

Datensatz Genitiv- und *von*-Attribute

Der Datensatz enthält 16.604 Korpusbelege aus Nominalphrasen mit Genitiv- und *von*-Attributen (*die Ideen zahlreicher Kinder, die Ideen von zahlreichen Kindern*), wobei die Genitivattribute prä- oder postnominal erscheinen können (*Mannheims Sehenswürdigkeiten, die Sehenswürdigkeiten Mannheims*).

Für jeden Beleg sind Informationen zu Land, Dekade und Medium enthalten. Hinzu kommen Angaben zu Kopf- und/oder Attributslemma (z. B. Namentyp, Flexionsklasse), Gesamtphrase (z. B. Definitheit, Kasus) und Attributsphrase (z. B. Kasusdistinktion, Länge). Zahlreiche Sonderfälle sind ebenfalls annotiert (z. B. Genitiv bei nichtflektiertem Adjektiv wie *Gebäck Mannheimer Bäckereien*, Phrasen mit adjektivisch flektierendem Attributsnomen wie *die Ideen Jugendlicher, die Ideen von Jugendlichen*).

Der Datensatz basiert auf einem Sample von Zeitungsbelegen aus dem KoGra-Untersuchungskorpus (basierend auf DeReKo, Release 2017-II; Kupietz et al. 2010, 2018), stratifiziert nach Land (AT, CH, DE) und Dekade (1990er, 2000er, 2010er) und Internetforenbelegen aus dem DECOW16B-Korpus (Schäfer & Bildhauer 2012). Enthalten sind je ca. 1.500 Zeitungsbelege für jede Kombination aus Land und Dekade sowie ca. 1.500 Belege für die Internetforen. Erhoben wurden Konstruktionen, bei denen ein Nomen (Appellativ oder Eigenname) im Genitiv syntaktisch abhängig von einem anderen Nomen ist, dem es voran- oder nachgestellt ist, oder bei denen eine Präpositionalphrase mit der Präposition *von* ein Nomen regiert und gleichzeitig von einem ihr vorausgehenden Nomen abhängig ist. Die Abhängigkeit wurde über die Dependenzannotation der Korpora bestimmt; dabei ist es unerheblich, wie viele Elemente zwischen den beiden Substantiven stehen. Alle Belege wurden im Anschluss manuell überprüft.

Der Datensatz enthält u.a. Teildatensätze für prä- und postnominale Genitivattribute, die gegeneinander austauschbar sind (GENVON), und Genitiv- und *von*-Attribute, die gegeneinander austauschbar sind (PREPOST). Er dient der Untersuchung der Genitivstellung (Kopf 2021a), der Ermittlung des Variationsbereichs der beiden Konstruktionstypen (Kopf 2021b) und der Ermittlung der variationsbedingenden Einflussfaktoren (Kopf & Bildhauer i. Vorb.). Für die Statistik und weitere Daten in Kopf (2021a) wird zusätzlich ein R-Skript zur Verfügung gestellt.

Literatur

- Kopf, Kristin. 2021a. Stellung des adnominalen Genitivs. In Marek Konopka, Angelika Wöllstein & Ekkehard Felder (Hgg.), *Determination, syntaktische Funktionen von Nominalphrasen und Attribution (Bausteine einer Korpusgrammatik des Deutschen 2)*, 103–134. Heidelberg: Heidelberg University Publishing. DOI: <https://doi.org/10.17885/heiup.bkgd.2021.1.24420>
- Kopf, Kristin. 2021b. Genitiv- und *von*-Attribute: Bestimmung des Variationsbereichs. In Marek Konopka, Angelika Wöllstein & Ekkehard Felder (Hgg.), *Determination, syntaktische Funktionen von Nominalphrasen und Attribution (Bausteine einer Korpusgrammatik des Deutschen 2)*, 135–172. Heidelberg: Heidelberg University Publishing. DOI: <https://doi.org/10.17885/heiup.bkgd.2021.1.24421>
- Kopf, Kristin & Felix Bildhauer. in Vorb. The genitive alternation in German.
- Kupietz, Marc, Cyril Belica, Holger Keibel & Andreas Witt. 2010. The German reference corpus DeReKo: A primordial sample for linguistic research. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, Daniel Tapias (Hgg.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC '10)*, 1848–1854. Valletta: European Language Resources Association (ELRA).
- Kupietz, Marc, Harald Lungen, Paweł Kamocki & Andreas Witt. 2018. The German reference corpus DeReKo: New developments – new opportunities. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, H el ene Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis & Takenobu Tokunaga (Hgg.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 4353–4360. Miyazaki: European Language Resources Association (ELRA).
- Sch afer, Roland & Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. In Nicoletta Calzolari et al. (Hgg.), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 12)*, 486–493. Istanbul: European Language Resources Association (ELRA).