

Mag. Philippa Maurer-Stroh

**Towards a Bilingual Adjective-Noun Collocation  
Dictionary of English and German**

**DISSERTATION**

zur Erlangung des akademischen Grades

**Doktorin der Philosophie**

**Universität Klagenfurt  
Fakultät für Kulturwissenschaften**

**1. Begutachter:** Univ.-Prof. Dr. Allan R. James

**Institut:** Anglistik und Amerikanistik

**2. Begutachter:** Univ.-Prof. Dr. Dušan Gabrovšek (Universität Ljubljana, Slowenien)

**Institut:** Anglistik und Amerikanistik, Fakultät für Kulturwissenschaften

Klagenfurt, Dezember 2004

## **EHRENWÖRTLICHE ERKLÄRUNG**

Ich erkläre ehrenwörtlich, dass ich die vorliegende Schrift verfasst und die mit ihr unmittelbar verbundenen Arbeiten selbst durchgeführt habe. Die in der Schrift verwendete Literatur sowie das Ausmaß der mir im Arbeitsvorgang gewährten Unterstützung sind ausnahmslos angegeben. Die Schrift ist noch keiner anderen Prüfungsbehörde vorgelegt worden.

Klagenfurt, am 15. Dezember 2004

# **Towards a Bilingual Adjective-Noun Collocation Dictionary of English and German**

by  
**Mag. Philippa Maurer-Stroh**

A Doctoral Thesis Submitted to the

**FACULTY OF ARTS  
UNIVERSITY OF KLAGENFURT**

In Partial Fulfilment of the Requirements for the Degree

**Doctor philosophiae**

At the Institute for English and American Studies

**Supervisor:** Univ.-Prof. Dr. Allan R. James

**Co-Supervisor:** Univ.-Prof. Dr. Dušan Gabrovšek (University of Ljubljana, Slovenia)

Klagenfurt, December 2004

*To my late grandmothers,*

*Christl Balling-Stroh*

*† Sept 2003*

*and*

*Eva-Maria Rahm*

*† Jan 2004*

# **Towards a Bilingual Adjective-Noun Collocation**

## **Dictionary of English and German**

### **Abstract (English):**

Collocations are recurrent, usage- rather than semantically based word combinations. The fact that they are largely language-specific – even with respect to languages as closely related as German and English – suggests the need for a bilingual collocation dictionary. This need is proven by analysing a representative selection of comprehensive present-day general bilingual dictionaries as well as by evaluating outputs from machine translation tools, both online and commercial. The doctoral thesis starts with an outline of the different concepts, methods, sources and applications as regards collocations in both the contemporary and historical perspectives. Next, collocations are discussed in the contrastive (English-German) light. Before embarking on the discussion of the methods I have tested with a view to compiling such an English-German adjective-noun dictionary, I outline the main principles of bilingual lexicography at large. By way of conclusion, the applicability of such a dictionary is tested along with the hypothesis as to the expansion possibility of the adjective-noun dictionary to other collocational structures.

### **Abstract (German):**

Kollokationen sind wiederholte, usuelle Wortkombinationen, deren Zusammensetzung sich meist nicht semantisch begründen lässt, sondern vielmehr lexikalischen und konventionalisierten Ursprungs ist. Kollokationen sind zudem größtenteils sprachenspezifisch, und auch ein engverwandtes Sprachenpaar wie das Deutsche und das Englische weist eine Vielzahl an semantischen und strukturellen Unterschieden auf. Anhand von Wörterbuchanalysen und Übersetzungsmaschinen (online oder kommerziell) wird die Notwendigkeit eines zweisprachigen Kollokationswörterbuchs deutlich gemacht. Zunächst werden die unterschiedlichen Begriffe, Methoden, Quellen und Anwendungen der Kollokationsforschung in historischer und zeitgenössischer Perspektive umrissen, darauf folgt eine Abhandlung über Kollokationen im Sprachenvergleich (Englisch-Deutsch). Danach werden die Grundlagen der zweisprachigen Lexikographie erläutert, bevor im Hauptteil der Arbeit auf die verschiedenen Methoden zur Erstellung eines Englisch-Deutschen Adjektiv-Nomen-Kollokationswörterbuchs näher eingegangen wird. Abschließend wird die Anwendbarkeit eines solchen Wörterbuchs diskutiert und dessen Erweiterung auf andere Kollokationsstrukturen besprochen.

## PREFACE

I occasionally work as a freelance translator. Whilst undertaking translation work, I have put in many hours of unnecessary work trying to find the correct equivalent of a multi-word unit. Having checked various bilingual and general monolingual dictionaries as well as monolingual collocation dictionaries and being finally convinced that my choice of a word combination is perfectly fine, it may nevertheless seem odd to a native speaker. Having a degree in English Studies and making extensive use of existing dictionaries is obviously not enough to know which words go together...

Accordingly, the aim of my thesis is to present an appropriate method of compiling a bilingual dictionary specialised on collocations.

*Language cannot be invented; it can only be captured.*

(John McH. Sinclair)

## ACKNOWLEDGEMENTS

I have been riveted on dictionaries and phrases ever since I was first introduced to dictionaries by a former English teacher of mine, way back in 1988, and I would like to thank my two supervisors, Univ. Prof. Allan James and Univ. Prof. Dušan Gabrovšek for greatly encouraging me with their support to engage in lexicography and phraseology – and, naturally, for supervising this doctoral thesis.

I would like to sincerely thank Professor James not only for introducing me to linguistics in general but also for finding me the best possible second supervisor ever. Apart from being my most valuable academic source concerning dictionaries and phraseology (and linguistics in general), Professor Gabrovšek is even more enthusiastic about dictionaries and phrases than I am – and if we had been able to do so, we would have spent (more) hours (in bookstores) indulging in “dictionariesms” and “collcoations”. LEXICOGRAPHY & PHRASEOLOGY GALORE – as Prof. Gabrovšek would say. I am deeply grateful to him for being there for me (in person, e-mail, phone and MS Messenger) whenever problems occurred or there was a danger of “losing track”. Also, I would like to thank him for helpful hints on and making available to me most of the literature on my topic. Finally, Prof. Gabrovšek provided me with the mental support I needed and without which I would never have finished this thesis. Thank you from the bottom of my heart!

Right after my graduation, Professor James made me aware of a number of relevant courses, workshops and other events to participate in. Without these hints I would not have been able to gain from the expertise of John Sinclair, Elena Tognini-Bonelli, Göran Kjellmer, Patrick Hanks, Sue Atkins, Adam Kilgarriff, Mike Rundell, Pernilla Danielsson and Frank Müller.

At this point I wish to express my sincere gratitude to the STUDIENREKTORAT at the University of Klagenfurt. It was their grant that made my participation in these courses possible. Additionally, I was able to use the grant for my journeys to Germany (Professor Michael Stubbs in Trier) and Britain (Professor Wolfgang Teubert in Birmingham).

I would like to thank Mike Stubbs for initiating me into corpus linguistics, for supporting me with advice and for inviting me to Trier to get to know the university and for hands-on sessions on various computer programs.

Prof. Teubert is a valuable contact as regards (contrastive) corpus linguistics and phraseology; he has ever since supported me wholeheartedly and I wish to personally thank him for that.

Furthermore, I would like to thank Ao. Univ. Prof. Günther Fliedl for his support. His enthusiasm and readiness to help were a driving force in setting up a project of international dimensions from a “small” computer program. Especially, I am deeply grateful to my brother Sebastian. He was the one who wrote that initial program – and set the ball rolling. Also, I wish to express my thanks to Ingrid Fernandez de Retana, Andreas Homa, Werner Sühs and Georg Weber, all members of the team.

With sadness, I wish to acknowledge the support of one of my deceased teachers, Ao. Univ. Prof. Hans Köberl. His translation courses were intriguing and illuminating. I wish I could have learned more from him concerning contrastive phraseology.

Researchers and professionals from other institutions also helped me gain better understanding of the subject or introduced me to specialist colleagues: Philipp Koehn, Cambridge (MA); Ann Lawson, Oxford University Press; Oliver Mason, Birmingham University; Matthias Richter, Leipzig University; Olaf Thyen, Duden-Redaktion, Mannheim.

I would like to thank all my close friends – Michael Dabringher, Ingrid Fernandez de Retana, Karolin Knoch, Daniel Koban, Philipp Loebell, Lotte Maurer, Nikolaus Neubauer, Philipp Nicolelli-Fulgenzi, Johannes Pagitz and Ursula Zager, in particular – for their support and help.

Finally, in the most private sphere of acknowledgements, Mum, Dad, Johnny, Basti and Fernanda, you are the best. Thank you for making all this possible! We all know that the situation is not easy, but together we can handle more than that!

## TABLE OF CONTENTS

<i>Preface</i>	i
<i>Acknowledgements</i>	ii
<i>Abbreviations</i>	vii
<i>List of Figures</i>	ix
<i>List of Tables</i>	xii

<b>1. Introduction</b>	1
1.1. An Adjective-Noun Collocation Dictionary	1
1.2. A Bilingual Adjective-Noun Collocation Dictionary	2
1.3. The Need for a Bilingual Collocation Dictionary	5
1.4. What Should an English-German Adjective-Noun Collocation Dictionary Look Like?	8
1.5. Monolingual Collocation Dictionaries vs. Bilingual Collocation Dictionaries	11
1.6. Outline of the Chapters	13
<b>2. Collocation: Concepts, Sources and Methods</b>	14
2.1. Different Approaches, Interpretations and Applications of the Concept of Collocation	15
2.2.1. The Pedagogical Approach to Collocations	16
2.2.2. The Russian Phraseological Approach to Collocations	16
2.2.3. The Phraseological Approach to Collocations	17
2.2.4. The Lexicographical Approach to Collocations	19
2.2.5. The Linguistic (Frequency-Based) Approach to Collocations	22
2.2.6. The Computational Approach to Collocations	24
2.3. Collocations in Phraseology	26

2.3.1. Collocations vs. Free Combinations	28
2.3.2. Collocations vs. Idioms	30
2.3.3. Collocations vs. Compounds	31
2.3.4. The Phraseological Cline	33
2.3.5. Collocations in Other Linguistic Branches	34
2.4. Collocations and Corpus Linguistics	34
2.4.1. Concordances	35
2.4.2. Tagging	36
2.4.3. Parsing	37
2.4.4. Corpus Linguistics, Collocations and Areas of Application	38
2.5. Collocations in Computational Linguistics	43
<b>3. Contrastivity</b>	<b>46</b>
3.1. Predictable and Structurally Congruent Combinations	47
3.2. Predictable and Structurally Incongruent Combinations	48
3.3. Unpredictable and Structurally Congruent Combinations	49
3.4. Unpredictable and Structurally Incongruent Combinations	50
3.5. Case Studies	51
3.5.1. English Collocations and Their Translations in Bilingual Dictionaries	51
3.5.2. English and German Adjective-Noun Combinations and Machine Translation	54
<b>4. Collocations and the Dictionary</b>	<b>59</b>
4.1. Monolingual Collocation Dictionaries	65
4.1.1. English Monolingual Collocation Dictionaries: General	66
4.1.2. English Monolingual Collocation Dictionaries: Specialised	75
4.2. Bilingual Collocation Dictionaries	80

4.2.1. Collocation Dictionaries of English and German	85
4.2.2. Proposals for Bilingual Collocation Dictionaries	88
4.3. Bilingual Lexicography with a Special Focus on English and German	90
4.3.1. Principles of Bilingual Lexicography	90
4.3.2. Collocations and Bilingual Lexicography	99
4.4. Towards a Bilingual Adjective-Noun Collocation Dictionary of English and German	106
<b>5. Methods of Compiling an Adjective-Noun Collocation Dictionary of English and German</b>	107
5.1. The Corpus-Driven Approach	113
5.1.1. ANCR – The Purely Statistical Approach	122
5.2. The Corpus-Based Approach	128
<b>6. Further Implications and Conclusions</b>	137
<b>7. References and Bibliography</b>	141
<b>APPENDICES</b>	161
<i>Appendix A – Corpus Linguistics</i>	162
(1) <i>General</i>	162
(2) <i>Corpus Software – Download and Online Use</i>	164
<i>Appendix B – Online Machine Translation Tools</i>	170
<i>Appendix C – Parallel Test Corpus</i>	171
(1) <i>Text Version of the Corpus</i>	171
(2) <i>Aligned and Annotated Parallel Corpus</i>	172
<i>Appendix D - A Sample of an English-German Adjective-Noun Collocation Dictionary</i>	177

## ABBREVIATIONS

ADJ	adjective
ALTD	<i>Advanced Learner's Trilingual Lexicon. Dictionnaire thématique anglais-français-allemand. Politique-Economie-Expression du temps et de la quantification</i> , 1997
ANCR	Adjective-Noun Collocation Retriever
ARCS	<i>Advanced Readers' Collocation Searcher</i> , 1997
BBI	<i>BBI Dictionary of English Word Combinations</i> , 1997
CALD	<i>Cambridge Advanced Learner's Dictionary</i> , 2003
COBCOLL	<i>Collins COBUILD Collocations on CD-ROM</i> , 1995
COBUILD	<i>Collins COBUILD English Language Dictionary</i> , 1987
COBUILD-CD	<i>COBUILD on CD-ROM</i> , 2001
COGER	<i>Collins German-English, English-German Dictionary Unabridged</i> , 1999
COLLINS-E	<i>Collins e-Großwörterbuch Englisch 4.0</i> , 2004
DAA	<i>Dictionary of Appropriate Adjectives</i> , 1994
DEC	<i>A Dictionary of English Collocations</i> , 1990
DEWC	<i>Dictionary of English Words in Context</i> , 1979
DICT.CC	English/German Online Dictionary by Paul Christian Hemetsberger. Available at <a href="http://www.dict.cc">http://www.dict.cc</a> [10/12/2004]
DUOX	<i>Duden Oxford Großwörterbuch Englisch: Deutsch-Englisch, Englisch-Deutsch</i> , 1999
DUOX-E	<i>Duden Oxford Großwörterbuch Englisch on CD-ROM [PC-Bibliothek 2001]</i> , 1999
DUST	<i>Duden – Das Stilwörterbuch</i> , 2001
ESP	English for Specific Purposes
fig.	figurative
GEW	<i>A Grammar of English Words</i> , 1938
HW	headword
ISED	<i>Idiomatic and Syntactic English Dictionary</i> , 1942
KDEC	<i>The Kenkyusha Dictionary of English Collocations</i> , 1995
L1	native language

L2	foreign language
LANGE	<i>Langenscheidts Handwörterbuch Englisch</i> , 2001
LDOCE	<i>Longman Dictionary of Contemporary English</i> , 2003
LEO	English/German Online Dictionary by the TU Munich. Available at <a href="http://dict.leo.org/">http://dict.leo.org/</a> [10/12/2004]
lit.	literal
lit. trans.	literal translation
LTP	<i>LTP Dictionary of Selected Collocations</i> , 1997
MAC	<i>Macmillan English Dictionary for Advanced Learners</i> , 2002
MW	<i>Merriam Webster's Collegiate® Dictionary</i> , 2003
N	noun
NOAD	<i>The New Oxford American Dictionary on CD-ROM</i> , 2001
NVELE	<i>Noun-Verb Expressions in Legal English</i> , 1975
OALD	<i>Oxford Advanced Learner's Dictionary of Current English</i> , 2000
OCD	<i>Oxford Collocations Dictionary for Students of English</i> , 2002
ODE	<i>Oxford Dictionary of English</i> , 2003
OED	<i>Oxford English Dictionary</i> , 1989
PIE	Phrases in English by William Fletcher. Available at <a href="http://pie.usna.edu/">http://pie.usna.edu/</a> [10/12/2004]
PONS	<i>PONS Wörterbuch für Schule und Studium Englisch-Deutsch</i> , 2003
POS	part of speech
RDWF	<i>Roget's Descriptive Word Finder</i> , 2003
RECDHB	<i>A Russian-English Collocational Dictionary of the Human Body</i> , 1996
REDVC	<i>Russian-English Dictionary of Verbal Collocations</i> , 1993
RH	<i>The Random House Dictionary of the English Language</i> , 1987
SL	source language (primarily when dealing with translation studies)
TL	target language (primarily when dealing with translation studies)
WNW	<i>Webster's New World® Collegiate Dictionary</i> , 2002

## LIST OF FIGURES

Fig. 1: Semantic and structural differences between English and German adjective-noun collocations

Fig. 2: A sample entry from an English-German adjective-noun collocation dictionary (1)

Fig. 3: A sample entry from an English-German adjective-noun collocation dictionary (2)

Fig. 4: Examples of the English *on a ... basis* pattern and its German translation equivalents

Fig. 5a: The entry for *comparison* in the English-German adjective-noun collocation dictionary

Fig. 5b: The entry for *comparison* translated from OCD

Fig. 5c: The entry for *comparison* translated from LTP

Fig. 5c: The entry for *comparison* translated from BBI

Fig. 6: Semantic tailoring

Fig. 7: Collocation vs. idiom

Fig. 8: The treatment of *foregone conclusion* in the English-German adjective-noun collocation dictionary

Fig. 9: Concordance of the noun *haunt*

Fig. 10: Example of a tagged text

Fig. 11: Example of a parsed sentence

Fig. 12: Concordance of *phone call*\*

Fig. 13: Parallel concordance of *winter*

Fig. 14: Examples of predictable and structurally incongruent combinations and their treatment in the English-German adjective-noun collocation dictionary

Fig. 15: Categories of correct translation of all five MT tools

Fig. 16: Sample page from GEW

Fig. 17: Sample page from DEWC

Fig. 18: Sample page from BBI

Fig. 19a: COBCOLL – first step

Fig. 19b: COBCOLL – second step

Fig. 19c: COBCOLL – stopwords

Fig. 19d: COBCOLL – KWIC concordance

Fig. 19e: COBCOLL – expanded example & source of the collocation

Fig. 20: Sample page from LTP

Fig. 21: Sample entry from OCD

Fig. 22: Sample page from DUST

Fig. 23: Sample from NVELE

Fig. 24: Sample entries from DAA

Fig. 25: Sample category from RDWF

Fig. 26: Sample entries from DEC

Fig. 27: Sample page from REDVC

Fig. 28: Extract from KDEC

Fig. 29: Extract from RECDHB

Fig. 30: Screenshot of an ARCS sample entry

Fig. 31: Extract from ALTD

Fig. 32: Active vs. passive dictionaries

Fig. 33: Bi-directional and bi-functional dictionaries

Fig. 34: Hausmann's typology of bilingual dictionaries

Fig. 35: The entry for *sample* in the English-German adjective-noun collocation dictionary

Fig. 36: Sample index of the adjective-noun collocation dictionary

Fig. 37: The semasiological approach to collocations in encoding and decoding entries

Fig. 38: The onomasiological approach to collocations in encoding entries

Fig. 39: An example of decoding convergence (En-De)

Fig. 40: An example of encoding divergence (De-En)

Fig. 41: Graphical user interface of an English-German adjective-noun collocation dictionary

Fig. 42: Sentence alignment and tokenising

Fig. 43: Lemmatising

Fig. 44: Tagging

Fig. 45: Parallel concordance of *rain*

Fig. 46: ANCR – Calling up components

Fig. 47: ANCR – Aligner

Fig. 48: ANCR - Starting the corpus analysis

Fig. 49: ANCR – Search for the English noun *basis*

Fig. 50: Adjective-noun collocations from ANCR

Fig. 51: Corpus-driven entry for *discovery*

Fig. 52a: Corpus-based entry for *discovery*

Fig. 52b: Corpus-driven entry for *discovery*

Fig. 53: The entry for *sample* in an English-German adjective-noun collocation dictionary

Fig. 54a: The corpus-based entry for *remedy*

Fig. 54b: The entry for *remedy* based on COLLINS-E and DUOX-E

## LIST OF TABLES

Table 1: Adjective + noun structures and headword policies as regards *foregone conclusion*

Table 2: Results of the bilingual extraction of word pairs

Table 3: English collocations and their distribution in English-German general bilingual dictionaries

Table 4: Inclusion policy of contrastive collocation categories in present-day bilingual dictionaries

Table 5: The compilation material for the noun *sample*

Table 6: Possible reasons for inclusion and exclusion

# Towards a Bilingual Adjective-Noun Collocation Dictionary of English and German

## 1. Introduction

When people from non-linguistic backgrounds ask me about the subject of my research and I tell them that I compile a bilingual dictionary of adjective-noun collocations, it is nothing unusual for me to get puzzled looks - what is a collocation? In a nutshell, I will say, a collocation can be defined as a typical co-occurrence of words. However, even this straightforward definition usually does not mean a thing to my conversation partners. Only when I illustrate the term collocation by using M.A.K. Halliday's (1966: 150) famous example and make them aware of the fact that the German adjective-noun combination *starker Tee* should be rendered into English as *strong tea* rather than *powerful tea* – although both *powerful* and *strong* are translation equivalents of the German *stark* – will they realise what I mean by the term collocation. Typically, they will agree that a bilingual collocation dictionary is a useful tool to enhance their skills in a foreign language.

### 1.1. An Adjective-Noun Collocation Dictionary

For lexicographical purposes in particular, collocations are normally divided into two categories: grammatical and lexical collocations (Benson et al. 1997: ix).

Grammatical collocations consist of a dominant (content) word, i.e. an adjective, a noun or a verb, and a preposition or a grammatical construction such as a clause or an infinitive. Grammatical collocations are, for example, *(to be) attracted to*, *(it is) delightful + to + infinitive*, *(to have) an interest in (something)*, *(it is) a myth that + clause* and *to vote against (someone/something)*.

Lexical collocations, on the other hand, consist of two content words and typically reveal the following structures:

(1) adjective + noun	<i>addled eggs, blithering idiot, heavy smoker, sheer folly</i>
(2) noun + verb	<i>(a) disaster strikes, negotiations collapse, prices fall</i>
(3) verb + noun	<i>you commit a crime, you give a lecture, you impose taxes</i>
(4) verb + adverb	<i>you deeply meditate, you defend something vigorously</i>
(5) adverb + adjective	<i>readily available, perfectly capable, utterly absorbed</i>

In English, nouns are often used attributively, i.e. pre-modifying another noun, and can thus be treated as adjectives, e.g. *book review* and *cough remedy*.

Given the fact that up to 80% of the words in discourse (spoken and written) are chosen according to the co-selection principle rather than to purely syntagmatic and grammatical factors (Sinclair 2000: 197), and that adjective-noun collocations are among the most common realisations of this lexical co-selection principle, the main focus of this thesis is on the analysis of adjective-noun combinations and combinations of attributively used nouns premodifying other nouns within the wider framework of phraseology, in order to develop a methodology for the compilation of a specialised collocation dictionary.

## 1.2. A Bilingual Adjective-Noun Collocation Dictionary

In both the relevant literature and lexicographical practice, the concept of collocation is still obscured by its numerous interpretations and varying terminology. The English language alone has more than 50 terms that can, in the broadest sense, be defined as 'typical co-occurrence of words' (cf. Wray and Perkins 2000: 3 and Cowie 1988: 7). I will give a brief overview of these in chapter 2 and delimit how I see collocations in the light of my specialised bilingual collocation dictionary.

With regard to this 'babel' it is, indeed, difficult to fulfil the expectations of each 'school of collocation' when writing a collocation dictionary, whether adjective-noun or

general, and this is certainly epitomised in the compilation of a bilingual collocation dictionary, specialised or not.

First of all, what kind of information will the user of a bilingual collocation dictionary of adjective-noun collocations expect to retrieve?

To begin with, many collocations are language-specific and, consequently, collocational patterns are different across language pairs. Even languages as closely related as English and German reveal considerable semantic and structural differences in their adjective-noun collocational patterning, as Fig. 1 clearly shows.

English	German
red-hot / odds-on favourite	haushoher Favorit (lit. trans. <i>house-high favourite</i> )
heavy smoker	starker Raucher (lit. trans. <i>strong smoker</i> )
(the patient is) in a critical condition	der Zustand des Patienten ist kritisch (lit. trans. <i>the condition of ... is critical</i> ) der Patient schwelt in Lebensgefahr (lit. trans. <i>the ... floats in danger of his life</i> )
to be employed as a teacher of French on a part-time basis	als Teilzeitlehrer(in) für Französisch angestellt sein (lit. trans. <i>angestellt sein als Lehrer(in) für Französisch auf Teilzeit-Basis</i> )

Fig. 1: Semantic and structural differences between English and German adjective-noun collocations

Second, in the contrastive framework, collocations have to be looked at from two different angles, i.e. decoding and encoding. Decoding a collocation means to comprehend a foreign-language (hereafter L2) collocation or translate it into one's native language (henceforth L1). Encoding a collocation means to produce an L2 collocation or to translate an L1 collocation into an L2.

Generally speaking, collocations do not cause many problems in the decoding process, as long as the two words that make up the collocation are known to the speaker or writer. The encoding process, however, can cause a lot of difficulties since speakers, in dealing with collocations, tend to rely on the hypothesis of transferability (Bahns 1993: 61), i.e. they take the L1 collocation and translate it word-by-word with the translation equivalents that spring to their minds most readily. The errors, consequently, are due to L1 interference. And it is particularly at the post-intermediate and advanced levels of learning a foreign language, i.e. the stages when the learners start creative writing, that these weaknesses come to the fore.

Collocations are passwords to native-like fluency in a foreign language (Cop 1991: 2776), and although an erroneous collocation may be still understood by the native speaker, it might cause laughter or even annoyance. Collocational errors should thus be minimised in translation and advanced writing in particular. This is why advanced learners and translators are the main target user groups of a collocation dictionary.

Research has shown that, in spite of the recent developments in monolingual learners' dictionaries (of English in particular), translators and learners still consult their bilingual dictionaries first before turning to the monolingual ones (for reassurance or because they were not able to find the information they needed in the bilingual dictionary). This and the fact that there are only few collocation dictionaries available in any language, monolingual or bilingual, justify my efforts to compile a bilingual collocation dictionary.

Since I am a learner of English and a native speaker of German, it would be obvious to make an encoding dictionary for German learners of English; however, in the compilation process, in the absence of an English native speaker, the best possible alternative is for me as the lexicographer to encode into my mother tongue, German, although this, but only at first sight, creates a somewhat artificial situation.

The result of this thesis will therefore be an adjective-noun collocation dictionary for the English user encoding into German. In essence, the dictionary will be monodirectional and monofunctional however, an index of German nouns in the back matter of the dictionary will make it useful also for the German speaker.

### 1.3. The Need for a Bilingual Collocation Dictionary

As mentioned in 1.2., users, particularly in the encoding process, still tend to first use bilingual dictionaries before they refer to monolingual ones. But how good are bilingual dictionaries when it comes to their provision of collocations?

In a recent paper of mine (Maurer-Stroh 2004b) I tested the four major comprehensive bilingual dictionaries of English and German: *Duden Oxford Großwörterbuch Englisch: Deutsch-Englisch, Englisch-Deutsch* (hereafter DUOX), *Collins German-English, English-German Dictionary Unabridged* (COGER), *PONS Wörterbuch für Schule und Studium Englisch-Deutsch* (PONS) and *Langenscheidts Handwörterbuch Englisch* (LANGE). With the help of these dictionaries, in my case study, I tried to translate 38 German adjective-noun collocations into English. The best result was achieved by using COGER, though this also only helped me translate 84% of the collocations in question.

Another test of mine using these four bilingual dictionaries was to translate 29 English adjective-noun collocations into German (Maurer-Stroh 2004c). The best result was only 79.3%.

Moreover, even if the collocation is in the dictionary, this does not mean that it will be easy to find. Bilingual dictionaries still lack a shared policy as to where they list adjective-noun collocations, i.e. under the adjective or under the noun.

Let us make an experiment. How many look-ups do the collocations *a poor comparison* and *a far-fetched comparison* need in DUOX, for example?

(1) *poor comparison*

- (a) you will not find the collocation in the entry for *comparison*
- (b) the entry for *poor* does not explicitly list the combination either
- (c) the next best thing would be to search the German translation of *comparison* since this noun is less polysemous than the adjective *poor* and, consequently, the entry to browse is simpler; and, indeed *poor comparison* can be found in the entry for *Vergleich*

(2) *far-fetched comparison*

- (a) the collocation is nor explicitly listed in the entry for *comparison*
- (b) nor is it listed in the entry for *far-fetched*
- (c) the entry for the German equivalent *Vergleich* does not contain the collocation either
- (d) as a next step we try to look for the collocation in the entry for the German translation equivalent of *far-fetched*, which is *weit hergeholt*; strangely, we do not find an entry for *weit hergeholt*
- (e) we do not find the collocation in the entry for *weit* either
- (f) in the entry for *herholen* (the infinitive of *hergeholt*) we eventually find *weit hergeholt* and its translation equivalent *far-fetched*; still, the collocation is not explicitly stated anywhere in the dictionary

While *poor comparison* needs three look-ups, not even a six-step look-up process in DUOX gives you the German translation of *far-fetched comparison*.

Electronic advances have made it possible to convert print dictionaries into machine readable databases and thus create electronic dictionaries on CD-ROM, which allow for full-text search in all entries. However, even in using the electronic version of DUOX, the translation of *poor comparison* needs three look-ups for successful retrieval since the full-text search lists three entries in which *poor* and *comparison* are included, not necessarily as a collocation though, and maybe scattered across the whole entry. However, the search for a translation equivalent of *far-fetched comparison* is highly facilitated with the help of the full-text search. Only one look-up is needed to find out that the German translation of the adjective in *far-fetched comparison* is *gesucht*. However, the exact collocation is not listed in its translation and the user would have to search for the German translation of *comparison* in another, second, look-up.

Yet, also with the full-text search tool in the CD-ROM version of DUOX you would search in vain for the collocation *apt comparison*.

A bilingual adjective-noun collocation dictionary would remedy this dilemma, as you can see in Fig. 2.

## comparison

appropriate schlagender / sehr überzeugender Vergleich | **apt** passender / treffender Vergleich | **close** eingehender / genauer Vergleich | **detailed** detaillierter / eingehender Vergleich | **direct** direkter Vergleich | **fair** fairer / gerechter Vergleich | **far-fetched** gesuchter / gekünstelter / weit hergeholt Vergleich | **favourable** vorteilhafte Gegenüberstellung, vorteilhafter Vergleich | **good** guter Vergleich | **inept** ungeeigneter / unangemessener / unpassender Vergleich | **invidious** ungerechter / unfairer / unpassender Vergleich | **meaningful** aussagekräftiger / sinnvoller Vergleich | **simple** einfacher Vergleich | **systematic** systematischer Vergleich | **unfair** ungerechter / unfairer Vergleich | **unfavourable** nachteiliger Vergleich | **useful** brauchbarer / praktischer Vergleich | **valid** treffender Vergleich

- **you couldn't have made a more apt comparison** man hätte es nicht treffender vergleichen können; **this is a feeble / misleading / poor comparison** der Vergleich hinkt  
**to make a point by point comparison** etwas Punkt für Punkt vergleichen  
**textual comparison** vergleichende Gegenüberstellung von Texten
- **price comparison** Preisvergleich m  
**textual comparison** (TECHN) Synopse f

Fig. 2: A sample entry from an English-German adjective-noun collocation dictionary (1)

Now one would ask oneself, could it be true that nobody else before has realised that there is a need for a collocation dictionary? And, even more importantly, has this theory been put into practice?

Interestingly enough, there are, in English, only four dictionaries available which explicitly state that they are collocation dictionaries: the *BBI Dictionary of English Word Combinations* (BBI), the *LTP Dictionary of Selected Collocations* (LTP) the *Oxford Collocations Dictionary for Students of English* (OCD) and *Collins COBUILD Collocations on CD-ROM* (COBCOLL). While the BBI and the LTP are available in print format only and COBCOLL is only available as a CD-ROM, the OCD is available in both print and electronic formats.

For the German language, only one combinatory dictionary is available, though it is not exclusively restricted to collocations, the *Duden – Das Stilwörterbuch* (DUST).

As far as bilingual collocation dictionaries are concerned, this type of specialised dictionary is rare in any language pair anywhere in the world, as far as I am aware; what you can find is a Chinese-English and a Japanese-English collocation dictionary, a Russian-English dictionary of verbal collocations and a Russian-English collocation dictionary of parts of the human body.

For English and German, at present, there are indeed two bilingual collocation dictionary available, the *Advanced Readers' Collocation Searcher* (ARCS) and *The Advanced Learner's Trilingual Lexicon. Dictionnaire thématique anglais-français-allemand. Politique-Economie-Expression du temps et de la quantification* (ALTD).

While ARCS is primarily an English collocation dictionary, which is only available on CD-ROM, that translates the English headword into German, ALTD, being a trilingual dictionary in the first place, offers only collocations from a highly restricted register.

Thus, the need for an English-German adjective-noun collocation dictionary is all the more evident.

#### **1.4. What Should an English-German Adjective-Noun Collocation Dictionary Look Like?**

The English-German adjective-noun collocation dictionary I envisage is a dictionary targeted at advanced English learners of German or English-to-German translators, since these two groups can benefit the most from a dictionary of this kind. It is primarily in the encoding process that each of this user group can take advantage of my dictionary in order to produce native-like texts in their L2.

An encoding adjective-noun collocation dictionary for English users necessitates an alphabetised list of headwords that consists of English nouns rather than adjectives, since it is the noun - semantically the core item - that is pre-modified by the adjective. In finding a translation equivalent of a collocation, the user will, therefore, start with the look-up of the noun.

On the microstructural level, the entries will be separated into four different sections.

In the first section, the user finds English adjectives that usually collocate with the headword noun and whose translation equivalents reveal the same adjective + noun structure. They are, for the sake of user-friendliness, listed in alphabetical order rather than by frequency.

Together with the translation of the respective adjectives, each German adjective is followed by the translation equivalent of the English headword. This is essential since some English nouns have, depending on the preceding adjective, more than one German translation equivalent.

The second section is devoted to English adjective-noun collocations whose German translations are structurally incongruent, i.e. not revealing an adjective + noun structure.

In the third section the user finds English patterns, e.g. preposition + adjective + noun, that are structurally incongruent in German.

Finally, in the fourth section, the user is made aware of English adjective-noun combinations that are German compounds, i.e. words that are orthographically written as a single word in German, or German simplex words.

With this layout the user might have to look at each of the sections to get the information they are looking for, but the division certainly makes sense with respect to the complexity the entry would show if all information was given in a single alphabetical stretch.

For polysemous English headwords this four-fold division is repeated for each sense. Sense divisions are marked by Roman numbers, and a synonym or paraphrase is given to find the needed sense easily.

Fig. 3 indicates what such an entry looks like.

## basis

**broad** breite Basis | **firm** feste Basis / Grundlage, stabile Basis | **scientific** wissenschaftliche Grundlage | **sound** solide / gesunde Basis | **theoretical** theoretische Grundlage

- **to work on a casual basis** Gelegenheitsarbeiter sein
- **to have a factual basis** auf Tatsachen beruhen
- **to have no firm / sound basis** (fig.) auf wackeligen Füßen stehen
- **to work on a commission basis** auf Provisions-/ Prozentbasis arbeiten, gegen Provision arbeiten
- **to put something on a different basis** etwas auf eine neue Grundlage stellen
- **to be engaged on a part-time basis to teach French** als Teilzeitlehrer/-in für Französisch angestellt sein /arbeiten
- **to be employed on a permanent / regular basis** fest angestellt sein
- **to do something on a professional basis** etwas [haupt]beruflich machen / betreiben
- **to approach a problem on a scientific basis** an ein Problem wissenschaftlich herangehen
- **to put a procedure on a scientific basis** ein Verfahren verwissenschaftlichen
- **to meet with someone on a social basis** mit jemandem außerdienstlich verkehren
- **to be on a sound basis (business)** auf feste Füßen stehen, (theory) auf einer soliden Basis ruhen
- **to help out on a temporary basis** als Aushilfe arbeiten
- **to be employed on a trial basis** auf Probe, probehalber/-weise beschäftigt sein
- **How does it average out on a weekly basis?** Wie viel ist das im Schnitt pro Woche?
- **on an annual basis** ein Mal im Jahr, jährlich
- **on a commercial basis** gewerblich
- **on a competitive basis** nach Leistung
- **on a daily / day-to-day basis** tageweise
- **on a daily basis** als Tagelöhner
- **on an equal basis** auf gleicher Basis, (bus) als [gleichgestellte(r)] Partner
- **on an experimental basis** versuchsweise
- **on a first-come first-served basis** (fig.) nach dem Prinzip "Wer zuerst kommt, mahlt zuerst"
- **on a full-time basis** ganztags, hauptamtlich, hauptberuflich
- **on a hand-to-mouth basis** (fig.) von der Hand in den Mund [leben]
- **on an hourly basis** stundenweise
- **on an interim basis** (form.) interimistisch, vorübergehend
- **on a long-term basis** auf lange Sicht, langfristig
- **on a monthly basis** ein Mal im Monat, monatlich
- **on a part-time basis** halbtags
- **on a pro-rata basis** anteilmäßig
- **on a random basis** stichprobenweise
- **on a regular basis** regelmäßig
- **on a sale-or-return basis** auf Kommission[sbasis]
- **on a selective basis** selektiv
- **on a systematic basis** systematisch
- **on a temporary basis** aushilfweise
- **on a trial basis** probe- / versuchsweise
- **on a voluntary basis** auf freiwilliger Basis, freiwillig
- **on a weekly basis** ein Mal in der Woche, wöchentlich
- **legal basis** Rechtsgrundlage f

Fig. 3: A sample entry from an English-German adjective-noun collocation dictionary (2)

## 1.5. Monolingual Collocation Dictionaries vs. Bilingual Collocation Dictionaries

As can be seen in Fig. 3, the third section of my entry for *basis* in the English-German adjective-noun collocation dictionary reveals a certain pattern, i.e. *on a ... basis*. However, looking at the German translations, you find that this pattern is not repeated.

Patterns, as it were, blur the traditional distinction between grammar and lexis as two different axes of language description, the syntagmatic and the paradigmatic axes (Sinclair 2000: 194-96). It is often the case, as with *on a ... basis*, for example, that “a particular grammatical or lexical choice correlates with one meaning of a word rather than another” (Sinclair 1991: 6).

In a monolingual collocation dictionary there will surely be a distinction between the senses of *basis* as ‘that from which something can be / is developed’ and ‘the way in which something is organised / done’; however, the *on a ... basis* pattern will not be treated exhaustively in a monolingual collocation dictionary since it is of little interest to list all the adjectives that are possible lexical choices in this pattern and the ‘etc. device’ may have a useful role to play.

