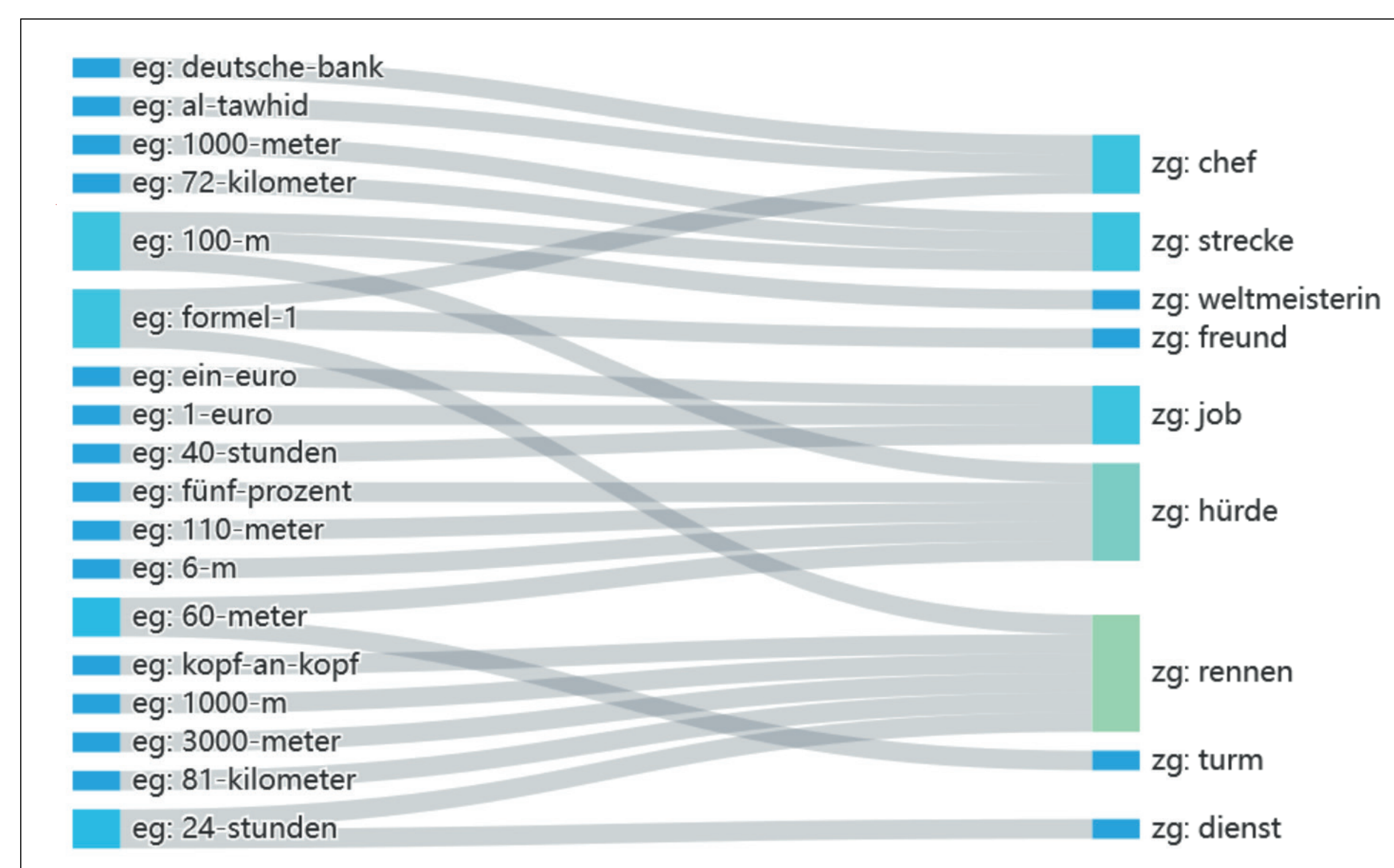
z. B. Komposita, deren Erst- und/oder Zweitglied ebenfalls ein Kompositum ist (*Autobahnraststätte*, *Frauenfußballweltmeisterschaft*)

z. B. ADJEKTIV-NOMEN-Komposita (*Kleinkind*)
z. B. NOMEN-NOMEN-Komposita (*Zeitpunkt*)

Komposita gruppiert nach semantisch-thematischen Mustern, z. B. ARTEFAKT-SUBSTANZ-Komposita (*Treibhausgas*)

Was ist KoMuX?

