

Hanna Kaczmarek

ZUM PROBLEM DER DIVERGENZ IN DER WORTAUSWAHL DER REPRÄSENTATIVEN WORTKORPORA

Die Frage nach dem adäquaten Umfang eines Grundwortschatzes ist von vielen Sprachwissenschaftlern verschiedenartig beantwortet worden. Die Differenzen im Wortlistenumfang sind grundsätzlich auf verschiedene Korpusgrundlagen und unterschiedlichen Einsatz der Auswahlkriterien zurückzuführen. Wie groß die Unterschiede in den angewandten Selektionskriterien und in dem Umfang der lexikalischen Lemmata sind, darauf soll in diesem Beitrag kurz eingegangen werden.

Inhaltsverzeichnis:

1. Auswahlkriterien zur Ermittlung der didaktisch orientierten Grundwortschatzlisten
2. Kurzpräsentation der repräsentativsten Wortkorpora
 - 2.1. Die Grundwortschatzliste von Pfeffer
 - 2.2. Die Grundwortschatzliste von Oehler
 - 2.3. Die Grundwortschatzliste von Plickat
 - 2.4. Die Grundwortschatzliste von Erk
 - 2.5. Die Grundwortschatzliste von Hoffmann
3. Die Differenzen zwischen den Wortschatzlisten
4. Literatur in Auswahl

1. Auswahlkriterien zur Ermittlung der didaktisch orientierten Grundwortschatzlisten

Die Ermittlung von Grundwortschatzen gründet sich prinzipiell auf der Frequenz/ Vorkommenshäufigkeit, die als objektiv nachprüfbares Selektionskriterium betrachtet wird. Die Grundwortschatzlisten¹ sind nicht immer

¹ hier nur auf Verblisten beschränkt.

explizit didaktisch orientiert, sondern dienen „ausschließlich der reinen zwecklosen Erkenntnis des zahlenmäßigen Gefüges des deutschen Sprachbaus sowie etwaiger statistischer Gesetzmäßigkeiten“. (Meier 1967:3).

Die Häufigkeitslisten basieren im Prinzip auf den Frequenzerhebungen. (Krohn 1992:31) Die Häufigkeit von lexikalischen Einheiten, die mit Hilfe repräsentativer Stichproben aus einem Textkorpus gewonnen werden, ist für die Erstellung von Frequenzlisten das entscheidende Kriterium.

Die didaktisch konzipierten Wortlisten modifizieren dagegen mehr oder weniger das Frequenzkriterium, das jedoch nicht als einziges Auswahlkriterium angesehen wird. Die frequentiellen Wortkorpora können mit Sicherheit für didaktische Wortlisten nutzbar gemacht werden, denn sie bieten den hohen Textdeckungsgrad und die daraus ableitbare hohe kommunikative Wirksamkeit der Lemmata. (Krohn 1992:56) Ein Argument gegen Frequenzuntersuchungen und ihre Auswertung für die Erstellung eines Grundwortschatzes ist die Tatsache, dass „Frequenzangaben von Textkorpus zu Textkorpus unterschiedlich ausfallen und deshalb immer nur für das jeweilige Korpus repräsentativ sind.“ (Kühn 1979:43) Bei näherem Hinschauen erweisen sich die bei der Wortselektion herangezogenen Auswahlprinzipien zur Ermittlung der didaktisch orientierten Wortlisten als eine komplexe Mischung von frequentiellen und didaktischmethodischen Komponenten. Die didaktisch motivierten Grundwortschatzlisten basieren auf der kollektiv kontrollierten und bestätigten sprachlichen Intuition von Experten. (Krohn 1992:63) Sie entscheiden darüber, welche Lexeme nach Vorgabe eines Kommunikationsrahmens genügend Nützlichkeit, Anwendungsbreite, Funktionsfähigkeit usw. besitzen. Die Gefahr der subjektiven Materialeinschätzung durch einzelne Mitglieder der Arbeitsgruppe glaubt man durch die objektivierende Wirkung von sachlicher Gruppendiskussion bannen zu können. (Krohn 1992:63)

Bei der Festlegung der didaktischen Grundwortschatzlisten bedient man sich auch der viel diskutierten Bestimmungsfaktoren der Brauchbarkeit, Verstehbarkeit und Lernbarkeit. Den genannten Kriterien wurde zunächst als Mangel angelastet, dass sie wenig objektiv sind. Man ist aber bald zu der Erkenntnis gekommen, dass die Bestimmung eines Grundwortschatzes nach rein objektiven Selektionsprinzipien nicht möglich ist. (Kühn 1981:172)

Die genaue Durchmusterung von Wortlisten bestätigt im gewissen Ausmaß den Verdacht, dass die Auswahlkriterien zur Erstellung eines Grundwortschatzes mehr oder weniger nachprüfbar sind.