In the bilingual framework, on the other hand, the repeated English pattern is of high significance since the German side reveals different translation equivalents for almost all instances for *on a ... basis* (cf. Fig. 3). Even within the separate sections of the entry different translations can be found:

English	German
to approach a problem on a scientific basis	an ein Problem wissenschaftlich herangehen
to put a procedure on a scientific basis on a weekly basis	ein Verfahren verwissenschaftlichen ein Mal in der Woche, wöchentlich
How does it average out on a weekly basis?	Wie viel ist das im Schnitt pro Woche?

Fig. 4: Examples of the English *on a ... basis* pattern and its German translation equivalents

What we learn from this observation is that one cannot simply translate a monolingual collocation dictionary in order to get a useful bilingual one, an argument which is further enhanced when you compare the entries for *comparison* in Fig. 5a to 5d:

<b>comparison</b>	
<p>appropriate schlagender / sehr überzeugender Vergleich   <b>apt</b> passender / treffender Vergleich   <b>close</b> eingehender / genauer Vergleich   <b>detailed</b> detaillierter / eingehender Vergleich   <b>direct</b> direkter Vergleich   <b>fair</b> fairer / gerechter Vergleich   <b>far-fetched</b> gesuchter / gekünstelter / weithergeholt Vergleich   <b>favourable</b> vorteilhafte Gegenüberstellung, vorteilhafter Vergleich   <b>good</b> guter Vergleich   <b>inapt</b> ungeeigneter / unangemessener, unpassender Vergleich   <b>invidious</b> ungerechter / unfaire Vergleich   <b>meaningful</b> aussagekräftiger / sinnvoller Vergleich   <b>simple</b> einfacher Vergleich   <b>systematic</b> systematischer Vergleich   <b>unfair</b> ungerechter / unfaire Vergleich   <b>unfavourable</b> nachteiliger Vergleich   <b>useful</b> brauchbarer / praktischer Vergleich   <b>valid</b> treffender Vergleich</p> <ul style="list-style-type: none"> <li>▪ <b>you couldn't have made a more apt comparison</b> man hätte es nicht treffender vergleichen können; <b>this is a feeble / misleading / poor comparison</b> der Vergleich hinkt</li> <li>▪ <b>to make a point by point comparison</b> etwas Punkt für Punkt vergleichen</li> <li>▪ <b>textual comparison</b> vergleichende Gegenüberstellung von Texten</li> </ul> <ul style="list-style-type: none"> <li>▪ <b>price comparison</b> Preisvergleich m</li> <li>▪ <b>textual comparison</b> (TECHN) Synopse f</li> </ul>	

Fig. 5a: The entry for *comparison* in the English-German adjective-noun collocation dictionary

<b>comparison</b> Vergleich m	
<p>favourable vorteilhaft, <b>unfavourable</b> unvorteilhaft   <b>broad</b> grob, <b>crude</b> primitiv, <b>simple</b> einfach   <b>close</b> genau, <b>detailed</b> detailliert, <b>direct</b> direkt   <b>fair</b> gerecht, <b>good</b> gut, <b>meaningful</b> sinnvoll, <b>useful</b>, nützlich, <b>valid</b> gültig   <b>unfair</b> ungerecht</p>	

Fig. 5b: The entry for *comparison* translated from OCD

<b>comparison</b> Vergleich m	
<p><b>close</b> genau, <b>unfair</b> ungerecht, <b>fair</b> gerecht, <b>far-reaching</b> weit reichend, <b>unfavourable</b> unvorteilhaft, <b>favourable</b> vortheilhaft, <b>instructive</b> aufschlussreich, <b>invidious</b> ungerecht, <b>true</b> wahrheitsgetreu, <b>unwelcome</b> unangenehm</p>	

Fig. 5c: The entry for *comparison* translated from LTP

<b>comparison</b>	
<p>a <b>favourable</b> ~ ein vorteilhafter Vergleich; an <b>unfavourable</b> ~ ein unvorteilhafter Vergleich</p>	

Fig. 5c: The entry for *comparison* translated from BBI

## 1.6. Outline of the Chapters

Chapter 2 starts with an overview of where the concept of collocation is to be placed within the wider framework of contemporary mainstream phraseology. In the subsequent sections I outline the different approaches and interpretations as seen from the historical angle, discussing the major research areas and the work done by the practitioners in the field of collocation studies. Particularly, I deal with corpus linguistics and its merits for collocation research, sketching out the contributions of computational linguistics.

Chapter 3 is devoted to collocations as seen in the contrastive light. I provide a model of classification of English and German adjective-noun collocations and demonstrate in a case study that the contrastive story does not end with bilingual dictionaries but has a much greater impact with regard to the numerous online translation tools which are available free of charge.

In Chapter 4 my main focus is on the lexicographical theory and practice with special reference to collocational considerations. I start by discussing existing collocation dictionaries, monolingual and bilingual, and continue with an outline of bilingual lexicography in general. Further to these theoretical issues, I deal with collocations as presented in modern bilingual dictionaries. In the closing section of Chapter 4 users' needs are examined closely in order to lay the foundation for Chapter 5 – compilation methods.

Chapter 5 first outlines two different methodologies in present-day (bilingual) lexicography and continues with the exemplification and comparison of these approaches.

In Chapter 6, I demonstrate that widening the scope of my research is indeed realistic in terms of both lexicographical and computational perspectives.

## 2. Collocation: Concepts, Sources and Methods

As pointed out already in the introduction (section 1.2.), there is no general consensus on what the term collocation stands for in phraseology, nor is there any universal classification in the relevant literature and practice to be found. What all theories do have in common, however, is that collocations are primarily syntagmatic relations the slots of which are not filled by grammar rules alone<sup>1</sup>.

Using grammaticality as the only constraint in language production is what Sinclair put forward as the ‘open-choice’ principle (1991: 109). Yet, since words do not necessarily occur at random in a text “the open-choice principle does not provide for substantial enough restraints on consecutive choices [and] we would not produce normal text simply by operating the open-choice principle.” (ibid: 110).

As a consequence, Sinclair proposed what he called the ‘idiom’ principle, i.e. that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices.” (ibid.)

The term ‘idiom’ principle is somewhat misleading in the wider framework of phraseology – as we will see later in this chapter – and Sinclair himself re-named the principle of the choice of words being affected by the choice of others in their vicinity (ibid: 173) to ‘co-selection’ principle (Sinclair 2000: 197).

In fact, this dichotomy expressing the two tendencies, the terminological and the phraseological tendency of language patterning (Sinclair 2004: 29), was observed already by Otto Jespersen in 1924 (Clas 1999:1) and Harold E. Palmer in 1933 (Cowie 1999: 132)<sup>2</sup> – although in different frameworks.

---

<sup>1</sup> In a way, one could say that collocations combine what Saussure (1916) has so cleverly separated into syntagmatic and associative (paradigmatic) relations.

<sup>2</sup> Note that by the time of their observations they were geographically wide apart – Jespersen was in Europe while Palmer taught in Asia.

Clearly, collocation is a concept which works on the level of co-selection, but this principle covers much more than that. What can also be subsumed under the superordinate co-selection are proverbs, clichés, idioms, compounds – just to mention a few examples.

One of the reasons why collocations are defined and looked at in a variety of ways is that it is notoriously difficult to classify them as a category in phraseology. Generally, they are defined according to the research area or practical purpose for which they are of immediate concern. Thus, a computational linguist will define collocation in a different way than a translatalogist or a lexicographer.

## 2.1. Different Approaches, Interpretations and Applications of the Concept of Collocation

Historically speaking, H.E. Palmer is said to have been the first to actually use the term ‘collocation’ in its modern linguistic sense in his report to the IRET (Institute for Research in English Teaching) Convention of 1927 in Tokyo (Cowie 1999: 53)<sup>3</sup>.

Yet, the father of phraseology in general is said to have been the Swiss structuralist Charles Bally<sup>4</sup> who distinguishes between *locutions phraséologiques* (phraseological units in the broadest sense) and *unités phraséologiques* (idioms in the strictest sense) (cf. Bally 1909, as quoted in Hecken 2004).

In principle, there are six major lines of ‘collocational thinking’: the pedagogical, the lexicographical, the early (Russian) phraseological, the (contemporary) phraseological, the linguistic (frequency-based) and the computational approaches. In practice, however, these approaches are all interrelated. In the following, I will outline the different approaches and identify the people most prominently associated with them.

---

<sup>3</sup> The OED states 1750 HARRIS *Hermes* II. iv. Wks. (1841) 197 as the first mention of collocation in the sense of ‘arrangement of words in a sentence’, while collocation as a technical linguistic term was, according to the OED, introduced by J.R. Firth in 1951.

<sup>4</sup> He was a pupil of Ferdinand de Saussure.

### 2.2.1. The Pedagogical Approach to Collocations

Harold E. Palmer's interest in collocations was primarily a pedagogical one since he was a teacher of English in Japan.

"It is not so much the words of English nor the grammar of English that makes English difficult, but that that vague and undefined obstacle to progress in the learning of English consists for the most part in the existence of so many odd *comings-together-of-words*."

(H.E. Palmer 1933: 13)

Thus, Palmer was fully aware of the need for a classification of these 'comings-together-of-words', and he defined collocation as "a succession of two or more words that may best be learned as if it were a single word." (Palmer 1938: x; cf. Cowie 1999: 54).

Although Palmer and his assistant and eventual successor A.S. Hornby used the term *collocation* for the whole spectrum of word combinations, their work laid the foundations for the treatment of collocations and idioms by later generations of phraseologists, both theorists and lexicographers<sup>5</sup>.

### 2.2.2. The Russian Phraseological Approach to Collocations

In Russia, linguists were strongly influenced by the work of Charles Bally (cf. section 2.1.), and it was in the time from the 1940s to the 1960s that phraseology started to develop there into a discipline in its own right (cf. Hecken 2004; Skandera 2004).

---

<sup>5</sup> Palmer and Hornby's awareness of collocations (in the broadest sense) was a result of their work on a controlled English vocabulary. The vocabulary control movement started in the late 1920s and arose from the pedagogical need to reduce the vocabulary for foreign learners of English to a manageable core – which the results of Palmer's earlier studies had shown to be 1,000 words as the required minimum in order to still be able to achieve naturalness. While Palmer, Hornby and Michael West (who joined them in the late 1930s) used personal judgment on the basis of their teaching experience to compile such a word list, the American linguist Edward L. Thorndike went for an objective-quantitative method in which he counted the most frequent words in a body of written text of four million words (Cowie 1999: 14-17).

Russian phraseologists, with V.V. Vinogradov and N.N. Amosova as their prime figures, saw phraseological units as different realisations on a phraseological continuum. Different degrees of opacity of the combinations were the key for they attempt to reach a classification of all phraseological units.

On the primary level of the Russian phraseologists' classification, sentence-like (or pragmatic) units were distinguished from word-like (or semantic) units. For the purpose of this thesis I will only deal with word-like units since adjective-noun combinations fall in this category.

Word-like units were classified along a cline of semantic motivation: (1) semantically opaque units, (2) partially motivated units and (3) phraseologically bound units – where collocations fall under the third category (Cowie 1998a: 4-7).

### **2.2.3. The Phraseological Approach to Collocations**

The phraseological approach to collocations has been strongly influenced by the Russian tradition in as much as it is an attempt to categorise all phraseological units along a cline of varying degrees of opacity (cf. Skandera 2004: 28).

#### **(1) A. P. Cowie**

Cowie takes this a step further, and categorises phraseological units not only on the basis of their semantic transparency, but also according to their grade of lexical and grammatical variability (cf. Nesselhauf 2004: 10).

First of all, Cowie (1994) distinguishes 'composites' from 'formulae' – a categorisation which goes hand in hand with the Russian tradition – the former being word combinations below the sentence level. According to semantic transparency and (lexical and grammatical) variability composites are further classified into the following categories which should be seen on a scale from (1) to (4):

- a) **pure idioms** – they are completely opaque and invariable, e.g. *red tape*
- b) **figurative idioms** – they are slightly variable and have one figurative and one literal interpretation, e.g. *green fingers*
- c) **restricted collocations** – they have one figurative and one literal element and there is an arbitrary combinability restriction on one of the elements with other elements outside the combination, e.g. *dry cow*
- d) **open collocations** – they consist of elements which are both used in their literal sense and they are freely combinable, e.g. *thunderous applause*.

## (2) I.A. Mel'čuk

Following the Russian classification model, Mel'čuk also distinguishes between sentence- and word-like units.

He refers to word-like units as semantic phrasemes (as opposed to pragmatic phrasemes) and subdivides them as follows (cf. Nesselhauf 2004: 12; Skandera 2004: 29):

- a) **idioms** – their meaning as a whole is not included in any of the constituents, e.g. *red tape*
- b) **quasi-idioms / quasi-phrasemes** – they have a literal meaning and implicitly express an additional one, e.g. *bacon and eggs* – where the meaning of ‘fried’ is part of either element (Nesselhauf 2004: 12)
- c) **collocations / semi-phrasemes**

Mel'čuk further classifies collocation in four major categories, which Nesselhauf (2004: 12) summarises in a less formalised language as (1) collocations including a delexical verb, e.g. *take a step*, (2) collocations in which the meaning of the dependent lexeme is expressed in combination with only a few others, e.g. *black* (‘without milk’) *coffee*, (3) collocations whose dependent lexeme can be used in the same sense in combination with a lot of other lexemes, e.g. *strong coffee*, and (4) collocations in which the

dependent lexeme includes the meaning of the other lexeme, e.g. the meaning ‘horse’ is included in *neigh*<sup>6</sup>.

What Mel’čuk is best known for is his model of ‘lexical functions’ (LFs), which is part of his ‘Meaning-Text-Theory’ and comprises well over 50 different types. He sees collocations as an oriented relationship where the so-called ‘keyword’ selects the dependent lexeme, the ‘value’, i.e. in a combination like *heavy smoker*, *smoker* is the keyword and selects the value *heavy* through the lexical function of Magn (‘magnitude’ = intensifier)<sup>7</sup>.

#### 2.2.4. The Lexicographical Approach to Collocations

The lexicographical approach deals with the question of how collocations should be presented in dictionaries. The prime practitioners are Morton Benson and Franz-Josef Hausmann, although also A.P. Cowie, I.A. Mel’čuk and H.E. Palmer must be mentioned here.

Cowie implemented his phraseological approach in the two volumes of the *Oxford Dictionary of Current Idiomatic English* (1975 and 1983), Mel’čuk compiled the *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques* (1984-99), so far in four volumes and, together with A.K. Zholkovsky, the *Explanatory Combinatorial Dictionary of Modern Russian* (1984) using his model of

---

<sup>6</sup> This conception of the term collocation can be compared to W. Porzig’s “Wesenhafte Bedeutungsbeziehungen” (1934) – this is basically an approach to meaning concerned with syntagmatic relations, e.g. *wiehern – Pferd* (viz. *neigh – horse*) where the semantic relevance of the syntactic relation is obvious since what can be connected with *wiehern* is implicit in it: “Es handelt sich dabei [...] um eine beziehung, die im wesen der gemeinten Bedeutungen selbst gründet. Ich nenne sie deshalb wesenhafte Bedeutungsbeziehungen.” (Porzig, 1934 as quoted in Bahns 1996: 3).

<sup>7</sup> Anecdotally, Mel’čuk invented his concept of “lexical functions to label systematic and re-current lexical-semantic relations” during a storm in the Russian countryside, when reflecting on the English language: the fact that “*rain* combines with *heavy* while *light* combines with *bright* led him to invent the first lexical function which he called Magn (from Magnitude) to refer to an intensifying meaning.” (cf. Evens 1998: 11, as quoted in Fontenelle 1997: 5)

LFs<sup>8</sup> and Palmer took his work on the core vocabulary of English as the basis for his dictionary *A Grammar of English Words* (1938).

**(1) M. Benson**

Benson, being an American Slavonic linguist, is also influenced by the Russian tradition in way that he defines collocations with respect to other word combinations. His approach to the presentation of collocations in a dictionary mainly had its basis in practical reasons for the purpose of compiling, together with E. Benson and R. Ilson, the BBI<sup>9</sup>.

As summarised in Nesselhauf (2004: 14-15), he divides lexical combinations (word-like units in the Russian tradition) into the following categories:

- a) **idioms** – they are defined as “relatively frozen expressions whose meanings do not reflect the meanings of their component parts” (Benson et al. 1986a: 252)
- b) **transitional combinations** – are more transparent than idioms but less variable than collocations
- c) **collocations**
- d) **compounds**

Collocations, for Benson, are “recurrent, semi-fixed combinations” and they can be further subdivided into grammatical and lexical collocations (Benson et al. 1997: ix; cf. section 1.1.).

---

<sup>8</sup> The four volumes (172, 332, 323 and 347 pp.) cover only a selection of some 500 entries scattered throughout the alphabet, of varying complexity (Gabrovšek 2004: E-mail communication). A full implementation of Mel'čuk's LFs is *A Russian-English Collocational Dictionary of the Human Body* compiled by Lidija Jordanskaja and Slava Paperno edited by Richard L. Leed, 1996. The online version is available at: [http://russian.dml.cornell.edu/russian.web/BODY/WIN\\_1251/INDEX.HTM](http://russian.dml.cornell.edu/russian.web/BODY/WIN_1251/INDEX.HTM) [10/10/2004]

<sup>9</sup> Interestingly, Benson recognised the need for a collocation dictionary for English learners when he compiled his Serbo-Croatian – English dictionary in the mid-to-late 1960s (Gabrovšek 2004: personal communication).

## (2) F.J. Hausmann

Hausmann, first of all, distinguishes fixed from non-fixed combinations. Fixed combinations, for him, are idioms, compounds, etc., while non-fixed combinations can be further classified as (Hausmann 1984, as quoted in Nesselhauf 2004: 16 and Bahns 1996: 23):

- a) **co-creations** – are free combinations that are creatively combined by the speaker, e.g. *une maison agréable* (viz. *a pleasant home*)
- b) **collocations** – are not creatively combined but put together out of some convention, e.g. *ton péremptoire* (viz. *peremptory tone*)
- c) **counter-creations** – are words that do not usually combine, they are mainly found in literature and advertisements to create a special effect, e.g. *jour fissuré* (viz. *cracking day*).

A collocation, according to Hausmann, can be defined as:

“la combinaison caractéristique de deux mots dans une des structures suivantes: a) substantif + adjective (épithète) b) substantif + verbe c) verbe + substantif (objet) d) verbe + adverbe e) adjective + adverbe f) substantif + (prép.) + substantif.”

(F.J. Hausmann 1989: 1010)

According to Nesselhauf (2004: 16-17), Hausmann's most important contribution to collocation theory is his view what the status of the two elements in the collocation is not the same. While one of the elements, the ‘base’, is semantically autonomous, the other one, the ‘collocator’ is chosen according to this base and only get its precise meaning from that combination<sup>10</sup> (cf. Hausmann 1989: 1010).

Lexicographically speaking, he put forward that collocations should be listed under their bases rather than their collocators, since “la fameuse recherche du mot propre est celle du collocatif.” (Hausmann 1989: 1010). He is of the opinion that listing the collocation in the collocator entry merely serves the testing of a hypothesis (“Can I say

---

<sup>10</sup> This phenomenon is referred to as ‘semantic tailoring’ by D.J. Allerton (1982 and 1984).

this?”), but listing the collocation in the base entry allows the dictionary user to hit upon an unknown collocation or a collocation they might have forgotten (*ibid.*).

### 2.2.5. The Linguistic (Frequency-Based) Approach to Collocations

In contrast to the Russian, the phraseological and the lexicographical approaches, the linguistic approach does not characterise collocations with respect to other word combinations on a certain cline. Collocations in the linguistic approach are mainly seen in terms of frequency and not so much in terms of semantics.

#### (1) J.R. Firth

Firth was the first to establish the expression *collocation* as a technical linguistic term: “I propose to bring forward as a technical term, meaning by ‘collocation’” (Firth 1951: 194)<sup>11</sup>.

The concept of collocation in Firthian terms developed in connection with his ‘Theory of Meaning’<sup>12</sup>, where he explains that “meaning by collocation is an abstraction at the syntagmatic level and it is not directly concerned with the conceptual or idea approach to the meaning of words [nor is it the same thing as contextual meaning]. One of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night*.” (Firth 1957b: 196).

However, Firth never clearly defines the term *collocation* (Lyons 1977: 612). As Nesselhauf (2004: 2-5) observes, he generally uses four different concepts. Collocations

---

<sup>11</sup> It may well be that Firth actually adopted the term from H.E. Palmer (Nesselhauf 2004: 2 and Mitchell 1975b: 134) – who decided to employ ‘collocation’ as the term for his most inclusive category of ‘comings-together-of-words’ since it had, in contrast to ‘idiom’, no settled status as a generic term in 1927 (although the sense ‘arrangements of words in a sentence’ dates back to 1750). (cf. footnote 3 and Cowie 1999: 54).

<sup>12</sup> Firth distinguished three levels of meaning: “meaning by collocation”, the “conceptual or idea approach to the meaning of words” and “contextual meaning”.

are thus either (1) co-occurrences of words in general (frequent or rare)<sup>13</sup>, (2) habitual (= frequent) co-occurrences of words, (3) a number of consecutive words or (4) an order of ‘mutual expectancy’.

Butler (1985: 7, as quoted in Bahns 1996), on the other hand, believes that Firth sees collocations exclusively as habitual co-occurrences of mutually predicting lexical items such as *dark + night* and *silly + ass*.

### (2) M.A.K. Halliday

Halliday was one of the first to build on Firth’s concept of collocation. He considers as collocations syntagmatic associations of lexical items of all probabilities which occur in a certain distance from each other.

He defines ‘probability’ as “the frequency of the item in a stated environment relative to its total frequency of occurrence.” (Halliday 1966: 156 and Nesselhauf 2004: 4), but he does not clearly define ‘distance.’

Like Firth, he sees collocation as a lexical phenomenon rather than a grammatical one – although he revised this conception later on, stating that

“it is essential also to examine collocational patterns in their grammatical environments, and to compare the descriptions given by the two methods, lexical and lexicogrammatical. This then avoids prejudging the answer to the question whether or not, and if so to what extent, the notion of ‘lexicalness’, as distinct from ‘lexicogrammaticalness’, is a meaningful one.”

(M.A.K. Halliday, 1966: 159).

### (3) T.F. Mitchell

Mitchell was a dedicated Firthian in as much as he saw ‘habitualness’ as the basis for his concept of collocations, but he also emphasised the necessity of studying collocations in their grammatical structure (Nesselhauf 2004: 6).

---

<sup>13</sup> Firth was analysing co-occurrence frequencies in letters dating from the 18<sup>th</sup> and 19<sup>th</sup> century and he called frequent co-occurrences, *inter alia*, ‘usual’ or ‘habitual’ collocations, while he called rare co-occurrences ‘personal’ or ‘idiosyncratic’ collocations.

He exemplifies this with the following example: While it is possible to say *heavy damage*, *to damage heavily* and *heavily damaged*, it is incorrect to say *\*heavy damager* or *\*heavy damaging* (Mitchell 1966: 337).

Furthermore, Mitchell was the first in the Firthian tradition “to attempt to draw a line between collocations and compounds and between collocations and idioms” (Nesselhauf 2004: 6). His concept can, therefore, be related to the phraseological and the lexicographical approaches.

Likewise, also Benson’s lexicographical-phraseological approach can be correlated with the Firthian frequency-based approach, since Benson sees collocations as ‘recurrent combinations’ (cf. section 2.2.4.).

### **2.2.6. The Computational Approach to Collocations**

The computational approach to collocations is, in essence, a continuation of the frequency-based methodology in collocation studies.

The most prominent figure in the computational approach is John McH. Sinclair. He is, principally, a follower of Firth’s tradition and his (early) work was strongly influenced by Halliday, Mitchell and McIntosh.

Apart from his computational efforts, Sinclair has, more than any other Firthian linguist, been “concerned with finding solutions to the practical problems that a Firthian view of collocation poses for analysis.” (Nesselhauf 2004: 8).

He helped disentangling the confusing definitions of collocations as being combinations of words (Firth), lexical items (Halliday) or roots (Mitchell) by putting forward that a collocation is a relationship between ‘words’, by which ‘lemmas’ are meant (cf. Sinclair 1991: 173). Lemmas are composite sets of word-forms, as are lexical items and

roots in Mitchell's terms, e.g. *sings*, *singing*, *sing* and *sang* can be summarised as the lemma SING<sup>14</sup>.

Sinclair consolidated the use of a special technical terminology in collocation research: The *node* is the word under study, the *collocate* is the word that enters into collocation with it and the *span* is the distance between the words. The set of all the collocates that can enter in collocation with the node is called (*collocational*) *range* of that particular word<sup>15</sup>.

Above all, Sinclair was well aware of the fact that with the help of computers, frequency counts in collocation studies could be highly facilitated since, in his approach to the study of lexis, he saw “problems which are not likely to yield to anything less imposing than a very large computer.” (Sinclair 1966: 410, cf. Bahns 1966).

Sinclair, together with S. Jones, first used a computer in the analysis of a text comprising 150,000 words. The results showed that 95% of the collocational influence of the node takes place within a span of four words to its left and to its right. (Jones and Sinclair 1974: 21, as quoted in Nesselhauf 2004: 8). Thus, a good move was made towards solving Halliday's problem with 'distance'.

Being a Firthian, Sinclair also distinguished between frequent and rare collocations, 'significant' and 'casual', in his terms (cf. Nesselhauf 2004: 8). With the help of the computer and statistical calculations also this distinction can be sharpened – where the threshold is to be set, however, remains a subjective matter.

In accordance with the phraseological approach, Sinclair is conscious of the fact that collocations and idioms may be structurally the same, but differ in their semantics. He even points out that “the line between them is not clear” (Sinclair 1991: 172) – an aspect which will be looked at in more detail in the subsequent section.

---

<sup>14</sup> It is a convention in linguistics to write lemmas in upper-case letters.

<sup>15</sup> Note that 'node' and 'collocate' are in par with the terms 'basis' and 'collocator' of the lexicographical approach. For the purpose of this thesis I will largely stick to the lexicographical terminology, but also occasionally use the linguistic expression, especially when referring to the computational approach.

### 2.3. Collocations in Phraseology

Gabrovšek (2000: 184) defines phraseology as “the investigation of the lexical phenomenon of (significant) word combinations”.

We have seen in section 2.2. that there is no common descriptive approach to word combinations, but what can be summarised from the above is that collocations are to be placed somewhere in between free combinations, idioms and compounds<sup>16</sup>.

Criteria which are frequently applied to distinguish collocations from free combinations, idioms and compounds are fixedness and semantic opacity (cf. Benson 1985; Gabrovšek 2000 and 2003; Heliel 1990; Moon 1998).

#### (1) Free combinations

Free combinations, in line with Sinclair’s open-choice principle, are sequences of words that follow rules of grammar and syntax of the language in question, and whose elements allow for free substitution. They are the least cohesive types of word combinations and the combination as a whole can be understood from the sum of the most prototypical literal meanings of the elements. Examples are *heavy bag*, *heavy basket*, *heavy stone* and *beautiful flower*, *yellow flower*, *red flower*.

#### (2) Collocations

Collocations are loosely fixed, typical and frequently co-occurring word combinations that are put together out of some arbitrary usage. Collocations are semantically transparent, i.e. their meanings can be derived from the denotative meaning of their constituents, but the elements cannot be freely substituted. When we refer to a solar eclipse, for example, we usually talk about a *total eclipse* and combinations with *absolute*, *complete*, *entire* or *whole* are usually not accepted, while *full eclipse* can be.

---

<sup>16</sup> For the purpose of this thesis I will not go into detail concerning other multi-word units such as allusions, clichés, proverbs, similes and the like.

Sinclair (2004: 29) observes that “many of these patterns seem almost purely linguistic ([...], where on semantic grounds the adjectives should be interchangeable, but on collocational grounds they are not).”

Charles Bally already (as quoted in Béjoint 1994: 213) observed that the links between words can be lexical (‘contrainte de signe’) rather than semantic (‘contrainte de contenu’) and he illustrated this with his famous example *gravement malade* vs. *\*grièvement malade*.

This idea is also discussed by Bo Svensén (1993: 98-102) who distinguishes lexical from semantic collocations. Semantic collocations are put together according to semantic restrictions. The unacceptability of the utterance *\*the driver of the car was badly damaged* results from the fact that the verb *to damage* needs its direct object to have the semantic feature [+inanimate] and not [+animate]. The bond between the elements in a lexical collocation, however, cannot be explained by either logic or semantic reasons, but “applies to the expressional aspect, and so relates to the choice of words; its basis is simply in the way that usage has evolved.” (ibid: 99). To demonstrate this, he states that it is not clear why it is correct to say *to make a trip*, but *\*to make a walk* is not acceptable. “Language simply seems to dictate, for no good semantic reason, that such-and-such a combination does, or does not, occur. The restriction is, thus, an arbitrary one.” (Allerton 1984: 28).

### (3) Idioms

Idioms are relatively fixed and semantically opaque word combinations. The criterion of semantic opacity of idioms was defined by Sweet as early as 1899: “The meaning of each idiom is an isolated fact which cannot be inferred from the meaning of the words of which the idiom is made up.” (Sweet 1899: 139, as quoted in Skandera 2004: 24). When we *give somebody the red carpet*, for example, we do not actually hand over a red carpet to them, but rather give them a special treatment as important visitors. Likewise, when someone *makes heavy weather of something*, this has nothing to do with an atmospheric condition, but they make things more complicated than they need to be.

In between collocations and idioms, there is yet another sub-category to be detected, that of semi-idioms (or ‘restricted collocations’, as some linguists call them). By definition, semi-idioms consist of one literal and one figurative element, e.g. *heavy heart* where the literal meaning of *heavy* is ‘tailored’ to the figurative meaning of *heart*.

#### (4) Compounds

Compounds are completely frozen combinations of two (or sometimes even more) words. They are referred to technically as ‘endocentric’ when the semantic head is inside the combination, i.e. the meaning of the whole combination characterises the determinatum, e.g. *wet suit* (= a special type of suit) and *prime minister* (= a special type of minister). ‘Exocentric’ compounds have their semantic head outside the combination and they denote an unknown variable, e.g. *white elephant* (= something that is completely useless although it might have cost a lot of money) and *wet blanket* (= a person who spoils the joy of others).<sup>17</sup>

#### **2.3.1. Collocations vs. Free Combinations**

Collocations can be distinguished from free combinations by the restricted commutability of the component elements as a typical result of semantic tailoring and their frequency of co-occurrence (cf. Aisenstadt 1979, Benson 1985, Heliel 1990 and Bahns 1996).

##### **(1) restricted commutability of the component elements as a typical result of semantic tailoring**

In the collocation *heavy rain* – where *heavy* means ‘lots of / more than usual’ – the noun cannot be substituted freely since only a small number of nouns are permissible without changing the meaning of *heavy*, e.g. *heavy frost / traffic / drinking / smoking*.

---

<sup>17</sup> It was Leonard Bloomfield who introduced these technical terms in 1933 already.

Likewise, *rain* can only go with a certain number of other adjectives to express a (near-) synonymous meaning to *heavy*, e.g. *pouring* / *lashing* / *driving* / *torrential rain*.

In free combinations like *heavy bag* both elements can be replaced by others, e.g. *light* / *brown* / *plastic* / *diplomatic* / *medical bag* and *heavy suitcase* / *basket* / *trunk* / *stone* / *boulder* / *box* / *umbrella*.

Restricted commutability can be inferred from what Allerton (1982 and 1984) calls semantic tailoring. This is especially the case with adjective-noun collocations where the meaning of the adjective is interpreted relative to the noun it modifies. This meaning is semantically tailored to the noun, i.e. it is “different from its meanings in more ‘neutral’ contexts” (Helie 1990: 130). To illustrate this let us look at the following examples.

heavy rain	=	‘a lot of’
heavy schedule	=	‘busy’
heavy furniture	=	‘large and solid’
heavy sigh	=	‘loud and deep’
heavy soil	=	‘wet and sticky’

Fig. 6: Semantic tailoring

In a way, all the meanings of *heavy* in these collocations could be united through the common denominator “emphasising that something is ‘more’ than usual”; they are, however, specialised from context to context, i.e. from noun to noun. They are semantically tailored to the noun.

As we will see later on, it is exactly this semantic tailoring that often results in interlingual unpredictability of (primarily) the adjectives in the translation of adjective-noun collocations.

## (2) frequency of co-occurrence

Another factor that distinguishes collocations from free combinations is their frequency of co-occurrence.

Frequency considerations in collocation studies date back to J.R. Firth and got a fresh input with John McH. Sinclair's computational considerations starting in the 1970s ( cf. 2.2.)

Over the past thirty years computer technology has steadily improved, making it possible to compile all sorts of texts and store them electronically to form huge text corpora which can be, with the help of specialised analysis tools, searched for patterns or statistical evidence (cf. 2.4. and 2.5.)

When we nowadays speak of frequency of co-occurrence in collocational analysis, we see this in terms of combined frequencies relative to the absolute frequencies each component of the collocation has in isolation and the overall combinatorial ability of the two given lexical item as compared to combinations of any other lexical items in a language.

Hence, the "strength" (Butler 1985: 7) of the collocation *heavy schedule* will be much greater than the strength of the free combination *heavy bag*.

### 2.3.2. Collocations vs. Idioms

Generally speaking, collocations can be distinguished from idioms by their semantic transparency. While the meaning of a collocation can be derived from the literal meanings of their constituent elements, idioms are semantically opaque, meaning that you no longer can deduce the overall meaning of the combination by simply adding up the literal/dictionary meaning of its component parts. The following examples illustrate this.

heavy rain (make) heavy weather	= lots of + water that falls from the clouds in drops ≠ lots of weight + the condition of the atmosphere at a particular place and time
------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------

Fig. 7: Collocation vs. idiom

Another criterion to differentiate between idioms and collocations is their (lexical and grammatical) fixedness (Svensén 1993: 109), e.g. while you can convert the adjective-noun collocation *heavy rain* into a verb + noun structure, viz. *to rain heavily*, you cannot use the comparative or superlative forms of the adjective in *to make heavy weather*, nor can passivise it.

### 2.3.3. Collocations vs. Compounds

Compounds are usually defined as “being one word (in the sense of lexeme) that is made up of two other words (in the sense of lexeme)” (Bauer 1998: 65).

While compounding is primarily a branch of word formation, compounds are sometimes difficult to tell from syntactic combinations like collocations (ten Hacken 2004: 53). So, what is it that differentiates a compound from a syntactic combination? This is a question which has not yet been answered satisfactorily in the literature and opinions throughout the linguistic world are divided. They range from the difficult-to-define ‘institutionalisation’ criterion to the fact that compounds are stored together in the mental lexicon.

In his article “When is a Sequence of Two Nouns a Compound in English?”, Laurie Bauer (1998) discusses six commonly found principles to distinguish between the two categories:

- (a) compounds are listed, i.e. idiomatic
- (b) compounds are spelled as one word
- (c) compounds have fore stress
- (d) the first element in a compound is syntactically isolated, i.e. variation is uncommon
- (e) compounds do not permit co-ordination
- (f) the head of a compound cannot be replaced by ‘one’

In fact, as ten Hacken observes (2004: 55), each criterion that is applied to prove the compound status of a given syntactic combination “will lead to a slightly different set of items [...], although there will of course be a large overlap between the sets”.

Also Bauer (1998: 83-84) remains unconvinced of a strict separation of compounds from other syntactic structures – a view which I entirely agree with.

Endocentric compounds, in my opinion, are on a par with collocations on the one hand and with semi-idioms on the other hand<sup>18</sup>, while exocentric compounds are idioms. This view is challenged by lexicographic praxis in which compounds are separated from collocations, semi-idioms and idioms (for a few well-known reasons such as institutionalisation, frequency and the ‘two-item convention’)<sup>19</sup> and are treated as lexical items in their own right and thus get headword status, while collocations, semi-idioms and idioms are listed within the entries<sup>20</sup>. What supports my claim, however, is that different dictionaries attribute headword status to different syntactic constructions. There is no common lexicographical policy, as the following example (Table 1) shows:

foregone conclusion	
<b>CALD</b>	headword
<b>COBUILD</b>	listed under the adjective, in sense 2 the combination is treated as if it was the headword
<b>LDOCE</b>	headword
<b>MAC</b>	headword
<b>MW</b>	headword
<b>WNW</b>	listed under the adjective, in sense 2 the combination is treated as collocation ('as said of conclusion')
<b>NOAD</b>	listed under the adjective and explicitly labelled 'phrase'
<b>OALD</b>	listed under the adjective and explicitly labelled 'idiom'
<b>ODE</b>	listed under the adjective and explicitly labelled 'phrase'
<b>RH</b>	headword

Table 1: Adjective + noun structures and headword policies as regards *foregone conclusion*

<sup>18</sup> Note that is far more common for an endocentric compound to be close to a collocation than to a semi-idiom.

<sup>19</sup> Gabrovšek 2004: private communication

<sup>20</sup> Note that, here, I am talking about combinations (including compounds) which are spelled as two words, hyphenated or not, which make the bulk of so-called English compounds – while in German compounds are almost always spelled as a single word.

As I will show in more detail in chapter 3-5, on the compilation of my dictionary, the distinction between collocations and compounds does not cause any major difficulties, since my policy is that any English adjective (or attributively used noun) + noun construction which can be semantically defined as a collocation (i.e. is not opaque) that is rendered into German by a single stretch of characters (I will call them ‘authentic’ German compounds), is listed in a special section in the entry (cf. 1.4.). Accordingly, the combination *foregone conclusion* would be presented in my dictionary as follows.

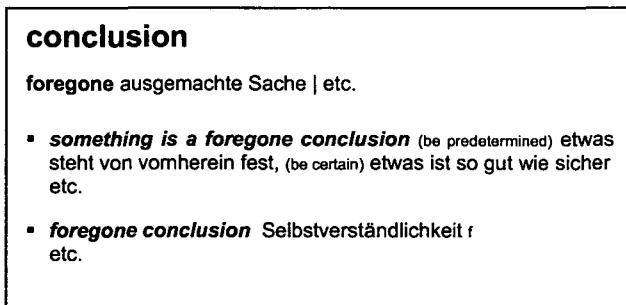


Fig. 8: The treatment of *foregone conclusion* in the English-German adjective-noun collocation dictionary

#### 2.3.4. The Phraseological Cline

As Gabrovšek (2003: 122) observes, most of the existing typologies of multi-word units recognise the fact that there are bound to be overlaps between the categories and that there are few discrete categories in the lexicon: things simply do not work like that (cf. Moon 1997: 48). This lends support to the notion of clines in lexical relations (cf. Bolinger 1976, as quoted by Wray and Perkins 2000: 5). Speaking in phraseological terms, one cannot say that a combination is either this or that, but rather that it is closer to one category than to another.

