

a **FREQUENCY** dictionary of
GERMAN

core vocabulary for learners

**Randall L. Jones
and Erwin Tschirner**

A Frequency Dictionary of German

A Frequency Dictionary of German is an invaluable tool for all learners of German, providing a list of the 4,034 most frequently used words in the language. Based on a 4.2-million word corpus which is evenly divided between spoken, literature, newspaper and academic texts, the dictionary provides the user with a detailed frequency- based list plus alphabetical and part of speech indexes.

All entries in the rank frequency list feature the English equivalent, a sample sentence plus an indication of major register variation. The dictionary also contains 21 thematically organized lists of frequently used words on a variety of topics as well as eleven special vocabulary lists.

A Frequency Dictionary of German aims to enable students of all levels to maximize their study of German vocabulary in an efficient and engaging way.

Randall L. Jones is Emeritus Professor of German at Brigham Young University, and **Erwin Tschirner** is Professor of German as a Foreign Language at the Herder-Institut, University of Leipzig.

Routledge Frequency Dictionaries

General Editors:

Anthony McEnery

Paul Rayson

Consultant Editors:

Michael Barlow

Asmah Haji Omar

Geoffrey Leech

Barbara Lewandowska-Tomaszczyk

Josef Schmied

Andrew Wilson

Other books in the series:

A Frequency Dictionary of Spanish: Core vocabulary for learners

hbk 0-415-33428-4

pbk 0-415-33429-2

Coming soon:

A Frequency Dictionary of Polish

A Frequency Dictionary of German

Core vocabulary for learners

Randall L. Jones and Erwin Tschirner

Contributing writers:

Agnes Goldhahn
Isabel Buchwald
Antina Ittner

First published 2006
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group

Transferred to Digital Printing 2006

© 2006 Randall L. Jones and Erwin Tschirner

Typeset in Parisine by Keystroke, Jacaranda Lodge, Wolverhampton

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data

Jones, Randall L.

A frequency dictionary of German/Randall Jones and Erwin Tschirner. — 1st ed.
p. cm. — (Routledge frequency dictionaries)

Includes bibliographical references and indexes.

1. German language—Word frequency—Dictionaries. I. Tschirner, Erwin P., 1956–
II. Title. III. Series.

PF3691.J66 2005

433'.1—dc22

2005012949

ISBN10: 0-415-31632-4 (hbk)

ISBN10: 0-415-31633-2 (pbk)

ISBN13: 9-78-0-415-31632-3 (hbk)

ISBN13: 9-78-0-415-31633-0 (pbk)

Contents

Thematic vocabulary lists | vi

Series preface | vii

Acknowledgements | ix

List of abbreviations | x

Introduction | 1

Frequency index | 9

Alphabetical index | 139

Part of speech index | 182

Thematic vocabulary lists

- | | |
|--|---------------------------------|
| 1 Definite articles 9 | 12 Sports 20 |
| 2 Forms of the verb <i>sein</i> 9 | 13 Weather 22 |
| 3 Forms of the verb <i>haben</i> 10 | 14 Colours 25 |
| 4 Personal pronouns 10 | 15 Plants and gardening 29 |
| 5 Forms of the verb <i>werden</i> 10 | 16 Months 32 |
| 6 Possessive pronouns 11 | 17 Seasons 37 |
| 7 Collocations with <i>Jahr</i> 12 | 18 Days of the week 39 |
| 8 Numbers 13 | 19 Appliances and utensils 47 |
| 9 Collocations with <i>Zeit</i> 14 | 20 Domesticated animals 48 |
| 10 Kinship terms 15 | 21 Articles of clothing 67 |
| 11 Collocations with <i>Fall</i> 17 | |

Series preface

There is a growing consensus that frequency information has a role to play in language learning. Data derived from corpora allows the frequency of individual words and phrases in a language to be determined. That information may then be incorporated into language learning. In this series, the frequency of words in large corpora is presented to learners to allow them to use frequency as a guide in their learning. In providing such a resource, we are both bringing students closer to real language (as opposed to textbook language, which often distorts the frequencies of features in a language, see Ljung 1990) and providing the possibility for students to use frequency as a guide for vocabulary learning. In addition we are providing information on differences between frequencies in spoken and written language as well as, from time to time, frequencies specific to certain genres.