2. Kurzpräsentation der repräsentativsten Wortkorpora

2.1. Die Grundwortschatzliste von Pfeffer

Zu den repräsentativen deutschen Wortkorpora, die didaktische Ziele verfolgen, gehört die Grundwortschatzliste von Pfeffer. (Pfeffer 1963)

Er entwickelte einen deutschen Grundwortschatz (für die Grundstufe) von 1.270 Lemmata, der auf einem Textkorpus von insgesamt 590.000 Wortformen basiert. (Pfeffer 1963:11) Seine Grundwortschatzliste wurde auf der Basis von der Frequenz, Streuung, Themenrelevanz und dem intuitiven Kriterium der Vollständigkeit ermittelt. Er entwickelt auch die Wortschatzliste für die Mittelstufe, die 1.536 Lemmata enthält.

Zwar betrachtet Pfeffer sein Wortkorpus „als zuverlässiges Sample der mündlichen Standardform“ (Pfeffer 1963:14), aber nur etwa 60% der lexikalischen Einheiten stellen tatsächlich Belege aus authentischen Texten der gesprochenen Gegenwartssprache dar. (Krohn 1992:32) Der übrig gebliebenen Teilmenge seiner Wortliste liegt kein Textkorpus zugrunde.

2.2 Die Grundwortschatzliste von Oehler

Oehlers „Grundwortschatz Deutsch“ (Oehler 1972) verzeichnet 2.300 Lemmata.

„Seine Wortliste ist nicht textorientiert, sondern das Ergebnis einer Kompilation unterschiedlichster Grundwortschätze und Häufigkeitswörterbücher.“ (Krohn 1992:36) Oehler hat sich nämlich bei der Zusammenstellung der Wortliste u.a. auf die deutschen Grundwortschatzkorpora von Pfeffer (Pfeffer 1963) und (Käding 1963) gestützt. Sein Basiswortschatz ist nicht ausschließlich frequenziell fundiert. Oehler betont auch die Wichtigkeit der nicht frequenziellen Selektionskriterien wie etwa der Allgemeingebräuchlichkeit und des Aussagewertsprinzips, wobei es eine ausführliche Darlegung der genannten Auswahl Faktoren fehlt. Aus dem didaktischen Gesichtspunkt ist Oehlers Liste nicht zu unterschätzen, weil sie als eine multilinguale Grundwortschatzliste für den Fremdsprachenunterricht vom großen Nutzen ist.

2.3. Die Grundwortschatzliste von Plickat

Die Grundwortschatzliste von Plickat (1983) wurde aus den Grundwortschatzkorpora von Pfeffer, Wängler (1963), Ortman (1975) und Steger – Keil (1971) exzerpiert. Seine Liste besteht aus einem Kernwortschatz und einem engeren Grundwortschatz. Der Kernwortschatz umfasst 101 Lemmata, die allen fünf Wortlisten gemeinsam sind. Der engere Grundwortschatz enthält 337 lexikalische Einheiten, die in vier oder in drei Wortlisten vorkommen. (Krohn 1992) Plickat glaubt bei der Ermittlung seiner Grundwortschatzliste

auf dem Vorkommenshäufigkeitskriterium aufzubauen. „Fragwürdig ist diese Schlussfolgerung, weil das Wortkorpus von Steger – Keil und Wendelmuth überhaupt nicht auf der Basis von Vorkommenshäufigkeiten ermittelt wurden (...)“ (Krohn 1992:38)