In practical lexicography, however, you scarcely have any other choice but to assign lexical relations to different classes. As already pointed out above, different dictionaries have different policies, and as long as they are applied consistently and spelled out in the user’s guide, nothing can be said against it, arbitrary as some groupings might seem.

### 2.3.5. Collocations in Other Linguistic Branches

As pointed out in the preceding sections, collocations are of major interest in (descriptive) phraseology and lexicography (notably Bahns 1996, Benson 1985, Cop 1990 and 1991, Cowie 1986, Gabrovšek 1998a, Hausmann 1989 and 1991, Svensén 1993 and Ter-Minasova 1992). Other areas which have recently started to pay particular attention to collocations are second language acquisition and pedagogy (Bahns 1993, Alexander 1987, Ellis 1997, Farghal and Obiedat 1995, Hoey 2000, Howarth 1998, Hussein 1990, Lontas 2002, Nattinger and DeCarico 1992 and Swan 1997), discourse analysis (Stubbs 1996 and 2001) and translation studies (Baker 1992, Gabrovšek 2000, Heliel 1990, Maurer-Stroh 2003 and Roos 1976).

With the advent of a new generation of computers and the ever-increasing storage and processing capacities collocations, these days, are generally looked at and dealt with computationally. What had started in the early 1970s with J. McH. Sinclair and S. Jones has turned into two interrelated disciplines<sup>21</sup>, corpus linguistics and computational linguistics.

## 2.4. Collocations and Corpus Linguistics

Formerly defined as 'the body of written or spoken material upon which a linguistic analysis is based'<sup>22</sup>, a corpus in its contemporary sense<sup>23</sup> is an electronically stored collection of text which is primarily used for linguistic analysis of any kind, be it for descriptive purposes, as a basis for lexicographical considerations or in search for forensic evidence, just to mention a few possibilities.

---

<sup>21</sup> Note that the question whether corpus and computational linguistics are disciplines in their own right or rather methodologies that can be applied to all linguistic branches is still a controversial matter in the linguistic community (cf. Borsley and Ingham 2003, Hunston 2002 and Meyer 2002). For the purpose of this thesis I shall see them as two methodologies developed to deal with collocations in a lexicographical framework.

<sup>22</sup> First mentioned in this sense in 1956 (source: OED).

<sup>23</sup> See Leech 2002 for further historical statements and a brief overview of corpus linguistics.

As Hunston (2002: 3) aptly puts it, “a corpus by itself can do nothing at all, being nothing other than a store of used language. Corpus access software, however, can rearrange that store so that observations of various kinds can be made.” Tools to generate frequency lists and other statistical data as well as concordance programs are essential to capture and analyse data from corpora<sup>24</sup>.

#### 2.4.1. Concordances

The vital feature of a concordance program is the KWIC (= key word in context) display. It shows the search term at the centre of a line that also contains the immediate context to the left and to the right of the search term (node). Such a list enables the linguist to look at recurrent patterns and it gives hints at the usage of the node word, e.g. typical collocations into which the word enters.

haunt/NN
<p>days birdwatching. A favourite haunt is Cuber reservoir in the mountains does all his own shopping. Favourite haunt is Shephards Bush market. [p] 3. filmed before. [p] Wigmore Hall, haunt of London's musical insiders, Hotel, the Rivetta is a favourite haunt of Venetian gondoliers. Tourists Why should they. The place is the haunt of off-duty policemen and the on-off Sloane Square is a favourite haunt of those in search of decorative Club in central London -- a regular haunt of many Conservative mps -- bar on Laurier Street. It's a regular haunt of young artists. Most of them don't friendly ambience, and is a regular haunt of foreign journalists and Dorchester in Park Lane and once the haunt of gambling-obsessed Arab switched to Ko Phangan as it is a haunt of young tourists as well as New America. The cemetery has become the haunt of drug addicts, tramps and freaks, a high-tech boom zone and a favourite haunt of disillusioned Californians. [p] at the fashionable swl club, haunt of London's brighter young things Speaking from the inn, a favourite haunt of the Duke and Duchess of York And at Signor Zilli's in Soho - the haunt of Tom Cruise, Naomi Campbell, and the restaurant is a popular local haunt. Tirolean specialities abound, from</p>

Fig. 9: Concordance of the noun *haunt*

<sup>24</sup> A list of useful websites related to corpus (and computational) linguistics can be found in Appendix A. For a detailed list of available corpora refer to Meyer (2002: 142-50).

The KWIC concordance of *haunt*<sup>25</sup>, as a noun, clearly shows that it usually collocates with the adjectives *favourite* and *regular*. The recurrent grammatical patterning reveals that 'a place can be *a / the (ADJ) haunt + OF someone*'.

Looking at concordance lines is only one way of detecting collocations in a corpus (and it can be very time-consuming and inefficient indeed when the node under scrutiny is a very frequent word!). For more sophisticated searches, the corpus can be linguistically annotated<sup>26</sup>.

#### 2.4.2. Tagging

One method of annotation is part-of-speech tagging. With the help of a tagging program (in short: tagger) word-class labels<sup>27</sup> are automatically assigned to the words in the corpus, and such a tagged text would look like this<sup>28</sup>:

a_ART	thick <sub>1</sub> ADJ	rain_N	cloud_N	hung_VER	over_PRP	the_ART
town_N	. <sub>2</sub> PM	the_ART	weather_N	forecast_N	was_AUX	for_PRP
severe <sub>3</sub> ADJ	gales_N	. <sub>4</sub> PM				
<i>ADJ</i>	<i>adjective</i>			<i>PM</i>	<i>punctuation mark</i>	
<i>ART</i>	<i>article</i>			<i>PRP</i>	<i>preposition</i>	
<i>AUX</i>	<i>auxiliary verb</i>			<i>VER</i>	<i>verb</i>	
<i>N</i>	<i>noun</i>					

Fig. 10: Example of a tagged text

<sup>25</sup> This concordance is an excerpt only and was generated by the free concordance and collocation sampler of Collins Wordbanks at <http://www.collins.co.uk/Corpus/CorpusSearch.aspx> [10/10/2004].

<sup>26</sup> Whether or not an annotated corpus is better than 'raw' text is still a controversial issue.

<sup>27</sup> Note that so-called 'tag-sets' vary considerably in different research centres.

<sup>28</sup> Those interested in an overview of different tagging methods should have a look at Linda Van Guilder's summary at [http://www.georgetown.edu/faculty/balcl/ling361/tagging\\_overview.html](http://www.georgetown.edu/faculty/balcl/ling361/tagging_overview.html); Exemplarily, you can try tagging your own text online at <http://www.comp.lancs.ac.uk/ucrel/claws/trial.html> (English) and <http://www.coli.uni-st.de/~thorsten/tnt/> (English and German) [all 10/10/2004].

Tagged corpora allow for concordances of a specific part of speech of a node which can take on different word classes. In Fig. 3, for example, I used such a tagged corpus and explicitly searched for instances of *haunt* as a noun.

### 2.4.3. Parsing

Another annotation method is parsing. In a nutshell, parsing is automatic syntactic analysis in which sentences are broken down into their constituents, i.e. words and phrases. With the help of parsed corpora, so-called 'treebanks',<sup>29</sup>, structures in any given language, can easily be identified.

As Fig. 6 shows, parsed corpora are also valuable as regards collocation analysis. In the two nominal phrases we can detect two adjective-noun collocations, viz. *laboratory assistant* and *brutal murder*, and one verb-noun collocation in the verbal phrase, viz. *to commit a murder*.

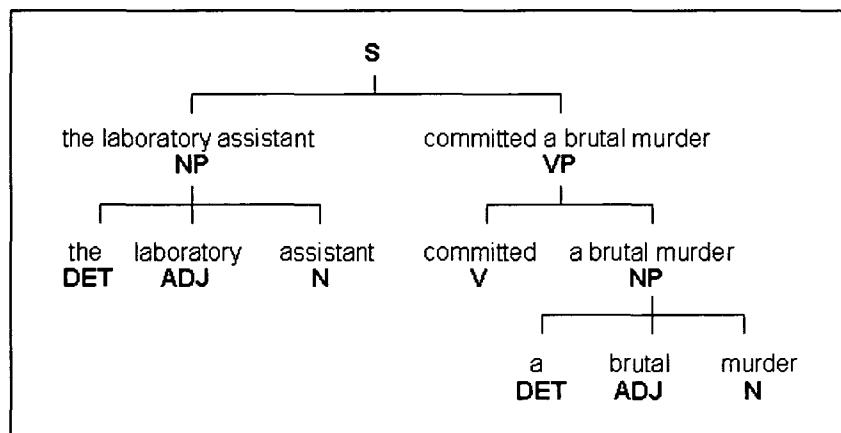


Fig. 11: Example of a parsed sentence

<sup>29</sup> Treebanks – this term was coined by Geoffrey Leech – are still very rare, the most important ones for English are the *Susanne Corpus*, the *Penn Treebank* and the *Lancaster Parsed Corpus*, while the most prominent German project is *TigerSearch* conducted at the IMS Stuttgart.

#### 2.4.4. Corpus Linguistics, Collocations and Areas of Application

In the following I will briefly outline two linguistic branches in which corpus linguistics has started to play a major role. I will restrict myself to areas which are associated with the topic of this thesis: lexicography and translation<sup>30</sup>.

##### (1) Lexicography

Although Sinclair may well be seen as the father of computerised lexicography in the corpus linguistic sense, the first use of computers in the dictionary-making process dates from the early 1970s – when the compilation of the third edition of the *Oxford Advanced Learner's Dictionary* included a computerised stage to turn the text into a print book (Cowie 1999: 118).

Sinclair's name is inseparably associated with the COBUILD<sup>31</sup> project which pioneered the corpus-based approach to produce widely acclaimed English monolingual dictionary for advanced learners, first published in 1987. They used computers to compile a corpus especially designed for lexicography<sup>32</sup>, applied the corpus frequency list as one of the criteria for the headword list, analysed concordance lines for sense discrimination and included 'authentic' example sentences in their dictionary (cf. Sinclair 1987b). Since then, every major (new) dictionary project followed their example and started using corpora, in one way or another, in the compilation process.

Sue Atkins broke new ground in 1994 when working on the new *Oxford-Hachette French Dictionary*. Her team used two comparable (in terms of size and contents) corpora, one English and one French, in order to produce their bilingual dictionary.

---

<sup>30</sup> Very good (introductory) textbooks on corpus linguistics and its applications are, inter alia, Hunston 2002, Kennedy 1998 and Meyer 2002.

<sup>31</sup> Collins Birmingham University International Language Database.

<sup>32</sup> The corpus consisted of over seven million words of written and spoken British English (cf. Cowie 1999: 118).

Although collocation dictionaries existed before (cf. 1.3.), the first (monolingual) collocation dictionary entirely based on an electronic corpus was G. Kjellmer's *A Dictionary of English Collocations*.

This dictionary comes in three large volumes and was published in 1994. What makes it problematic as a collocation source, however, is that their policy included any sequence of words that was grammatically correct and came up at least once in the corpus<sup>33</sup> was included, thus comprising also combinations like *the diamond, did I, did all, did not look at him, be retrieved from the dictionary*.

## (2) Translation

In her introductory textbook on corpus linguistics, Hunston (2002: 123) states that the use of corpora in translation and translation studies is becoming more and more important. Indeed, corpora can be a useful reference tool for translators.

Monolingual corpora can be used like monolingual dictionaries, that is, the translator knows the target language (TL) equivalent of the source language (SL) term. The corpus can then be searched for the phraseological behaviour of the term by looking at the concordance lines.

Imagine the following scenario: You are a German translator and have to translate *ein Telefonat führen* (literally *to conduct a phone call*) into English, but you are not sure about the verb. You search an English monolingual corpus for a concordance of *phone call*.

With an output as shown in Fig. 12, you will surely be able to find the correct translation equivalent of the phrase.

---

<sup>33</sup> Kjellmer's dictionary is based on the one-million-word Brown Corpus of American English.

the front door and <u>make</u> a	phone call	dialling with a pencil
or a drought by <u>making</u> a	phone call	. His rail competitor had to
ay night he had received a	phone call	from their daughter, Zenani,
is. Finally we got another	phone call	. [p] Mr. Hart, we're worried
[p] Yeah, don't <u>make</u> silly	phone calls	. Maybe they have too much in-
attending meetings, <u>making</u>	phone calls	, writing reports, or list
gossiping, <u>making</u> private	phone-calls	, doing personal jobs
of Paul's given to <u>making</u>	phone-calls	to his wife and mistresses
. Then we started getting	phone calls	that there was an invasion
round in my nightie <u>making</u>	phone-calls	and coffee and occasionally
loved to watch TV and <u>make</u>	phone calls	, Boffins at the French

Fig. 12: Concordance of *phone call*\*

While monolingual corpora are a valuable help in the translation process, the typical corpus in translation research is a parallel corpus consisting of two subcorpora, one containing texts originally written in the SL and the other one containing the translation of those texts into the TL.

These parallel corpora have to be aligned – at least at sentence level (so that sentence 22 in the SL sub-corpus is equivalent to sentence 22 in the TL sub-corpus) – in order to run parallel concordance programs, such as *ParaConc*<sup>34</sup>, on the corpus.

When the concordancer searches through the text, the only information the program has about the links between the different languages is the alignment. No use is made of bilingual dictionaries or of any kind of language-particular information. (Barlow 2002: 1).

In an earlier study of mine (Maurer-Stroh 2003), I compiled a parallel corpus of German and English to demonstrate the use of this type of corpus for translation. The corpus consists of original German tourism texts and their translations into English, together amounting to 11,004 words (tokens).

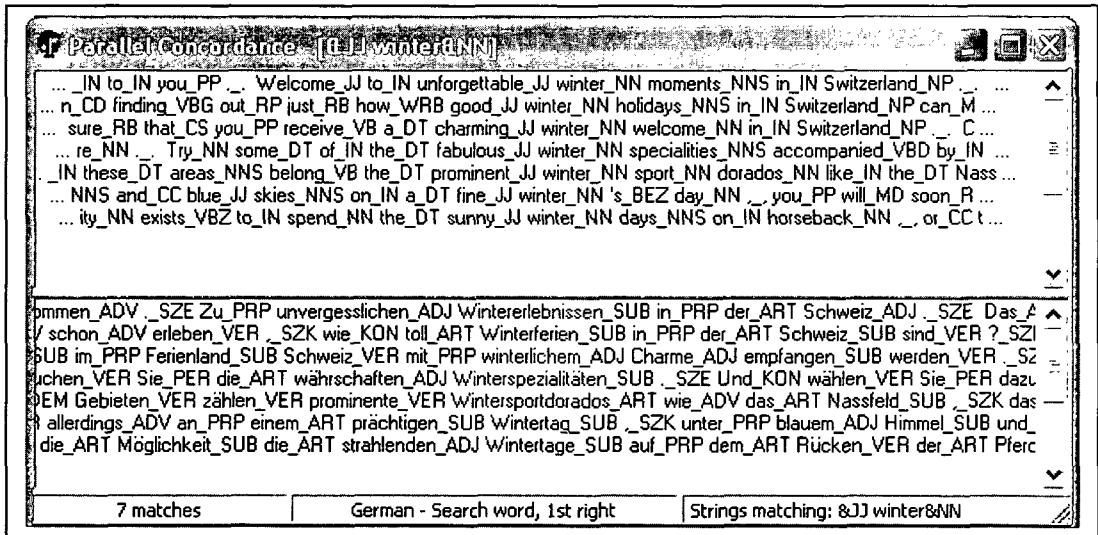
The translator can use the parallel corpus as a bilingual dictionary in terms of looking for word combinations. If, for example, the translator wants to know which adjectives go with *winter* (as noun or part of a composite noun compound), *ParaConc* run on my (manually) tagged corpus would produce the following useful information (Table 2):

<sup>34</sup> Michael Barlow's *ParaConc* can be purchased from <http://www.athel.com/para.html> [10/10/2004].

German	English
unvergessliche Wintererlebnisse	unforgettable winter moments
tolle Winterferien	good winter holidays
währschaften Winterspezialitäten	fabulous winter specialities
prominente Wintersportdorados	prominent winter sport dorados
prächtiger Wintertag	fine winter's day
strahlende Wintertage	sunny winter days

Table 2: Results of the bilingual extraction of word pairs

The bilingual adjective + noun (phrase) structures as seen in Table 2 can be obtained by analysing the parallel concordance lines resulting from the query ‘retrieve all occurrences of *winter* as a noun which are preceded by an adjective from the English sub-corpus together with their German equivalent structures in the German sub-corpus; KWIC-sort words in German which have *winter* as one of their components’.

Fig. 13: Parallel concordance of *winter*

Hence we can see that parallel corpora can come in very handy in practical translation, translation study and contrastive linguistics. However, there are relatively few corpora of this type available<sup>35</sup>.

<sup>35</sup> Secondly, (language-independent) taggers still lack the desired accuracy for more sophisticated queries.

One of the major drawbacks to using monolingual reference corpora in practical translation is that translators, in general, do not have time to study and analyse all these concordance lines (online), nor are there enough suitable (in size, design and language) corpora available to meet the translators' needs, e.g. collocation retrieval requires a huge collection of data as well as appropriate software packages – which are normally not available free of charge.

A less complicated method for translation-related information retrieval is to use the World Wide Web (= WWW) as a corpus: “A haphazard accumulation of machine-readable texts, the World Wide Web is unparalleled for quantity, diversity and topicality. This ever-expanding body of documents now encompasses at least 10 billion ( $10^9$ ) webpages publicly available via links, with several times that number in the “hidden” Web accessible only through database queries or passwords.” (Fletcher 2004: 2).

If, for example, the translator is not sure whether the adjective *fastidious* goes with the preposition *about* or *on*, the easiest way to find out via the WWW is using a search engine (e.g. Google) and querying “fastidious about” as opposed to “fastidious on”. The number of hits will show which construction is more frequent, e.g. *fastidious about* = 15,900 times vs. *fastidious on* = 214 times<sup>36</sup>.

Since Jones and Sinclair in the early 1970s, corpus linguistics, in both the monolingual and bi- or multilingual frameworks, has moved towards statistical methods for the extraction of collocations. These methods can be summarised within a section on computational linguistics.

---

<sup>36</sup> These figures are from 13/11/04. In my 2003 paper the same query resulted in 4,430 “about” and 18 “on” cases!

## 2.5. Collocations and Computational Linguistics

The projects in which computational methods are used for collocation analysis and processing are almost countless. In this section, therefore, I will briefly outline only some historical aspects and provide terminological preliminaries while also focusing on projects which are of immediate interest to my research.

As summarised in McKeown and Radev (2000), early work on statistical extraction of collocations from corpora was done by Choueka et al. in 1983. Their team used raw frequencies to identify typical sequences of two to six adjacent words. Their work pioneered the statistical approach to collocations in as much as they used a (for that time) relatively large electronic corpus, one consisting of 11 million words from texts of the *New York Times*. The drawback of their method, however, was that because of the use of raw frequency only, the results were highly dependent on the corpus size.

Another disadvantage of a list of raw frequencies, like the one in the Choueka et al.'s approach, is that it is impossible to attach a degree of relative significance to the combinations.

Accordingly, Church et al. (1991) used a correlation-based statistical method to extract collocations from corpora. They saw collocations as two (not necessarily adjacent) words that appeared together more often than would be expected by chance. Their significance measure was mutual information (MI)<sup>37</sup>, which, very generally speaking, compares the actual co-occurrence of the two words with their expected co-occurrence if the words in the corpus were to occur in a totally random order. In other words, the MI-score indicates the strength of a collocation (Hunston 2002: 71).

Although Church et al.'s method was less sensitive to corpus size, it was restricted to extracting only collocations consisting of two words.

---

<sup>37</sup> Another measure of significance, usually found in the literature, is the t-score (or t-test). The t-score uses "a calculation of standard deviation which takes into account the probability of co-occurrence of the node and its collocate and the number of tokens in the designated span in all lines." (Hunston 2002: 70). In short, the t-score measures the certainty of a collocation (*ibid*: 73). Yet other statistical formulas are z-score, log-likelihood ratio, chi-square, cubic association ratio, common birthday, etc. For a detailed overview see Manning and Schütze 1999 and Oakes 1998.

Frank Smadja (1993) and his team developed a system which not only uses significant measures (here, the z-score), but also draw on filters based on linguistic properties, such as syntax and semantics. Smadja had the results of the different stages of their tool *Xtract* tested by a lexicographer and this evaluation clearly showed the necessity of combining statistical and linguistic information, in syntax in particular.

On the basis of *Xtract*, Smadja et al. (1996) developed a bilingual collocation extractor called *Champollion*<sup>38</sup>. *Champollion* uses a statistical method to translate collocations between French and English using the Canadian Hansard Corpus (which contains bilingual reports from the proceedings of the Canadian parliament). *Champollion*'s output is a bilingual list of collocations which can be implemented in machine translation systems.

More recent approaches of bilingual collocation extraction<sup>39</sup> include the *DECIDE* project<sup>40</sup> which ran in 1994-1996 and drew on data from raw and tagged corpora as well as bilingual dictionaries.

Another such project was *KoKs* (= Korpusbasierte Kollokationssuche, viz. corpus-based collocation search)<sup>41</sup>, which was carried out between 2000 and 2002. The aim of *KoKs* was to build a lexicon for collocations from parallel German and English corpora and existing bilingual dictionaries.

There is yet another project which I would like to present, the *ANCR* (= The Adjective-Noun Collocation Retriever). *ANCR* developed from a discussion between my brother Dr. Sebastian Maurer-Stroh, then at the Institute of Molecular Pathology in Vienna,

---

<sup>38</sup> Jean-François Champollion (1790 – 1832), a French egyptologist, deciphered the famous Rosetta stone. The Rosetta stone is, amongst corpus linguists, usually referred to as the first 'parallel corpus' since the writings on it were carved in two languages (Egyptian and Greek) using three different scripts (hieroglyphic, demotic and Greek).

<sup>39</sup> Prominent researchers in bilingual computational and corpus linguistics are, inter alia, Thierry Fontenelle, Philipp Koehn, Franz-Josef Och, Wolfgang Teubert and Jean Véronis.

<sup>40</sup> Details about *DECIDE* can be found on their website at <http://engdep1.philo.ulg.ac.be/decide/> [10/10/2004].

<sup>41</sup> The homepage of *KoKs* is available at [http://www.cogsci.uni-osnabrueck.de/~koks/main/index\\_e.html](http://www.cogsci.uni-osnabrueck.de/~koks/main/index_e.html) [10/10/2004].

and myself. The project was launched as a joint undertaking of different departments at the University of Klagenfurt in summer 2003. The aim of *ANCR* is to automatically retrieve German and English adjective-noun collocations from a parallel corpus without having to linguistically pre-process or annotate the two corpora. The bilingual collocation retrieval is based on probability counts and uses several filters, e.g. German capitalisation (Fliedl et al. 2004). Our method and evaluation of the tool will be discussed in more detail in chapter 5.

Statistical extraction of German and English collocations is also done within the framework of Projekt Deutscher Wortschatz. This project does not, however, combine the two languages but offers online collocation tools for both languages<sup>42</sup>.

Without wishing to belittle major achievements in (bilingual) computational linguistics / collocation research, let me conclude this chapter by a quotation which should make us remember that we are dealing with ‘language’ not ‘strings of characters’ before we turn to collocations in the contrastive light.

“Phraseological significance means something more complex and possibly less tangible than what any computer algorithm can reveal.”

(P.A. Howarth 1998: 27)

---

<sup>42</sup> See their web page at <http://wortschatz.uni-leipzig.de/> [10/10/2004]

### 3. Contrastivity

As already pointed out in section 1.2., collocations are largely language-specific, i.e. interlingually, “the structural and semantic ‘images’ of such phrases will often be different” (Gabrovšek 2000: 212). While in English, a person who smokes a lot is referred to as *heavy smoker*, in German we refer to them as *starker Raucher* (lit. trans. *strong smoker*). Or, the English *main entrance* is rendered into German as an “authentic” compound (cf. 2.3.3.), *Haupteingang*, thus being incongruent in structure, yet semantically 100% predictable, since *main* = *Haupt* and *entrance* = *Eingang*. However, there are also semantically predictable and structurally congruent translation equivalents to be found in the language pair English and German, e.g. *rancid butter* vs. *ranzige Butter* or *rich harvest* vs. *reiche Ernte*.

Depending on (in-)congruence and/or (un-)predictability of the combinations in the interlingual context, “one should really try to work out a more detailed contrastive ‘difficulty scale’” (Gabrovšek 2000: 214), since “there have been so far few attempts at identifying or categorizing them.” (ibid: 215).

But first things first. While collocations in the monolingual framework are nowadays usually seen as a statistical, i.e. frequency-based, concept (cf. Chapter 2), the analysis and treatment of collocations in the bilingual framework call for criteria other than sheer numbers.

Contrasting two languages is like dealing with two sides of the same coin. On the one side you decode (= comprehend) a foreign text (written or spoken), while on the other side you encode (= produce) a text in a foreign language. It is particularly in the encoding process that problems can arise since speakers tend to rely on the hypothesis of transferability (cf. section 1.2.).

Most collocational errors are due to L1 interference, where the speaker simply translates an L1 collocation word-by-word into the L2 (cf. 1.2.). This is a consequence of the native speakers’ unawareness of the arbitrary, i.e. not semantically but rather usage-based, nature of collocations in their mother tongue (cf. 2.3.). However, the fact that

there are indeed – interlingually – collocations that are ‘exact’ word-by-word translations (e.g. *starless night* vs. *sternlose Nacht*) underscores the need for a special treatment of collocations in the contrastive light.

For the purpose of this thesis, I shall see the term contrastivity as an umbrella term for structural (in-)congruence and semantic (un-)predictability (cf. Gabrovšek 2000: 214-17 and 223).

### 3.1. Predictable and Structurally Congruent Combinations

Semantically predictable and structurally congruent collocations are in principle the easiest-to-encode category of interlingual collocations. They account for the fact that, although the collocational ranges of L1 and L2 items might differ, there are certain overlaps to be found which result in direct, i.e. word-by-word, translation equivalents (cf. Gabrovšek 1998b: 129 and Heliel 1990: 131), e.g. *wet nappy* vs. *nasse Windel* and *heavy scent* vs. *schwerer Duft*. In addition to their semantic predictability, they reveal the same structure, namely adjective + noun, and belong to the same category, i.e. collocation (as defined in section 2.3.).

Some linguists argue that in the contrastive pedagogical and lexicographical framework these combinations should play a minor role as compared to ‘more difficult-to-encode combinations’ like unpredictable and structurally incongruent ones (cf. Bahns 1993 and Gabrovšek 2000: 212-18 and 222-23).

The problem with this view is where to draw the dividing line in practical lexicography, especially also with respect to the notion of ‘reassurance’ on the part of the dictionary user. Let me illustrate this with an example.

While *wet nappy* and *nasse Windel* have semantically predictable constituents and are structurally congruent, how can the English speaker be sure that the somewhat semantically tailored sense of *wet* as ‘soaked with urine’ is the same in German, i.e. that the core sense of the adjective *nass* can be tailored to the noun *Windel*.

To overcome this problem, in view of my English-German adjective-noun collocation dictionary, I base the selection of included adjective (and attributively-used-noun) collocators not only on monolingual dictionaries and corpora, but also on existing bilingual dictionaries. Consequently, the collocations listed include those that are important, i.e. frequent, on the English monolingual side but that do not show difficulties in the bilingual light since they are predictable and congruent, e.g. *important discovery* vs. *wichtige Entdeckung*. Furthermore, contrastive collocations that show minor monolingual significance, i.e. that are less frequent than other items in the collocational range in question, but are of importance from the contrastive angle, will likewise be included, e.g. *on an equal basis* vs. *auf gleicher Basis* or *als [gleichgestellte(r)] Partner*. The inclusion policy is entirely based on the sources mentioned.

### 3.2. Predictable and Structurally Incongruent Combinations

This class comprises contrastive combinations like *main entrance* vs. *Haupteingang* and *prime minister* vs. *Premierminister*, where both constituents of the combination are 100% predictable, but the combinations as such exhibit a different structure (and what one frequently finds are compounds on the German side). Other examples of predictable but structurally incongruent combinations are *red pencil* vs. *Rotstift* and *used car* vs. *Gebrauchtwagen*. Combinations of this type need special attention since they can also be rendered into German as structurally congruent translations, viz. *roter Stift* and *gebrauchtes Auto*<sup>1</sup>.

I deliberately use the term ‘combination’ here rather than ‘collocation’ since it is arguable whether *prime minister*, *main entrance*, *used car* and *red pencil* are collocations, or, for the sake of institutionalisation or something else, compounds. As pointed out in section 2.3.3., for my English-German adjective noun collocation dictionary it matters very little which phraseological category the four combinations

---

<sup>1</sup> *Auto* and *Wagen* are both translation equivalents of *car*; *Auto* is more frequently used with adjectives, while *Wagen* predominates in compounding.

belong to, because the fact that they are all rendered into German with a different structure makes them perfect candidates for inclusion (see Fig. 1 below).

<b>car</b>
used gebrauchtes Auto   etc.
▪ <b>used car</b> Gebrauchtwagen m etc.
<b>entrance</b>
▪ <b>main entrance</b> Haupteingang m etc.
<b>minister</b>
▪ <b>prime minister</b> Premierminister m etc.
<b>pencil</b>
red roter Stift   etc.
▪ <b>red pencil</b> Rotstift m etc.

Fig. 14: Examples of predictable and structurally incongruent combination and their treatment in the English-German adjective-noun collocation dictionary

### 3.3. Unpredictable and Structurally Congruent Combinations

Examples of unpredictable yet structurally congruent combinations in English and German include *heavy smoker* vs. *starker Raucher* (lit. trans. *strong smoker*), *barefaced lie* vs. *faustdicke Lüge* (lit. trans. *a lie as thick as a fist*) and *magic carpet* vs. *fliegender Teppich* (lit. trans. *flying carpet*).

Note that the unpredictable element in adjective-noun combinations is almost exclusively the adjective. One of the few exceptions is (the idiom or compound) *green fingers* – which is rendered into German as *grüner Daumen* (lit. trans. *green thumb*).

### 3.4. Unpredictable and Structurally Incongruent Combinations

Unpredictable and structurally incongruent collocations are a class which needs special attention in the contrastive analysis of collocations. According to Bahns (1997: 108) there is no such thing as 100% unpredictability and structural incongruence in contrastive collocations since the fulfilment of the two criteria would make them fall under the category of idioms.

As part of the material on which my dictionary entries are based is taken from English collocation dictionaries and general bilingual ones, the question where to draw the dividing line between collocations and idioms (and semi-idioms) is not an issue since I include those combinations which are recorded in the dictionaries in question and show corpus evidence. Consequently, I include the English *on a first-come first-served basis*, although it is translated into German as the idiomatic expression *nach dem Prinzip 'Wer zuerst kommt, mahlt zuerst'* (lit. trans. *on the principle that 'Who comes first, grinds first'*; cf. the English proverb *the early bird catches the worm*).

Another such example that would be included in the English-German adjective-noun collocation dictionary is *to have no sound basis* which can be rendered into German as the figurative expression *auf wackeligen Füßen stehen* (lit. trans. *to be a bit shaky on one's feet*).

Likewise, the English semi-idiom<sup>2</sup> *in broad daylight* would be included. Its German equivalent is *am helllichten Tag(e)*: We cannot possibly equate *broad* with *helllichten* (which is, above all, only used in the combination with *Tag(e)* and does not exist elsewhere), nor can we relate *daylight* (lit. trans. *Tageslicht*) to *Tag* (lit. trans. *day*).

It goes without saying that those instances of idiomatic, highly restricted or figurative uses are meta-lexicographically labelled in the entries.

---

<sup>2</sup> The expression is idiomatic in the sense that the adjective *broad* is only used as a kind of intensifier in this combination (Gabrovšek 2004: private communication).

### 3.5. Case Studies

In the following, two case studies will demonstrate the need for a bilingual collocation dictionary.

#### 3.5.1. English Collocations and Their Translations in Bilingual Dictionaries

By way of investigating the provision of adjective-noun collocations in existing bilingual dictionaries, the first case study is designed, first, to show that there is indeed a need for an English-German adjective-noun collocation dictionary since translating requires information on collocational possibilities, and in the absence of such information, the translator or learner has to rely on dictionaries (Heliel 1990: 134). Second, this case study examines whether bilingual dictionaries of German and English use any contrastive filters in their inclusion policies.

I took 18 random adjective-noun collocations from each of the three English collocation dictionaries, the BBI, LTP and OCD. From the resulting 54 collocations 14 were found in all three English collocation dictionaries; these are: *irresistible attraction, unshakeable belief, insurmountable difficulty, strenuous exercise, invaluable help, complete ignorance, false impression, pressing matter, selfish motive, unsolved mystery, rough outline, enormous pressure, sweeping reform and deep respect*. Another 15 were recorded in two of the collocation dictionaries in question; these include: *established author, gentle breeze, insatiable curiosity, nice guy, vivid illustration, overwhelming joy, grand opera, doubtful origin, utter perfection, abandoned property, close shave, crisp snow, stubborn stain, distorted vision and sharp wit*.

I checked these 29 collocations in the four major comprehensive bilingual dictionaries of English and German: DUOX, COGER, PONS and LANGE. The main criterion was that the collocation is mentioned, i.e. given explicitly.

As can be seen in Table 1, the best result is achieved with PONS. 51.7% of the tested collocations were found in either the adjective or the noun entry. Making use of all of

the bilingual dictionaries in the translation task would result in 79.3% explicitly mentioned collocations.

Six collocations out of the 29, however, are not recorded in any of the dictionaries in the study<sup>3</sup>. These are *insurmountable difficulty*, *pressing matter*, *utter perfection*, *enormous pressure*, *deep respect* and *distorted vision*. The reason why they are not listed in the dictionaries could be that they are all semantically predictable (i.e. the most core sense of the adjective can be translated predictably into German) and structurally congruent.

COLLOCATIONS	OCD	BBI	LTP	DUOX	COGER	PONS	LANGE
established	author	-	x	x	ADJ	-	-
unshakeable	belief	x	x	x	-	-	ADJ
gentle	breeze	x	x	-	ADJ	ADJ	ADJ
insatiable	curiosity	x	-	x	ADJ	ADJ	-
insurmountable	difficulty	x	x	x	-	-	-
strenuous	exercise	x	x	x	-	-	ADJ
nice	guy	x	x	-	-	ADJ	ADJ
invaluable	help	x	x	x	ADJ	ADJ	ADJ
complete	ignorance	x	x	x	ADJ	-	-
vivid	illustration	x	-	x	ADJ	-	-
false	impression	x	x	x	-	-	ADJ
overwhelming	joy	x	x	-	-	-	ADJ
pressing	matter	x	x	x	-	-	-
selfish	motive	x	x	x	-	-	ADJ
unsolved	mystery	x	x	x	-	ADJ	ADJ
grand	opera	x	x	-	HW	HW	HW
doubtful	origin	x	-	x	ADJ	-	-
rough	outline	x	x	x	ADJ	-	-
utter	perfection	x	-	x	-	-	-
enormous	pressure	x	x	x	-	-	-
abandoned	property	-	x	x	ADJ	-	ADJ
sweeping	reform	x	x	x	ADJ	-	N
deep	respect	x	x	x	-	-	-
close	shave	x	x	-	N, ADJ	N	N, ADJ
crisp	snow	x	x	-	ADJ	ADJ	ADJ
stubborn	stain	x	x	-	-	ADJ	ADJ
distorted	vision	x	-	x	-	-	-
sharp	wit	x	x	-	ADJ	ADJ	-
<b>COLLOCATIONS FOUND</b>				<b>14</b>	<b>10</b>	<b>15</b>	<b>1</b>
<b>PERCENTAGE</b>				<b>48.3%</b>	<b>34.5%</b>	<b>51.7%</b>	<b>3.4%</b>

Table 3: English collocations and their distribution in English-German general bilingual dictionaries

Interestingly, only one collocation can be found in all bilingual dictionaries, namely *close shave*. DUOX, COGER and PONS all record both the literal and the figurative use

<sup>3</sup> Note that I used the print dictionaries for this case study, so no full-text search was possible.

of the combination while LANGE only lists the figurative sense. In its literal sense, *close shave* has to be rendered into German as the “authentic” compound *Glattrasur* (lit. trans. *smooth shave*). It is thus unpredictable (due to the adjective) and structurally incongruent. The figurative reading of *close shave*, namely ‘a situation in which you only just avoid an accident, a disaster, etc.’, is 100% unpredictable and structurally incongruent in German:

I had a close shave.	Ich bin gerade noch / mit knapper Not davongekommen.
That was a close shave!	Das war knapp!

Similarly, four collocations are recorded in three of the four bilingual dictionaries I tested: *gentle breeze*, *invaluable help*, *grand opera* and *crisp snow*. While *invaluable help* can be translated into German word-by-word, the meaning of the adjective *crisp* is tailored to the noun *snow*. This makes the collocation, although structurally congruent, semantically unpredictable since the translator needs to know with which word the specialised sense of the adjective *crisp* is rendered into German, the direct translation equivalent of the most core sense of *crisp* (viz. *knusprig*) or a more specialised adjective that attributes the property of *crisp* to *snow* (viz. *verharscht*).

The combination *grand opera* is another interesting example. Although its German equivalent, *große Oper*, is 100% predictable and structurally congruent, the combination as such can be seen as a semi-idiom (or an endocentric compound of course) in as much as the adjectives *grand* / *groß* have the meaning ‘only sung, no spoken parts’ only in combination with *opera*. Furthermore, DUOX, COGER and PONS all assign headword status to the combination.

The collocation *gentle breeze* can be translated into German as either a predictable and congruent collocation, viz. *sanfte* / *leichte* / *schwache Brise*, or as a structurally incongruent and unpredictable German simplex word, viz. *Lüftchen*.