Why should one do this? Nation (1990) has shown that the 4,000–5,000 most frequent words account for up to 95 per cent of a written text and the 1,000 most frequent words account for 85 per cent of speech. While Nation's results were for English, they do at least present the possibility that, by allowing frequency to be a general guide to vocabulary learning, one task facing learners – to acquire a lexicon which will serve them well on most occasions most of the time – could be achieved quite easily. While frequency alone may never act as the sole guide for a learner, it is nonetheless a very good guide, and one which may produce rapid results. In short, it seems rational to prioritize learning the words one is likely to hear and use most often. That is the philosophy behind this series of dictionaries.

The information in these dictionaries is presented in a number of formats to allow users to access the data in different ways. So, for example, if you would prefer not to simply drill down through the word frequency list, but would rather focus on verbs, the part of speech index will allow you to focus on just the most frequent verbs. Given that verbs typically account for 20 per cent of all words in a language, this may be a good strategy. Also, a focus on function words may be equally rewarding – 60 per cent of speech in English is composed of a mere 50 function words.

We also hope that the series provides information of use to the language teacher. The idea that frequency information may have a role to play in syllabus design is not new (see, for example, Sinclair and Renouf 1988). However, to date it has been difficult for those teaching languages other than English to use frequency information in syllabus design because of a lack of data. While English has long been well provided with such data, there has been a relative paucity of such material for other languages. This series aims to provide such information so that the benefits of the use of frequency information in syllabus design can be explored for languages other than English.

We are not claiming, of course, that frequency information should be used slavishly. It would be a pity if teachers and students failed to notice important generalizations across the lexis presented in these dictionaries. So, for example, where one pronoun is more frequent than another, it would be problematic if a student felt they had learned all pronouns when

they had learned only the most frequent pronoun. Our response to such issues in this series is to provide indexes to the data from a number of perspectives. So, for example, a student working down the frequency list who encounters a pronoun can switch to the part of speech list to see what other pronouns there are in the dictionary and what their frequencies are. In short, by using the lists in combination a student or teacher should be able to focus on specific words and groups of words. Such a use of the data presented here is to be encouraged.

Tony McEney and Paul Rayson
Lancaster, 2005

References

Ljung, M. (1990)

A Study of TEFL Vocabulary. Stockholm: Almqvist & Wiksell International.

Nation, I.S.P. (1990)

Teaching and Learning Vocabulary. Boston: Heinle and Heinle.

Sinclair, J.M. and Renouf, A. (1988)

"A Lexical Syllabus for Language Learning". In R. Carter and M. McCarthy (eds) *Vocabulary and Language Teaching* London: Longman, pp. 140–158.

Acknowledgements

The authors wish to express their gratitude to Brigham Young University and Leipzig University for their generous support in the preparation of this book. Appreciation is also acknowledged for the many students at both universities who assisted with the work.

Abbreviations

The following abbreviations are used in this dictionary.

adj	adjective	prep	preposition
adv	adverb	pron	pronoun
art	article	sb	somebody
aux	auxiliary verb	sich	reflexive verb
conj	conjunction	sth	something
inf	infinitive marker	A	Academic subcorpus
interj	interjection	I	Instructional subcorpus
num	number	L	Literature subcorpus
part	particle	N	Newspaper subcorpus
pl	plural	S	Spoken subcorpus

Introduction

Dictionaries and frequency dictionaries

The word “dictionary” usually brings to mind a large book with definitions or second language glosses. This book is precisely what the title promises: a dictionary of German word frequencies. The 4037 entries represent the most commonly occurring words in a four million word corpus of German and are listed according to the frequency of their occurrence. Listed with each entry is information about the part of speech, an English translation, and a brief example of how the word is used. Some words, especially among the highest frequencies, have more than one meaning and in some cases can belong to more than one part of speech.