- ▶ Webanwendung erlaubt musterbasierte Suche nach abstrakten oder lexikalisch teilspezifizierten Komposita-Mustern
- ▶ Visualisierungen helfen dabei, die Ergebnisse näher zu analysieren
- ▶ Alle Daten im CSV-Format herunterladbar
- ▶ API erlaubt direkten Zugriff auf die Datenbank



Ausschnitt aus dem Verknüpfungsdiagramm für die Ergebnismenge mit dem Muster [PHRASE_{Erstglied}+NOMEN_{Zweitglied}]

MUSTERPERSPEKTIVE

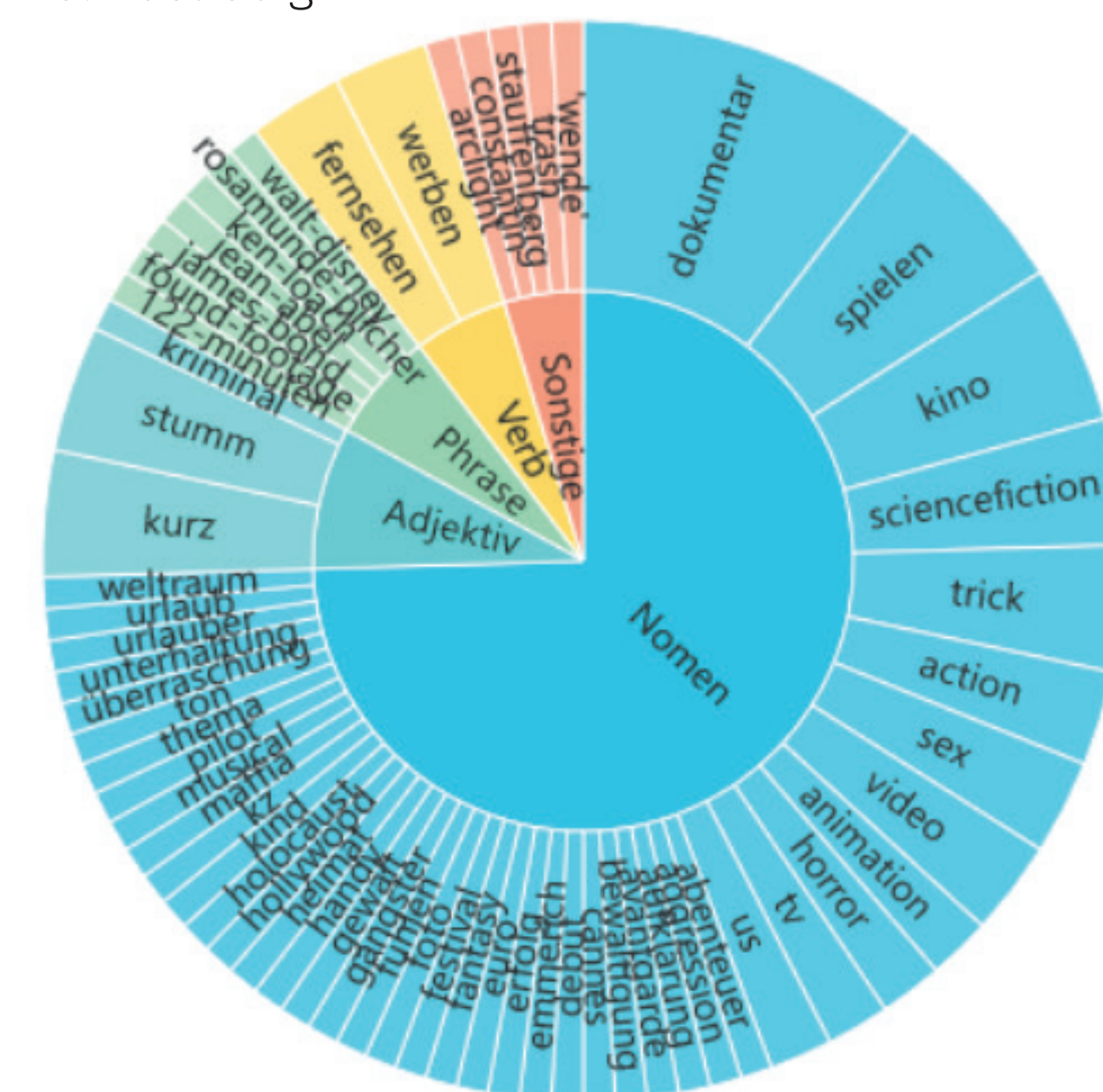
- ▶ Muster-Zugang zum Komposita-Inventar ist auch linguistisch motiviert: Komposita als konkrete sprachliche Realisierungen von zugrundeliegenden Mustern
- ▶ Muster als Grundprinzip, das erklärt, wie die Komposition grundsätzlich funktioniert → Anknüpfungspunkte zur Konstruktionsgrammatik (KxG)
- ▶ Mit KoMuX machen wir auch einen Teil der Datengrundlage zugänglich, auf der unsere eigenen Forschungen beruhen, z. B.