Summarising, if we take minimum occurrence in three of the four dictionaries as our threshold, collocations in present-day bilingual dictionaries are included whether they are predictable and congruent or unpredictable and congruent or unpredictable and

incongruent (see Table 2), so there seem to have been no specific contrastive filters implemented.

predictable / congruent	(3) <i>invaluable help, grand opera, gentle breeze</i>
unpredictable / congruent	(3) <i>crisp snow</i>
unpredictable / incongruent	(4) <i>close shave, close shave, (3) gentle breeze</i>

Table 4: Inclusion policy of contrastive collocation categories in present-day bilingual dictionaries

However, if this policy had been consistently applied, also the six predictable and congruent collocations which are not recorded in any of the bilingual dictionaries I tested would have been listed (*insurmountable difficulty, pressing matter, utter perfection, enormous pressure, deep respect* and *distorted vision*).

At this point I should mention once again that the English-German adjective-noun collocation dictionary will consistently include all four contrastive categories (cf. sections 3.1. to 3.4.). The criteria I use are monolingual corpus frequency as well as monolingual collocation dictionary data combined with full-text search in bilingual dictionaries.

### 3.5.2. English and German Adjective-Noun Combinations and Machine Translation

Machine Translation (hereafter MT) is (the process of) automatic translation from one natural language to another by a computer. Although patent applications of ‘translation machines’ started as early as the 1930s, commercial computer-assisted tools only became available first in the 1980s (cf. Gómez 2003).

Recently, the internet with its multilingual coverage fuelled a real boom in MT – which has been made possible by ever-better computational linguistic methods. While search

engines like *Google* or *Altavista*<sup>4</sup> offer automatic translation of whole web pages, there are also some pages where the user can type in their own texts and have them translated into various languages for free. The leading tools among them are *Free2ProfessionalTranslation* (PROF), *Promt-Online* (PROM), *Softissimo* (SOFT), and *Systran* (SYST)<sup>5</sup>.

So, how good are such MT tools when it comes to the automatic translation of adjective-noun collocations?

The second case study serves to illustrate that the automatic translation of collocations (and other word combinations in this respect) is the poor relation of collocational research in computational linguistics.

I randomly chose 30 English and 30 German adjective-noun combinations (including free combinations, endocentric compounds, collocations, semi-idioms, idioms and exocentric compounds) and evaluated the outputs of both the four online MT tools mentioned above (PROF, PROM, SOFT, SYST) and the commercial MT product from Langenscheidt, *Langenscheidt's T1* (Version 5.0).

The following adjective-noun combinations were translated from German into English:

*rote Blume, hölzerne Tür, gemütliches Sofa, silberner Knopf, offenes Fenster Rotstift, Schnellstraße, Gebrauchtwagen, (im Geschäft herrschte) Hochbetrieb, Geheimbund, schwere See, reiche Ernte, starker Raucher, schwere Last, haushoher Favorit, faustdicke Lüge, blinder Passagier, kalter Krieg, schwerer Schlag, trockener Wein, (etwas auf die lange Bank schieben), der kleine Mann, (jem.) grünes Licht (geben), kalte Füße (bekommen), (bei jem. ein) offenes Ohr (finden), Junggeselle, Schöngest, Blondschoß, Milchgesicht und (ein) Dickkopf (sein).*

The English adjective-noun collocations I tested in MT tools included the following:

*yellow flower, bus ticket, interesting book, poor actor, sunny day, prime minister, magic carpet, yellow pages, wet suit, general strike, stale bread, dry country, main entrance, heavy smoker, poor actor, naked truth, (see something with the) naked eye, white Christmas, (in) broad daylight, dry cow, (give the) red carpet, (appear in its) true colours / (show one's) true colours, (have) green fingers, full steam (ahead), smart money (is on somebody / something), wet blanket, musical chair, blackmail, lame duck, sitting duck.*

---

<sup>4</sup> Both search engines use *Systran* technology.

<sup>5</sup> The relevant links can be found in Appendix B.

To be able to judge which phraseological category is the 'easiest' for the MT tools to handle correctly; for starters, I only take into consideration combinations which have been properly translated by all five tools.

Regardless of the translation direction, 12 of the 60 combinations (20%) are correctly translated by all machines. These include five free combinations (41.67%), two endocentric compounds (16.67%), one collocation (8.33%), two semi-idioms (16.67%), no idioms (0%) and two exocentric compounds (16.67%): *bus ticket*, *sunny day*, *interesting book*, *rote Blume*, *offenes Fenster*, *Schnellstraße*, *Gebrauchtwagen*, *main entrance*, *kalter Krieg*, *trockener Wein*, *Junggeselle* and *blackmail*.

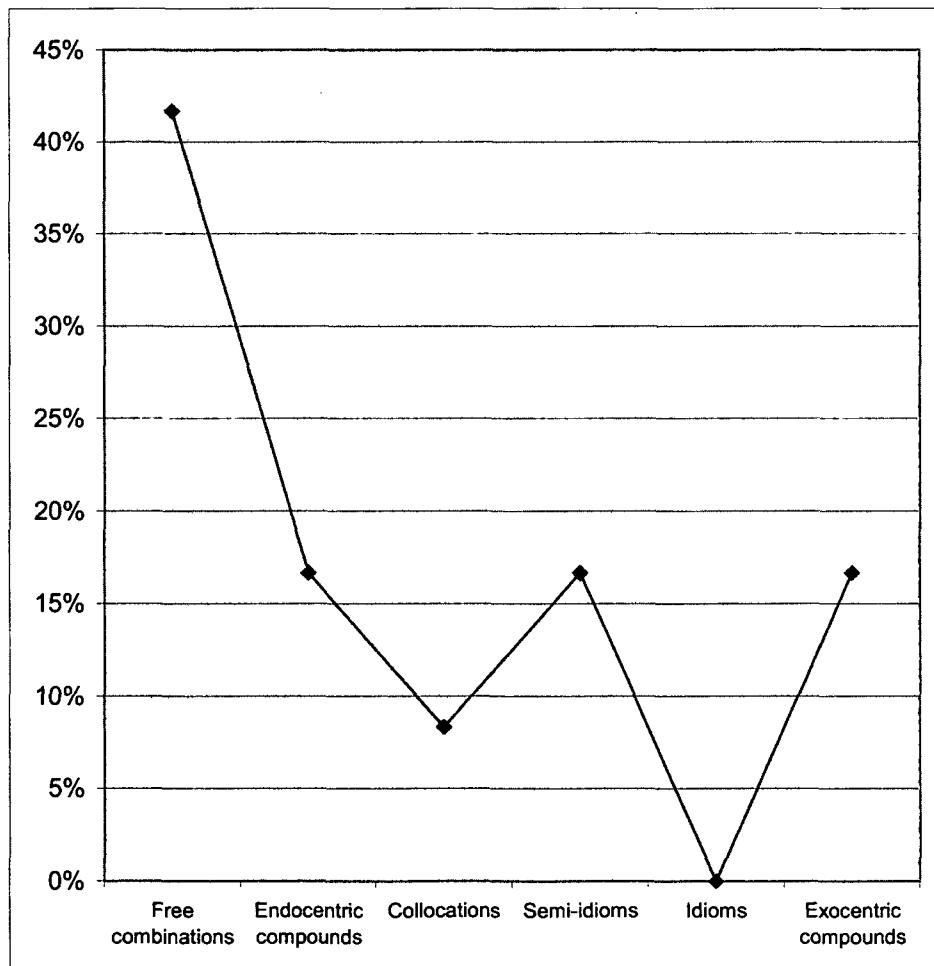


Fig. 15: Categories of correct translation of all five MT tools

When we look at Fig. 2, a certain trend can be observed, namely that the more 'restricted' combinations become semantically and structurally, the less likely are they to be correctly translated by the MT tools (free combinations – endocentric compounds – collocations – idioms). What is striking, however, is the unexpected results of semi-idioms and exocentric compounds in this graph.

The reason why exocentric compounds are more often correctly translated than collocations, as I see it, is, first, due to the fact that they are more striking, overall, semantically institutionalised than collocations. That means that in MT tools where existing bilingual lexicons and grammars are implemented, they can often be readily translated.

Secondly, what with statistical programming techniques, the fact that endocentric compounds are semantically unpredictable and, for the most part, structurally incongruent, i.e. that they are spelled as one word in German, makes them easier to deal with in MT than collocations. In other words, when looking at endocentric compounds from the contrastive angle, they are more salient in the computational view than are collocations – which the computer has difficulties distinguishing from free combinations.

Also the digression of semi-idioms can be explained by these two reasons. First, semi-idioms, having one highly restricted, viz. unpredictable, element are more likely to be found in existing bilingual dictionaries which are implemented in MT tools. Secondly, in statistical terms, the fact that contrastive semi-idioms are very unlikely to be word-by-word translations from one language into the other and that their combined significance with respect to their translations as opposed to the frequency of their constituents is rather high, makes them easier to deal with than collocations.

What is also interesting to observe from this case study is the following ranking of the MT tools<sup>6</sup>:

---

<sup>6</sup> Note that, in this study, the commercial product, T1, is only in the middle and SYST – which is used by many search engines – is next to last.

1. PROM (32 out of 60 correct = 53,33%)
2. SOFT (30 correct = 50%)
3. T1 (27 correct = 45%)
4. SYST (25 correct = 41,67%)
5. PROF (24 correct = 60%)

By way of concluding this chapter, let me point out that the results in these case studies only go to show that, despite the recent advances in computation and the increased awareness of phraseology in bilingual lexicography, there is still a lot to be done to fulfil users' needs.

The process of covering one part of phraseology, namely adjective-noun collocations, may well be smoothed out, in both computational and lexicographical terms, with the methods I put forward in Chapter 5.

## 4. Collocations and the Dictionary

As summarised in Hausmann (1989: 1011), the tradition of collocation dictionaries can be traced back as far as the 16<sup>th</sup> century. Towards the end of the Renaissance the main focus was on so-called epithets (adjectives or participles that premodify nouns); this type of collocation dictionaries only came to an end with the beginning of the 20<sup>th</sup> century – when modern language lexicographers introduced a new tradition of collocation dictionaries, then better named combinatorial dictionaries.

In 1900, Albert Heintze invented the term *Stilwörterbuch* (viz. dictionary of style) for a general monolingual dictionary with a major combinatorial component. The term was soon adopted, though with a slightly different meaning, by Albrecht Reum who published, *inter alia*, his *Dictionary of English Style* in 1931. Other ‘early’ collocation, viz. combinatorial, dictionaries of English include H.E. Palmer’s *A Grammar of English Words* (1938), J.I. Rodale’s *Word Finder* (1947) and Friederich and Canavan’s *Dictionary of English Words in Context* (1979).

In the following I will briefly discuss H.E. Palmer’s and Friederich and Canavan’s dictionaries – two representative early works from two different traditions.

### (1) *A Grammar of English Words* (GEW)

The title of *GEW* suggests that this dictionary is essentially a grammar<sup>1</sup> that lists, so the subtitle, “one thousand English words and their pronunciation, together with information concerning the several meanings of each word, its inflections and derivatives, and the collocations and phrases into which it enters.” (Palmer 1938).

As already pointed out in Chapter 2, H.E. Palmer was an English teacher in Japan and his research into vocabulary control recognised the need to focus on a core vocabulary

---

<sup>1</sup> Palmer himself said that the book might also be entitled *A Grammatical Dictionary of English Words* (1938: iii).

and the difficulties it presented, not so much in terms of the meaning of the words, but of their constructions (cf. Cowie 1999: 36).

In the introduction to *GEW* Palmer (1938: iv-v) points out that his work should fill the gap between traditional grammars and dictionaries<sup>2</sup>. While dictionaries at the time when *GEW* was published (viz. native-speaker monolingual ones) normally only exemplified the use of uncommon or rare expressions and left usage and sentence building of ‘normal’ words to grammarians, grammars only treated constructions which were of interest to the grammarians themselves, elaborating on theories and merely touching on lexis.

However, as Palmer saw it, of the 20,000 words which were in fairly current use in English, “about 1,000 present considerable difficulty to the foreign student of English; the remainder [...] present little or no difficulty.” (*ibid*: iii). Furthermore, he points out that it is in connection with these 1,000 words – which form the headwords of *GEW* – that a great majority of mistakes in grammar and composition are made, and that it is not knowing about the usage of these words that prevents the learner from producing correct English in speech and writing (*ibid*.)

Principally, Palmer (*ibid*: iv) suggested six categories of difficulty along which he decided on inclusion in *GEW*:

- (1) A word belongs to more than one part of speech, e.g. *for*, which belongs to the class of prepositions and the class of conjunctions.
- (2) A word has more than one meaning (sometimes the number is considerable), and each meaning may be represented with a different word in the student’s mother tongue, e.g. *to mean* (Ger. *beabsichtigen, meinen, bedeuten*, etc.)
- (3) A word enters more than one ‘sentence-pattern’, e.g. *wish for something, wish to do something, wish somebody to do something*, etc.
- (4) A word has several inflected forms and derivatives (esp. irregular ones), e.g. *just* in *(in)justice, justify, justification, (un)just(ly)*, etc.

---

<sup>2</sup> “It has been said that there is ‘a vast uncharted territory lying between the respective domains of the dictionary-maker and the grammarian’, a sort of no-man’s land in which reside the great majority of those points that perplex those to whom English is a foreign language.” (Palmer 1938: v)

(5) A word enters into a large number of collocations and phrases – and some of them again have more than one meaning, e.g. *all at once* (= ‘everyone at the same time’ / ‘suddenly’)

(6) A word is a component part of another, thus forming a compound whose overall meaning is sometimes not deducible from the constituents e.g. *blackboard*, etc.

What do entries in *GEW* look like? Fig. 16 is a scan from *GEW* and exemplifies the organisation of the dictionary’s microstructure (viz. internal organisation of the entries).

free	64	from
<p>3. = not busy, not occupied, at leisure Are you free? I shall not be free until five o’clock. He had no free time.</p> <p>4. = without restraint a free movement. be free to do sg. You are free to do what you like. free and easy</p> <p>5. = not costing anything The books were given away free. You need not pay—it is free. You may have free use of my house during the summer.</p> <p>Δ <b>freedom</b> ['frɪdəm], <i>n. Uncountable</i> He was given his freedom. freedom of thought.</p> <p>Δ <b>freely</b> ['frɪli], <i>adv.</i> speak and act freely.</p> <p><b>FRESH</b></p> <p><b>fresh</b> [freʃ], <b>fresher</b> ['freʃə], <b>freshest</b> ['freʃɪst], <i>adj.</i></p> <p>1. = newly produced or made or grown fresh flowers [eggs, bread]. fresh paint [still wet].</p> <p>2. = sound, clean, not spoilt This meat is not very fresh. Everything in the house still looks quite fresh.</p> <p>3. <i>Said of the weather</i> a fresh wind. fresh air.</p> <p>4. = novel I’d like to see something fresh. Is there any fresh news? make a fresh start.</p> <p>5. = not salt fresh water. fresh butter.</p> <p>Δ <b>freshly</b> ['frɛʃli], <i>adv.</i> freshly-caught fish.</p> <p>Δ <b>freshness</b> ['freʃnɪs], <i>n. Uncountable</i></p> <p><b>FRIEND</b></p> <p><b>friend</b> [frend], <b>friends</b> [frendz], <i>n.</i> He has a large number of friends. Where’s your friend? He’s a friend of mine. My friend Smith called yesterday. A boy [girl, man, woman] friend of ours. We are great [good] friends. One of my business friends. A friend of the poor. He has been a good friend to me. The dog is a friend of man. I make friends (with sy.) = become the friend of... I make friends again = become friendly after a quarrel or disagreement I be friends with sy. = be in the relation of a friend to... Δ <b>friendly</b> ['frendli], <b>friendlier</b> ['frendliə], <b>friendliest</b> ['frendliəst], <i>adj.</i> He spoke in a friendly way. The dog was quite friendly. He’s a friendly sort of fellow.</p>	64	<p>Δ <b>unfriendly</b> [ʌn'frendli], <i>adj.</i> Δ <b>friendship(s)</b> ['frendʃɪp(s)], <i>n.</i> How long will the friendship last? The friendship between England and France. feelings of friendship. friendship for [with] sy.</p> <p>¶ <b>in friendship</b></p> <p>Δ <b>friendliness</b> ['frendlɪnɪs], <i>n. Uncountable</i></p> <p>Δ <b>unfriendliness</b> [ʌn'frendlɪnɪs], <i>n. Uncountable</i></p> <p><b>FRIGHT(EN)</b></p> <p>I. <b>frighten</b> ['frایtn], <b>frightens</b> ['frایtnz], <b>frightened</b> ['frایtnɪd], <b>frightening</b> ['frایtnɪŋ], <i>v.</i> frighten sy. <i>See V.P.</i> The noise frightened me. I was very much frightened. Are you frightened by earthquakes? ¶ <b>be frightened of</b> Are you frightened of earthquakes? ¶ <b>frighten</b> <i>sy. into doing sg. See V.P.</i></p> <p>6</p> <p>¶ <b>frighten</b> <i>sy. out of doing sg. See V.P. 6</i></p> <p>¶ <b>be frightened out of one’s life</b> = be very frightened</p> <p>II. <b>fright</b> [frایt], <b>frights</b> [frایtnz], <i>n.</i> What was the cause of his fright? die of fright. get [have] a fright.</p> <p>¶ <b>give</b> <i>sy. a fright</i> = <b>frighten</b> <i>sy.</i></p> <p>¶ <b>take</b> <i>fright</i> <i>at sg.</i> = <b>be frightened</b> <i>at sg.</i></p> <p>III. <b>frightful</b> ['frایtful], <i>adj.</i> a frightful place [accident, etc.]. It was a frightful storm.</p> <p>Δ <b>frightfully</b> ['frایtfuli], <i>adv.</i> It was frightfully dangerous.</p> <p><b>FROM</b></p> <p><b>from</b> (from, fram), <i>prep.</i></p> <p>1. <i>Marking starting point, followed by point of arrival, expressed or understood.</i> go from London to Birmingham. carry sg. from one place to another. count from one to ten. from the beginning to the end. from one end to the other. from beginning to end [start to finish, morning to night, top to bottom, etc.].</p> <p>¶ <b>from hand to mouth</b></p> <p>2. <i>Marking point of origin</i> Start from here. Where does he come from? I have received a letter from a friend. We get these goods from foreign countries. Can you see it from here? From this point of view. It was hanging from a branch. I took it from under these books. He was looking at me from over his glasses.</p>

Fig. 16: Sample page from *GEW*

Palmer uses the term ‘caption words’ for what we normally refer to as headwords; they are printed in bold capitals, e.g. *fresh*, *friend*, *fright(en)*, *from*. He distinguishes caption words from so-called ‘working units’ which are specific realisations of caption words.

They are printed in bold and are followed by their pronunciation, their word-class label, and their inflected forms<sup>3</sup>, e.g. under the caption word *fright(en)* we find the three working units *frighten*, *fright* and *frightful* – where each working unit stands for one ‘grammatical function’.

Grammatical functions, which are ordered by frequency, are set out in separate paragraphs and are preceded by Roman numerals (I, II, etc.), while shifts of meaning (viz. different senses) are indicated by Arabic numerals (1, 2, etc.), e.g. *fresh* = 1. ‘newly produced or made or grown’, 2. ‘sound, clean, not spoilt’, 3. ‘said of the weather’, etc.

When working units are derivatives formed with regular affixes they are marked by  $\Delta$ , e.g. *friend* – *friendly* – *friendship* – *unfriendly* – *friendliness* – *unfriendliness*.

When a word (viz. working unit) forms an important element of a collocation, namely “a succession of two or more words that may best be learnt as if it were a single word” (Palmer 1938: x), the collocation is shown in bold type and preceded by the conventional sign  $\P$ , e.g. **fright** –  $\P$  give somebody a **fright**,  $\P$  take **fright**.

Palmer distinguishes phrases from collocations. “While collocations are comparable in meaning and function to ordinary single ‘words’ (and indeed are often translated by single words in the student’s mother-tongue), phrases are more in the nature of conversational formulas, sayings, proverbs, etc.” (ibid: xi). Phrases are marked *Phr.*, e.g. in the entry for *serve* we find *Phr. serve two masters*.

Square brackets in the entries indicate alternatives, e.g. *fresh flowers* [eggs, bread], while round brackets enclose optional omissions, e.g. *to make friends (with somebody)*.

---

<sup>3</sup> Note that irregular inflections are printed in bold, e.g. **take**, **takes**, **took**, **taken**, **taking**, **v.**

Patterns into which the working units usually enter are distinguished from examples by letter, e.g. *go from London to Birmingham* vs. *Where does he come from?*.

The ‘see V.P. + number’ device refers to Appendix A of *GEW* in which the 27 most important English verb patterns can be found, e.g.:

Verb-Pattern 1	verb	x	0	<i>Birds fly. I know.</i>
Verb-Pattern 4	verb	x	direct object	<i>I read the book.</i>
Verb-Pattern 19	verb	x	gerund	<i>He stops doing it.</i>

As I already touched upon it in Chapter 2, Palmer uses the term collocation in slightly different terms than current mainstream linguistics would do. He chose to name collocation the category of any ‘odd coming-together-of-words’ since, in the 1930s, collocation as a generic term had no settled status yet, and he refused to employ ‘idiom’ as a cover term for this category as, as he saw it, this would have meant broadening his scope to include also proverbs, sayings and figurative expression (cf. Cowie 1999: 54).

In *GEW* he explicitly states that “in this book no need has been found for the term idiom since what are usually called idioms are (a) collocations (b) phrases and sayings (c) rarer semantic varieties of words and collocations (d) peculiar construction patterns – in short, any form that is likely to puzzle a foreign student.” (Palmer 1938: xii).

There is yet other information to be found in *GEW* which is not mentioned in the user’s guide. First, compounds, like *sister ship*, *warship*, *merchant ship*, *shipbuilding*, *ship-owner* in the entry of the caption word *ship*, are marked by *Comp.* or *Comps..*

Furthermore, the use of *Fig.* as a label is occasionally made. However, the distinction between *Phr.* as in *No smoke without fire*. and *Fig.* as in *Fortune has always smiled on him*. is not made clear.

Summarising, let me say that *GEW*, although it is specialised in the sense that it only focuses on the core vocabulary of English, definitely constitutes a landmark in learner and combinatorial lexicography, but its success was overshadowed by the publication in 1942 of a general-purpose dictionary for foreign learners of English with rich

phraseology, namely A.S. Hornby, E.V. Gatenby and H. Wakefield's *Idiomatic and Syntactic English Dictionary* (ISED)<sup>4</sup>.

(2) *Dictionary of English Words in Context* (DEWC)

In the Preface to *DEWC*, Wolf Friederich (1979: i) addresses once again the need for learners of English to have a good command of fixed expressions and collocations:

“Wir wissen, daß der schwierigste Baustein zusammenhängender Texte nicht die Einzelwörter oder die Regeln der Grammatik sind, sondern die Redewendungen oder Kollokationen – die Verbindungen, die Wörter miteinander eingehen. [...] Die Verbindungen, die Substantive, Verben, Adjektive und Adverbien miteinander reingehen, sind nicht frei, sondern liegen fest; sie unterliegen keinen Regeln oder Gesetzmäßigkeiten.“

(W. Friederich 1979: i)

Friederich further states that it would have been impossible to capture the phraseological behaviour of the entire English lexicon, which is why they excluded ‘matter of course’ combinations (viz. predictable and free combinations) as well as archaic and rare expressions. They also excluded words whose usages could be explained in purely syntactic terms and could be looked up in A. Leonhardi and B.W.W. Welsh's *Grammatisches Wörterbuch Englisch*. (ibid.).

Headwords in *DEWC* are adjectives, adverbs (other than those ending in -ly), nouns and verbs. They are listed in alphabetical order where adjective-noun collocations can be found in the noun entry and adverb-verb as well as noun-verb collocations are listed in the verb entry.

Each headword is followed by its part-of-speech label. Different senses are indicated by Arabic numerals and short glosses (in italics) are provided for those senses whose meaning could not be sufficiently explained by the examples, e.g. *free* – (3) ‘relieve’, as can be seen in Fig. 17.

---

<sup>4</sup> ISED was photographically reprinted and published by Oxford University Press in 1948 as *A Learner's Dictionary of Current English*. In 1952 it was retitled *The Advanced Learner's Dictionary of Current English* and got ‘Oxford’ added in its title with the third addition, (The Oxford Advanced Learner's Dictionary of Current English).

free — fright
<i>ing to be paid for) ~ tickets for the theatre; a ~ copy of a book; the refreshments are ~; ~ education; ~ admission; there's a ~ offer of a toothbrush with every tube of toothpaste you buy this week; a ~ pass (for rail travel); ~ accommodation; ~ on rail Munich; carriage ~; delivered ~; the price is quoted ~ German frontier; 'F.O.B.' means ' ~ on board'. (13 <i>generous</i>) he is ~ with his money; give ~ly: give things away with a ~ hand; be ~ with one's praise / advice; a ~ flow of water; ~ bloomers (<i>plants with many flowers</i>) (14 <i>spontaneous</i>) a ~ offer; ~ conversation / laughter (at a party) (15 <i>not faithful to the original</i>) a ~ translation; a ~ adaptation; ~ly adapted from Goethe; a ~ interpretation of the text; a ~ copy / likeness (16 <i>not consistent with rules</i>) ~ verse; a ~ fight.</i>
<b>free</b> <i>adv</i> (→ <b>free</b> <i>adj</i> ) (12) children are admitted ~; get in ~; travel ~ (on the railways); he has / gets the accommodation ~.
<b>free</b> <i>vt</i> (1) the prisoners were ~d; ~ a man from prison / slavery / the wreckage of a train; ~ an animal from a trap; she tried to ~ herself from his embrace. (2) ~ s.o. from debt / anxiety / fear / an obligation; ~ the mind from prejudice (3 <i>relieve</i> ) ~ s.o. of a task / a responsibility for s.th; try to ~ s.o. of prejudices; ~ s.o. of a burden / worry.
<b>freedom</b> <i>n</i> (1 — liberty 1) take up arms for ~; give prisoners their ~; win / gain / lose / enjoy (religious / political / intellectual) ~; live in ~; the basic ~s; ~ of conscience / action / speech / thought / will; rob s.o. of his ~; the four ~s — ~ of speech, ~ of religion, ~ from fear, ~ from want; ~ from pain / disease / fear / hunger / blame / debt / tax / obstruction / oppressors / restrictions / controls / etc; ~ of speech / communication / the press / the air ( <i>broadcasting</i> ) / the seas; academic ~. (2) (give s.o.) the ~ to do s.th; speak with ~; the question was discussed with great ~; treat s.o. with the ~ and familiarity of an equal; give a friend the ~ of the house / library ( <i>allow him to use it freely</i> ); give s.o. / receive the ~ of a city ( <i>honorary rights of citizenship</i> ) (3 → <i>liberty</i> 2) take / use ~s with s.o.
<b>freeze</b> <i>vi</i> (1) put your overcoat on — it's ~ing outside; it was ~ing last night; ~ hard; it is going to ~ tonight. (2 <i>cover with ice</i> ) the puddles are frozen; the roads
were frozen; the waterpipes are frozen; the lake was frozen over. (3 <i>become ice, become covered with ice</i> ) the lake froze over in January; all the window-panes froze over; the radiator froze (up); the pipes have frozen; water ~s at a temperature of 0° C; the oil is ~ing; the ground was frozen hard; ( <i>fig</i> ) a frozen smile; the smile froze on his lips / face. (4 <i>become very cold</i> ) — to death; his hands froze to the oars; I'm ~ing ( <i>inform</i> ). (5a <i>turn into s.th. like ice</i> ) my fingers are quite frozen; we were frozen stiff ( <i>inform</i> ); the news made her blood ~ (in her veins) ( <i>curdle</i> ). — <i>vt</i> (5b) the sight of the monster froze the blood in his veins. (6 <i>make into ice</i> ) ~ ice-cream; frozen meat / vegetables (7) late frosts sometimes ~ the apple blossom. (8) ~ capital / assets / credits; ~ prices / wages.
<b>friend</b> <i>n</i> (1) he is a ~ of mine / of my brother's; ~s and relatives; ~ and foe; a close / fast / bosom / intimate / faithful / stalwart / unsailing / consistent / constant / dependable / desirable / familiar / discreet / invaluable / new / old / lost ~; a dubious / inconstant / fair-weather ~; they are inseparable / warm / near ~s; a common / mutual ~ of ours; he has shown himself / proved a true / loyal ~; we are old ~s of his family; we have been ~s with his family for years; I have been (great) ~s with him ever since our schooldays; make ~s with s.o.; they became ~s; they are ~s again; he is ~s with everyone; I have no ~s here; be a good ~ to s.o.; have a good ~ in s.o.; make a ~ of s.o.; this has made him many ~s; a soldier / actor / etc ~ of his; they parted good ~s. (2) the dog is man's best ~; a vacuum-cleaner is the housewife's best ~; night is the ~ of smugglers and thieves; who goes there — ~ or foe ( <i>militi</i> )? (3) he was a ~ of the people; a ~ of / to the poor; he was a ~ to many charities / all good causes; a ~ to those in need.
<b>friendship</b> <i>n</i> a durable / lasting / lifelong / close / intimate / unstable / etc ~; a ~ of twenty years; bonds / ties of ~; form / strike up a ~ with s.o.; win s.o.'s ~; break off a ~; live together in ~; help s.o. out of ~; never forget old ~s; swear eternal ~; the two countries signed a ~ pact.
<b>fright</b> <i>n</i> (1) be filled with ~; die of ~; the deer took ~ at the shot; get / have a (sud-

Fig. 17: Sample page from DEWC

DEWC is undeniably more comprehensive than GEW and can be seen as a 'true' forerunner of what we understand today as a collocation dictionary.

#### 4.1. Monolingual Collocation Dictionaries

Monolingual collocation dictionaries in the modern sense only came into being in the 1980s. So, in this section I will briefly discuss existing general and specialised English monolingual collocation dictionaries.

#### 4.1.1. English Monolingual Collocation Dictionaries: General

General English monolingual dictionaries include Ch.D. Kozłowska and H. Dzierżanowska's *Selected English Collocations* (1982/88), M. Benson et al.'s *The BBI Combinatory Dictionary of English* (1986/1997), J.McH. Sinclair's *Collins COBUILD Collocations on CD-ROM* (1995), J. Hill and M. Lewis's *LTP Dictionary of Selected Collocations* (1997) and M. Deuter et al.'s *Oxford Collocations Dictionary for Students of English* (2002).

##### (1) The BBI Combinatory Dictionary of English (BBI)

As already pointed out in Chapter 2, Morton Benson recognised the need for a collocation dictionary for English learners when he was working on his Serbo-Croatian – English dictionary in the mid-to-late 1960s (Gabrovšek 2004: personal communication).

Since its first publication in 1986, the *BBI* has been considered as THE English collocation dictionary (but this reputation is now being challenged by the *Oxford Collocation Dictionary for Students of English*, which I will deal with later in this section).

The *BBI*, in its 2<sup>nd</sup> edition, covers 18,000 entries and 90,000 collocations. Collocations in the *BBI* are subdivided into lexical and grammatical collocations (see Chapter 1), where lexical collocations precede grammatical ones. Some linguists (cf. Cowie 1999: 79) claim that the *BBI* failed in this respect since grammatical constructions are better listed in valency dictionaries, for example, than in a collocation dictionary.

Another reported deficiency of the *BBI* is the fact that although the authors explicitly rule out free combinations and idioms (Benson et al. 1997: xxxiv), lots of those can be found in the dictionary (and they are marked with double quotation marks as such!). Likewise, many frequent and salient collocations have been left out (cf. Gabrovšek 1998: 133).

What is unique to the *BBI* is its comprehensive user's guide (31 pages!) in which the authors explain to the users how to find the collocation they need and provide detailed information on the types of collocations listed in the dictionary<sup>5</sup>.

141	frugal
free hand <i>n.</i> ["freedom of action"] 1. to give smb. a ~ 2. to get: have a ~ 3. a ~ to + inf. (she had a ~ to do whatever she wanted)	sm. 3. a nasty; sudden ~ 4. stage ~ 5. in, with ~ (to scream with ~)
free throw <i>n.</i> (basketball) to make a ~	frighten <i>v.</i> 1. (d; tr.) to ~ into (to ~ smb. into submission) 2. (d; tr.) to ~ out of (to ~ smb. out of doing smth.) 3. (misc.) to ~ smb. to death
free will <i>n.</i> 1. to exercise one's ~ 2. of one's own ~	frightened <i>adj.</i> 1. ~ about, at, by, of (~ at the very thought; ~ of the dark) 2. ~ to + inf. (she was ~ to see a stranger approach) 3. (misc.) to be ~ out of one's wits; to be ~ to death
freeze I <i>n.</i> ["frost"] ["freezing"] 1. a deep, hard ~ ["freezer"] (BE) 2. a deep ~ ["fixing at a certain level"] 3. to impose a ~ 4. a nuclear; price; wage ~ 5. a ~ on	frightening <i>adj.</i> 1. ~ to + inf. (it's ~ to contemplate such a possibility) 2. ~ that + clause (it's ~ that a war could break out at any time)
freeze II <i>v.</i> 1. to ~ hard, solid (it froze hard last night) 2. (D; tr.) to ~ out of (to ~ smb. out of a conversation) 3. (D; intr.) to ~ to (his exposed skin froze to the metal; to ~ to death)	frighthful <i>adj.</i> see frightening
freezer <i>n.</i> 1. to defrost a ~ 2. a home ~ 3. (misc.) a ~ compartment (in a refrigerator)	fringe <i>n.</i> 1. the lunatic ~ 2. on the ~s (of society)
freight <i>n.</i> ["goods, cargo"] 1. to carry; handle; load; ship ~ ["freight train"] (colloq.) (AE) 2. to hop, jump ("board") a ~	fringe benefits <i>n.</i> to get; provide ~
French <i>n.</i> Canadian ~ (natives of Quebec speak Canadian ~)	fritter away <i>v.</i> (D; tr.) to ~ on (to ~ one's time away on trifles)
French toast <i>n.</i> to make ~	frivolous <i>adj.</i> ~ to + inf. (it was ~ of him to make such an accusation)
frenzy <i>n.</i> 1. a wild ~ 2. in a ~ (in a ~ of despair) 3. (misc.) to work oneself up into a ~	frog <i>n.</i> 1. a grass; green; wood ~ 2. ~s croak; jump 3. an immature ~ is a tadpole 4. (misc.) to have a ~ in one's throat ("to be hoarse")
frequency <i>n.</i> ["number of repetitions"] 1. alarming; great, high; low ~ 2. with ~ (with alarming ~) ["number of periodic waves per unit of time"] (physics) 3. high; low; medium; ultrahigh ~ 4. a radio ~ 5. on a ~	front I <i>adv.</i> to face ~
fresco <i>n.</i> to paint a ~	front II <i>n.</i> ["front line"] (mil.) 1. at, on the ~ (the war correspondents spent two days at the ~; there has been no activity on this ~) ["area of activity"] 2. the home; political ~ 3. on a ~ (on a broad ~; on the home ~) ["advanced part"] 4. at the ~ of; in ~ of ["movement"] ["campaign"] 5. a popular; united ~ (to present a united ~) ["boundary"] (meteology) 6. a cold; occluded; stationary; warm ~ ["walk, road along a body of water"] (BE) 7. a river; sea ~ 8. along a ~ (to walk along the sea ~) 9. on a ~ (is there a hotel on the sea ~?) ["behavior"] 10. to put, put up a ~ 11. a bold, brave, brazen ~ (to put on a bold ~) ["facade"] 12. a ~ for (the store was a ~ for illegal drug sales) ["misc."] 13. up ~ ("in advance"; "frankly")
fresh I <i>adj.</i> ["recent"] ["new"] ~ from, out of (~ out of school)	front III <i>v.</i> 1. (d; intr.) to ~ for (to ~ for the mob) 2. (d; intr.) to ~ on, onto (our building ~s on the main road)
fresh II <i>adj.</i> (colloq.) (AE) ["bold"] ["impudent"] ~ with (don't get ~ with me)	frontage <i>n.</i> lake; ocean; river ~ 2. ~ on
freshener <i>n.</i> an air, room ~	frontier <i>n.</i> 1. to advance, extend, push back, roll back a ~ (to extend the ~s of science) 2. to cross a ~ 3. on a ~ 4. a ~ between
fret <i>v.</i> 1. (D; intr.) to ~ about, over 2. (misc.) to ~ and fume	front line <i>n.</i> at, in, on the ~
friction <i>n.</i> 1. to create, generate, produce ~ 2. ~ among, between; with (there has been some ~ between the union and management)	frost <i>n.</i> 1. a bitter, hard, heavy, severe; light, slight ~ 2. eternal ~, permafrost 3. ~ forms 4. a touch of ~
friend <i>n.</i> 1. to be; make a ~ 2. to be; make ~s (with smb.) 3. a bosom, close, good, intimate, old; faithful, fast, loyal, staunch, strong, true; lifelong ~ 4. a mutual; personal; special ~ 5. a fair-weather; false ~ 6. inseparable ~s 7. a pen ~ (BE; CE has <i>pen pal</i> ) 8. a ~ to (she was a good ~ to us) 9. (misc.) my good ~	froth <i>v.</i> to ~ at the mouth
friendliness <i>n.</i> ~ to, towards	frown I <i>u.</i> 1. to wear a ~ 2. an angry; perpetual ~
friendly <i>adj.</i> 1. ~ of (that was ~ of you) 2. ~ to, towards, with 3. ~ to + inf. (it was ~ of him to offer his help) 4. (misc.) user-friendly	frown II <i>v.</i> 1. (D; intr.) to ~ at ("to look with displeasure at") (the teacher ~ed at the noisy children) 2. (d; intr.) to ~ on, upon ("to disapprove of") (they ~ on all forms of affection in public) 3. (misc.) to ~ with displeasure
friendship <i>n.</i> 1. to cement, develop, form, make, strike up a ~ 2. to cherish, cultivate a ~ 3. to promote (international) ~ 4. to break up, destroy a ~ 5. a close, firm, intimate, lasting, strong, warm; lifelong; long; special ~ 6. the bonds of ~ 7. (a) ~ among, between; with	frozen <i>adj.</i> ~ hard, solid, stiff
fright <i>n.</i> 1. to give smb. a ~ 2. (esp. BE) to take ~ at	frugal <i>adj.</i> (formal) ~ of (esp. BE), with (~ of one's

Fig. 18: Sample page from BBI

<sup>5</sup> Note that there also exists an accompanying workbook with exercises of how to use the *BBI*, namely *Using The BBI – A Workbook with Exercises* (1991).