This book is not intended to replace a conventional German–English dictionary; rather it is to be used as a supplementary tool for learning vocabulary. The frequency dictionary can help the learner focus on the most commonly used German words. A conventional dictionary can then be consulted to provide additional useful information about meaning, usage, etc. for each entry.

In learning a second language such as German, vocabulary can be learned randomly, i.e. as it occurs in a natural authentic setting, or systematically, as is usually the case in a structured language learning environment. Each approach has its advantages and disadvantages. In a structured learning environment selection, quantity, and sequencing of the individual vocabulary items are important considerations. How many words should be introduced at each stage along the way and which ones should they be? Which words should be introduced at the beginning stages and which ones at later stages? For the learner and instructor alike, lexical frequency information can be useful in selecting and sequencing vocabulary items.

German word frequency

Word frequency analysis in German is not new; indeed it goes back over 100 years to 1898, when F.W. Kaeding published his *Häufigkeitwörterbuch der deutschen Sprache* (Kaeding 1898). His interest was

not in the area of language learning, rather he was interested in developing a new stenographic shorthand system for German. In spite of his intentions, his frequency list has enjoyed pedagogical application for many years (see esp. Morgan 1928, Ortman 1975). More recently, J. Alan Pfeffer developed a spoken German frequency list with the specific goal of assisting in the learning of vocabulary (Pfeffer 1964). There have been other frequency lists developed for German, based on newspaper texts, literature, and other registers (e.g. Scherer 1965, Meier 1967, Swenson 1967, Rosengren 1972, Ruoff 1981). This book differs from earlier published frequency lists in at least three ways. First, it is a balanced, structured, and integrated corpus, meaning that it was carefully planned to achieve representation of genre, register, style, geography, and age group. The samples for each part are balanced and are large enough to be characteristic of each specific type of language. Second, the frequency list has been processed to reduce ambiguity and uncertainty. More will be said about this later in the Introduction. Finally, the language is contemporary, spanning the past fifteen years but concentrating on the last five.

What is a corpus?

The basis of our lexical frequency list is a corpus, i.e. a structured collection of language texts that is intended to be a rational sample of the language in question. A corpus should be large enough to contain a sufficient number of words to provide a useful basis from which to work, although it has never been established what a threshold level should be. The Brown Corpus of English was produced in 1960 and contained 1 million words of written American English and was thought at the time to be more than adequate in size (Francis and Kučera 1964). The British National Corpus, completed in 1994, has 100 million words of spoken and written British English (Aston and Burnard 1998). A similar corpus for American English is now completed (Reppen and Ide

2004) and numerous other large corpora for a variety of languages are in the works or are already in use.

The Leipzig/BYU Corpus of Contemporary German

The Leipzig/BYU Corpus of Contemporary German contains 4.2 million words of spoken and written German. It is a balanced, structured, and integrated corpus, meaning that it was carefully planned to achieve representation of genre, register, style, geography, and age group. It consists of one million words each of spoken language, literature, newspapers, and academic texts, and 200,000 words of instructional language.

Spoken language

The spoken component consists of 700,000 words of spontaneous conversation and 300,000 words of television material. The conversation texts are the same as the BYU Corpus of Spoken German (Jones 1997) and consist of 402 12 to 15 minute conversations between native German speakers. The conversations took place between 1989 and 1993 in 60 localities in Germany (East and West), Austria, and Switzerland. The speakers reflect a balanced representation of age, gender, and social class. The topics are varied and include current events, personal interests, family, local tourist attractions, politics, weather, reminiscences of childhood, future plans, etc. The conversations were recorded and transcribed using a broad orthographical transcription system.

The 300,000 words of television material consist of transcriptions of three types of programmes. The first is family oriented semi-serious light dramas known in Germany as *Vorabendserien* because they are generally broadcast in the early evening for family viewing. They treat a broad range of topics of current interest and have a wide viewing audience. The language is scripted but it is written to sound like natural spoken language as opposed to bookish German. Furthermore, the actors often take liberties and improvise in order to be more natural.

