

men in einer Sprache darstellen. Zum anderen sieht man, dass es sich bei den häufigsten Nomen und Adjektiven um relativ allgemeine Begriffe handelt, bei den Nomen etwa um Maßeinheiten, die Zeit, Geld oder Mengen beschreiben. Funktionswörter führen typischerweise die Frequenzlisten an – und das unabhängig von der Textsorte. Da Funktionswörter den Kitt bilden, der einen Text grammatisch zusammenhält, finden sie sich in jedem Korpus in großer Zahl, wohingegen die Anzahl und Art der Inhaltswörter wie Nomen, Adjektive oder Verben stärker vom Inhalt und der Textsorte abhängt.

Wortlisten können nach unterschiedlichen Kriterien gebildet und geordnet werden. Zum einen gibt es Wortlisten, die wie die Wortliste des IDS Wortform-Types auflistet, andere Wortlisten wiederum fassen die einzelnen Wortformen zu Lemma-Types zusammen. Beide Arten von Listen können alphabetisch, nach Frequenz, nach Wortlänge oder nach dem ersten Auftreten eines Lemmas sortiert werden. Rückläufige Wortlisten, d.h. Wortlisten, die Wörter nach ihrem Wortende ordnen, können dazu dienen, charakteristische Wortendungen wie z.B. *-el* oder *-itis* zu identifizieren.

Aufgabe 17: Bitte erstellen Sie eine alphabetisch geordnete Frequenzliste (Wortform-Types) für den Textausschnitt aus dem Mainzer Zeitungskorpus (vgl. Abbildung 3, Kapitel 3.1).

Enthalten Wortlisten nicht nur Wortform- oder Lemma-Types, sondern auch Angaben zu deren Häufigkeit, so sind neben einer qualitativen Analyse des Wortschatzes (Welche Wörter finden sich überhaupt im Korpus?) auch quantitative Aussagen möglich (Wie häufig sind die einzelnen Wörter? Welche Wörter kommen besonders häufig vor?). Wortlisten mit Angaben zur Frequenz einzelner Lemma- oder Wortform-Types lassen sich vielfältig nutzen.

Frequenzangaben können z.B. dabei helfen, die Schlag- und Schlüsselwörter eines Korpus bzw. der darin enthaltenen Texte zu identifizieren, da diese im Text überdurchschnittlich häufig auftreten. Dies ist insbesondere dann von Interesse, wenn es wie bei der Analyse von Wahlkampfreden oder Werbetexten darum geht, sprachliche Mittel herauszuarbeiten, die der Verfasser einsetzt, um die Adressaten in seinem Sinne zu lenken. Wortlisten mit den häufigsten 500 oder 1.000 Wörtern einer Sprache erlauben es, den Kernwortschatz einer Sprache oder Varietät zu bestimmen. Die so entstandenen Wortlisten können wie der Grundwortschatz Deutsch von Pfeffer (1970) im Fremdsprachenunterricht eingesetzt werden, um den Lernenden zuerst jene Wörter beizubringen, denen sie mit

der größten Wahrscheinlichkeit begegnen werden (vgl. Kapitel 1.5). Dass die Intuition eines Sprechers nicht unbedingt zu zuverlässigen Frequenzeinschätzungen führt, zeigte die Wortliste des IDS.

Wortlisten auf der Grundlage von Wortform-Types können zwar relativ leicht erstellt werden, bringen jedoch eine Reihe von Problemen mit sich. So unterscheiden Wortform-basierte Wortlisten wie die des IDS nicht zwischen homographen Wortformen (vgl. das Beispiel *Regen* in Kapitel 3.4). Auch die unterschiedlichen Bedeutungsvarianten eines polysemen Wortes wie *Absatz* können in Wortlisten nicht voneinander getrennt werden (vgl. Kapitel 3.5).

3.8 Zusammenfassung

Korpuslinguistische Beschreibungsebenen wie Textwort, Type und Token müssen von sprachlichen Beschreibungsebenen wie Lexem und Wortform unterschieden werden.

Korpusdaten können sowohl qualitativ als auch quantitativ ausgewertet werden. Beim Vergleich von Daten muss die Korpusgröße berücksichtigt werden.

Verbreitete Verfahren zur Analyse von Korpusdaten sind die Stichwortsuche, das Erstellen von Konkordanzen und Wortlisten sowie die Kollokationsanalyse.

Grundbegriffe: Frequenz, Hapax Legomenon, Kollokation, Konkordanz, KWIC, Normalisierung, Produktivität, qualitative Korpusanalyse, quantitative Korpusanalyse, Textwort, Token, Type, Type-Token-Verhältnis, Wortliste

Weiterführende Literatur

Einen knappen Überblick über methodische Fragen bieten Biber *et al.* (1998, Teil IV). Ausführlich mit der Korpusanalyse befasst sich Kennedy (1998, Kapitel 4), speziell mit quantitativen Analysen McEnery/Wilson (2003², Kapitel 3). Kollokationen thematisieren Steyer (2002) und Hausmann (2004), Konkordanzen und Kollokationen Sinclair (1991). Mit der Abgrenzung des Wortes in der Linguistik bzw. Korpuslinguistik befassen sich Meibauer *et al.* (2002, Kapitel 2) und Haß-Zumkehr (2002). Nähere Informationen zu statistischen Verfahren bietet Schlobinski (1996).