The headwords in the *BBI* are listed alphabetically and sense divisions are marked by brief definitions in square brackets, e.g. *front*– ['front line'] vs. ['area of activity'].

The *BBI* was compiled chiefly on the basis of the intuition of three American lexicographers, and no corpus was used for collocation retrieval or validation – which distinguishes the *BBI* from the other collocation dictionaries I discuss in this section.

## (2) *Collins COBUILD Collocations on CD-ROM (COBCOLL)*

What distinguishes *COBCOLL* from the other dictionaries in this chapter is that it comes solely on a CD-ROM. It was published in 1995 under the editorship of John McH. Sinclair whose name is inseparably associated with the COBUILD project – a pioneering dictionary project in which headwords as well as examples and collocations were drawn from a corpus, now known as the Bank of English<sup>6</sup>.

Basically, *COBCOLL* is a concordance program (see Chapter 2) on CD-ROM which contains 10,000 nodes (= keywords that can be investigated through concordancing and statistical collocation tools) that are linked with up to 20 examples per collocation from the Bank of English. Altogether 140,000 collocations can be found and analysed in their 'natural' context.

The user's guide makes clear that the selection of common nodes was done by a computer rather than a human being and that, therefore, odd examples could be found as well as offensive ones.

The fact that *COBCOLL* was compiled along computational lines rather than lexicographical or pedagogical ones, has led to constant criticism as far as the use of *COBCOLL* as a collocation dictionary goes<sup>7</sup>. However, it is an invaluable source in illustrating the corpus approach to collocations.

---

<sup>6</sup> For a detailed overview of the COBUILD project see Sinclair (ed. 1987b)

<sup>7</sup> See Dirk Siepmann's review at: [http://www.dirk-siepmann.de/Publications/Rezension\\_Sinclair/REVSINCL.htm](http://www.dirk-siepmann.de/Publications/Rezension_Sinclair/REVSINCL.htm).

When you open the program, you simply type in the node word you are interested in. What you can immediately see is its overall frequency.

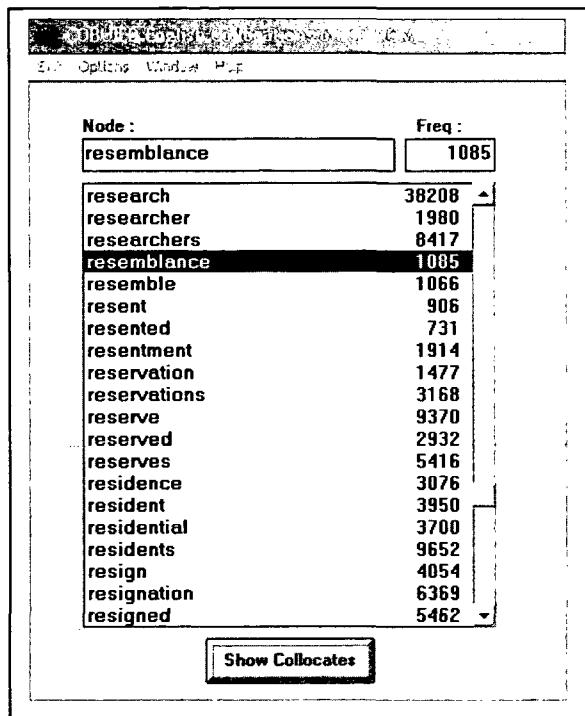


Fig. 19a: COBCOLL – first step

Next, you click on 'Show Collocates'.

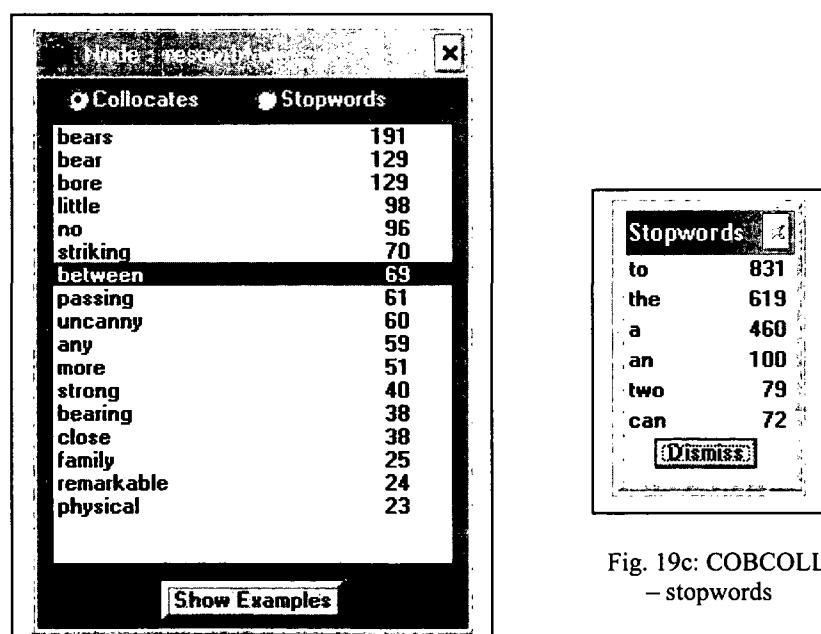


Fig. 19c: COBCOLL – stopwords

Fig. 19b: COBCOLL – second step

As Fig 19b shows, with just the second step you can find all collocates of the node word in descending frequency. It shows, however, that the input data was not lemmatised, that is, different word-forms were not assigned to their lemma (e.g. *bears*, *bear*, *bore*, *bearing* – BEAR).

The feature ‘stopwords’ (Fig. 19c) has been implemented for reference purposes only. The user can click on the button and have a look at so-called stopwords (viz. about one hundred very frequent English function words such as *a*, *an*, *the*, etc.) which were omitted in the search for collocates of the node under scrutiny.

In step three, you click on ‘Show Examples’ and a KWIC concordance of the node pops up. You can align the concordance according to the collocate or the node and sort it alphabetically by either the right- or the left-hand side.

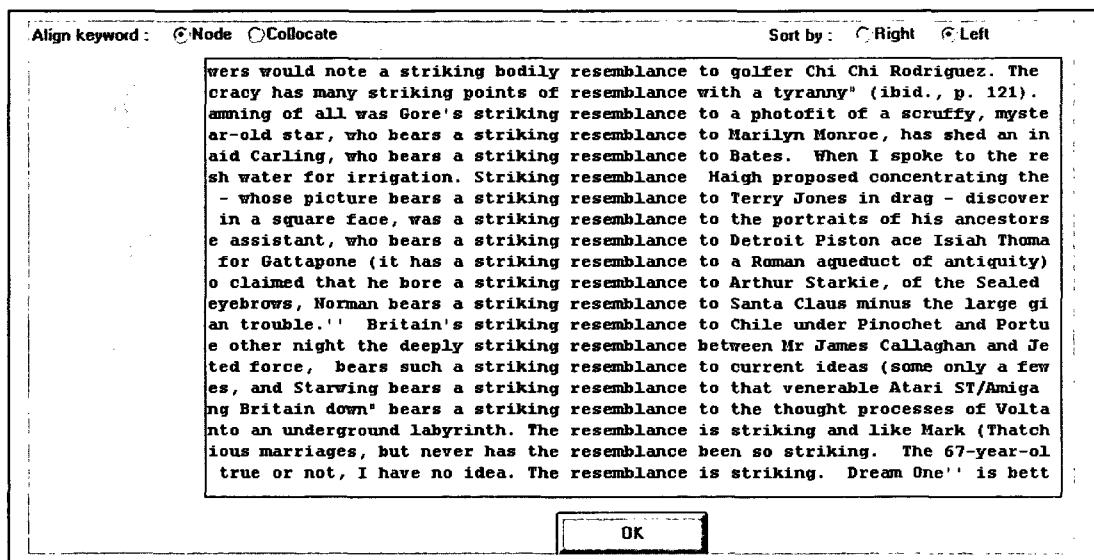


Fig. 19d: COBCOLL – KWIC concordance

If you need an expanded context for your research you just double click on the concordance line in question (Fig. 19e). Here, you can also see which source the collocation comes from, e.g. books, journals, newspapers, etc., British or American.

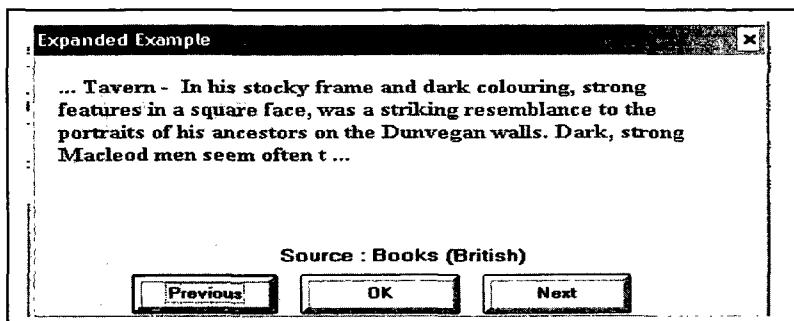


Fig. 19e: COBCOLL – expanded example & source of the collocation

### (3) LTP Dictionary of Selected Collocations (LTP)

Hill and Lewis's *LTP* is based on Ch.D. Kozłowska and H. Dzierżanowska's *Selected English Collocations* (1982/88) and Ch.D. Kozłowska's *English Adverbial Collocations* (1991), both of which are based on a corpus of post-1960 British English.

*LTP* covers 55,000 collocations under 3,200 alphabetically listed headwords. In the entries collocations are subdivided into word classes. Roughly two thirds of the *LTP* are devoted to collocations starting from a noun (2,000 headwords), while the remaining part is dedicated to adverbial collocations starting from a verb or adjective (1,200 headwords).

The intended user of *LTP*, according to the editors, is any of those well past the intermediate level of language learning. No glosses are provided, so the user has to have prior semantic knowledge about the words that make up the collocation (glosses are only provided when the headword was editorially determined to have 'clearly distinct meanings', e.g. *talk* (= 'lecture'), *talks* (= 'negotiations') (Cueto 1998: 1)

A drawback of the *LTP* is that the collocation compilation process is not explicitly stated, although the authors claim to have excluded collocations that are too common, too technical, too colloquial, or too difficult. "Included are those which are in strong relation, the judgment of strength presumably resting upon the editors' expertise and their native speaker inclinations." (ibid: 2).

FOUNDATION	Starting from a noun
<b>FOUNDATION</b>	
V: build (on), create, destroy, establish, form, lay, make, provide, reinforce, rest on, serve as, shake, undermine ~	
V: ~ collapsed, crumbled	
A: firm, lasting, permanent, reliable, secure, shaky, solid, sound, steady, strong, weak ~	
<b>FRAMEWORK</b>	
V: build, construct, create, destroy, erect, fit into, lay down, make, place sth in, provide, put sth in, set up, work within ~	
A: balanced, basic, conceptual, convenient, loose, proper, regular, reliable, rigid, solid, strong, suitable, tight ~	
<b>FRAUD</b>	
V: accuse sb of, be guilty of, commit, convict sb of, deny, detect, discover, engage in, expose, perpetrate, prevent, suspect, uncover ~	
A: large-scale, petty ~	
<b>FREEDOM</b>	
V: abuse, achieve, cherish, curtail sb's, deprive sb of, desire, endanger, enjoy, ensure, fight for, forfeit, gain, get, give, grant, guarantee, guard, jeopardize, limit, lose, preserve, promise, restore, restrict, secure, take away, threaten, value, win ~	
V: ~ has been eroded, is at risk/in danger/threatened ~	
A: absolute, academic, comparative, complete, hard-won, individual, lasting, (un)limited, newfound, perfect, personal, political, relative, religious, total ~	
<b>FRICITION</b>	
V: aggravate, avoid, cause, create, generate, give rise to, lead to, produce, reduce, remove ~	
<b>FRIEND</b>	
V: keep, lose, make, win ~	
A: best, close, dangerous, fair-weather, faithful, false, family, firm, good, intimate, lifelong, loyal, mutual, new, old, personal, school, special, sympathetic, true ~	
<b>FRIENDSHIP</b>	
V: break off, cement, cultivate, destroy, develop, form, promote, spoil, strike up, value ~	
A: abiding, close, deep, firm, intimate, lifelong, long, personal, sentimental, unstinted, warm ~	
P: act of, bonds of, ties of, token of ~	
<b>FRONT (defence)</b>	
V: adopt, display, form, make, present, set up, show ~	
A: common, solid, united, wide ~	
<b>FRONTIER</b>	
V: alter, change, close, create, cross, (re)define, determine, draw, establish, fix, guard, mark, move, open, push back, recognise, redraw, violate ~	
V: ~ has changed, moved, runs/stretches from ... to ...	
<b>FRUIT</b>	
V: bear, grow, pick ~	
A: bitter, citrus, dried, exotic, firm, fleshy, fresh, juicy, luscious, ripe, rotten, soft, succulent, tinned, tropical, wild ~	
<b>FRUSTRATION</b>	
V: bottle up, give vent to, suffer, vent ~	
A: continual, deep, great, immense, inevitable, pent-up, understandable ~	
<b>FULFILMENT</b>	
V: achieve, bring, lead to, prevent, reach, seek, strive for ~	
A: complete, immediate, instant, partial, personal, sexual, total ~	

96

Fig. 20: Sample page from LTP

#### (4) *Oxford Collocations Dictionary for Students of English (OCD)*

The *OCD* is the most recent achievement in collocational research and dictionary-making since it was published in 2002 and constitutes a real landmark in learners' lexicography in terms of both user-friendliness and comprehensiveness.

In contrast to the other dictionaries under discussion, the *OCD* is available in book form as well as on CD-ROM as *Oxford Phrasebuilder Genie* (together with the *Oxford Advanced Learner's Dictionary*, 6<sup>th</sup> edn. [2000]).

*OCD* lists 150,000 collocations of nouns, verbs and adjectives and provides over 50,000 examples of collocations in context.

Headword bases are in bold, and where a word has different senses, each has a different sub-heading. Collocators<sup>8</sup> are grouped by word classes and further grouped by the meaning of the collocation in question (cf. Patterson 2003):

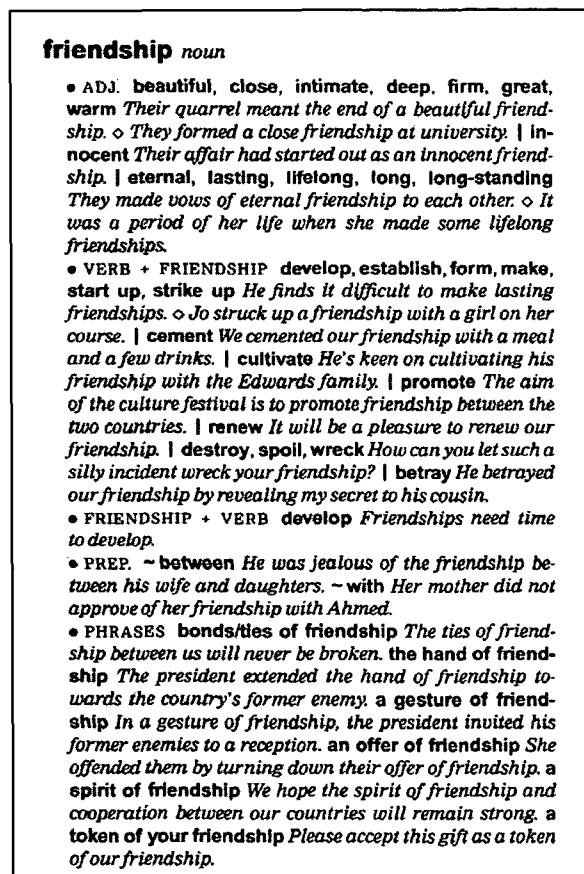


Fig. 21: Sample entry from *OCD*

*OCD* is based on the 100-million-word British National Corpus; however, no clear (statistical) statement is made as to the actual compilation process in collecting the collocations. What the editors outline (Introduction, pp. viii-ix), though, is that three questions were vital in the elaboration of their inclusion policy: (1) Is this a typical use of language? (2) Might a student of English want to express this idea? And (3) Would a student look up this entry to find this expression?

<sup>8</sup> They actually use the collocate-node terminology (Deuter et al., eds. 2002: xii).

### (5) Duden – Das Stilwörterbuch (DUST)

There is, as far as I know, no German collocation dictionary. The one work which comes nearest to the general English monolingual collocation dictionaries is Wermke et al.'s *Duden – Das Stilwörterbuch* (2001), which, however, records a variety of diverse word combinations, from free combinations to idioms as well as.

First of all, what is a *Stilwörterbuch* (cf. page 59)? Bahns (1996: 93) gives a brief survey of the history of collocation dictionaries and characterises a *Stilwörterbuch* as a monolingual dictionary of style that has a significant collocational component.

The subtitle of *DUST* is somewhat misleading as it reads “Idiomatisches Deutsch – Bedeutung und Verwendung der Wörter im Satz”, viz. ”Idiomatic German – Meanings and Uses of Words within Sentences”.

From the foreword, however, it is clear that *DUST* indeed lists not only idiomatic expressions but also collocations: “Idiomatic language use does not arise from simply knowing what words mean in isolation, but from the knowledge of how to (grammatically and semantically) correctly combine them in a sentence. The *Stilwörterbuch* lists common words in German together with the combinations they can enter into. Furthermore, stereotyped phrases as well as sayings and proverbs are listed – and, therefore, the *Stilwörterbuch* presents the enormous wealth of expressive power the German language has to offer” ([translated by the author]; cf. Wermke et al., eds. 2001: 5)<sup>9</sup>.

My analysis of *DUST* (Maurer-Stroh, P. 2004b) has revealed that adjective-noun collocations are, in essence, listed in both types of entries, the one for the noun and the one for the adjective.

This can be also seen from Fig. 22, where we can find adjective collocators in the entry for the noun *Freund* as well as retrieve noun bases from the entry of the adjective *freundschaftlich*.

---

<sup>9</sup> The dictionary in the Duden series ‘really’ concerned with idiomatic German is Volume 11, *Duden – Redewendungen*.

329	freundschaftlich
<p>schon sein F.?: (derb) Von menschlicher Nahrung; das war ein elendes F.; * ein [gefundenes] Fressen für jmdn. sein (ugs.; jmdn. sehr gelegen kommen).</p> <p>Freude, die: 1. <b>das Frohsein</b>: eine große, riesige, unerwartete, stille, heimliche, diebische (<i>heimliche</i>) F.; die F. des Wiedersehens; die F. an der Natur, über das Geschenk; F. erfassst, überwältigt, erfüllt jmdn.; (geh.:) F. strahlt aus seinen Augen; in der Stadt herrsche große F. über den Sieg; es ist eine [rechte, wahre] F., ihm zuzuschauen; sie tanzte, dass es eine [wahre] F. war; das ist keine reine F. (<i>das tut, das sieht man nur ungern</i>); es wird mir eine F. sein, Sie zu begleiten (<i>ich werde Sie gerne begleiten</i>); jmdn. [eine] F. bereiten, machen, spenden, schenken, jmdn. die/eine F. gönnen, rauben, vergällen, verderben; etw. trübt, stört jmds. F.; F. zeigen; an etw. viel, wenig, keine [große] F. haben; F. über etw. empfinden; sie hatte ihre helle F. daran; sie erlebte viel F. mit ihren Kindern; etw. aus F. an der Sache, aus Spaß an der F. (ugs. scherzh.) tun; mit kindlicher, naiver F. zuschauen; voll[er] F. stimmt er zu; sie weinte vor F., war vor F. außer sich; die Kinder sprangen vor F. [fast] an die Decke (<i>haben sich sehr gefreut</i>); zur F. der Eltern wurde das Kind bald gesund.</p> <p>2. <b>etw., was jmdn. erfreut</b>: die Freuden des Sommers, der Liebe; die kleinen Freuden des Alltags; sie will die Freuden des Lebens genießen; sie leben herrlich und in Freuden (<i>es ging ihnen sehr gut</i>); * Freud und Leid (geh.: <i>Glück und Unglück des Lebens</i>) - mit Freuden (<i>beglückt, erfreut</i>).</p> <p>freudig: a) <b>froh</b>: ein freudiges Gefühl; voller freudiger Erwartung; mit freudiger Stimme; jmdn. f. begrüßen; f. erregt, überrascht sein; b) <b>Freude bereitend</b>: eine freudige Nachricht, Überraschung.</p> <p>freuen: 1. a) (sich f.) <b>Freude empfinden</b>: sich ehrlich, aufrichtig, herzlich, (ugs.): mächtig f.; sich heimlich, diebisch (<i>verstohlen</i>), königlich (<i>sehr</i>), im Stillen f.; sie kann sich f. wie ein Kind; er freut sich wie ein Schneekönig (ugs.; <i>sehr</i>); (iron.): da hast du dich zu früh gefreut; ich freue mich sehr, dass es dir gut geht; wir freuen uns, Ihnen helfen zu können; (sich etw. (Gen.) f.) er freut sich seines Lebens (geh.; <i>genießt sein Leben</i>); (sich über jmdn., etw. f.) sich über einen Erfolg, über ein Geschenk f.; ich freue mich darüber, dass du mithäst; (sich mit jmdn. f.) wir freuen uns sehr mit Ihnen; b) (sich an jmdn., etw. f.) <b>seine Freude haben</b>: sich an [den] Blumen, an seinen Kindern f.; ich freue mich an eurem Glück; c) (sich auf jmdn., etw. f.) <b>freudig erwarten</b>: wir freuen uns auf unser Kind, auf deinen Besuch, auf die Ferien; d) (sich für jmdn. f.) <b>herzlich gönnen</b>: ich freue mich für dich, dass du diese Stellung bekommen hast.</p> <p>2. (jmdn. f.) <b>Freude bereiten</b>: das freut mich [auf- richtig]; das Geschenk freut mich sehr; es freut mich [für dich], dass du mitkommen kannst; es soll[te] mich f. Sie recht bald bei uns zu sehen; (iron.): das freut einen denn ja auch!</p> <p>Freund, der: 1. a) ein guter, vertrauter, intimer, treuer, bewährter F.; er ist mein väterlicher F.; falsche Freunde; ein alter F. unseres Hauses; unser junger F.; mein lieber, bester F.; mein F. Klaus; er ist mein ältester, einziger F.; sie sind unzertrennliche, (ugs.): dicke Freunde; griechische Freunde von uns; (ugs. iron.): du bist mir ein schöner F.; mein F. ist er gewesen! (ugs.): die Freundschaft ist vorbei); Freunde werden; er hat, besitzt, findet keinen F.; sich jmdn. zum F. machen; unter Freunden sein; wie gehts, alter F.? (ugs. als Anrede); [mein] lieber F.! (scherzh. oder drohend; als Anrede); ② Freunde in der Not gehn hundert/Lausend auf ein Lot (<i>in Notzeiten sind die Freunde selten</i>); b) <b>Gesinnungsgenosse, Parteifreund</b>: seine politischen Freunde; c) <b>Partner</b>: sie hat einen neuen F.; sie hat noch keinen festen F.</p> <p>2. <b>Förderer, Liebhaber</b>: er ist ein F. guter Musik; ein F. der Tiere; er ist ein F. von guten Weinen; * Freund Hein (verhüllt; <i>der Tod</i>) - Freund und Feind (<i>jedermann</i>): er war angesehen bei F. und Feind - kein Freund von etw. sein (<i>etw. nicht schätzen</i>); ich bin kein F. großer Worte.</p> <p>freundlich: a) <b>liebenswürdig</b>: eine freundliche Miene; ein freundliches Gesicht; wir fanden freundliche Aufnahme; ein freundliches Wort an jmdn. richten; jmdn. einen freundlichen Empfang bereiten; besten Dank für ihr freundliches Anerbieten; mit freundlicher Genehmigung des Verlages; <i>Briefschluss</i>: mit freundlichen Grüßen Ihr ...; (f. zu jmdn. /seltener:) gegen jmdn., sie war immer f. zu mir / (seltener:) gegen mich; würden Sie so f. sein, mir zu helfen?; f. lächeln; jmdn. f. ansehen; jmdn. f. gesinnt sein; /Aufforderung des Fotografen: bitte recht f.; b) <b>heiter (stimmend)</b>: freundliches Wetter; eine freundliche Wohnung, Stadt, Gegend; die Farben des Kleides sind sehr f. (<i>hell und ansprechend</i>); die Wohnung ist f. und sauber; die Haltung, Stimmung an der Börse ist f. (<i>Wirtsch.; günstig</i>).</p> <p>Freundschaft, die: eine treue, herzliche, innige, langjährige, (ugs.): dicke F.; eine F. fürs Leben; die F. zwischen Schülern, zwischen den Völkern; das ist echte, wahre F.; unsere F. ist in die Brüche gegangen, ist (ugs.): eingeschlafen, vertieft sich; uns verbindet eine tiefe F.; mit jmdn. F. schließen, die [alte] F. erneuern; (geh.): jmdn. die F. aufkündigen; alte Freundschaften bewahren; etw. aus [reiner] F. tun; (geh.): sie war uns in F. verbunden; ich sage es dir in aller F.; sie trennten sich in F.</p> <p>freundschaftlich: eine freundschaftliche Gesinnung; freundschaftliche Ermahnungen; die beiden Länder unterhalten freundschaftliche Beziehungen; mit jmdm. auf freundschaftlichem Fuß stehen; sie verkehren f. miteinander; jmdn. f. be-</p>	

Fig. 22: Sample page from DUST

#### 4.1.2. English Monolingual Collocation Dictionaries: Specialised

In addition to general-purpose English monolingual collocation dictionaries, also a few specialised English collocation dictionaries have been published, notably A. Spencer's *Noun-Verb Expressions in Legal English* (1975), Ch.D. Kozłowska's *English Adverbial*

*Collocations* (1991), E.H. Mikhail's *Dictionary of Appropriate Adjectives* (1994) and B.A. Kipfer's *Roget's Descriptive Word Finder - A Dictionary / Thesaurus of Adjectives* (2003).

**(1) *Noun-Verb Expressions in Legal English (NVELE)***

Only few people seem to have ever seen Arnold Spencer's innovative work (Gabrovšek 2004: private communication), so I have to base my comments and the sample page in Fig. 23 on A.P. Cowie's (1985) survey of collocation dictionaries.

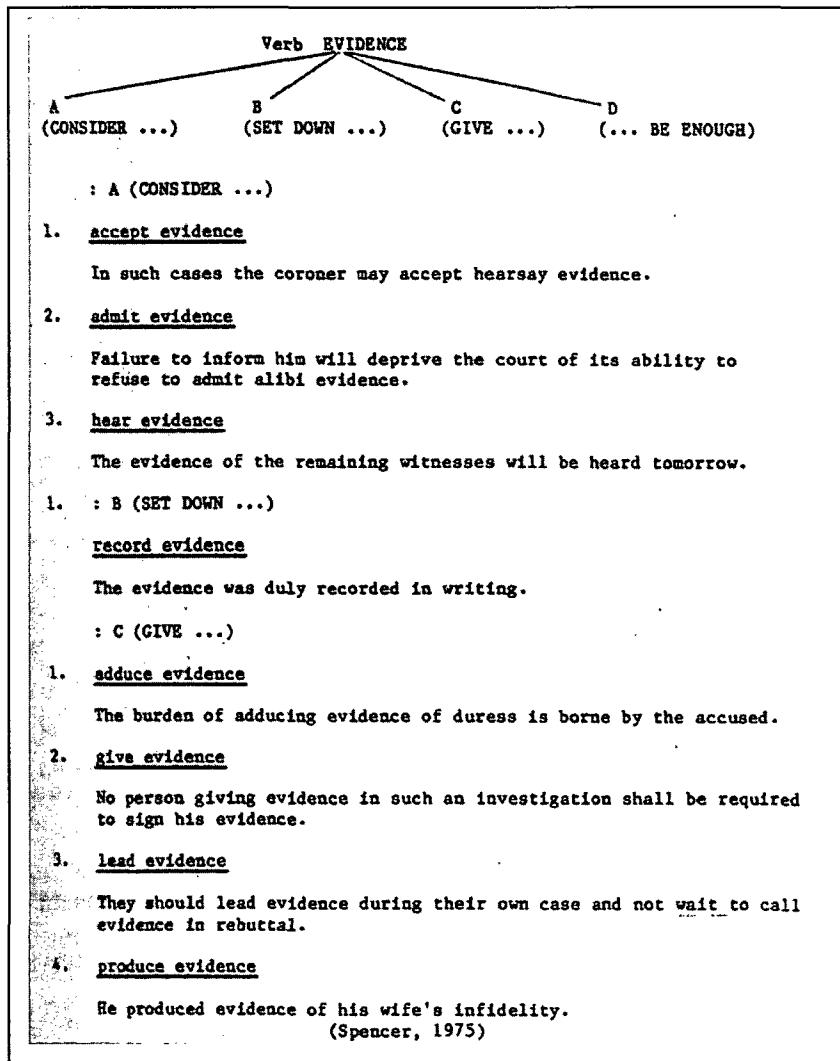


Fig. 23: Sample from NVELE

Spencer's approach is to specify a number of 'core' verb meanings which are compatible with the noun *evidence*. They are set out as keywords (rough translations of the Arabic words) in a simple diagram at the head of the entry. The selected keyword leads the user on to a set of verbs (Cowie 1985: 67-68).

According to Cowie, this is the most original and effective arrangement, though not the most economical one; for this reason, it is only suitable for ESP glossaries with limited word coverage (*ibid.*).

## (2) *Dictionary of Appropriate Adjectives (DAA)*

E.H. Mikhail's DAA lists some 4,000 nouns (in alphabetical order) and provides in 342 pages<sup>10</sup> numerous adjectives (arranged in thematic sequences) that are used to describe them.

According to what is stated on the back cover and in the Preface, the *DAA* will help you (1) find a specific adjective to describe a noun, e.g. *a guttural consonant, a temporal ruler*, (2) make an exact technical description of a noun; an *arch*, for example, could be described by a variety of adjectives, such as *vaulted, scalloped, corbelled* or *elliptical*, (3) discover a suitable classical or mythological label, e.g. *a Cadmean victory, Orphic mystery, Spartan endurance* and (4) find the appropriate adjective that applies to a certain animal, e.g. *cooing pigeon, hissing snake*.

The *DAA* is essentially an adjective-noun dictionary which does not only include collocational information of this particular structural type but also list technical terms and archaic, stilted and technical expressions which serve for creative writing more than as a collocational aid for foreign learners of English.

Only words that are listed in standard-language dictionaries as adjectives are included in this dictionary, so, all other formally / etymologically related forms, e.g. participles ending in *-ed* or *-ing*, are excluded. Also included are nouns used idiomatically as adjectives, e.g. *crocodile tears*. These 'adjectives' are listed at the end of the entry

---

<sup>10</sup> Included in the 342 pages are three appendices: (A) Good Personal Qualities (pp. 333-34), (B) Bad Personal Qualities (pp. 335-37) and (C) Similes with Adjectives (338-42).

together with technical adjectives. A user looking for an adjective to describe less common nouns should look for it under more common ones, e.g. *for boutique* see *shop*, etc. (Mikhail 1994: vii).

<b>frankness</b>	refreshing, bracing, disarming, breathtaking, surprising, astonishing, astounding, amazing, unusual, blunt, utter, sincere, warm, taut, evident, remarkable, obsessive, acid, brutal, naïve
<b>fraud, frauds</b>	deliberate, callous, blatant, barefaced, outright, gross, absolute, extreme, all-out, gigantic, unqualified, unmitigated, widespread, rampant, actual, apparent, constructive, electoral, criminal
<b>freedom, freedoms</b>	flowering, unencumbered, real, perilous, (ir)responsible, giddy, gypsy, rascally, untrammelled, new-found, (hard-)won, elusive, short-lived, lost, basic, fundamental, relative, comparative, limitless, (un)limited, partial, total, absolute, moral, religious, political, academic, intellectual
<b>frequency, frequencies</b>	startling, astounding, astonishing, amazing, growing, rising, increasing, alarming, frightening, maddening, significant, common, habitual, persistent, nightmarish; low, reduced, medium, high, radio, angular, base, audio
<b>freshness</b>	delicious, dewy, crisp, vivid, youthful, vernal, open-air
<b>friend, friends</b>	intimate, beloved, bosom, close, well-matched, congenial, firm, delightful, affectionate, constant, loyal, staunch, (stead)fast, attached, lifelong, stalwart, steady, inseparable, sincere, loving, trusty, devoted, dear, constant, warm, faithful, unfailing, tender, (e)special, exceptional, sympathetic, near, (much-)valued, valuable, favoured, select, particular, unsycophantic, like-minded, reassuring, familiar, candid, judicious, level-headed, (un)reliable, (in)considerate, honest, hospitable, influential, powerful, high-powered, high-profile, high-placed, well-placed, posh, illustrious, useful, former, (long-)lost, quondam, erstwhile, long-ago, long-time, old-time, one-time, sometime, unworthy, false, so-called, dubious, fair-weather, deceitful, treacherous, faithless, perfidious, disloyal, unfaithful, officious, personal, childhood, school, college, mutual, political

Fig. 24: Sample entries from DAA

(3) Roget's Descriptive Word Finder - A Dictionary / Thesaurus of Adjectives (RDWF)

Another specialised dictionary devoted to adjectives is Barbara Ann Kipfer's *RDWF*. It is a combination of a dictionary and a thesaurus listing in 457 pages both common and unfamiliar adjectives to add nuance and originality to your piece of writing. (Kipfer 2003: back cover).

In *RDWF* adjectives are organised by theme, bringing related words together under specific (572!) category headings.

Kipfer, as she herself states (2003: 1), included every adjective described in other publications about adjectives, e.g. print dictionaries, dictionaries of descriptive terms, thematic dictionaries and dictionaries of obscure words.

*RDWF* is NOT a collocation dictionary (although some collocations might have found their way into it), and Kipfer herself suggests the users employ other dictionaries for context clarification:

"Whenever in doubt of for further clarification, do consult an authoritative college or unabridged print dictionary. Sometimes the user will need a more detailed definition to be sure that the chosen word is correct for the intended context."

(B.A. Kipfer 2003: 1)

<b>Perfume</b>	
<b>ambrosial:</b> divinely fragrant	<b>redolent:</b> diffusing a pervasive odor; fragrant
<b>antibromic:</b> deodorant	<b>rose:</b> sweetly scented like a rose
<b>aromatic:</b> having a pleasing scent or odor	<b>rosewatered:</b> delicate; sentimental; perfumed
<b>balmy:</b> fragrant like aromatic balm	<b>savory:</b> agreeable in smell; especially fragrant
<b>euodic:</b> aromatic	<b>scented:</b> smelling
<b>fragrant:</b> having a sweet or pleasing odor	<b>spicy:</b> having a pungent odor
<b>libanophorous:</b> scented; producing incense	<b>suaveolent:</b> sweet-smelling
<b>muscadine:</b> having the fragrance of the Southern fox grape	<b>sweet:</b> fragrant, perfumed, or fresh to the smell
<b>olent:</b> fragrant; having an odor	<b>sweet-scented:</b> pleasing to the smell
<b>perfumatory:</b> yielding a pleasant odor	<b>sweet-smelling:</b> pleasing to the smell
<b>perfumed:</b> sweet-scented	<b>thuriferous:</b> producing or bearing frankincense

Fig. 25: Sample category from RDWF

I included *RDWF* in this survey of collocation dictionaries since my primary concern is adjectives preceding nouns and, as I see it, any dictionary having to do with adjectives is in essence a combinatorial one – and it definitely is of relevance in a thesis on how to create an adjective-noun collocation dictionary.

## 4.2. Bilingual Collocation Dictionaries

Bilingual collocation dictionaries, both specialised and general, of any two languages are very rare. The ones I am aware of are W. Weng-chang's *A Dictionary of English Collocations* (1990)<sup>11</sup>, M. Benson and E. Benson's *Russian-English Dictionary of Verbal Collocations* (1993), S. Ichikawa et al.'s *The Kenkyusha Dictionary of English Collocations* (1995)<sup>12</sup> and L. Iordanskaja and S. Paperno's *A Russian-English Collocational Dictionary of the Human Body* (1996) (cf. section 2.2.4),

### (1) *A Dictionary of English Collocations* (DEC)

*DEC* is another dictionary I have never held in my hands; however, a review by C.-Y. Tsai (1997) provides the necessary data for a brief outline of this bilingual English-Chinese collocation dictionary intended for students of English as a foreign language.

*DEC* is a product of a team of 20 faculty members (English Department) at Suzhou University, China, and is based on the (English-Japanese) *A New Dictionary of English Collocations* edited by S. Katsumata in 1958. It comprises 15,000 entries and 300,000 collocational examples. In addition to the main dictionary part, *DEC* also includes a list of compounds (pp. 1861-1928).

The fact that *DEC* not only includes closed collocations, but also open ones, viz. free combinations in the Bensonian sense, suggests why it is considerably bigger than the *BBI* (with 90,000 collocations) – which is not true when we compare the number of entries, i.e. 14,000 vs. 15,000.

---

<sup>11</sup> This is an English-Chinese collocation dictionary.

<sup>12</sup> English-Japanese.

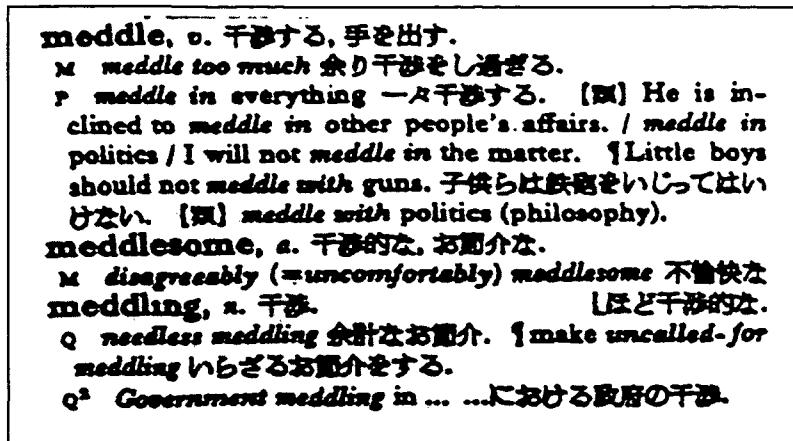


Fig: 26: Sample entries from DEC

Tsai observes that there are quite a few misplacements and that *DEC* contains some unacceptable phrases (due to the compilers' not being English native speakers). However, most of the mistakes can be traced back to the 1958 Kenkyusha dictionary.