The second type of television material is talk shows. The hosts are professional television personalities but the guests include a variety of people, e.g. politicians, sports personalities, actors, business people, and average people who have something interesting to say. Often there is more than one guest, i.e. the host interacts randomly with

any of three or four people. With the possible exception of a brief introduction by the host, the language is completely spontaneous.

The third type of television material is spontaneous broadcasting and may be a report of a sports event or an interview with an athlete or other personality. Most of these programmes have been broadcast since 2000.

Literature

For the literature sub-corpus, 10,000 words were selected from each of 100 different works, including *Hohe Literatur*, *Jugendliteratur*, *Bestseller*, *Humor*, *Reiseliteratur*, *Gesellschaftsroman*, and *Abenteuer/Krimi*. Approximately one-third of the words were taken respectively from the beginning, middle, and end of each work. With one exception the books were published in Germany, Austria, and Switzerland between 1990 and 2000.

Newspapers

The million words of newspaper text were taken from 50 editions each of local and national newspapers in Germany, Austria, and Switzerland published between September 2001 and February 2002. Texts were selected from *Politik*, *Wirtschaft*, *Kultur*, *Sport*, and *Kommentar*. Complete articles were selected.

Academic

The academic section consists of one million words of material from 100 different sources, including university-level course books, *Gymnasium* second-level books, popular science journals, and technical journals. The subject matter includes virtually all topics treated at the *Gymnasium* and university, e.g. natural and social sciences, technology, humanities, art, music, law, and medicine. The books and journals were published in Germany, Austria, and Switzerland.

Instructional

The 200,000 words of instructional texts or *Gebrauchstexte* consist of five types: *Anleitungen* (e.g. How to build a birdhouse), *Anzeigen* (misc. wanted-ads), *Ratgeber* (e.g. Tips on gardening), *Kundeninformation* (e.g. Product information), *Gesetzestexte* (e.g. a rental agreement). Many of these sources were relatively short and all of them used some specialized vocabulary.

Some of the texts were taken from the Internet and others were scanned. All texts were carefully proofread several times by native German speakers. Because the time span of the texts in the corpus covers the period of time before and after the inception of the German Spelling Reform, the entire corpus has been standardized to conform to the *Neue Rechtschreibung*. In addition, words from Swiss sources that use *ss* instead of *ß* have also been changed.

Processing the corpus data

After the corpus had been assembled and proofread, it was ready to be processed. The ultimate goal is of course to produce a word frequency list, but there were numerous steps that had to be accomplished before this was possible.

The first step was to provide additional information to each word in the corpus in order to reduce ambiguity. For example, the German words *liebe* and *Liebe* would be recognized by text processing software as one word, even though the first example could be an adjective or a verb and the second a noun. (Case is usually ignored in order to avoid confusion arising from capitalizing the first word in a sentence.) Likewise, the word *sein* can be a form of a verb or a possessive pronoun. To deal with this problem, sophisticated software programs known as “taggers” have been developed for several languages, including German. A tagger examines the context of each word in a corpus, and then assigns a part of speech (POS) tag to it. Subsequent software can then distinguish among otherwise ambiguous forms, e.g. LIEBE[SUBS], LIEBE[VERB] and LIEBE[ADJE]. For our project we used the Stuttgart Tree-Tagger (Schmid 1995) and the extended Stuttgart-Tübingen Tag-Set (STTS). If it is uncertain about a tag, it assigns an asterisk after the tag, an indication that the user needs to have a closer look. By checking the uncertain forms and carefully analysing the others, it is possible to achieve a relatively high degree of accuracy.

