

einer Filmkritik zu Lars von Triers Weltuntergangsfilm *Melancholia* nannte ich die Mutter der Braut einen *Selbstverwirklichungsdrachen*. Beide Wörter sind nie in den Duden gelangt; die überwältigende Mehrheit der deutschen Muttersprachler hat sie nie gehört oder gelesen. Dennoch kann sie jeder auf Anhieb verstehen.

In der Literatur gibt es solche Gelegenheitsbildungen massenhaft. Während ich dies schreibe, lese ich Heinrich Heines *Ideen. Das Buch Le Grand* und Bertolt Brechts erstes Theaterstück *Baal*. In beiden Texten stoße ich auf jeder Seite auf Wörter, die nicht im Duden stehen.

Auf der ganz willkürlich

aufgeschlagenen Seite 63 der Großen Frankfurter und Berliner Ausgabe der Werke von Bertolt Brecht finde ich beispielsweise die Ausdrücke *fruchttragende Ährenmeere*, *weißstaubige Straßen*, *maitoller Bursche*. Weder *Ährenmeer* noch *weißstaubig* noch *maitoll* stehen im Duden. Die letztgenannten finden sich noch nicht einmal im *Grimm*, während sich *Ährenmeer* dort immerhin bis zum Dichter Barthold Heinrich Brockes zurückverfolgen lässt, dessen Wirken mal dem Spätbarock und mal der Frühaufklärung zuzurechnen ist.

Bei Heinrich Heine finde ich auf der ebenfalls ganz zufällig aufgeschlagenen Seite 72 des Reclamhefts *lotosgeblümte*

*Pantalons* und ein Paar *Nankinghosen*. Während der Duden mich überrascht, weil dort *Pantalon* verzeichnet und erklärt ist – »lange Männerhose mit röhrenförmigen Beinen« –, kennt er *lotosgeblümt* natürlich so wenig wie die *Nankinghose*. Aber da er immerhin *Nanking* (»ein Baumwollgewebe«) und *Lotos* (»eine Seerose«) erklärt, kann ich mir beide Zusammensetzungen erschließen. Solche Wörter sind es, aus denen sich die Millionen-Legionen des deutschen Wortschatzes zusammensetzen.

Die zweite Erklärung für die überraschend hohe Zahl deutscher Wörter, die Professor Klein und seine Kollegen ermittelt haben, ist das

Bemühen der Berliner Wissenschaftler um eine größtmögliche Datenbasis. Sie schätzten nicht, indem sie – wie bisher meist geschehen – nach der Methode Pi mal Daumen die Stichwortzahlen der umfangreichsten Wörterbücher hochrechneten. Sondern sie wollten von Computern anhand einer repräsentativen digitalen Textsammlung – eines sogenannten Korpus – mit verlässlichen statistischen Methoden ausrechnen lassen, wie viele Wörter es im Deutschen tatsächlich gibt.

Kleins Korpus bestand aus einer über das ganze 20. Jahrhundert gleichmäßig gestreuten sorgfältigen Auswahl repräsentativer Texte aus

folgenden Bereichen: Belletristik (also Romane, Erzählungen und andere fiktionale Literatur), Zeitungen, Gebrauchstexte (wie Ratgeber, Kochbücher und Rechtstexte) und wissenschaftliche Texte aus verschiedenen Gebieten.

Nachdem potenzielle Fehlerquellen einigermaßen ausgeschlossen bzw. mit statistischen Methoden herausgerechnet waren, ergab sich: »In einem Textkorpus der deutschen Gegenwartssprache, das eine Milliarde Textwörter lang ist, kommen etwa 5,3 Millionen lexikalische Einheiten – also Wörter, so wie sie im Wörterbuch stehen – vor.«

Die Gegenwart, von der Klein an