## (2) Russian-English Dictionary of Verbal Collocations (REDVC)

Morton and Evelyn Benson's *REDVC* is another bilingual collocation dictionary; in contrast to *DEC*, it is a specialised dictionary since it lists Russian and English verbal collocations (but also miscellaneous verbal phrases including selected idioms and figurative expressions like *to laugh up one's sleeve*) and translates them into English; whenever possible, corresponding English collocations are used in these translations. Most of the collocations listed in it are grammatical (see Chapter 1) in nature; still, also lexical collocation of the type *to argue heatedly* and *to be seriously ill* are included. (Benson & Benson 1993: viii).

An entry consists of the following elements which are always in the same order (ibid: xvi):

- aspectual forms
- translation of the meaning(s) of the verb
- the collocation proper presented with the help of Russian question words
- an illustrative example or examples within parentheses translated into English

Different collocational constructions and their variants are marked by Arabic numerals (1, 2, etc.) while different meanings of verbs are marked by capital letters (A, B, etc.). Idioms and figurative expressions are preceded by an asterisk and whenever possible an appropriate English idiom is used in the translation.

## A

**АГИТИРОВАТЬ** (несов.) A. to agitate, campaign, push 1. А за кого/что (А за кандидата to campaign for a candidate; они агитируют за реформу they are agitating/campaigning/pushing for reform) 2. А против кого/чего (мы должны агитировать против этого кандидата we must campaign against this candidate; А против чьего предложения to campaign against smb.'s proposal) 3. (rare) A + неопр. ф. (А провести референдум to push for a referendum) 4. А (за то, чтобы с придат. (они агитировали за то, чтобы выборы были свободными they were pushing for free elections) 5. misc. А среди студентов to campaign among students B.

**САГИТИРОВАТЬ** (сов.) *Note:* The translation depends on the aspect; in the imperfective, the verb means to urge; in the perfective, it is usually translated by to persuade, talk into 1. А кого за что (агитировали нас за новый проект they urged us to support the new project) 2. А кого на что (она сагиттировала нас на поход в горы she talked us into going on a hike in the mountains) 3. А (кого) чем (сагиттировать кого всемозможными средствами to win smb. over with all sorts of inducements) 4. А + неопр. ф. (её агитировали подписать обращение they urged her to sign the petition; её сагиттировали подписать обращение they persuaded her to sign the petition)

**АДАПТИРОВАТЬ** (сов. и несов.) see ПРИСПОСАБЛИВАТЬ

**АДАПТИРОВАТЬСЯ** (сов. и несов.) see ПРИСПОСАБЛИВАТЬСЯ

**АДРЕСОВАТЬ** (сов. и несов.) to address, direct, send 1. А что кому/чему (А письмо директору музея to address a letter to the director of a museum; А вопрос докладчику to direct a question to a speaker; А послание правительству to send a message to the government) 2. (formal) А что куда на чьё имя (А телеграмму в институт на имя ректора to direct a telegram to the head of an institute)

**АККЛИМАТИЗИРОВАТЬ** (сов. и несов.) see ПРИСПОСАБЛИВАТЬ

**АККЛИМАТИЗИРОВАТЬСЯ** (сов. и несов.) see ПРИСПОСАБЛИВАТЬСЯ

**АККОМПАНИРОВАТЬ** (несов.) to accom-

pany 1. А кому (А скрипачу to accompany a violinist) 2. А на чём (А певцу на рояле to accompany a vocalist on a piano)

**АККРЕДИТОВАТЬ** (сов. и несов.) to accredit 1. А кого где/куда (дипломат был аккредитован в Париже the diplomat was accredited to Paris; корреспондент был аккредитован на чемпионат мира по футболу the correspondent was accredited to cover the world football BE/soccer AE championship) 2. Аккредитоваться (to be accredited)

**АКЦЕНТИРОВАТЬ** (сов. и несов.) to concentrate 1. А что на чём (А внимание на какой проблеме to concentrate one's attention on a certain problem)

**АЛИКАТЬ** (несов.) (obsol., lit.) to thirst for A

чего (А познаний to thirst for knowledge)

**АМПУТИРОВАТЬ** (сов. и несов.) to amputate, remove 1. А что кому (ампутировали ему ногу he had his leg amputated or his leg was amputated) 2. misc. А хирургическим ножом to remove surgically; А с помощью лазера to remove by laser

**АПЕЛЛИРОВАТЬ** (сов. и несов.) (formal) to appeal 1. А к кому/чему (А к здравому смыслу to appeal to common sense; А к общественному мнению to appeal to/enlist public opinion; А к чьей совести to appeal to smb.'s conscience) 2. А во что (А в высшую инстанцию to appeal to a higher court; А в областной суд to appeal to a circuit court)

**АПЛОДИРОВАТЬ** (несов.) *Note:* The perf.

**ЗААПЛОДИРОВАТЬ** is used to denote the beginning of an action) to applaud 1. А кому/чему (зрители бурно аплодировали артистам the audience applauded the performers loudly; or, — colloq.: the audience gave the performers a big hand) 2. misc. А бурно/горячо/громко to applaud loudly; они дружно заапплодировали they all started to applaud

**АРГУМЕНТИРОВАТЬ** (сов. и несов.)

(formal) to back up, support A что чем (А

выступление цифрами to back up a proposal

with figures/statistics)

**АРЕНДОВАТЬ** (сов. и несов.) to lease, rent

1. А что у кого (мы арендовали помещение

у них we leased office space from them) 2. А

что кому/чему (А театр организации to rent

a theater to an organization) 3. А что за что (А

помещение за пять тысяч рублей to lease

Fig. 27: Sample page from REDVC

### (3) *The Kenkyusha Dictionary of English Collocations (KDEC)*

*KDEC*, an English-Japanese collocation dictionary, is an extensive revision and vast expansion of Katsumata's *A New Dictionary of English Collocations* (1958). While the older edition lists 200,000 examples, *KDEC* comprises 380,000 collocations (Tsai 2000: 205).

According to Tsai compilation methods changed in the new version since a corpus was used for the revision, but some of the old problems still linger in this new edition, the most conspicuous of which is the inclusion of free combinations. Furthermore, there are still misplacements of collocators in the new edition, e.g. *apply for a patent* in the entry for *patent* (ibid: 207).

Another shortcoming of *KDEC* is that different senses of the headwords are separated by semicolons only, and users would have to browse through long entries while being very careful to get the information they need (ibid.).

To conclude the review, Tsai observes that using a corpus in the compilation of a collocation dictionary does not guarantee that the dictionary is useful (cf. *COBCOLL*, p. 68).

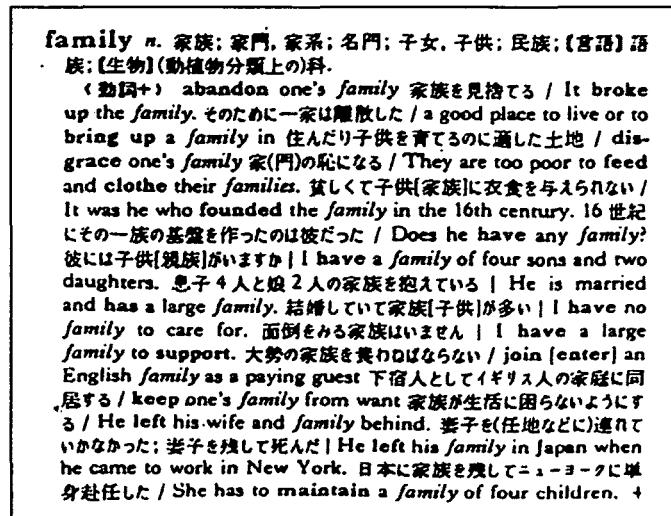


Fig. 28: Extract from KDEC

(4) A Russian-English Collocational Dictionary of the Human Body (RECDHB)

*RECDHB* is a thematic Russian-English dictionary which describes 63 parts of the body, plus two organs (heart, stomach), certain emissions (sweat, tears, blood), some physical manifestations of emotional states (laughter, smile), and, finally, voice<sup>13</sup>. This information is presented in a user-friendly format based on the format of the *Explanatory Combinatorial Dictionary of Modern Russian* (1984-) proposed by Igor A. Mel'čuk and Alexander K. Zholkovsky (cf. chapter 2)<sup>14</sup>.

<b>бедро hip</b>	
<a href="#">Semantics</a> <a href="#">Morphology</a> <a href="#">Syntax</a> <a href="#">Lexical Relationships</a> <a href="#">Sample Texts</a> <a href="#">Entire Entry</a>	
<b>Lexical Relationships</b>	
<ul style="list-style-type: none"> <li>• <a href="#">Synonyms</a></li> <li>• <a href="#">Diminutives</a></li> <li>• <a href="#">Syntactic derivatives</a></li> <li>• <a href="#">Parts</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Size and shape</a></li> <li>• <a href="#">Sustaining injuries</a></li> <li>• <a href="#">Movements</a></li> <li>• <a href="#">Clothing</a></li> <li>• <a href="#">Other expressions</a></li> </ul>
<b>Size and shape</b>	
ширина бёдер,	width of the hips
Чтобы худеть надо брать, и для этого знать ширину бёдер.	<i>I'll have to know the width of your hips to buy you pants.</i>
окруженность бёдер,	hip measurement
широкие бёдра	broad hips
узкие бёдра	narrow hips
округлые бёдра	rounded hips
полные бёдра	full hips
расплазаться/расплзаться в бёдрах,	to grow wider in the hips
тяжёлые бёдра	heavy hips
круглые бёдра	curvaceous hips
худые бёдра	bony hips
мальчишеские бёдра	boyish hips
<i>NB The above expression is said of narrow hips on a woman</i>	
жёночные бёдра	feminine hips
<i>NB The above expression is said of broad hips on a man</i>	

Fig. 29: Extract from RECDHB

<sup>13</sup> See their webpage at [http://russian.dmlc.cornell.edu/russian\\_web/BODY/WIN\\_1251/INDEX.HTM](http://russian.dmlc.cornell.edu/russian_web/BODY/WIN_1251/INDEX.HTM) [10/10/2004].

<sup>14</sup> The user of RECDHB, so the authors observe in their Preface, need know nothing about lexical functions themselves, but it is important to note that this dictionary is organized according to some reasonable principle — it is not merely a miscellaneous collection of phrases. Each entry in the dictionary contains a section entitled 'Lexical Relationships' and the headings and sub-headings under this rubric represent lexical functions (or their classes), viz. Mel'čuk's oriented lexical relationships (cf. p. 19).

Each entry consists of sections on semantics, morphology, syntax, lexical relationships and sample texts. The section on lexical relationships (see Fig. 29)<sup>15</sup> is the one containing collocations: “They are grouped together semantically under various headings, often with further subdivisions: one group consists of expressions describing the appearance of the body part, another is devoted to sensations, another to movements, etc.” (Lordanskaja and Paperno 1996<sup>16</sup>). *RECDHB* is available online, in CD-ROM format and in printed form.

There are yet another two bilingual collocation dictionaries which are conspicuous by their absence both on the internet and in western libraries. Benson et al. (1997: vii) explicitly state among their sources for the revised edition of the *BBI* two bilingual versions of the *BBI*, namely the Maruzen edition (Tokyo, 1993) and the *Longman Dictionary of English Collocations* (Hong Kong, 1995).

#### 4.2.1. Collocation Dictionaries of English and German

There are indeed two collocation dictionaries of English and German available, H. Bogatz’s *The Advanced Reader’s Collocation Searcher* (1997) and D. Siepmann’s *Advanced Learner’s Trilingual Lexicon. Dictionnaire thématique anglais-français-allemand. Politique-Economie-Expression du temps et de la quantification* (1997). While Bogatz’s dictionary is only available in CD-ROM format and merely translates the headword bases into German, Siepmann’s work is actually trilingual and covers only a very specialised area, namely politics and economy.

##### (1) *The Advanced Reader’s Collocation Searcher (ARCS)*

*ARCS*, as Horst Bogatz describes it<sup>17</sup>, is an electronic word combination searcher. So, like *COBCOLL* it only comes as a CD-ROM version. *ARCS* is primarily an English

---

<sup>15</sup> I deliberately left out the Russian side here since my browser does not support the Russian orthography.

<sup>16</sup> See the online version at [http://russian.dml.cornell.edu/russian.web/BODY/WIN\\_1251/INDEX.HTM](http://russian.dml.cornell.edu/russian.web/BODY/WIN_1251/INDEX.HTM) [10/10/2004].

<sup>17</sup> See his homepage at: <http://www.geocities.com/Athens/Acropolis/7033/> [7/12/2004]

monolingual collocation dictionary in as much as it lists English node words and collocates and provides the German translation of the headwords only.

Within its electronic format *ARCS* covers more than 40,700 nodes which are linked with their collocates to roughly 1.4 million examples; moreover, *ARCS* can be searched globally starting from either the English headword (adjective, noun, adverb, verb) or the German translation. This search facility indeed makes *ARCS* a useful tool for learners of English. However, the learner has to have prior semantic knowledge of the English combinations since no glosses (or translations) are provided for the meaning of the collocations.

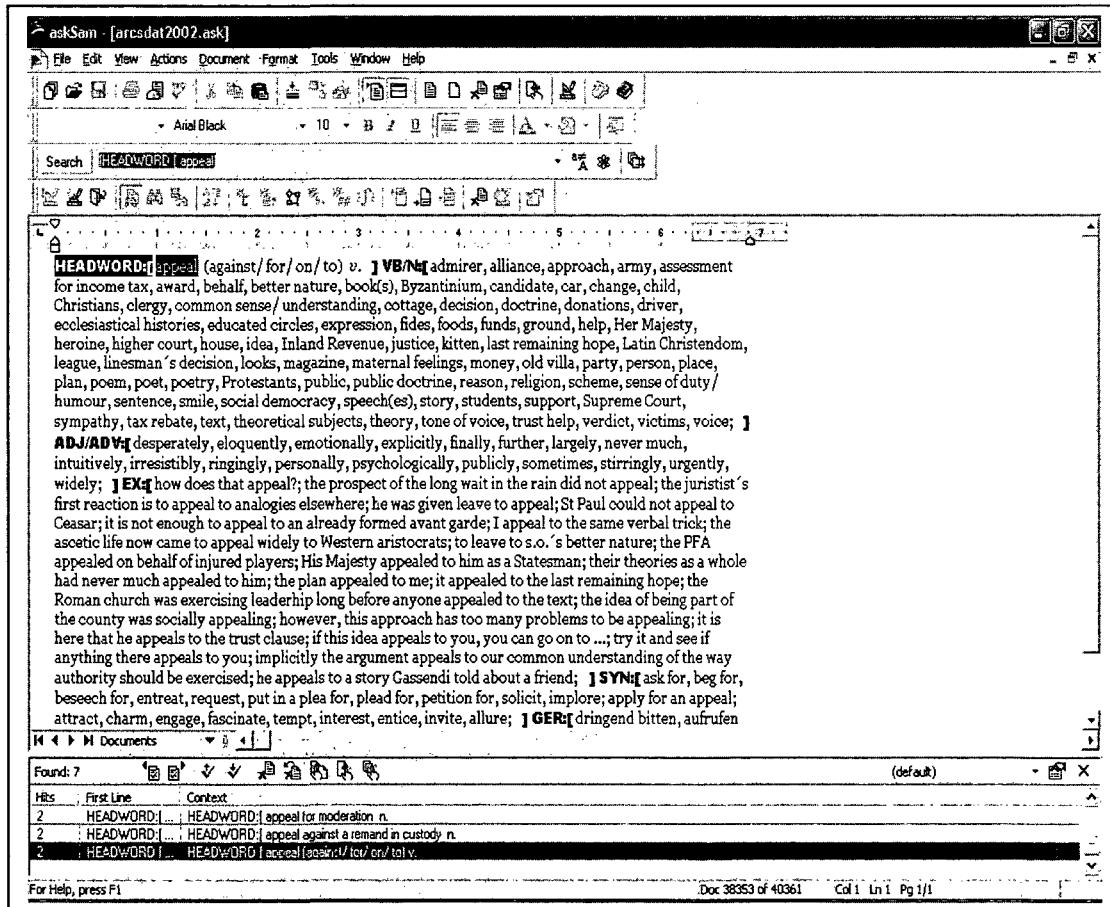


Fig. 30: Screenshot of an ARCS sample entry

(2) Advanced Learner's Trilingual Lexicon. Dictionnaire thématique anglais-français-allemand. Politique-Economie-Expression du temps et de la quantification (ALTD)

Dirk Siepmann's *ALTD* is a trilingual (print) dictionary of English, German and French and its intended target groups are post-intermediate learners of any of the three languages as well as translators, interpreters and people from political, economic and sociology backgrounds.

*ALTD* is based on a specialised corpus of political and economic texts running into millions of words – and it is this corpus-based approach that made it possible to focus on useful word combinations. Further, it is organised into three chapters: (1) politics, (2) economics and the environment and (3) time and quantification. These chapters are further divided into sections, i.e. the vocabulary in *ALTD* is categorised by topics, e.g. The Executive Power, Banking, etc<sup>18</sup>.

Excerpt from the chapter on <i>conflict</i> :		
people's dissatisfaction / disgruntlement (with) general dissatisfaction	le mécontentement populaire / des masses (de) un ras-le-bol général	die Unzufriedenheit in der Öffentlichkeit (mit) eine allgemeine (Staats-)Verdrossenheit der soziale Unmut
social discontent / dissatisfaction mounting frustration(s) // tension	le mécontentement / malaise social // la grogne des masses des frustrations // tensions croissantes	wachsende Frustration(en) // Spannung(en) die aufgestaute Verzweiflung die innere Wut die aufgestaute Gewalt eine Welle der Unzufriedenheit / Klagewelle
accumulated despair inner rage pent-up violence an outpouring of discontent	l'accumulation de désespoir la colère intérieure la violence intérieure un flot de mécontentement	starken Unmut auslösen
to create great dissatisfaction	susciter un sérieux mécontentement	seinen Unmut zum Ausdruck bringen
to voice one's dissatisfaction	exprimer son mécontentement	seinen Klagen Luft machen (über)
to air one's grievances (against)	manifester / exposer ses griefs // exprimer ses doléances (à l'égard de)	
to deepen dissatisfaction the general unrest in the town a flammable // tense situation	aggraver le mécontentement le bouillon général qui agite la ville une situation explosive // tendue	die Unzufriedenheit verstärken die allgemeine Unruhe in der Stadt eine explosive // angespannte Situation
a divided / sundered nation / society a nation split between the rich and the poor the social fabric is being torn apart social divisions a fractured country to sharpen class divisions	une nation divisée // société éclatée une nation divisée entre les riches et les pauvres le tissu social se déchire / se délite la fracture sociale un pays coupé en deux agraver les divisions sociales	eine geteilte Nation / gespaltene Gesellschaft eine in Arm(e) und Reich(e) gespaltene Nation das Sozialgefüge bricht auseinander der Bruch in der Gesellschaft ein gespaltenes Land die Klassengegensätze verschärfen eine gelähmte Gesellschaft
a crippled / paralysed society	une société bloquée / paralysée / en panne	

Fig. 31: Extract from ALTD

<sup>18</sup> For further details see Siepmann's homepage at <http://www.dirk-siepmann.de/Publications/publications.html> [7/12/2004]

It is interesting to note that the number of monolingual and bilingual collocation dictionaries, both general and specialised, is rather small when compared to the huge number of dictionaries of idioms or proverbs available almost all over the world.

#### **4.2.2. Proposals for Bilingual Collocation Dictionaries**

As far as I am aware, there are only two proposals for specific bilingual collocation dictionaries actually formulated. These are A.F. Abu-Ssaydeh's suggestion for an Arabic English collocation dictionary (1995) and D. Gabrovšek's proposal for an interlingual collocation dictionary of English and Slovene (1998a).

D. Gabrovšek considers 12 factors which should be taken into account when creating a bilingual collocation dictionary (Gabrovšek 1998a: 137-138):

- (1) A bilingual collocation dictionary is essentially a dictionary for encoding, viz. text production.
- (2) The intended user groups are advanced language learners as well as translators, especially for the written mode since, here, mother-tongue interference problems are much more likely to occur (i.e. when the translator sees his or her L1 collocation and simply translates the collocation word-by-word into the L2) than in spontaneous speech.
- (3) Slovene collocations are placed on the left-hand side – where Slovene users would typically start, while the English equivalents are to be found on the right-hand side.
- (4) The inclusion policy should be based upon contrastive considerations, i.e. included should be collocations that do not readily translate into English.
- (5) Headwords should be for bases since it is normally the base from which users start looking at collocations – in text production. Furthermore, collocators could be listed as an index in the back matter, with cross-references to the base entries in the main part.
- (6) In the base entries, collocators should be group according to a clear scheme that has to be explained in the front matter, i.e. user's guide.

- (7) Borderline cases should be included.
- (8) Collocations should be illustrated in brief passages rather than lengthy examples.
- (9) The starting point for the organisation of the dictionary should be the user's mother tongue.
- (10) Headword lists can draw upon existing dictionaries.
- (11) The dictionary should illustrate the lexicographer's awareness of the open-choice vs. idiom principle dichotomy.
- (12) The compilation of the dictionary should involve active participation of both English and Slovene native speakers.

A.F. Abu-Ssaydeh started to compile data for his Arabic-English dictionary in the late 1980s; however, this process of compilation later encountered various obstacles – which is why the production of the dictionary has been delayed for an indefinite period (Abu-Ssaydeh 1995: 12).

In contrast to all the other collocation dictionaries I discussed so far, Abu-Ssaydeh's *Arabic-English Collocation Dictionary* would be the only one to include not only lexical and grammatical collocations but also idioms, such as *to have one's back to the wall*. Furthermore, he took a personal interest in synonymous and near-synonymous collocations, such as *bidaya – dawn (of an age, civilization, hope, love)*, *inception (of a project)*, *onset (of fever, winter)*, *outset (of a career)*. However, this feature in particular proved very tricky in the compilation process (ibid: 17).

From the organisational point of view, the *Arabic-English Collocation Dictionary* lists Arabic nouns as headwords. These are followed by an explanation of their meanings in both English and Arabic. Then, verbs and adjectives as well as prepositions that go with the headword noun are listed. Finally, idioms and idiosyncratic combinations are entered. If different verb and adjective collocators are synonyms or near-synonyms, their distribution is given in English (see paragraph above) (ibid: 20-21)

As the main source of data for the *Arabic-English Collocation Dictionary* Abu-Ssaydeh used English monolingual dictionaries as well as English-Arabic dictionaries, the *BBI*, Rodale's *Word Finder* and Cowie et al.'s *Oxford Dictionary of Current Idiomatic English*.

Having outlined collocation dictionaries in the monolingual and the bilingual framework, I now turn to a discussion of bilingual lexicography in more detail – in order to form the basis for a sound methodology of the compilation of a bilingual adjective-noun collocation dictionary of English and German.

### **4.3. Bilingual Lexicography with a Special Focus on English and German**

Our knowledge of the history of bilingual lexicography – and German-English lexicography in particular – is mainly based on the work of Hans-Peder Kromann, Theis Riiber, Poul Rosbach; Gabriele Stein; Franz Josef Hausmann and Margaret Cop (Kromann, Riiber and Rosbach 1991, Stein 1985 and Hausmann and Cop 1985).

The first language reference works with a Germanic component which can be compared to modern bilingual dictionaries date back to the 8<sup>th</sup> century. These so-called glossaries explained Latin words for ‘English’ speakers.

The earliest ‘genuine’ bilingual dictionary, a Latin-English one, originated in the 15<sup>th</sup> century, while the first English-German dictionary appeared only in 1706. The first German-English dictionary was published in 1716. Both dictionaries were compiled by Christian Ludwig (see Stein 1985 and Hausmann and Cop 1985 for a detailed account of the history of English-German lexicography).

#### **4.3.1. Principles of Bilingual Lexicography**

Bilingual dictionary compilation, according to Kromann, Riiber and Rosbach (1991: 2713-14), should include consideration of three vital aspects, namely the user aspect, the linguistic aspect and the empirical aspect.

The user aspect is concerned with WHO will be the user of the dictionary. The linguistic aspect focuses on WHAT kind of relations will be dealt with, i.e. equivalence

relations as well as relations on both the paradigmatic and the syntagmatic axes, etc. Finally, the empirical aspect unites the former two in order to tell the lexicographer HOW to establish relevant corpora and other material and best retrieve the information from them (in the light of the requirements set out by the analysis of both the user and the linguistic aspects).

### (1) The User Aspect

It was the L.V. Shcherba who first identified and reported on the important point of the user aspect resulting from his (and Matusevich's) work on a Russian-French dictionary in 1936. His famous work, first published in Russian in 1940, was not translated into English until 1995 (Farina 1995)<sup>19</sup>. In a nutshell, he arrived at the conclusion that one has to distinguish between translation, and therefore dictionaries, from the native language into a foreign language on the one hand, and translation from a foreign language into one's native language on the other. Keeping in mind that the creation of a bilingual dictionary actually involves two language pairs, thus, two different target groups, this results in the need for four distinct dictionaries when compiling a single bilingual reference work. In the words of Smolik (1969 [as quoted by Kromann et al. 1991]), who used the terms active vs. passive dictionary, for English and German speakers these are:

(a) For German speakers

1. active dictionary = translation from German into English ('encoding')
2. passive dictionary = translation from English into German ('decoding')

(b) For English speakers

1. active dictionary = translation from English into German ('encoding')
2. passive dictionary = translation from German into English ('decoding')

Fig. 32: Active vs. passive dictionaries

<sup>19</sup> Originally published in Russian as "Oypt obshchei teorii leksikografii" in the Journal of the Soviet Academy of Sciences, *Izvestiia Akademii Nauk SSSR, Otdelenie literatury i iazyka*, 1940, No. 3: 89-117

More methodically than Shcherba and Smolik, Williams (1959) had called attention to the functions along which bilingual dictionaries can be classified. Bilingual dictionaries can be either mono-functional, i.e. decoding OR encoding for only one language group, or bi-functional, which means that the bilingual dictionary is intended to be used by a target group both in encoding AND decoding. Likewise, a bi-directional (= two-way) and bi-functional dictionary would serve two language groups in both decoding and encoding.

To put it differently and perhaps more accurately, general-purpose bilingual dictionaries of German and English can (or should) have two functions, and two directions respectively, as regards the user aspect:

<p><b>Function 1: ENCODING</b></p> <ul style="list-style-type: none"> <li>- direction 1 = English – German (for English native speakers)</li> <li>- direction 2 = German – English (for German native speakers)</li> </ul> <p><b>Function 2: DECODING</b></p> <ul style="list-style-type: none"> <li>- direction 1 = English – German (for German native speakers)</li> <li>- direction 2 = German – English (for English native speakers)</li> </ul>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Fig. 33: Bi-directional and bi-functional dictionaries

In practical lexicographical terms encoding means that the users look up lexical items in their native language in order to get foreign-language equivalents. Consequently, metalexicographical information such as style labels and sense discriminators should be given in the users' L1<sup>20</sup>.

On the other hand, decoding suggests that unfamiliar foreign words are listed and explained in the users' mother tongue. Here, users start from the L2.

<sup>20</sup> Williams (1959: 251, as quoted in Kroman et al. 1991: 2715) argued that for the purpose of a bi-directional and bi-functional dictionary sense discriminators should be introduced for both languages, but he also admitted that that solution would not be realistic in the light of the space constraint in the print dictionaries – a problem which could be overcome with recent electronic versions of bilingual dictionaries, e.g. PC-Bibliothek on CD-ROM.

In my opinion, Shcherba's typology definitely constitutes a considerable advance in terms of user-friendliness and lexicographical practice; however, there were still lexicographers trying to take this typology of four "implicit" dictionaries even further. Zgusta et al. (1971: 299-300), for example, put forward the consideration of the dimension of 'purpose'. In this context, 'purpose' refers to whether the bilingual dictionary is addressed to the public, i.e. general translation, or whether it is intended for use in literary translation, business contracts, etc.

Similarly, Hausmann (1988: 138-39) is not convinced either of the four-part typology. For him, EACH language group ought to have four dictionaries; consequently, each language pair should cover as many as eight dictionaries:

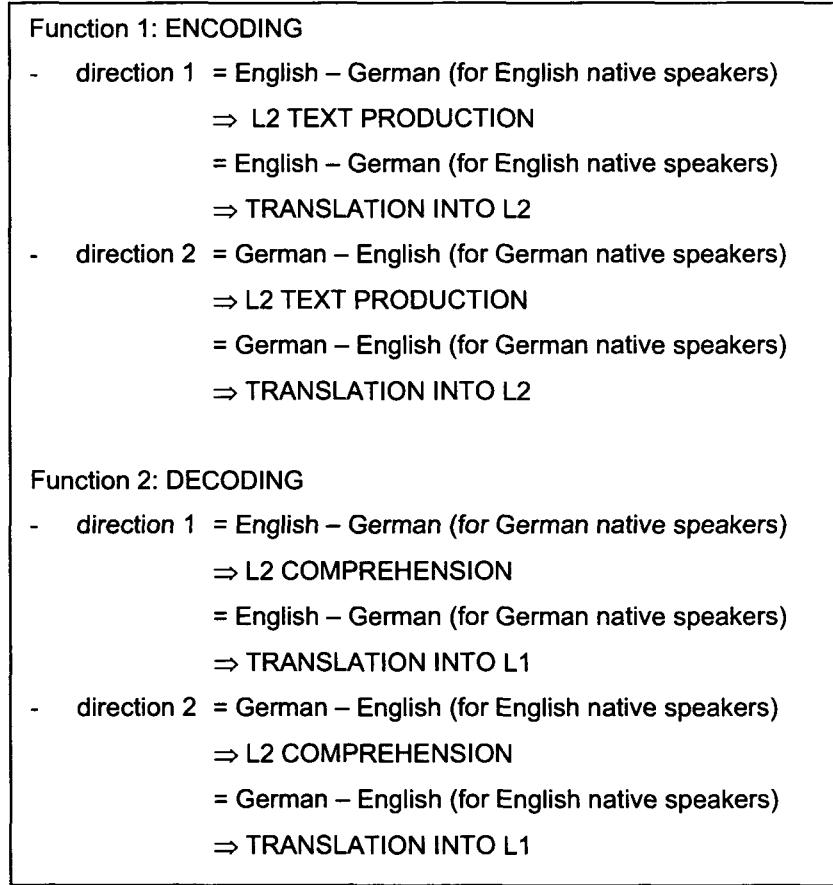


Fig. 34: Hausmann's typology of bilingual dictionaries

Further angles of the user aspect include recent studies of user sociology of the type “How and why do users consult dictionaries?”. Sue Atkins and Krista Varantola conducted such a survey which also comprised a section on bilingual dictionary use (Atkins and Varantola 1998: 111-14)<sup>21</sup>. Their results<sup>22</sup> revealed that out of 1,000 look-ups 71% were done in bilingual dictionaries while only 28% of the look-ups were observed in L2 monolingual dictionaries.

Interestingly, the percentages of the bilingual look-ups for an L1-L2 word translation do not vary considerably from the percentage of look-ups for an L2-L1 translation, although, naturally, L1-L2 translation, viz. encoding, is in the lead (L1-L2 = 72% vs. L2-L1 = 63%).

However, when the subjects were given an L1-L2 collocation task, only 40% used a bilingual dictionary while 60% consulted an L2 monolingual dictionary. However, it became clear that users started from bilingual dictionaries and only later turned to monolingual ones. Some users even went back to bilingual dictionary consultation after the monolingual look-up, either because they thought they must have missed something in the previous bilingual look-up or because they needed reassurance. It is especially this finding that was the most striking for Atkins and Varantola, “the amount of reassurance sought from [the] dictionaries, particularly about L2 collocation, by even the most skilled non-native speakers, however experienced in translation they may be” (*ibid*: 115).

Summarising the discussion of the user aspect, let me point out that my adjective-noun collocation dictionary is, in essence, a mono-functional English-to-German encoding dictionary. However, as I see it, a dictionary of this kind might also be used as a decoding help for German speakers, especially during a translation task when the German translator might be ‘too immersed in the English language’ to quickly find the correct L1 equivalent of the collocation in question.

---

<sup>21</sup> For another survey on user aspects as regards bilingual lexicography in general see Hartmann (1987). Research on collocations and dictionary use will be dealt with in section 4.3.2.

<sup>22</sup> They tested their methodology on 71 participants in 15 different language communities at the EURALEX *Workshop on Dictionary Use* held in Oxford in 1991.

Furthermore, an index of German nouns in the back matter of the dictionary could help both English decoding users as well as German encoding users. Let me illustrate this with a sample entry from my dictionary (Fig. 35):

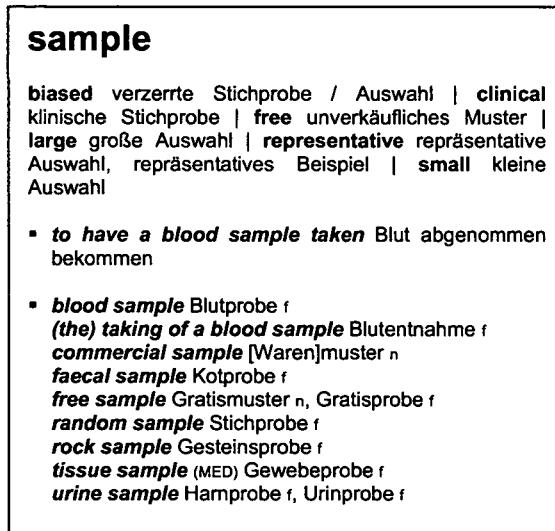


Fig. 35: The entry for *sample* in the English-German adjective-noun collocation dictionary

The English user encoding into German finds adjective collocators in the entry for *sample*. The collocators are listed in alphabetical order right after the headword, e.g. *biased sample* = *verzerrte Stichprobe*. These are followed by a section on English adjective-noun collocations whose German translations do not reflect the adjective + noun structure, e.g. *to have a blood sample taken* = *Blut abgenommen bekommen* (noun + verbal passive structure). In the subsequent section, the English user finds adjective-noun collocations that are rendered into German as ‘authentic’ compounds, i.e. spelled as one word, or German single words, e.g. *blood sample* = *Blutprobe* (*Blut* + *Probe*) or *commercial sample* = *Muster*.

From the grammatical information provided in Fig. 35 we can see that this is indeed an entry intended for English users rather than German ones since each German compound comes with its grammatical gender (which a German user would not need)<sup>23</sup>.

<sup>23</sup> On the other hand, they would not disturb German users in a German-English dictionary, and there is no such thing as grammatical gender in English anyway.

Furthermore, in the adjective-noun section, the adjective collocators are listed in their inflected form relative to the German noun.

The reason why the noun is listed in the German translation is that some English nouns, according to their senses, can be rendered into German with different nouns, e.g. *free sample* = *unverkäufliches Muster* vs. *large sample* = *große Auswahl*.

The index in the back matter created from the entry in Fig. 35 and the entry for *tendency*<sup>24</sup> would look like this:

GERMAN NOUN	ENGLISH ENTRY
Angewohnheit	tendency
Auswahl	sample
Beispiel	sample
Blut	sample
Blutentnahme	sample
Blutprobe	sample
Drang	tendency
Gesteinsprobe	sample
Gewebeprobe	sample
Gratismuster	sample
Gratisprobe	sample
Hang	tendency
Harnprobe	sample
Kotprobe	sample
Linksdrall	tendency
Mörder	tendency
Muster	sample
Neigung, -en	tendency
Rechtsdrall	tendency
Stichprobe	sample
Tendenz, -en	tendency
Trend	tendency
Urinprobe	sample
Veranlagung	tendency
Warenmuster	sample

Fig. 36: Sample index of the adjective-noun collocation dictionary

An index like the one in Fig. 36 can be used by both an English user decoding from German into English and a German user encoding into English – all it takes is a double look-up.

<sup>24</sup> For illustrative purposes I added the index for *tendency* here as well.

In an electronic version of the English-German adjective-noun collocation dictionary, queries would be possible not only starting from English and German nouns but also starting from English and German adjectives as well as collocations proper. So, the query for *free sample* would result in *unverkäufliches Muster / Gratismuster / Gratisprobe* and searches for both *repräsentatives Beispiel* and *repräsentative Auswahl* would list *representative sample*.

## (2) The Linguistic Aspect

Central to the linguistic aspect of bilingual lexicography is the notion of equivalence which I have dealt with in Chapter 3 (Contrastivity), already, but which I will discuss in more detail in section 4.3.2.

Further to equivalence, the linguistic aspect also covers issues like the degree of specialisation of the dictionary. This, in turn, slightly overlaps with the user aspect insofar as the specialisation definitely determines the target group<sup>25</sup>.

For the purpose of this thesis I chose only English adjective-noun or noun-noun collocations to be considered for inclusion in the bilingual collocation dictionary. I base my decision on the following factors:

- (1) Adjective-noun / noun-noun collocations are, together with verb-noun/noun-verb collocations, commonest types of lexical collocations. It is true that both types can be semantically unpredictable and, thus, can cause difficulties in encoding; however, adjective-noun collocations and nominal collocations are far more often subject to structural incongruence, viz. variation on the syntagmatic level, than are verbal collocations (at least in the English-German direction), e.g. *homicidal tendency* = *Drang zum Töten* (lit. trans. *tendency towards homicide*) which is a nominal phrase "turned" into the noun *Drang* and a post-modifying prepositional phrase *zum Töten* vs. to *commit suicide* = *Selbstmord begehen* (noun + verb). As I see it, difficulties in encoding begin with structural

---

<sup>25</sup> The target user groups in my case are advanced English-speaking learners of German who have reached creative-writing level as well as English-German translators who might run the risk of relying too much on the hypothesis of transferability (cf. 1.2.).

incongruence and then increase in semantic unpredictability<sup>26</sup>, e.g. *[this is] a feeble / misleading / poor comparison = der Vergleich hinkt* [noun + verb] (lit. trans. *the comparison walks with a limp*).

- (2) The fact that some English noun-noun combinations are regarded as compounds while others are lexicographically categorised as free combinations or collocations also makes topical the idea of an adjective-noun collocation dictionary. Naturally, combinations of nouns pre-modifying nouns have to be included in the collocation dictionary when they are not classified into the group of compounds or are free combinations. Chapter 5 will address this in more detail.
- (3) Yet another factor for choosing the adjective-noun structure was that it is easy to compare along different approaches, e.g. simple corpus linguistic vs. computational approaches, since in both German and English adjectives normally immediately precede the nouns they modify.