For subsequent processing we used a program called WordSmith Tools (Scott 1999). As the plural form “tools” suggests, it accomplishes a number of tasks. The first step was to create a raw frequency list, i.e. rearrange the words in the corpus to generate a list that shows rank ordering and absolute

1	DIE[ARTI]	119,126
2	UND[KONJ]	119,100
3	DER[ARTI]	113,709
4	IN[APPR]	70,216
5	ICH[PPER]	46,937
6	DEN[ARTI]	44,923
7	IST[AVER]	42,962
8	SIE[PPER]	42,932
9	NICHT[PTNG]	40,251
10	MIT[APPR]	36,714
11	VON[APPR]	36,690
12	ES[PPER]	35,098
13	DAS[ARTI]	33,331
14	SICH[PREF]	32,878
15	EIN[ARTI]	30,795
16	AUCH[ADVB]	30,309
17	AUF[APPR]	28,368
18	EINE[ARTI]	28,240
19	IM[APPR]	27,880
20	FÜR[APPR]	26,641

Above are listed the first twenty entries of the Leipzig BYU Corpus after it had been processed by the wordlist generator. Note that the words are listed in order of frequency, and each word is followed by a value that represents its absolute frequency. Note also that these entries are in some cases forms of the word or lemma and not the base word or dictionary form. For example, it was useful for our purpose to combine the various forms of the definite article (1, 3, 6, 13) into one entry. The entry *ist* (7) is not a lemma but part of the verb *sein*. The entries *ein* (15) and *eine* (18) belong together, and *im* (19) is really a contraction of *in dem*.

The task of “re-mapping” word forms is accomplished in WordSmith using a semi-automatic lemmatizing tool. The base word is identified, and then each form which belongs to it is subsequently marked. Finally, they are all joined as one entry with a simple click of the mouse. The procedure appears to be quite straightforward, and for the most part it is, but difficulties arise from having to make decisions about what should be joined. Participial forms of verbs should belong to the infinitive, but sometimes present and past participles have become lexicalized in their own right. For example, the adjective *überwiegend* ultimately derives from the verb *überwiegen*, but to list it as a form of the verb and not as a separate word would be like the tail wagging the

dog. Should *verheiratet* be listed as a separate adjective or is it part of the verb *verheiraten*? More on this later.

One of the most labour intensive tasks in the processing of entries was that of recognizing and reconstituting verbs with separable prefixes. When a verb such as *ausmachen* appears in a text, there is a good chance that *aus* will occur separated from *machen*. WordSmith is not able to recognize this, although the prefix is recognized and tagged as such. WordSmith can then provide information about which verb prefixes occur in the same environment as potential verbs. Thus, with the assistance of WordSmith and a great deal of patience, it was possible to locate separated verb prefixes and their respective base verbs with a high degree of accuracy.

The decision to limit the number of entries to 4,034 was not entirely arbitrary. In addition to limitations of space in the book, it was felt that this represented a useful number for the beginning and intermediate student of German. It is interesting to note that the first ten words in the frequency list account for approximately 27 per cent of all the words in the corpus. Add the next ten and the coverage increases to approximately 35 per cent. The 4,034 words contained in this book account for between approximately 80 per cent and 90 per cent of the words in the corpus, depending on the register (Jones 2005).

Why the odd number 4,034? Entry 4,000 is one of 87 words that occur 16 times per million words of text, but it is by no means the last. Within this frequency level the words are listed alphabetically, and it seemed only proper to continue past *nirgendwo* and on to *zweifellos*.

Contents of the book

The first list in the book consists of the 4,034 words ordered by relative frequency, i.e. from the most frequent to the least. The structure of each entry is as follows:

- 1 Rank frequency
- 2 Word
- 3 Part of speech
- 4 Translation
- 5 Sample sentence
- 6 Occurrences per million
- 7 Range information (does not occur with all words)
- 8 Multi-word units and dominant word forms

The word *wenn* can serve as an example.

43 *wenn* conj if, when

- Wenn es regnet, bleiben wir zu Hause.

3051

This word is number 43 in the frequency ranking; it is a conjunction and means “if” or “when” in English. The sample sentence means, “If it rains, we’ll stay at home.” The word *wenn* occurs 3,051 times in a million words of the corpus.