2.4. Die Grundwortschatzliste von Erk

Die Wortliste von Erk (Erk 1975) basiert auf einem Textkorpus von 250.000 Wörtern. Für die Rangordnung der Lemmata werden die Kriterien der Frequenz und Streuung verwendet. Seine Grundwortschatzliste ist nach Wortklassen aufgeschlüsselt und umfasst 833 Substantive, 589 Verben und 704 Adjektive, Adverbien und Partikel. „(...) Erk verfolgt mit der rein frequentiellen Wortliste von 2.126 Lemmata ein explizites didaktisches Anliegen, nämlich den Aufbau einer primär rezeptiven lexikalischen Kompetenz, die das Leseverständniswissenschaftlicher Texte zum vorrangigen Lernziel erhebt.“ (Krohn 1992) Das zugrunde liegende Korpus besteht aus einer Sammlung wissenschaftlicher Fachtexte, „in der 34 Fächer, die traditionsgemäß zum Fächerkanon deutscher Universitäten gehören, mit jeweils drei Texten von genormter Länge vertreten sind.“ (Erk 1975:7) Das Wortkorpus von Erk ist in drei Kategorien aufgeteilt: in allgemeinen, allgemeinwissenschaftlichen und speziellen Wortschatz.

2.5. Die Grundwortschatzliste von Hoffmann

Einen explizit frequenziellen Ansatz vertritt Hoffmann, für den die Häufigkeit von lexikalischen Einheiten das entscheidende Kriterium der Wortauswahl ist. (Hoffmann 1984:225) Hoffmann hat die Fachlexik aus den Gebieten der Medizin, Physik, Chemie, Mathematik, Tierproduktion, Gesellschaftswissenschaften und Bauwesen in Form von dreisprachigen Häufigkeitslisten herausgegeben. Sie umfassen die 1.100 – 1.200 Lexeme der jeweiligen Fachdisziplin in russischer, englischer und französischer Sprache und ihre deutschen Äquivalente. Das Ziel, welches Hoffmann verfolgt, ist nicht „einen konsistent definierten Fachwortschatz zu erfassen, sondern (...) den in fachsprachlichen Texten vorkommenden Wortschatz insgesamt zu dokumentieren“ (Krohn 1992:47).

3. Die Differenzen zwischen den Wortschatzlisten

Die kurz besprochenen Wortschatzlisten erheben den Anspruch, mit ihrem Wortschatz den deutschen Kern-/Grundwortschatz zu repräsentieren, der gewissermaßen eine Grundlage für den lernzielspezifischen Aufbau der lexikalischen Kompetenz darstellt. Die zwischen den einzelnen Wortlisten auftretenden Wortschatzdivergenzen ergeben sich nicht ausschließlich aus den

unterschiedlichen Auswahlkriterien, die den Wortlisten zugrunde liegen, sondern sie sind auch auf die Tatsache zurückzuführen, dass das thematische Profil der Korpora oft verschiedenartig ist. Im direkten Vergleich verschiedener Grundwortschatzkorpora erweist sich, dass umfangmäßig die Grundwortschatzlisten meist an der magischen 2000-Wort-Grenze festhalten. Die Anzahl der lexikalischen Einheiten beträgt am häufigsten etwas mehr oder weniger als 2000 Lemmata. Der Grundwortschatz von Pfeffer liegt sowohl unterhalb als auch oberhalb dieser 2000-Wort-Grenze, je nachdem, ob nur die Grundstufe oder die Grund- und die Mittelstufe (Pfeffer 1970) als notwendiges und tragfähiges Minimum für Gesprächskontakte aufgefasst werden. Oehler mit seinen 2.300 Lemmata und Erk mit 2.126 separaten lexikalischen Einheiten weichen von diesem statistischen Durchschnittswert nicht ab. Die Tendenz, die Grundwortschatzlisten auf ungefähr 2.000 Lemmata zu beschränken, ist nicht nur in den deutschen Wortschatzkorpora zu beobachten. Die Grundwortschatzliste für das Englische von West (West:1953) enthält rund 2000 Wörter, zu denen in einem Anhang noch 425 Wörter aus dem wissenschaftlichen und technischen Bereich kommen. Es gibt aber auch Wortschatzlisten, die von diesem Durchschnittswert in beiden Richtungen abweichen. So enthält der „Grundwortschatz Russisch“ (Brandt 1986) sogar 4000 Lemmata, dagegen „Français Fondamental“ (Gougenheim 1959) nur 1.350 lexikalische Exponenten.