### (3) The Empirical Aspect

One of the major drawbacks bilingual dictionaries suffer from is their taking over inventories of entries from monolingual dictionaries – and merely translating them (Kromann, Riiber, Rosbach 1991: 2714). However, assessing collocations in the bilingual framework is a lot different from dealing with collocations from the monolingual point of view (cf. Gabrovšek 1998) – as we have seen already in Chapters 1 and 3.

Bilingual dictionary compilation has improved a lot over the last few decades; e.g. Atkins (1994) pioneered the use of two (comparable) monolingual electronic corpora as the reference for the compilation of the *Oxford-Hachette* (Corréard and Grundy eds. 1994). As far as I know, *Collins German* (Terrell et al. eds. 1999) and the two-volume English-Slovene *Oxford veliki angleško slovenski slovar* [forthcoming 2005-06] are the only dictionaries compiled along these lines.

---

<sup>26</sup> cf. Chapter 3

Using corpora in the dictionary compilation process means that what is displayed in the dictionary is not only the introspection of a team of lexicographers, but also actual language in use. It is true, as John Sinclair stated, that “language cannot be invented; it can only be captured” (Sinclair 1997: 31).

Some of the recent studies in contrastive phraseology highlight the use of parallel corpora, i.e. texts and their translations, from which one can easily extract multi-word units and their translation equivalents. For brief surveys see Maurer-Stroh (2003) and Fliedl et al. (2004)<sup>27</sup>. The drawbacks of this method lie, *inter alia*, in the limited availability of such corpora and the lack of appropriate software for analysis as well as in the fact that translations retrieved from a parallel corpus are only as good as the translations themselves; personal factors like exhaustion, pressure of time, boredom, etc. can highly influence the quality of the translation. This goes hand in hand with Dušan Gabrovšek’s “danger of taking parallel corpora as automatically the most reliable and useful sources for cross-linguistic analysis” (Gabrovšek, e-mail communication 2004; cf. Toury 1995: 275).

#### **4.3.2. Collocations and Bilingual Lexicography**

Collocations and their status in lexicography have been discussed a lot in the past years. Although bilingual and monolingual lexicographical approaches to collocations are hardly comparable along the same lines of thinking, they do have a common denominator, namely the question where to place a collocation in a dictionary, bilingual or monolingual.

Hausmann (1988), for example, sees the representation of collocational information as one of the three main problems bilingual lexicography is confronted with. Since adjective-noun collocations consist of two elements, should collocations be listed under the base (in this case the noun) or under the collocator (the adjective)?

---

<sup>27</sup> cf. sections 2.4. and 2.5.

### (1) Different Approaches to the Representation of Collocations in Bilingual Dictionaries

Cop (1990) distinguishes two approaches, the semasiological and the onomasiological ones. In a nutshell, the semasiological approach to the representation of collocations in the bilingual dictionary means that bases are listed in the collocator entries while the onomasiological approach suggests the representation of collocations in the base entry<sup>28</sup>.

Cop (1990: 38) sees the semasiological approach to collocations as an aid to meaning discrimination of the adjective. The meanings of the adjectives and, consequently, the different translations of them, are tailored to the respective base nouns. This approach is of equal relevance to decoding and encoding<sup>29</sup>. Let me illustrate this with the following examples:

<i>Decoding from English into German (for German speakers):</i>	<i>Encoding from German into English (for German speakers):</i>
<b>heavy</b> ADJ (Himmel) bedeckt   (Arbeit, Boden, Parfum) schwer   (Regen, Raucher) stark   (Seufzer) tief	<b>schwer</b> ADJ (work, soil, perfume) heavy   (disappointment, shock, storm) severe   (gold) solid
<i>Decoding from German into English (for English speakers):</i>	<i>Encoding from English into German (for English speakers):</i>
<b>schwer</b> ADJ (work, soil, perfume) heavy   (disappointment, shock, storm) severe   (gold) solid	<b>heavy</b> ADJ (Himmel) bedeckt   (Arbeit, Boden, Parfum) schwer   (Regen, Raucher) stark   (Seufzer) tief

Fig. 37: The semasiological approach to collocations in encoding and decoding entries

Note that for decoding purposes the meaning-discriminating information (in brackets) should be in the user's native language, while for the encoding process the bases are best given in the target language so that the user gets the full target-language collocation at once (cf. Pätzold 1991: 2963).

<sup>28</sup> cf. section 2.2.4.

<sup>29</sup> But is simply much more needed for encoding tasks (Gabrovšek 2004: e-mail communication).

Cop's onomasiological approach suggests that collocations should be represented in the base entry. For her, this is the perspective encoding dictionaries should adopt (Cop 1990: 40):

<i>Encoding from English into German (for English speakers):</i>	<i>Encoding from German into English (for German speakers):</i>
<b>perfume</b> N Parfum, Duft (angenehm) fragrant   (schwer) heavy   (zart) subtle	<b>Parfum</b> N perfume (fragrant) angenehm   (heavy) schwer   (subtle) zart

Fig. 38: The onomasiological approach to collocations in encoding entries

Heliel (1990: 134-35) points out that the collocator is the unpredictable element in a collocation and that the dictionary user, in encoding, usually knows the equivalent of the base (or they can easily be found in the bilingual dictionary). He or she is mainly concerned with finding the collocates and thus needs the onomasiological approach.

In contrast to Cop, Heliel (ibid.) sees the semasiological approach relevant only for decoding purposes since in comprehending a foreign-language collocation it is primarily the collocator that causes difficulties – which is, again, due to semantic tailoring.

Hausmann (1988: 149; cf. Bahns 1996: 47) notes that, basically, it makes no difference to the translator where he or she will find the collocation, under the base or under the collocator. Still, when the collocation is only listed in one entry, this often results in time-consuming double look-ups. Hausmann thus arrives at the conclusion that a collocation should ideally be found in all four entries: base and collocator of the source language and base and collocator of the target language. However, limited space in printed dictionaries makes this an impossible option since “general bilingual dictionaries have a much larger job than just presenting collocations” (Cop 1991: 276)<sup>30</sup>.

<sup>30</sup> The space problem can be easily overcome by the implementation of the dictionary on a CD-ROM. Furthermore, I am of the opinion that electronic versions of bilingual dictionaries should have a special collocation feature, such as the one in the *Longman Dictionary of Contemporary English on CD-ROM*.

Looking at the four major German-English / English-German dictionaries on the market (cf. sections 1.3. and 3.5.1.) we can observe that, first of all, on a purely quantitative basis, in the best of all cases in a German-English encoding translation task of adjective-noun collocations, six German collocations out of 38 are listed under their noun bases on the German side. 16 collocations were found in the adjective, the collocator, entry though. For the encoding translation of 29 English-German collocations only one collocation was listed explicitly in the noun entry on the English side of the bilingual dictionaries in question, while the best dictionary managed 15 collocations given in the adjectives entries.

Thus, we might assume that in German and English lexicographical practice the semasiological approach to the representation of collocations prevails; however, if we look at the different user's guides, we will learn otherwise.

With regard to adjective-noun collocations the *DUOX* user's guide informs us that "with adjectives, collocators are nouns which the translations typically qualify. They are normally placed after the translation." (Thyen et al. eds. 1999: 25). In the German-English part, collocators are in English; in the English-German part they are in German.

We see that *DUOX* treats adjectives as bases and nouns as collocators. This could lead to the incorrect assumption that in *DUOX* collocations are listed under their bases – a vital prerequisite for an encoding dictionary as suggested by Hausmann (1988), Heliel (1990) and Cop (1991). Without wishing to belittle Thyen et al.'s work, I would prefer them to have applied standard terminology – since, in adjective-noun collocations, the noun IS the base specifying the meaning (and translation) of the pre-modifying adjective in the first place.

Likewise, in the user guide to *COGER*, Terrell et al. (eds. 1999: xiv-xv) point out, under the heading Explanatory Material, that collocators and typical complements are, in contrast to sense indicators, not in brackets but in italics. In adjective entries the user will find typical nouns that the adjective modifies. They are usually found before the translation. Again, adjectives are treated as bases and nouns as collocators.

In *PONS*'s user's guide, we find that "context elements, also called collocates, are given in italics" (MacKenzie, ed. 2003: xiii). Further, we find in adjective entries nouns that are typically modified by them, while in noun entries typical genitive complements can be found (ibid: xiv). Consequently, *PONS* too lists adjective-noun collocations under their ('real') collocates and not bases.

Interestingly, *LANGE*, which is an abridged version of *Der Kleine Muret-Sanders*<sup>31</sup> – which would better compared to the three other 'Großwörterbücher'<sup>32</sup> – does not state anything about collocational treatment in neither of the two user's guides (*LANGE* is actually composed of two different dictionaries, edited by S. Brough and H. Messinger).

For my English-German adjective-noun collocation dictionary I adopted the onomasiological approach, i.e. adjective-noun collocation are presented in the base entry.

## (2) Equivalence

As already hinted at in section 4.3.1., equivalence is "the axis about which the activity of translation turns" (Kromann et al. 1991: 2717). There are three types of equivalence: full equivalence, partial equivalence and zero equivalence.

Full equivalence of adjective-noun collocations in the interlingual light is indicative of the fact that, although the collocational ranges of source- and target-language items might differ, there are overlaps to be found, which results in the easiest-to-encode or easiest-to-decode category: predictable and lexically congruent collocations (cf. Heliel 1990: 131, Maurer-Stroh 2004a: 5), e.g. *heavy sea* = *schwere See* and *unexpected discovery* = *unerwartete Entdeckung* (all four elements in the collocations are literal translations).

---

<sup>31</sup> While the revision of the English-German part was published in 2001, the German-English part of *Der Kleine Muret-Sanders* has not been recently revised (the latest edition dates back to the 1980s) – the information I obtained from [www.langenscheidt.de/deutsch/produkte/wb/ms-gesamt.pdf](http://www.langenscheidt.de/deutsch/produkte/wb/ms-gesamt.pdf) [12/10/2004]

<sup>32</sup> However, an "inter-Langenscheidt" comparison carried out by myself revealed a similar treatment of collocations.

Partial equivalence can be observed when one of the components of the adjective-noun collocation cannot be literally translated into another language, e.g. *heavy smoker* vs. *starker* (= ‘strong’) *Raucher* and *feverish condition* = *fieberhafte Erkrankung* (= ‘illness’).

In adjective-noun collocations the item which cannot be rendered directly into the other language is usually the adjective – which can be seen as a consequence of the semantic tailoring of the adjective to the noun respectively in both languages in question<sup>33</sup>.

Finally, zero equivalence means that neither of the two elements of the adjective-noun collocations can be LITERALLY translated, in an adjective-noun structure or otherwise. Examples of zero equivalence include idioms (also: endocentric compounds, cf. 2.3.3.) as well as structurally incongruent collocations, e.g. *red tape* = *Bürokratie* and *a recent discovery* = *eine Entdeckung neueren Datums* (lit. trans. *a discovery with a new date*).

Equivalence, as a technical term in general bilingual lexicography, is discussed in more detail in Kromann et al. (1991: 2717-18) and Svensén (1993: 140-62). I will, however, not go into too much detail here since the different concepts of partial equivalence, viz. interlingual anismorphism<sup>34</sup>, convergence and divergence, are of little relevance to my work as, with collocations, collocators act as sense discriminators.

Put briefly, in bilingual general lexicography, convergence and divergence call for decisions in the meta-lexicographical light, i.e. concern sense discriminators.

We talk about encoding convergence when the L1 lexical item has more lexical realisations than the L2 item, e.g. (for English speakers encoding from English into German) *stag night* + *hen night* = *Polterabend* (= the term used in German for both the bride- and the groom-to-be to celebrate their farewell to singleness).

Here, no sense discriminators have to be given as *hen night* and *stag night* constitute two different entries (in an adjective-noun / noun-noun collocation dictionary like the

---

<sup>33</sup> cf. Chapter 3.

<sup>34</sup> Full equivalence is referred to as (interlingual) isomorphism.

one I have in mind the two combinations will be put together under the headword *night*, listed alphabetically in the ‘authentic’ [i.e. spelled as one word] compound section).

In a general-purpose bilingual dictionary, decoding convergence has to be marked metalexicographically since one L2 translation corresponds to two L1 entries, e.g. (for German speakers decoding from English into German):

**news**

1. (generelle neue[re] Ereignisse, lit. trans. *recent events in general*) *Neuigkeiten*
2. (neue[re] Ereignisse, die in der Zeitung oder dem Fernsehen behandelt werden, lit. trans. *recent events in the newspapers or broadcast*) *Nachrichten*

Fig. 39: An example of decoding convergence (En-De)

Likewise, divergence means that the L2 lexical item has more sense realisations than the L1 item has. Therefore, in decoding divergence, the English speaker decoding from German into English will get two entries for *news*, i.e. *Nachrichten* and *Neuigkeit(en)*.

In encoding divergence, the German speaker encoding into English would get one entry for *Polterabend* with sense discriminators:

**Polterabend**

(of the bride) *hen night*; (of the groom) *stag night*

Fig. 40: An example of encoding divergence (De-En)

Nonetheless, as I have pointed out before, for my English-German adjective noun collocation dictionary I do not have to bother with divergence, convergence and sense discrimination since collocators (and their translations) normally serve as ‘natural’ meaning discriminators, e.g. *forensic examination* (= thorough investigation) vs. *oral examination* (= test of one’s knowledge).

#### **4.4. Towards a Bilingual Adjective-Noun Collocation Dictionary of English and German**

So, what needs to be considered in the compilation process of a bilingual adjective-noun collocation dictionary of English and German?

- (1) Linguistic aspect – general-purpose vs. specialised
- (2) User aspect / target group considerations – which function and direction will the dictionary have?
- (3) Linguistic aspect – headword selection, e.g. which headwords, in my case nouns, will I deal with in the dictionary?
- (4) Empirical aspect – where will I get the collocations and their translations from? What will I include or exclude?
- (5) User aspect – how will I present the information?

Each of these questions will be answered in the next chapter.

## 5. Methods of Compiling an Adjective-Noun Collocation Dictionary of English and German

As Svensén (1993: 4) points out, the work on a dictionary cannot start until suitable materials to form the basis of the compilation process have been collected and selected.

Before deciding on the material, however, the lexicographer has to determine which type of dictionary he or she is going to produce. Principally, there are two decisions to be made: (1) monolingual vs. bilingual and (2) general-purpose vs. specialised.

A collocation dictionary is a specialised dictionary<sup>1</sup> since it only deals with one aspect of the headwords, namely their combinability patterns. Yet, an adjective-noun collocation dictionary is even more focused.

As already hinted at in sections 1.2. and 1.3. and discussed in Chapter 4, there are very few monolingual collocation dictionaries available while I am aware of even fewer bilingual ones. Although general bilingual dictionaries do contain collocational information, collocations as a whole are far from being comprehensively covered there since studies I carried out revealed that only up to about 80% of the adjective-noun collocations tested could be translated with the help of the four major German and English dictionaries on the market. Hence my decision to create a bilingual adjective-noun collocation dictionary.

Another reason for choosing the adjective-noun structure was that, in this way, I am able to compare different compilation methods including a fairly simple probability-based retrieval of adjacent (in this case adjective + noun) words from a German and English parallel corpus.

Next, the lexicographer needs to decide which function and direction the dictionary will have. For my English-German adjective-noun collocation dictionary I came to the

---

<sup>1</sup> There are, basically, two types of specialisation: linguistic (etymology, usage, proper names, pronunciation, etc.) and field (e.g. astronomy, cooking, advertising) (Gabrovšek 2004: e-mail communication).

decision that, in the absence of an English native-speaker lexicographer, I should compile an encoding English-into-German dictionary (for English users). In this way I, as the lexicographer, can decode into my mother tongue, thus not only relying on the corpus information or material taken from existing dictionaries (English and German) but also using my native-speaker intuition.

Naturally, an encoding English-German dictionary for speakers of English is a mono-functional and mono-directional work, yet, as I pointed out in section 4.3.1., it could also be used for decoding purposes by German speakers (since it is this process that actually takes place in the compilation process).

Then again, such a dictionary provided with an index of German nouns in the back matter could also be valuable in decoding for English speakers and encoding for German speakers, the only drawback being that a double look-up is necessary<sup>2</sup>.

This brings us to think of the format the dictionary will have, viz. will it be a print dictionary, an electronic one, or both? For the moment, I decided on a print version only, but to convert it into CD-ROM or online format is an easy step that could be done within days, since the contents and layout remain the same and all that has to be done in order for the dictionary to be searchable via various queries is to link all the adjectives and nouns in it with a searchable database.

As a result, the English user wanting to encode into German could perform their searches starting from an English noun or an English adjective; when wishing to decode from German into English, the English user simply switches the language in the interface. Likewise, the German user wishing to encode into English can start his or her query from a German adjective or noun, or, in decoding, simply switches the language and starts from the English nouns and adjectives.

The interface of my English-German adjective-noun collocation dictionary could look like this:

---

<sup>2</sup> Although, naturally, “encoding and decoding in a collocational dictionary can never really be on a par with each other, not even roughly, whichever way you look at it.” (Gabrovšek 2004: e-mail communication).

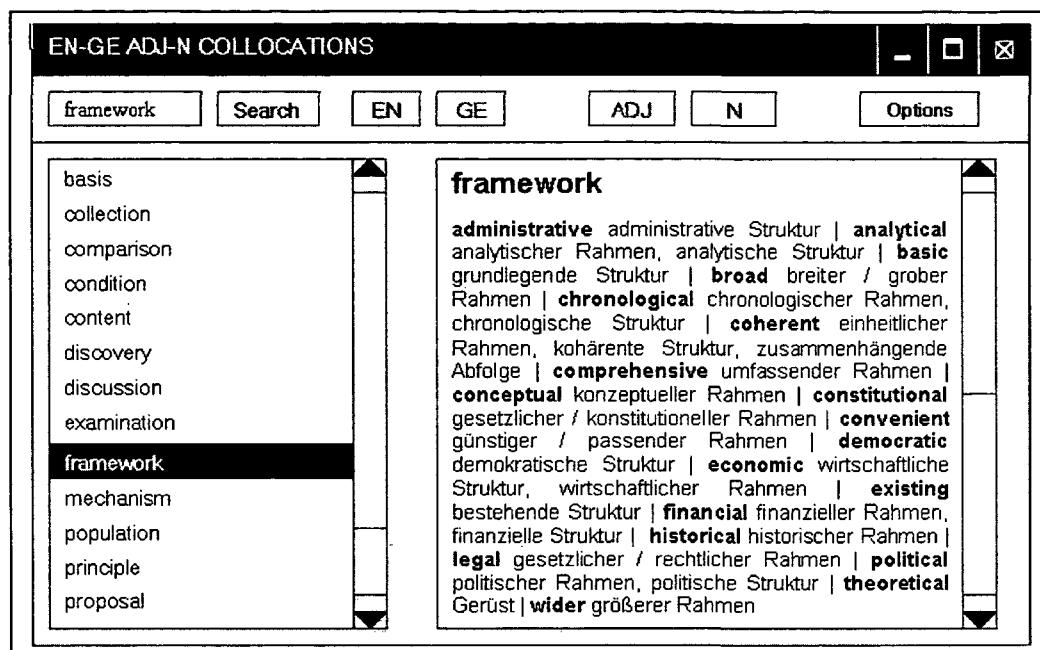


Fig. 41: Graphical user interface of an English-German adjective-noun collocation dictionary

The next question to be addressed before the actual compilation process is which headwords will be included in the dictionary. As we have seen in section 4.3.2., an adjective-noun collocation dictionary should, if the collocations are not listed under both constituents, best list collocations in the base entry, viz. the noun entry. So, the point at issue for my dictionary is where to get a list of nouns from. For illustration purposes in this thesis I compared Michael West's *General Service List* from 1953<sup>3</sup>, the *Longman Defining Vocabulary* from 1987<sup>4</sup>, and a list of nouns frequently used in academic writing (Hayes 1999: 53-55).

<sup>3</sup> As published, the *General Service List* is a medium-sized book organized like a dictionary. It is an inventory of 2,000 words selected to be of the greatest "general service" to learners of English. They are not the commonest 2,000 words, though frequency was one of the factors taken into account in making the selection. The version I used had been amended with word frequency information by John Bauman and Brent Culligan in 1995 and made available at <http://jbauman.com/aboutgsl.html> [13/10/2004].

<sup>4</sup> The *Longman Defining Vocabulary* is a list of roughly 2,100 words which the *Longman Dictionary of Contemporary English* 1987 uses to express all its definitions. As such, it is a useful source of basic vocabulary. This version is taken from <http://www3.cmp.uea.ac.uk/~irk/conlang.dir/LongmanVocab.html> [12/12/2004]. Although the website states 1988 as the original date they seem to refer to the 1987 version of the dictionary since there was no edition published in 1988. Note that the defining vocabulary is amended in each edition of the dictionary and listed in the back matter.

Since the *General Service List* and the *Longman Defining Vocabulary* contain not only nouns, I first had to find a way of retrieving only these nouns. What I did was have the two word lists tagged for parts of speech by Oliver Mason's *QTAG*<sup>5</sup>. I deleted all other word classes and came up with 1,137 nouns from the *Longman Defining Vocabulary* and 1,395 nouns from the *General Service List*. The intersection of the two lists and Hayes's list of nouns used in academic writing resulted in 22 nouns. These are: *basis, collection, comparison, condition, content, discovery, discussion, effect, examination, framework, influence, inquiry, measures, mechanism, population, principle, proposal, range, remedy, sample, study and tendency*. From these I randomly chose 16 nouns from which I created dictionary entries for the English-German adjective-noun collocation dictionary (see Appendix D).

In practice, that is, when compiling a dictionary of this kind that is to be published, the selection of headwords should be carefully made. I am, however, convinced that a combination of the nouns from the *General Service List* and the *Longman Defining Vocabulary*, together amounting to 1,692 nouns, would be a good starting point. In addition, the most frequent nouns from the British National Corpus, for example, could be compared to these 1,692 nouns. In this way, the headword list would contain not only 'easy' nouns from learners' lists but also more sophisticated, maybe technical, and up-to-date material.

Next on the list of decisions prior to the actual compilation process is the question of where I get the collocations and their translations from.

As Svensén (1993: 40) points out, the material to be taken into consideration should be authentic. Well, authenticity can be achieved by drawing on corpora and existing dictionaries which are based on a sizeable corpus of naturally occurring written and spoken text. For this purpose I decided to retrieve my collocational material from the four major English monolingual learner's dictionaries, *COBUILD-CD*, *LDOCE*, *MAC* and *OALD* – all corpus-based dictionaries. Additionally, I added the collocations listed in the three leading English monolingual collocation dictionaries, viz. the *BBI*, the *LTP*

---

<sup>5</sup> I downloaded the tagger from <http://web.bham.ac.uk/o.mason/software/tagger/> in May 2003, it seems, however, that it is no longer available from this address free of charge.

and the *OCD*. Further, I searched the British National Corpus for adjective-noun combinations through William Fletcher's *Phrases in English*<sup>6</sup>. All collocations that came up at least twice (e.g. in at least two dictionaries or one dictionary and the corpus) were candidates for inclusion.

As a next step, I conducted full-text searches on *COLLINS-E* and *DUOX-E* for the English nouns. All explicitly listed adjective-noun combinations were retrieved from these two bilingual dictionaries and compared with my list of English collocations. Again, the inclusion criterion was that the collocation had to be listed at least twice, i.e. in at least two dictionaries or one dictionary and the corpus.

For collocations which were not readily translatable with the full-text search in the two bilingual dictionaries I searched *DUST* and a German corpus that could be compared in both size and design with the British National Corpus, namely the corpus from *Digitales Wörterbuch der Deutschen Sprache*<sup>7</sup>. This procedure will be discussed in more detail and illustrated in section 5.2.

Finally, in the planning stage already, the lexicographer has to decide on the way the information should be presented to the user. The microstructure should demonstrate a clear organisation and help the user find the information they want in a straightforward way.

Having done a lot of research in the field of German and English adjective-noun collocations, I knew from the beginning that I would have to deal with not only structurally congruent collocations but also structurally incongruent ones (cf. Chapter 3). This makes it obvious that the entries should be given in more than one section, so that the user could see at once which is not rendered as adjective + noun into German. I decided on four sections:

- (1) The structurally congruent English-German adjective + noun section, e.g. *sparse population* = *schwache Besiedelung*.

---

<sup>6</sup> Available at: <http://pie.usna.edu/> [09/12/2004].

<sup>7</sup> Available at <http://www.dwds.de> [09/12/2004].

- (2) The structurally incongruent English-German adjective + noun section, e.g. a *cold remedy* = *ein Mittel gegen Erkältung* [lit. trans. *a remedy for cold*]. Also listed in this section are collocations which are different in their translation from section (1) when a particular verb goes with them, e.g. *drastic remedies* = *drastische Mittel* vs. *to apply drastic remedies* = (lit.) *drastische Mittel ergreifen*, (fig.) *mit eisernen Besen kehren* [lit. trans. *do the sweeping with an iron broom*].
- (3) Section 3 lists English collocations which occur in a particular frame, with a ‘slot’, e.g. within a prepositional phrase, such as *on a ... basis* or *in ... condition*.
- (4) Section 4 lists all English adjective-noun collocations which are to be rendered into German as ‘authentic’ compounds or simplex words, e.g. *general tendency* = *Grundrichtung* and *commercial sample* = *Muster*.

Note that collocations can be listed in more than one section, e.g.

(1) <i>drastic remedies</i>	drastische Mittel
(2) <i>to apply drastic remedies</i>	(lit.) zu drastischen Mitteln greifen
	(fig.) mit eisernem Besen kehren
(4) <i>drastic / kill-or-cure remedy</i>	Radikalkur, Rosskur
(1) <i>free sample</i>	unverkäufliches Muster
(4) <i>free sample</i>	Gratismuster, Gratisprobe

The methodology I have outlined in the above passages can be referred to as the corpus-based approach. The corpus-based approach to lexicography means that the corpus acts as a sort of complementary help in the compilation process, e.g. to check frequencies of words, look for their collocational behaviour, extract sample sentences, etc. The corpus-driven approach, on the other hand, is one where, in the compilation process, the corpus is the only source for information retrieval (cf. Tognini-Bonelli 2001: 65 and 84).

In the following sections I will present these two approaches in more detail.

### 5.1. The Corpus-Driven Approach

As already pointed out, the corpus-driven approach means that the corpus is the only source for information retrieval. So, which type of corpus do we need for the retrieval of bilingual adjective-noun collocations? The best suitable material for this task is a parallel corpus, i.e. texts and their translations, and suitable software.

How can the computer help in generating translation equivalents of adjective-noun collocations?

Shocking (and incomprehensible) as it may seem, the computer does not know that the German word *Amok*, for example, is the translation equivalent of the word *amok* in English. Tragically, the computer does not even know that, given the two languages, English and German, *Amok* has to be German because of the capitalisation of nouns.

Given that no dictionary or grammar whatsoever is implemented in the process of collocation retrieval, the minimum requirement for a bilingual extraction of word combinations is, as I have pointed out above, to have a parallel corpus.

Keeping in mind that the computer still has no idea how to correlate these two files, we have to provide small, then smaller, parallel chunks that the computer can deal with.

Starting from paragraph alignment (paragraph 1 in file 1 is paragraph 1 in file 2 – which is easy to detect by paragraph markers) and telling the computer that .!? end a sentence, logical algorithms concerning sentence length (if sentence 1 in file 1 has 30 characters it cannot be sentence 1 in file 2 with 5 characters) and approximate matching string similarity (which would generate *Parlament* and *parliament* in one sentence) can be applied to an instance of alignment on sentence level. The most widely used sentence alignment algorithm nowadays is the one proposed by K.W. Church and W.A. Gale<sup>8</sup>, which has been implemented, *inter alia*, in P. Danielsson and D. Ridings's *Vanilla Aligner* and Fieldl et al.'s *ANCR* – which I will discuss later on in more detail.

---

<sup>8</sup> See <http://acl.ldc.upenn.edu/J/J93/J93-1004.pdf> [9/12/2004].

When the two corpora are aligned on sentence level, there are, principally, two corpus-driven methodologies available for the retrieval of bilingual adjective-noun collocations, namely (1) making use of parts of speech and (2) making use of statistics.

To exemplify the part-of-speech approach to how to compile a bilingual collocation dictionary, I constructed an annotated specialised parallel corpus consisting of roughly 30 sentences<sup>9</sup>.

When writing the corpus texts I took care to use some adjective-noun combinations more often than others. I manually sentence-aligned the texts (the basic requirement) and annotated them: I tokenised them (as a first step to making the computer work more efficiently), lemmatised them (to avoid distortion by the German inflection system), and POS-tagged them (I assigned a part-of-speech to each lemma in order to have clues for parallel concordancing).

#### (1) Sentence Alignment and Tokenising

As pointed out before, in order for a researcher to be able to computationally work with texts and their translations, the texts have to be, in a first step, sentence-aligned. Another necessary precondition for the (linguistic) processing of texts is the 'tokenising', which means that words have to be separated from punctuation marks (see Fig. 42).

This can be done by simple replace-commands. The computer will, then, no longer take *rain.* and *rain* as two different occurrences of the lemma RAIN.

Problems occur when tokenising English abbreviated forms (such as *isn't* and *I'd*) and German colloquial forms (such as *Was gibt's?*) as well as dates. I will, however, not go into detail here since this is only done for the purpose of illustrating the corpus-driven part-of-speech approach.

---

<sup>9</sup> The texts can be found in Appendix C.

<p>Ein schlechtes Zeichen , denn der Wetterbericht sagt schwere Stürme voraus .</p> <p>Manchmal gibt es auch schwere Stürme .</p> <p>Ein schlechtes Zeichen , denn der Wetterbericht sagte schwere Stürme in dieser Gegend voraus .</p>	<p>A bad sign . The weather forecast is for severe gales .</p> <p>Sometimes there are even severe gales .</p> <p>A bad sign . The weather forecast was for severe gales in this region .</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Fig. 42: Sentence alignment and tokenising

## (2) Tagging & Lemmatising

To tag a corpus means to assign a part-of-speech label to each token (= word) in the text. Theoretically, the corpus can then be simply searched for patterns of, for example, adjectives preceding nouns.

However, an automatic tagger can only work accurately on a text file if it knows all the words in the file along with their inflectional behaviour (which can be quite demanding for German!). This is why taggers either have to be trained on existing already tagged corpora of the respective language (to proceed to do what they have learned), or a comprehensive grammar of the language in question has to be implemented – which can be a daunting task.

The process of tagging goes hand in hand with a preliminary lemmatisation of the tokens in a text. Lemmatising means that each token in a text is ‘transferred back’ to its root form, meaning that nouns are set to nominative singular, verbs to infinitives, etc.

<p>ein schlecht Zeichen , denn der Wetterbericht sagen schwer Sturm voraus .</p> <p>manchmal geben ich auch schwer Sturm .</p> <p>ein schlecht Zeichen , denn der Wetterbericht sagen schwer Sturm in dieser Gegend voraus .</p>	<p>a bad sign . The weather forecast be for severe gale .</p> <p>sometimes there be even severe gale .</p> <p>a bad sign . The weather forecast be for severe gale in this region .</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Fig. 43: Lemmatising

<p>1. ein_ARTIND schlecht_ADJ Zeichen_SUB ,_SZE denn_KONNEB der_ARTDEF Wetterbericht_SUB sagen_VER schwer_ADJ Sturm_SUB voraus_ADV ,_SZE</p>
<p>2. manchmal_ADV geben_VER ich_PERPRO auch_ADV schwer_ADJ Sturm_SUB ,_SZE</p>
<p>3. ein_ARTIND schlecht_ADJ Zeichen_SUB ,_SZE denn_KONNEB der_ARTDEF Wetterbericht_SUB sagen_VER schwer_ADJ Sturm_SUB in_PRP dieser_DEMPRO Gegend_SUB voraus_ADV ,_SZE</p>
---
<p>1. a_ARTIND bad_ADJ sign_SUB ,_SZE The_ARTDEF weather_SUB forecast_SUB be_VERAUX for_PRP severe_ADJ gale_SUB ,_SZE</p>
<p>2. sometimes_ADV there_ADV be_VERAUX even_ADV severe_ADJ gale_SUB ,_SZE</p>
<p>3. a_ARTIND bad_ADJ sign_SUB ,_SZE The_ARTDEF weather_SUB forecast_SUB be_VERAUX for_PRP severe_ADJ gale_SUB in_PRP this_DEMPRO region_SUB ,_SZE</p>

Fig. 44: Tagging

Collocations, in a monolingual corpus, are usually detected via several statistical measurements (cf 2.5.) When, for example, the actual co-occurrence of words in a corpus is higher than the random co-occurrence, the word combination can be called a collocation.

Quite in contrast to this, I will, for the most part of this approach to bilingual collocation retrieval, stick to simple absolute frequency counts.

Starting with a frequency list of the words in the German corpus, all lemmas which have not been tagged as a noun are excluded. A threshold should be set for the minimum number of occurrences of the source-language (= SL) node (in our case: the noun).

As a next step, each valid node will be searched for its preceding adjectives. This can be done by a POS-sensitive query: We take each node and attach to it the preceding lemma that has been annotated as adjective.

If we take the number six as the threshold for the minimum occurrence of the nouns to be taken into account, we get the following output for adjective-noun collocations in German:

<b>sentence</b>	<b>adjective</b>	<b>noun</b>
21	tobend_ADJ	Sturm_SUB
29	tobend_ADJ	Sturm_SUB
8	tobend_ADJ	Sturm_SUB
19	schwer_ADJ	Sturm_SUB
28	schwer_ADJ	Sturm_SUB
7	schwer_ADJ	Sturm_SUB
11	böig_ADJ	Wind_SUB
31	böig_ADJ	Wind_SUB
20	böig_ADJ	Wind_SUB
18	eisig_ADJ	Wind_SUB
27	eisig_ADJ	Wind_SUB
6	eisig_ADJ	Wind_SUB
30	stürmisch_ADJ	Wind_SUB
9	stürmisch_ADJ	Wind_SUB
20	stürmisch_ADJ	Wind_SUB

As a next step, we take the English adjective-noun concordances and connect them with the respective sentences in the German corpus.

As an optimum, each of the respective sentences in English and German only has one adjective-noun combination – as is the case with sentences 7, 19 and 28, for example. Consequently, the resulting adjective-noun collocations are translation equivalents:

German			English		
19	schwer_ADJ	Sturm_SUB	19	severe_ADJ	gale_SUB
28	schwer_ADJ	Sturm_SUB	28	severe_ADJ	gale_SUB
7	schwer_ADJ	Sturm_SUB	7	severe_ADJ	gale_SUB

If, however, a sentence has more than one adjective-noun combination, the computer cannot know which is the translation equivalent of which combination, as in sentence 21, for example:

21. eigentlich\_ADV mögen\_VERMOD ich\_PERPRO beide\_DEMPRO ,\_SZK  
ein\_ARTIND tobend\_ADJ Sturm\_SUB und\_KONNEB ein\_ARTIND  
angenehm\_ADJ Brise\_SUB ,\_SZE

21. i\_PERPRO actually\_ADV like\_VER both\_DEMPRO ,\_SZK a\_ARTIND  
fierce\_ADJ gale\_SUB and\_KONNEB a\_ARTIND pleasant\_ADJ breeze\_SUB  
,\_SZE

As a solution, one might divide the process of bilingual collocation retrieval into the following steps: (1) First, only sentences containing just one adjective-noun combination are validated. The resulting translation equivalents of this process have to be stored for the next step. (2) Second, the validated translation equivalents from the first step are used as a reference dictionary for cases where more than one adjective-noun collocation is found in a sentence. If the sentence contains two adjective-noun combinations, for example, and one of them has already been stored in the first process, then the remaining adjective-noun combination of the SL can be assigned to the remaining adjective-noun combination in the target language (= TL) as its translation equivalent. Again, these results have to be stored for later use. (3) The process of the

second step has to be continued until every ‘known’ adjective-noun pair has been generated and all stored and validated translations of adjective-noun combinations from step one and step two have been split into translation equivalents of adjective lemmas and noun lemmas. With the help of these, the remaining sentences can be processed.

Let me briefly exemplify these steps. Sentence 20, for example, contains three adjective-noun pairs:

```
wenn_KONUNT draußen_ADV ein_ARTIND Sturm_SUB tobten_VER ._SZK
bleiben_VER der_ARTDEF Leute_SUB zu_PRP Haus_SUB und_KONNEB
beobachten_VER der_ARTDEF stürmisch_ADJ Wind_SUB ._SZK der_ARTDEF
böig_ADJ Wind_SUB und_KONNEB der_ARTDEF stark_ADJ Regen_SUB
._SZE
```

```
when_KONUNT i_PERPRO blow_VER a_ARTIND gale_SUB outside_ADV ._SZE
people_SUB stay_VER at_PRP home_SUB and_KONNEB watch_VER
the_ARTDEF blustery_ADJ wind_SUB ._SZK the_ARTDEF gusty_ADJ
wind_SUB and_KONNEB the_ARTDEF heavy_ADJ rain_SUB ._SZE
```

As a result from the first step – where obvious translation equivalents were extracted from sentences only containing one adjective-noun collocation – we know that *stürmisch* + *Wind* is the translation of *blustery* + *wind* (sentence 9), which leaves us only two adjective-noun pairs to deal with:

```
._SZK der_ARTDEF böig_ADJ Wind_SUB und_KONNEB der_ARTDEF
stark_ADJ Regen_SUB ._SZE
```

```
._SZK the_ARTDEF gusty_ADJ wind_SUB and_KONNEB the_ARTDEF
heavy_ADJ rain_SUB ._SZE
```

Since *stürmisch*\_ADJ + *Wind*\_Sub and *blustery*\_ADJ + *wind*\_SUB are already stored in our database as translation equivalents, we can split these combinations and their translation into: *stürmisch*\_ADJ = *blustery*\_ADJ and *Wind*\_SUB = *wind*\_SUB.

As a result, we can disambiguate the above excerpt of sentence 20 by claiming that *böig* + *Wind* has to be the translation of *gusty* + *Wind*, because *Wind* is the translation of *wind* and not of *rain*. Consequently, we also get the translation equivalents *stark* + *Regen* and *heavy* + *rain*.