The next list is an alphabetical list of all the words in the list, together with the part of speech, the English translation, and the rank value. The remaining lists consist of proper names, abbreviations, multi-word units, and individual parts of speech, e.g. nouns, verbs, and prepositions, arranged by rank value and including the English translation.

Main entries

In addition to the inflected forms that belong to a headword, the following derived forms are also included:

- neuter nouns derived from infinitives, e.g. *das Gehen* (“walking”) from *gehen* (“to walk”)
- neuter nouns derived from adjectives, e.g. *das Gute* (“the good”) from *gut* (“good”)
- masculine and feminine nouns referring to people derived from adjectives, e.g. *die Schöne* (“the beautiful one”) from *schön* (“beautiful”)
- adjectives derived from present or past participles, e.g. *verkaufte Waren* (“sold goods”) from the verb *verkaufen* (“to sell”) or *lächelnde Gesichter* (“smiling faces”) from the verb *lächeln* (“to smile”).

Derived forms that have become lexicalized, i.e. have taken on a new meaning, are not listed in this way.

Multi-word units and dominant word forms

Multi-word units (MWU) are phrases such as *zum Beispiel*, *Guten Tag* that occur with sufficient frequency (at least 16 occurrences per million words) that they are equivalent in number to the 4,000 most frequent words. They are listed on a new line under the entry of the key word in the MWU (e.g. *Beispiel*, *Tag*) followed by an English translation, a German sample sentence, and a number indicating the occurrences per million words of text.

Dominant word forms are special forms of the entry such as participles of verbs and superlatives of adjectives which constitute more than 20 per cent of all examples within the entry and have a frequency per million words of at least 16. Only transparent word forms are listed, i.e. forms that have not assumed a new meaning. Words such as *geboren* (from the verb *gebären*) and *das Essen* (from the verb *essen*) are separate entries. Dominant forms are listed on a new line followed by a part of speech code, an English translation, a sample sentence in German, and a number indicating the occurrences per million words of text.

Abbreviations

German abbreviations are counted together with the corresponding full forms, e.g. *Universität, Uni*. If the abbreviation is the most commonly used form, e.g. *Aids, BSE*, it becomes the headword followed by the full form if it is not obvious, e.g. *LKW, Lastkraftwagen* (“truck”).

Additional lists

In addition to the rank frequency list, the following special lists are included:

- an alphabetical list of the same words found in the rank frequency list. The number after each word corresponds to the first number of each entry in the rank frequency list. Also included is part of speech information as well as a translation
- the most frequent 100 nouns, verbs, adjectives, and adverbs
- all prepositions and conjunctions contained in the 4,037-word list
- all articles, pronouns, and irregular verbs contained in the 4,037-word list
- all abbreviations and proper names contained in the 4,037-word list

Infoboxes

Scattered through the frequency list are “Infoboxes”, which contain smaller lists and rank value of related words, e.g. days of the week, numbers, colours, etc. Some of the Infoboxes contain grammatical information, e.g. paradigms of auxiliary verbs.

Parts of speech

The translation and ordering of parts of speech differ from language to language and can even differ within a language, depending on who is making the decisions. For our list, the following parts of speech are included:

Adjective

Inflected adjective forms have been lemmatized with the base form. Many German adjectives can also be adverbs, e.g. *persönlich: Das ist eine persönliche Sache* (adjective), *ich persönlich verstehe es nicht* (adverb). In such cases only the part of speech *adj* is listed. Some German adjectives occur only in an inflected form, i.e. always with an ending. These are listed with the three possible nominative endings, e.g. *letzte* (r, s).

Adverb

Adverbs are not inflected in German. As mentioned above, adverbs that can also be adjectives are not listed separately.

Article

German has two classes of articles: definite (*der, die, das*) and indefinite (*ein, eine*). The indefinite article can also function as the number “one”. The German sentence *Ich habe eine Schwester* can mean both “I have a sister” and “I have one sister”.

Both definite and indefinite articles can be used as pronouns, e.g. *Der kann heute nicht kommen, Einer der Besucher kommt später*. The word *kein* presents a special problem, as it is really nothing more than the negative form of *ein* but is generally considered to be a pronoun, as it is listed here.