Die Tatsache, dass die Grundwortschatzlisten nicht nur auf dem frequentiellen Prinzip, sondern auch auf einer Reihe anderer qualitativer Kriterien erfaßt werden, soll als einer der Gründe der unterschiedlichen Wortschatzmenge angesehen werden. Die einzelnen Wortschatzlisten sind miteinander nur dann zu vergleichen, wenn die Auswahlkriterien, die zur Aufstellung der konkreten Grundwortschatzlisten dienen, identisch sind.

Trotz der bedeutsamen Divergenzen zwischen der einzelnen Wortschatzkorpora ist ihre Rolle bei der Erstellung der Fremdsprachenlehrwerke unbestritten.

Literatur

- Brandt, B.(1986): Grundwortschatz Russisch – Seine linguistischen und fremdsprachenmethodischen Grundlagen. [in:] Linguistische Studien, Reihe A. Arbeitsberichte 147. Berlin: Akademie der Wissenschaften der DDR
- Erk, H. (1975): Hochfrequenter Wortschatz aus wissenschaftlichen Texten, (Hrsg.) vom Goethe-Institut, Arbeitsstelle für wissenschaftliche

- Didaktik, Projekt Wissenschaftsdeutsch. München: Goethe Institut
- Gougenheim, G. (1959): *Français Fondamental, 1er degré, élargie*. Paris: Didier
- Hoffmann, L.(1984): *Fachwortschatz – Grundwortschatz – Minimum*, [in:] *Deutsch als Fremdsprache*. Leipzig: Herder-Institut
- Kaczmarek, H. (2000): *Exkurs zum Grundwortschatz*. Częstochowa: Wydawnictwo WSP
- Käding, F.W. (Hrsg.) (1963): *Häufigkeitwörterbuch der deutschen Sprache*. Quickborn bei Hamburg: Schnelle
- Krohn, D.(1992): *Grundwortschätze und Auswahlkriterien. Metalexikographische und fremdsprachendidaktische Studien zur Struktur und Funktion deutscher Grundwortschätze*. Surte: Acta Universitatis Gothoburgensis
- Kühn, P.(1979): *Der Grundwortschatz: Bestimmung und Systematisierung*, Tübingen: Niemeyer
- Kühn, P.(1981): *Notwendigkeit, Verwendung und Gebrauchswert von Grundwortschatzbüchern*. [in:] *Zeitschrift für Germanistische Linguistik*. H.9. Berlin: de Gruyter
- Meier, H.(1967): *Deutsche Sprachstatistik*. 2.Bd.. Hildesheim: Olms
- Oehler, H.(1972): *Grundwortschatz Deutsch*. Stuttgart: Klett
- Ortmann, W.D.(1975): *Hochfrequente deutsche Wortformen I., 1995 Wortformen der Käding-Zählung, rechnerorientiert in alphabetischer und rückläufiger Folge, nach Häufigkeit und nach Hauptwortarten*. München: Kemmler
- Pfeffer, J.A.(1963): *Basic (Spoken) German Word List, Level I*. Eglewood Cliffs: Prentice Hall
- Pfeffer, J.A.(1970): *Basic (Spoken) German Word List, Level II*, Pittsburgh: University of Pittsburg Press
- Plickat, H.-H.(1983): *Deutscher Grundwortschatz*, Weinheim: Beltz
- Steger, H./Keil, M.(1971): *Wortliste für das VHS-Zertifikat Deutsch als Fremdsprache*. Frankfurt: Typoscript
- Wängler, H.-H.(1963): *Rangwörterbuch hochdeutscher Umgangssprache*. Marburg: Elwert
- West, M.(1953): *General Service List of English Words*. London: Longman

STRESZCZENIE

Artykuł jest wynikiem wnikliwych badań autorki nad słownictwem podstawowym (Grundwortschatz), mających na celu wskazanie rozbieżności w kryteriach doboru słownictwa, stosowanych przy zestawianiu poszczególnych minimów leksykalnych. Analiza porównawcza list słownictwa podstawowego wyekscerpowanych przez różnych badaczy wykazała, że dyferencje dotyczą nie tylko kryteriów doboru, lecz także zasobu ilościowego słownictwa uznawanego przez autorów za podstawowe. Choć listy ze słownictwem podstawowym nie są jednoznacznie przeznaczone do celów dydaktycznych, to jednak częstokroć stanowią one bazę dla powstających podręczników do nauki języka.