Hein/Brunner (2020): Why do some lexemes combine more frequently than others? In: Online Proceedings of the 12th Mediterranean Morphology Meeting. DOI: 10.26220/mmm.3247

Brunner/Engelberg/Hein (2021): The distribution of constituent words in nominal compounds and its impact on semantic interpretation: an empirical study. In: Journal of Word Formation. DOI: 10.3726/zjw.2021.01.01

Erstglied

Lemma: <beliebig>



Zweitglied

Lemma: film

film

Quantitative Verteilungen (Suchebenen ‚Wortart‘ und ‚Lemma‘) beim Erstglied für Komposita mit dem Zweitglied *Film*.

5599 unterschiedliche Konstituenten		Tabelle durchsuchen	
Konstituente	Häufigkeit als Erstglied	Häufigkeit als Zweitglied	
abend	13	41	
abenteuer	5	3	
Konstituente als Erstglied: Abenteuerlust, Abenteuerland, Abenteuersport, Abenteuerfilm, Abenteuerreise		Konstituente als Zweitglied: 007-Abenteuer, Agenten-Abenteuer, Knet-Abenteuer	
abitur	2	2	
abo	1	1	

Ausschnitt aus der Konstituenten-Tabelle für die Ergebnismenge mit dem Muster [SIMPLEX + SIMPLEX]

DATENGRUNDLAGE

- ▶ Beruht auf dem KoGra-Untersuchungskorpus (ca. 7 Milliarden Tokens) [<https://grammis.ids-mannheim.de/korpusgrammatik/6615>]
- ▶ 100.000 nominale Komposita-Tokens, zufällig gewählt aus einer Gesamtmenge von ca. 489 Millionen Tokens → über 50.000 Komposita-Typen
- ▶ Automatische morphologische Analysen, manuell nachkorrigiert
- ▶ Anreicherung mit GermaNet-Kategorien für die Erst- und Zweitglied-Lemmata (grobe semantisch-thematische Zuordnung)

Kontakt:
Dr. Annelen Brunner / Dr. Katrin Hein
Abteilung Lexik
Leibniz-Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim

Tel: +49 621 1581-433 (Brunner)
+49 621 1581-421 (Hein)
Fax: +49 621 1581-200
brunner@ids-mannheim.de
hein@ids-mannheim.de

Hausanschrift:
Leibniz-Institut für Deutsche Sprache
R 5, 6-13
68161 Mannheim

Tel: +49 621 1581-0
Fax: +49 621 1581-200
info@ids-mannheim.de
www.ids-mannheim.de

© 2023 IDS Mannheim/ÖA