Conjunction

There are relatively few conjunctions in German and their meaning is fairly straightforward.

Noun

Nouns are identified by the appropriate definite article, *der, die, das*, which reflects the gender. Only the nominative singular form is listed. All forms that have been inflected for number and case have been lemmatized.

A few German nouns occur exclusively or almost always as a plural form, e.g. *Leute, Ferien, Schulden, Unterlagen*. These are assigned the part of speech designation “die (pl)”.

Number

Cardinal numbers (*zwei*) and ordinal numbers (*zweit*) have been combined as have close derivatives, e.g. *zweiten*. The frequency count for *eins* is not accurate because in most contexts it is indistinguishable from the indefinite article.

Particle

German has several classes of particles, including a class known as modal particles. Modal particles, e.g. *denn, eben, eigentlich*, usually convey an attitude of the speaker and are therefore difficult to translate. An approximate English translation is given for most of them.

Preposition

German prepositions dictate the case of the following noun, i.e. accusative, dative, and genitive. This information is not given. Furthermore, German prepositions generally have a variety of meanings, but it is not possible to show this in a simple entry.

Pronoun

There are several classes of pronoun in German but they are not distinguished in the entry. The sample sentences make most of the distinctions clear.

Verb

Only the infinitive form of the verb is listed. All forms that have been inflected for person, number, tense, case, and mood have been lemmatized with the infinitive form. Many verbs in German are irregular, i.e. their conjugation is not predictable. Infoboxes and special lists are provided for the highest frequency verbs, which show the individual forms plus their rank order and frequency per million.

Range

Most words in the list occur in all of the text registers (spoken, newspaper, literature, academic, instructional) and are fairly evenly spread throughout the corpus. These words are said to have a broad lexical range. Some words, however, occur numerous times in just a few texts and therefore have a narrow range. Words that occur 90 per cent or more in just one register are considered over-represented in that register and are marked with a +, e.g. +A. Words that occur less than 5 per cent in a register are considered under-represented and are marked with a –, e.g. –S. Any word marked as + automatically infers – for the

other registers. Words that occur exclusively in just one register and words that do not occur in at least five different texts in a second register have been excluded. Examples of words such as this are *Milzbrand* (anthrax), *Substrat* (substratum), and *Enzym* (enzyme). In many cases these words have English cognates and are easily recognizable.

In some cases it was difficult to determine a range value for a word, especially for the separable prefix verbs and special forms of entries. In cases of doubt no range information is given.

A final caveat

This German frequency dictionary is intended to be used as a learning tool in conjunction with other learning tools, i.e. a good German/English dictionary, a good textbook, and other support material that can be helpful for learning German. The principal information the book can provide is to know which vocabulary items to concentrate on at various stages of the learning process. Simply selecting a certain number of words and memorizing them may not be as productive as selecting those words and using them as the basis for a discovery experience. Look them up in a German/English dictionary and read the entries. Find them in the index of your textbook and see how they are used there. Develop a small corpus of your own using Internet material and find the words there as they are used in authentic contexts. As you achieve mastery in more and more of the words in the dictionary, the better you will be able to understand and speak the German language.

References

- Aston, Guy and Lou Burnard (1998). *The BNC Handbook*. Edinburgh: Edinburgh University Press.
- Francis, W.N. and Kučera H. (1964). *Manual of Information to Accompany "A Standard Sample of Present-day Edited American English, for Use with Digital Computers"* (revised 1979). Providence, RI: Department of Linguistics, Brown University.
- Jones, Randall L. (1997). "Creating and Using a Corpus of Spoken German". In Anne Wichmann et al. (eds) *Teaching and Language Corpora*. London: Longman, pp. 146–156.

- (2005).
“An Analysis of Lexical Text Coverage in Contemporary German”. In Andrew Wilson, Paul Rayson, and Dawn Archer (eds) *Corpus Linguistics around the World*. Amsterdam: Rodopi, pp. 115–120.
- Kaeding, F.W. (1898).
Häufigkeitswörterbuch der deutschen Sprache. Steglitz bei Berlin: self-published.
- Meier, Helmut (1967).
Deutsche Sprachstatistik. Hildesheim: Georg Olms.
- Morgan, B.Q. (1928).
German Frequency Word Book. New York: Macmillan.
- Ortmann, Wolf Dieter (1975).
Hochfrequente deutsche Wortformen. Munich: Goethe Institut.
- Pfeffer, J. Alan (1964).
Basic (Spoken) German Word List. Englewood Cliffs, NJ: Prentice-Hall.
- Reppen, R. and Ide, N. (2004).
“The American National Corpus: Overall Goals and the First Release”. *Journal of English Linguistics*, 32, 2, 105–113.
- Rosengren, Inger (1972).
Ein Frequenzwörterbuch der deutschen Zeitungssprache, Lund, Sweden: Gleerup.
- Ruoff, Arno (1981).
Häufigkeitswörterbuch gesprochener Sprache. Tübingen: Niemeyer.
- Scherer, George A.C. (1965).
Final Report of the Director on Word Frequency in the Modern German Short Story. Boulder, CO: self-published.
- Schmid, Helmut (1995).
“Improvements in Part-of-speech Tagging with an Application to German”. *Proceedings of the ACL SIGDAT-Workshop*. Dublin.
- Scott, Michael (1999).
Wordsmith Tools version 3, Oxford: Oxford University Press.
- Swenson: Rodney (1967).
“A Frequency Count of Contemporary German Vocabulary Based on Three Current Leading Newspapers”. *Dissertation Abstracts*, 28: 2222A–2223A.

This page intentionally left blank

References

Series preface

Ljung, M. (1990)

A Study of TEFL Vocabulary. Stockholm: Almqvist & Wiksell International.

Nation, I.S.P. (1990)

Teaching and Learning Vocabulary. Boston: Heinle and Heinle.

Sinclair, J.M. and Renouf, A. (1988)

"A Lexical Syllabus for Language Learning". In R. Carter and M. McCarthy (eds) Vocabulary and Language

Teaching London: Longman, pp. 140-158.

Introduction

- "An Analysis of Lexical Text Coverage in Contemporary German". In Andrew Wilson, Paul Rayson, and Dawn Archer (eds) *Corpus Linguistics around the World*. Amsterdam: Rodopi, pp. 115-120.
- Kaeding, F.W. (1898). *Häufigkeitwörterbuch der deutschen Sprache*. Steglitz bei Berlin: self-published.
- Meier, Helmut (1967). *Deutsche Sprachstatistik*. Hildesheim: Georg Olms.
- Morgan, B.Q. (1928). *German Frequency Word Book*. New York: Macmillan.
- Ortmann, Wolf Dieter (1975). *Hochfrequente deutsche Wortformen*. Munich: Goethe Institut.
- Pfeffer, J. Alan (1964). *Basic (Spoken) German Word List*. Englewood Cliffs, NJ: Prentice-Hall.
- Reppen, R. and Ide, N. (2004). "The American National Corpus: Overall Goals and the First Release". *Journal of English Linguistics*, 32, 2, 105-113. *Ein Frequenzwörterbuch der deutschen Zeitungssprache*, Lund, Sweden: Gleerup.
- Ruoff, Arno (1981). *Häufigkeitwörterbuch gesprochener Sprache*. Tübingen: Niemeyer.
- Scherer, George A.C. (1965). *Final Report of the Director on Word Frequency in the Modern German Short Story*. Boulder, CO: self-published.
- Schmid, Helmut (1995).

"Improvements in Part-of-speech Tagging with an Application to German". Proceedings of the ACL SIGDAT-Workshop. Dublin. Scott, Michael (1999). Wordsmith Tools version 3, Oxford: Oxford University Press. Swenson: Rodney (1967). "A Frequency Count of Contemporary German Vocabulary Based on Three Current Leading Newspapers". Dissertation Abstracts, 28: 2222A-2223A. This page intentionally left blank