Both the monolingual and the bilingual collocation extractions have to be done starting from both languages, meaning that we have to go through both processes with either German or English as the SL. This is absolutely vital when considering the following case:

As a threshold, the minimum occurrence of the node nouns was set to six. Starting from German as the SL, we would get the following output:

German	English
strömend_ADJ Regen_SUB	pouring_ADJ rain_SUB
strömend_ADJ Regen_SUB	pouring_ADJ rain_SUB
strömend_ADJ Regen_SUB	pouring_ADJ rain_SUB
stark_ADJ Regen_SUB	heavy_ADJ rain_SUB
stark_ADJ Regen_SUB	heavy_ADJ rain_SUB
stark_ADJ Regen_SUB	heavy_ADJ rain_SUB

If, however, English was the SL, the output would be the following:

English	German
pouring_ADJ rain_SUB	strömend_ADJ Regen_SUB
pouring_ADJ rain_SUB	strömend_ADJ Regen_SUB
pouring_ADJ rain_SUB	strömend_ADJ Regen_SUB
heavy_ADJ rain_SUB	stark_ADJ Regen_SUB
heavy_ADJ rain_SUB	stark_ADJ Regen_SUB
heavy_ADJ rain_SUB	stark_ADJ Regen_SUB
torrential_ADJ rain_SUB	wolkenbruchartig_ADJ Schauer_SUB
torrential_ADJ rain_SUB	wolkenbruchartig_ADJ Schauer_SUB
torrential_ADJ rain_SUB	wolkenbruchartig_ADJ Schauer_SUB

The problem here is that some words have more than one translation equivalent and some translations are more frequently used than others. We see that *rain* can be translated as *Regen* or *Schauer*. *Regen* is used six times and *Schauer* only comes up three times (and as the threshold was set to six, it did not appear at all in the list of candidates taken from the German corpus).

In more practical terms, dictionary compilation along these lines would, if done strictly automatically, need a huge corpus and, most importantly, an accurate linguistic preprocessing.

An example of a program which could help with the semi-automatic corpus-driven compilation of an English-German adjective-noun collocation dictionary is a parallel concordancer, Michael Barlow's *Paraconc* (cf. 2.4.4.). With the help of *Paraconc*, the lexicographer could search for all noun occurrences of *rain* preceded by an adjective and validate them manually:

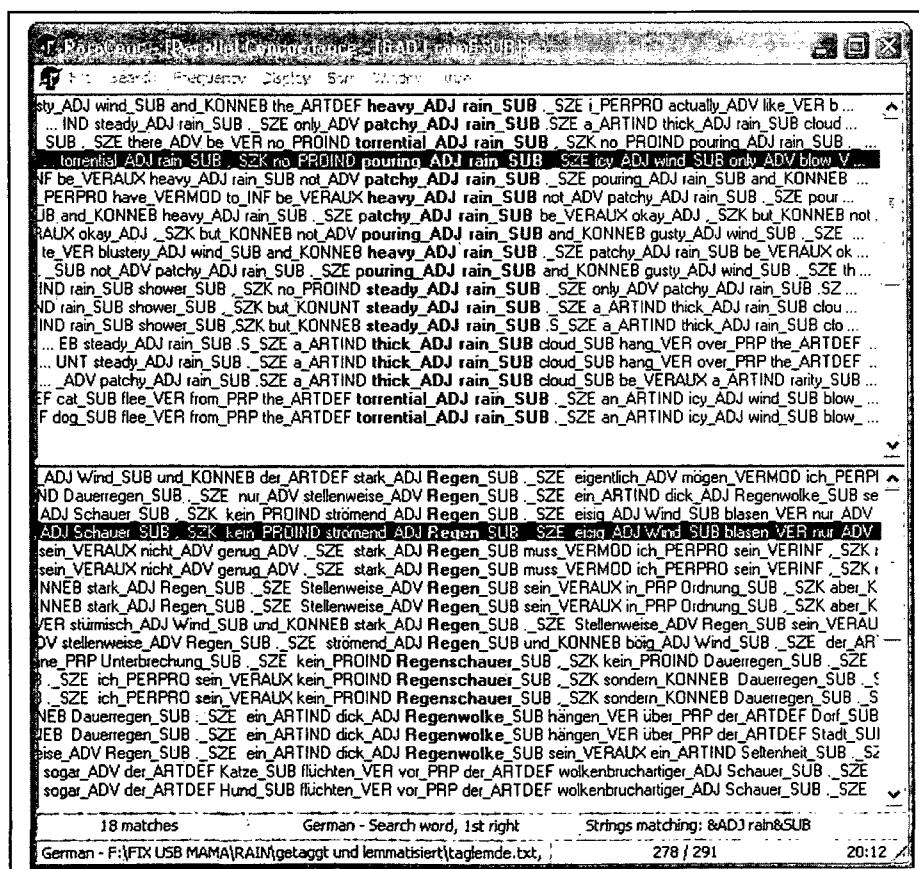


Fig. 45: Parallel concordance of *rain*

This methodology seems to work quite well with a small test corpus, where only 18 matches were found. Imagine, however, the workload the lexicographer would have to cope with when analysing a large corpus.

The fact that the corpus-driven approach based on parts of speech is highly dependent on the preprocessing software and that manual validation would be a tedious task, made me think of another methodology, namely a purely statistical approach without prior linguistic annotation.

It was my brother, Dr. Sebastian Maurer-Stroh, then at the Institute of Molecular Pathology in Vienna, who implemented this initial idea into a perl script. In summer 2003, the project *ANCR* (= Adjective-Noun Collocation Retriever) was thus launched as a joint undertaking of the Department of English Studies and the Department of Computer Science at the University of Klagenfurt, Austria.

### 5.1.1. ANCR - The Purely Statistical Approach

*ANCR* starts from English or German texts and their translations. With an in-built aligner based on the Church and Gale algorithm (cf. 5.1.) you can, in the *ANCR* environment, sentence-align your texts for further processing<sup>10</sup>.

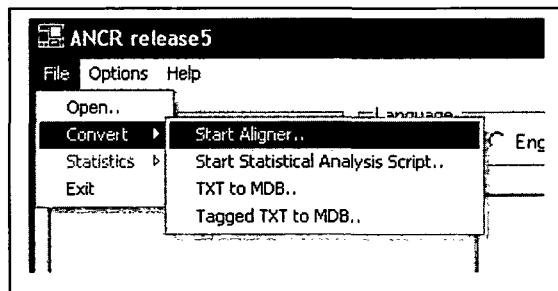


Fig. 46: ANCR – Calling up components

---

<sup>10</sup> For a detailed discussion of our aligner and the ANCR environment see A. Homa and W. Sühs (2004).

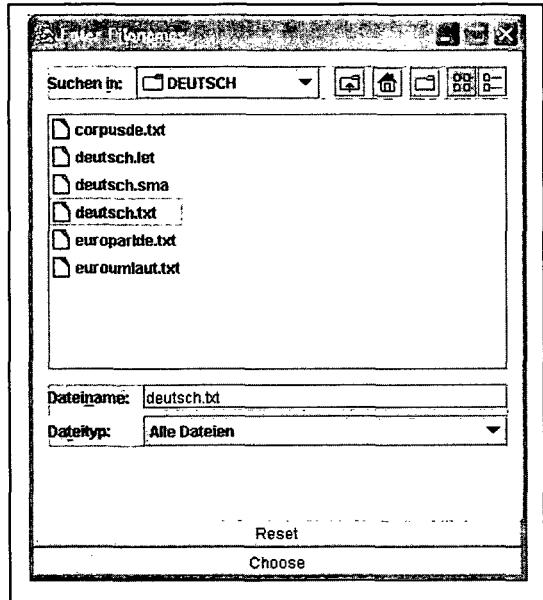


Fig. 47: ANCR – Aligner

At the moment we are using Philipp Koehn's Europarl and 'de-news' (both of which were already sentence-aligned and together amount to approximately 34 million words), but we plan to compile our own, more heterogeneous, bilingual corpus which should contain not only EU Proceedings and news, but also fiction, tourism texts, etc.

Having sentence-aligned the files, you call up the statistical analysis script – which was developed by Georg Weber.

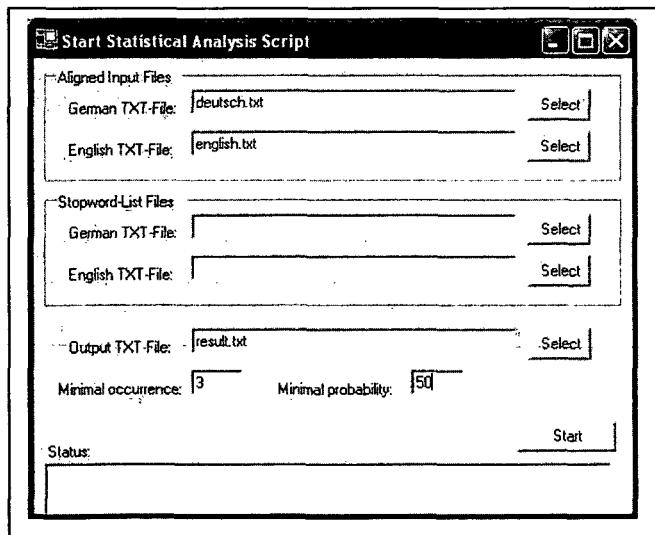


Fig. 48: ANCR - Starting the corpus analysis

Put simply, the program performs the following steps:

- (1) If you have not set the parameter for your own .txt stopword list, the program automatically generates a list of the 100 most frequent words from each corpus. N-grams (at the moment bi-grams and tri-grams, viz. two or three adjacent words) including stopwords will not be counted.
- (2) The program then reads in sentence-per-sentence in both files, i.e. sentence 1 in file 1, sentence 2 in file 2 etc. For each sentence only ‘allowed’ n-grams are retrieved. The filters we are currently using exclude n-grams with stopwords, numerals, symbols, special characters, strings with less than three characters and n-grams including capitalised English words – as these can be taken to be proper nouns. Additionally, only German bi-grams with the second word starting with the capital letter are included for further processing – this filter has been implemented since in German nouns – in contrast to other word classes – are capitalised; in this way, we can secure that only nouns and their immediately preceding elements are extracted<sup>11</sup>.
  - a. German and English words which are part of an n-gram are saved in temporary files (viz. a German and English one).
  - b. Also German and English n-grams are stored in temporary files.
- (3) In step 3 the program generates n-gram pairs by comparing each valid n-gram from step 2b with the respective target file (sentence-per-sentence). All parallel n-grams are counted and stored in a temporary file (sorted by their parallel frequency).
- (4) Now the probability is calculated for each parallel, i.e. bilingual, n-gram. For this, the absolute frequency of each German n-gram and German word in the German corpus is contrasted with the absolute frequency of each English n-gram and English word in the English file and compared to the parallel frequency list from step 3.

---

<sup>11</sup> Here, we work from the assumption that articles and pronouns have been filtered out by the stopword list already.

(5) In the final step, n-gram pairs with a lower probability than that set in the starting window will be excluded and the result is saved as a .txt file in the directory you chose.

We increased the calculated n-gram size from only bi-grams (adjective + noun) to also tri-grams since more than 60% of the errors in the first release of *ANCR* were due to the difference in German and English compounding, that is, that in German compounds are spelled as one word while in English they are usually spelled separately, e.g. *schlechte Arbeitsbedingungen* = *poor working conditions* (2:3 correspondence).

With *ANCR – release 5* we have achieved a 76.4% performance as opposed to 56.3% in the first version.

When the program has created the result file, it has to be converted into a database that can be read by the graphical user interface. For this, you simply use the ‘file – convert txt to MDB’ command. All you have to do now to start your analysis is to load the database into the interface.

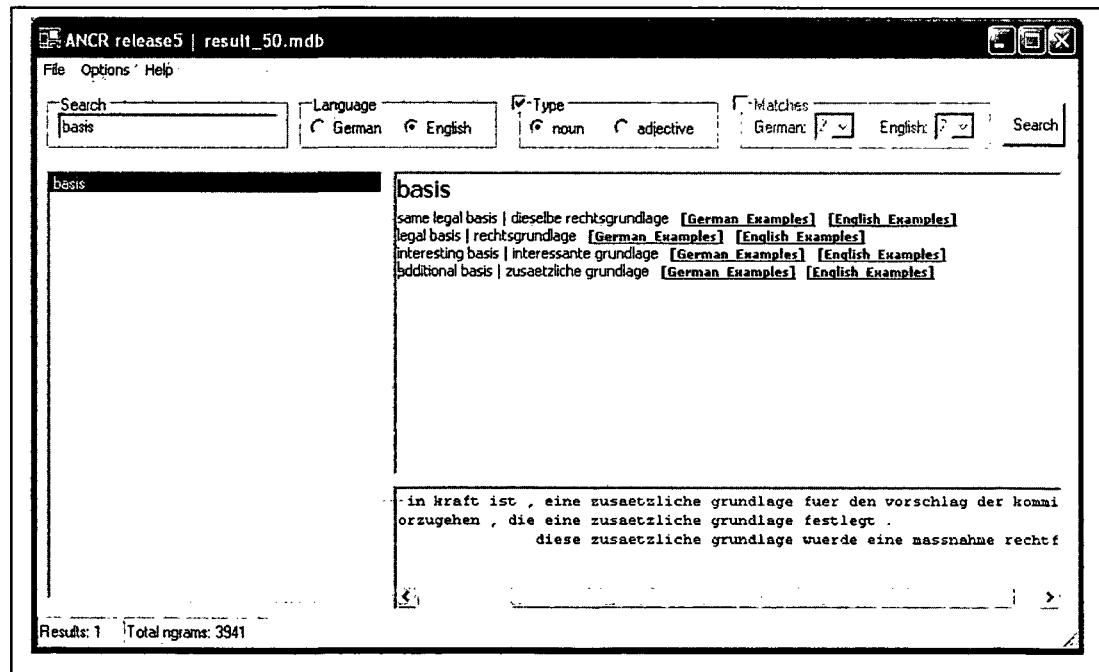


Fig. 49: ANCR – Search for the English noun *basis*

*ANCR* can be searched starting from English or German adjectives or nouns. Additionally, for purposes of contrastive analysis, *ANCR* can also list all 1-2, 2-1, 2-2, 2-3, 3-2, 1-3 and 3-1 occurrences in the analysed corpus (these correspondences have been saved during the probability analysis).

I tested *ANCR* for the 22 English nouns I extracted from the *General Service List*, the *Longman Defining Vocabulary* and Hayes's (1999) list of nouns used in academic writing and found out that, with a minimum probability of 20%, *collection*, *comparison*, *discovery*, *effect*, *examination*, *influence*, *range*, *remedy* and *sample* revealed no or incorrect results.

The remaining 13 nouns would result in the following corpus-driven English-German adjective-noun 'dictionary'.

<b>BASIS</b> additional zusätzliche Grundlage   clear klare Grundlage   good gute Grundlage   important wichtige Grundlage	<b>MECHANISM</b> 1. SG <b>adjustment</b> Anpassungsmechanismus   <b>allocation</b> Zuteilungsmodus   <b>defence</b> Abwehrmechanismus   <b>flexibility</b> Flexibilitätsmechanismus 2. PL <b>control</b> Kontrollmechanismen   <b>decision making</b> Entscheidungsmechanismen   <b>market</b> Marktmechanismen
<b>CONDITIONS</b> acceptable akzeptable Bedingungen   actual tatsächliche Bedingungen   American amerikanische Bedingungen   clear eindeutige / klare Bedingungen   <b>ecological</b> ökologische Bedingungen   <b>environmental</b> Umweltbedingungen   <b>licensing</b> Lizenzverfahren   <b>market</b> Marktbedingungen   <b>political</b> politische Bedingungen   <b>safety</b> Sicherheitsbedingungen   <b>transport</b> Transportbedingungen	<b>POPULATION</b> Austrian österreichische Bevölkerung   <b>black</b> schwarze Bevölkerung
<b>CONTENT</b> 1. SG <b>political</b> politischer Inhalt 2. PL <b>audiovisuell</b> auso visuelle Inhalte	<b>PROPOSAL</b> 1. SG <b>ambitious</b> ehrgeizige Vorschläge   <b>balanced</b> ausgewogener Vorschlag   <b>basic</b> Grundsatzvorschlag   <b>clear</b> klarer Vorschlag   <b>draft</b> Vorschlagsentwurf   <b>full</b> vollständiger Vorschlag   <b>important</b> wichtiger Vorschlag   <b>legislative</b> Legislativvorschlag 2. PL <b>amended</b> geänderte Vorschläge   <b>balanced</b> ausgewogene Vorschläge   <b>compromise</b> Kompromissvorschläge   <b>constructive</b> konstruktive Vorschläge   <b>counter</b> Gegenvorschläge   <b>excellent</b> ausgezeichnete Vorschläge
<b>DISCUSSIONS</b> difficult schwierige Diskussionen	<b>STUDY</b> 1. SG <b>case</b> Fallstudie 2. PL <b>case</b> Fallstudien   <b>comparative</b> vergleichende Studien
<b>FRAMEWORK</b> financial Finanzrahmen	<b>TENDENCY</b> dangerous gefährliche Tendenz
<b>INQUIRY</b> administrative administrative Untersuchung	
<b>MEASURES</b> administrative measures administrative Maßnahmen   <b>bilateral</b> bilaterale Maßnahmen   <b>control</b> Kontrollmaßnahmen   <b>environmental</b> Umweltmaßnahmen   <b>restructuring</b> Umstrukturierungsmaßnahmen	

Fig. 50: Adjective-noun collocations from *ANCR*

The resulting dictionary is clearly not overwhelming in size, but, judging from the results in Fig. 50 and the overall performance, we are definitely on the right track in further pursuing the *ANCR* approach.

Before I embark on the corpus-based approach, which I chose as the most appropriate method of compiling an English-German adjective-noun collocation dictionary, let me illustrate with a further example the fact that a corpus-driven approach to collocations is simply not enough for the time being – what with the software and corpora currently available (we do not know, however, what lies in store for us in the future in computational linguistics).

From *Projekt Deutscher Wortschatz* (cf. 2.5.) I collected all adjective collocators for the English noun *discovery* which their (common-birthday) statistical calculation retrieved from an English corpus of 266,799,551 tokens. Likewise, I used the same tool to retrieve adjective collocators for the German equivalent *Entdeckung* run on the 518,209,793-token German corpus<sup>12</sup>.

I translated each of the adjectives, English and German, with the help of their corpus-driven dictionary and based on my decisions of correct translation equivalents on pure quantitative findings, i.e. which equivalents occurred more often.

As a result, I was able to compile the following entry:

<b>discovery</b> Entdeckung f
accidental zufällige   archaeological archäologische   grisly grässliche, grausige   gruesome grässliche, grausige, furchtbare, schreckliche   important bedeutende, entscheidende, wichtige   interesting interessante, merkwürdige, seltsame, spannende   latest neueste   major bedeutende, größte, wichtig   most important wichtigste   most recent jüngste   recent jüngste, neueste   significant bedeutende, wichtige   startling aufsehenerregende   unexpected überraschende, unerwartete

Fig. 51: Corpus-driven entry for *discovery*

<sup>12</sup> These numbers are taken from an e-mail communication (October 2004) with Matthias Richter who is a member of the *Projekt Deutscher Wortschatz* team.

The result looks quite acceptable. If we compare it with the entry for *discovery* which I compiled along my – corpus-based – methodology, the drawbacks become obvious.

<b>discovery</b>				
amazing	erstaunliche Entdeckung			
archaeological	archäologische Entdeckung, archäologischer Fund   great	große Entdeckung		
grim	grauenvolle / grausige Entdeckung			
important	wichtige Entdeckung / wichtige Erkenntnis   incredible	erstaunliche /, (coll.) unglaubliche Entdeckung   major	wichtige Entdeckung   new (PL)	neue Entdeckungen   recent die jüngste / neueste Entdeckung   remarkable bemerkenswerte / verblüffende Entdeckung   revolutionary revolutionierende Entdeckung   scientific wissenschaftliche Entdeckung   significant bedeutsame / bedeutsame Entdeckung   startling aufregende / sensationelle Entdeckung   surprising überraschende Entdeckung   terrible (coll.) fürchterliche / schreckliche Entdeckung   unexpected unerwartete Entdeckung
<ul style="list-style-type: none"> <li>▪ <b>an inspired discovery</b> (fig.) das Ei des Kolumbus</li> <li>▪ <b>a recent discovery</b> eine Entdeckung neueren Datums</li> <li>▪ <b>new discovery</b> (lit. a. fig.) Neuentdeckung f</li> </ul>				

Fig. 52a: Corpus-based entry for *discovery*

<b>discovery</b> Entdeckung f	
accidental	zufällige   archaeological archäologische   grisly grässliche, grausige   gruesome grässliche, grausige, furchtbare, schreckliche   important bedeutende, entscheidende, wichtige   interesting interessante, merkwürdige, seltsame, spannende   latest neueste   major bedeutende, größte, wichtig   most important wichtigste   most recent jüngste   recent jüngste, neueste   significant bedeutende, wichtige   startling aufsehenerregende   unexpected überraschende, unerwartete

Fig. 52b: Corpus-driven entry for *discovery*

First of all, based on a merely quantitative inspection, we can see that Fig. 52a has more collocations listed in the entry. Second, in Fig. 52b, variation in the translation equivalent of the noun does not show, e.g. *archaeological discovery* = *archäologische Entdeckung* or *archäologischer Fund*. Next, German compounds could not be retrieved with the corpus-driven method, e.g. *new discovery* = *Neuentdeckung*. Finally, register labels can not be extracted from a corpus, but from existing corpus-based dictionaries.

## 5.2. The Corpus-Based Approach

The corpus-based approach (i.e. one using [comparable] corpora and existing, corpus-based, dictionaries) is the methodology which I found best suited for the compilation of an adjective-noun collocation dictionary of English and German.

In contrast to the corpus-driven approach where the corpus is the only source for information retrieval, the corpus-based methodology in lexicography uses the corpus as a kind of complementary help in the compilation process.

What kind of complementary help can be envisaged from a corpus? Well, when writing a bilingual collocation dictionary, the two corpora<sup>13</sup> can be searched for the statistical significance of collocations, for example. Furthermore, concordance lines generated for the keywords can be looked at for patterns, or different senses of those words can be studied.

In the compilation process of my sample entries I made extensive use of corpora for the simple reason of reassurance on the part of the lexicographer. Being a native speaker of German creating encoding English-German entries basically meant putting myself in the position of an English native speaker; however, I could only benefit in the translation process since this was actually a decoding task being done into my mother tongue, from English into German.

As I have pointed out earlier in this chapter, I drew my headwords from the intersection of Michael West's *General Service List*, the *Longman Defining Vocabulary* and M.L. Hayes's (1999) list of nouns that are frequently used in academic writing. I randomly chose 16 nouns to illustrate my methodology (the sample dictionary section including the user's guide and a back-matter index can be seen in Appendix D)<sup>14</sup>. In the following I will comment on the creation of the entry for *sample*.

Having laid down preliminaries, i.e. my dictionary policy<sup>15</sup>, I started with the collection of the material.

---

<sup>13</sup> The two corpora do not have to be parallel ones, but they should be comparable in terms of size, design and text genres represented in it. For the compilation of a dictionary that is not restricted to special jargon, the corpus should be of a general nature: "The value of a general corpus as a place of reference is very great [...]. More and more people in every branch of information science are coming to realize that a corpus as a sample of the living language, accessed by sophisticated computers, opens new horizons." (Sinclair 1991: 14).

<sup>14</sup> The nouns I covered are: *basis, collection, comparison, condition, content, discovery, discussion, examination, framework, mechanism, population, principle, proposal, remedy, sample, and tendency*.

<sup>15</sup> cf. Section 4.4. and pp. 107-12

I collected all adjective-noun collocations from the four major English monolingual learner's dictionaries based on corpora, viz. *COBUILD*, *LDOCE*, *MAC* and *OALD*. Next, I searched English monolingual collocation dictionaries, namely the *BBI*, *LTP* and *OCD*<sup>16</sup> for adjectives that collocate with *sample*. Then I searched the British National Corpus through PIE for adjectives preceding *sample*. Finally, I conducted a full-text search in my electronic version of two comprehensive bilingual English-German dictionaries, *COLLINS-E* and *DUOX-E*<sup>17</sup>, for *sample* and retrieved from it all adjective-noun collocations and their translations.

I saved all this information in a file, re-ordered it so that the adjective collocators were in first position, and attached to it the source where I got them from, i.e. bilingual dictionary, collocation dictionary, corpus, learner's dictionary. Then I sorted the list alphabetically. In this way, I was able to count the occurrences of the particular combinations. The resulting table can be seen below (Table 5):

ADJECTIVE	SOURCE	TRANSLATION
advertising	in one bilingual dictionary	Werbemuster
biased	in one collocation or learner's dictionary	
biased	in PIE (singular)	
blood	in more than one collocation or learner's dictionary	
blood		(MED) Blutprobe
blood		taking of blood sample = Blutentnahme
blood		Blutprobe
blood		Blutprobe
blood		have a blood sample taken = Blut abgenommen bekommen
blood	more than once in both bilingual dictionaries	taking of a blood sample = Blutentnahme / to go to the doctor to have a blood sample taken = zur Blutentnahme zum Arzt gehen
blood		Blutprobe
blood		take a blood example from somebody = jemandem Blut / eine Blutprobe entnehmen
carbon-bearing	in PIE (plural)	
clinical	in PIE (plural)	
clinical	in one collocational or learner's dictionary	
commercial	in one collocation or learner's dictionary	
commercial	in one bilingual dictionary	Muster
core	in one bilingual dictionary	Bohrprobe
different	in PIE (plural)	
entire	in PIE (singular)	
environmental	in PIE (plural)	

<sup>16</sup> *OCD* is the only collocation dictionary available based on a computerised corpus.

<sup>17</sup> Both *COLLINS-E* and *DUOX-E* are said to be corpus-based, but I only know for sure (since the front matter of the print dictionary says so) that *COLLINS-E* is based on at least one computerised corpus, *Collins Wordbanks* (a corpus in lexicography can be any collection of text, citations etc. used for compilation purposes).

faecal	in one collocation or learner's dictionary	
faecal	in PIE (plural)	
fair	in one bilingual dictionary	that's a fair sample of ... = das ist ziemlich typisch für ...
formal	in PIE (singular)	
free	in more than one collocation or learner's dictionary	
free	in PIE (singular)	
free	in PIE (plural)	
free	more than once in both bilingual dictionaries	Gratisprobe
free		unverkäufliches Muster
free		Gratisprobe
free		Probe
free		Gratismuster, Gratisprobe
free		unverkäufliches Muster
free		
further	in PIE (singular)	
good	in PIE (singular)	
individual	in PIE (plural)	
initial	in PIE (singular)	
knitted	in PIE (plural)	
known	in PIE (singular)	
large	in one collocation or learner's dictionary	
large	in PIE (plural)	
lunar	in PIE (plural)	
male	in PIE (plural)	
medical	in one bilingual dictionary	Ärztemuster
minute	in PIE (plural)	
mucosal	in PIE (plural)	
national	in PIE (singular)	
new	in PIE (plural)	
normal	in PIE (plural)	
original	in PIE (singular)	
other	in PIE (plural)	
present	in PIE (singular)	
random	in more than one collocation or learner's dictionary	
random	in PIE (singular)	
random	in PIE (plural)	
random	more than once in both bilingual dictionaries	Stichprobe
random		[Zufalls]stichprobe
random		Stichprobe
representative	in more than one collocation or learner's dictionary	
representative	in PIE (singular)	
representative	in PIE (plural)	
representative	more than once in both bilingual dictionaries	(sample) repräsentativ
representative		repräsentative Auswahl
rock	in one collocation or learner's dictionary	
rock	more than once in both bilingual dictionaries	Gesteinsprobe
rock		Gesteinsprobe
selected	in PIE (plural)	
single	in PIE (singular)	
small	in one collocation or learner's dictionary	
small	in PIE (singular)	
soil	in one bilingual dictionary	Bodenprobe
solid	in PIE (plural)	
stratified	in PIE (singular)	
sub	in PIE (singular)	
tissue	in one collocation or learner's dictionary	
tissue	in one bilingual dictionary	(MED) Gewebeprobe
total sample	in PIE (singular)	
trade	in one bilingual dictionary	Warenmuster, Warenprobe
typical	in one bilingual dictionary	that's a typical sample of her cooking = genau so kocht sie immer
unrepresentative	in one bilingual dictionary	(sample) nicht repräsentativ
urine	in more than one collocation or learner's dictionary	
urine	more than once in both bilingual dictionaries	Harnprobe
urine		Urinprobe
various	in PIE (plural)	
water	in one bilingual dictionary	to take water sample = Wassersproben nehmen
whole	in PIE (singular)	

Table 5: The compilation material for the noun *sample*

All the highlighted collocators in Table 5 are candidates for inclusion since they come up at least twice in my sources, i.e. English monolingual collocation dictionaries, an English general corpus, and English-German general-purpose comprehensive dictionaries searched in their full-text version.

Since I used the POS search tool (here: adjective + noun) when consulting PIE, I was, naturally, not able to retrieve noun + noun combinations from the British National Corpus. Consequently, if the corpus data had confirmed it, also *core*, *soil*, *trade* and *water* could have found their way into my dictionary. However, the dictionary I have in mind, for the purpose of this thesis, is primarily focused on “real” adjective + noun structures, even if in English nouns can pre-modify nouns. Still, the material I chose gives examples of that: *blood sample*, *rock sample*, *tissue sample* and *urine sample*.

My candidates for inclusion were, therefore, *biased*, *blood*, *clinical*, *commercial*, *faecal*, *free*, *large*, *random*, *representative*, *rock*, *small*, *tissue* and *urine*.

In the absence of an English native speaker to consult for the contextual meaning of the collocations, as the next step, I consulted the leading English monolingual learner’s dictionaries and the British National Corpus<sup>18</sup> for reassurance purposes. As Dušan Gabrovšek (private communication: 2004), a practical lexicographer himself, told me, the reassurance factor is indeed a major issue in practical lexicography, but there are, so far, not many - well, almost no - studies on that matter.

What made me seek reassurance was, clearly, the fact that I had to start from a language which is not my L1. Consequently, as to the meaning of *sample* and the collocations it can form, I felt the urge to not only consult English monolingual learner’s dictionaries, English monolingual dictionaries for native speakers (*MW* and *ODE*), but also the British National Corpus and the Internet itself, as a corpus (via Google search engine).

Only armed with this knowledge from dictionaries and corpora was I able to put myself in the position of an English learner of German to create the entry.

---

<sup>18</sup> This time through the simple online concordance at <http://sara.natcorp.ox.ac.uk/lookup.html> [10/12/2004].

Then again, when actually translating the English adjective-noun collocations, I did not only consult the (German) corpus of *Digitales Wörterbuch der Deutschen Sprache* and check Google (sites in German) for the translation of the combinations which had no translation equivalent in my first collection (Table 5), but also double-checked the 'given' translations, which are *blood sample*, *commercial sample*, *free sample*, *random sample*, *representative sample*, *rock sample*, *tissue sample* and *urine sample*.

In addition to the two German corpora (the general one and the German sites on the Internet), I also consulted two online dictionaries, DICT.CC and LEO, and, finally, also checked with *DUST* (the German near-equivalent to a monolingual collocation dictionary).

Besides, whenever 'phrases' (in the sense of extended adjective-noun collocations) occurred more than once in the bilingual dictionaries, I took them over but shortened them to a minimum of information, e.g.

- a) *have a blood sample taken* = Blut abgenommen bekommen
- b) *taking of a blood sample* = Blutentnahme
- c) *to go to the doctor to have a blood sample taken* = zur Blutentnahme zum Arzt gehen

⇒ *to have a blood sample taken* = Blut abgenommen bekommen  
*taking of a blood sample* = Blutentnahme

The resulting entry looks like this:

sample
<p><b>biased</b> verzerzte Stichprobe / Auswahl   <b>clinical</b> klinische Stichprobe    <b>free</b> unverkäufliches Muster   <b>large</b> große Auswahl    <b>representative</b> repräsentative Auswahl, repräsentatives Beispiel    <b>small</b> kleine Auswahl</p> <ul style="list-style-type: none"> <li>▪ <b>to have a blood sample taken</b> Blut abgenommen bekommen</li> <li>▪ <b>blood sample</b> Blutprobe f  <b>(the) taking of a blood sample</b> Blutentnahme f  <b>commercial sample</b> [Waren]muster n  <b>faecal sample</b> Kotprobe f  <b>free sample</b> Gratismuster n, Gratisprobe f  <b>random sample</b> Stichprobe f  <b>rock sample</b> Gesteinsprobe f  <b>tissue sample</b> (MED) Gewebeprobe f  <b>urine sample</b> Hamprobe f, Urinprobe f</li> </ul>

Fig. 53: The entry for *sample* in an English-German adjective-noun collocation dictionary

For reasons of comparison, and to further enhance my methodology, I compiled the entry for *remedy* along two lines, my corpus-based approach (Fig. 54a) and simple retrieval from *COLLINS-E* and *DUOX-E* (Fig. 54b).

remedy	remedy
<p><b>adequate</b> adäquates Mittel   <b>alternative</b> alternatives Mittel   <b>appropriate</b> geeignetes Mittel   <b>common</b> allgemein bekanntes / verbreitetes Mittel   <b>desperate</b> äußerstes Mittel   <b>drastic</b> PL drastische Mittel   <b>effective</b> wirksames Mittel   <b>good</b> gutes Mittel   <b>homoeopathic</b> homöopathisches Mittel   <b>palliative</b> schmerzstillendes Mittel   <b>proposed</b> vorgeschlagene Lösung   <b>simple</b> einfaches Mittel   <b>sure</b> sicheres / zuverlässiges Mittel   <b>traditional</b> traditionelles / herkömmliches Mittel   <b>unfailing</b> unfehlbares Mittel   <b>universal</b> universales / universelles Mittel   <b>usual</b> übliches Mittel</p> <ul style="list-style-type: none"> <li>▪ <b>to take Bach's flower remedy</b> (MED) eine Bachblütentherapie machen <b>the best remedy for</b> das beste Mittel gegen / für <b>a cold remedy</b> ein Mittel gegen Erkältung <b>a desperate disease requires a dangerous / drastic remedy</b> (fig.) extreme Situationen erfordern extreme Maßnahmen <b>to apply drastic remedies</b> (lit.) drastische Mittel ergreifen, (fig.) mit eisernem Besen kehren</li> <li>▪ <b>civil remedy</b> (JUR) Rechtsbehelf n, Rechtsmittel m <b>cough remedy</b> Hustenmittel n <b>contractual remedy</b> Vertragsentschädigung f <b>folk remedy</b> Volksmittel n <b>legal remedy</b> (JUR) Rechtsbehelf n, Rechtsmittel m <b>drastic / kill-or-cure remedy</b> Radikalkur f, Rosskur f <b>herbal remedies</b> PL Kräuterheilmittel n <b>home / household remedy</b> Hausmittel n <b>judicial</b> (JUR) Rechtsbehelf n, Rechtsmittel m <b>natural remedy</b> (medication) Naturheilmittel n, (therapy) Naturheilverfahren n <b>patent remedy</b> (lit. a. fig.) Spezialrezept n, Patentrezept n, Patentlösung f <b>popular remedy</b> Hausmittel n <b>recuperative remedies</b> (PL) Heilmittel <b>rough and effective remedy</b> (sl.) Rosskur f <b>quack remedy</b> (sl. derog.) Mittelchen n <b>universal remedy</b> (lit. a. fig.) Allheilmittel n, Universalmittel n</li> </ul>	<p><b>diuretic</b> hamreibendes Mittel   <b>helpful</b> hilfreiches Mittel   <b>palliative</b> linderendes Mittel   <b>speedy</b> schnell wirkendes Mittel   <b>sure</b> zuverlässiges Mittel   <b>unfailing</b> unfehlbares Mittel   <b>universal</b> universales, universelles Mittel   <b>useless</b> unwirkliches, wirkungsloses Mittel  </p> <ul style="list-style-type: none"> <li>▪ <b>to take Bach's flower remedy</b> (MED) eine Bachblütentherapie machen <b>a cold / flu remedy</b> ein Mittel gegen Erkältung / Grippe <b>a desperate disease must have a desperate remedy</b> (fig.) extreme Situationen erfordern extreme Maßnahmen <b>apply drastic remedies</b> (fig.) mit eisernem Besen kehren</li> <li>▪ <b>cough remedy</b> Hustenmittel n <b>herbal remedy</b> Kräuterheilmittel n <b>household remedy</b> Hausmittel n <b>drastic / kill-or-cure remedy</b> (inf) Radikalkur f, (sl) Rosskur f <b>legal remedy</b> (JUR) Rechtsbehelf n, Rechtsmittel m <b>natural remedy</b> (medication) Naturheilmittel n, (therapy) Naturheilverfahren n <b>patent remedy</b> (lit. a. fig.) Spezialrezept n, Patentrezept n, Patentlösung f <b>popular remedy</b> Hausmittel n <b>recuperative remedies</b> (PL) Heilmittel <b>rough and effective remedy</b> (sl) Rosskur f <b>quack remedy</b> (sl. derog.) Mittelchen n <b>sovereign remedy</b> (lit. a. fig.) Allheilmittel n <b>universal remedy</b> (lit. a. fig.) Allheilmittel n, Universalmittel n</li> </ul>

Fig. 54a: The corpus-based entry for *remedy*

Fig. 54b: The entry for *remedy* based on *COLLINS-E* and *DUOX-E*

The highlighted collocators in each column are the ones which a single approach was not able to retrieve. Obviously, the corpus-based approach starting from the English collocation and learner's dictionaries has outdone the best-bilingual-dictionary approach by 13:4 in the first section alone.

Except for the first adjective collocator, *diuretic*, which is rather specialised, I wonder why either of the bilingual dictionaries lists *helpful remedy*, *speedy remedy* and *useless remedy* while these combinations are not to be found in the British National Corpus, nor the learner's dictionaries and the collocation dictionaries.

Not adhering to my clear policy, I would, judging from Fig. 54b and Fig. 54a, include/exclude these combinations for the following reasons:

COLLOCATOR	INCLUSION	EXCLUSION	REASON
diuretic		x	too specific
helpful		x	the English user could look for <i>adequate</i> , <i>appropriate</i> and <i>effective</i>
speedy	x		none of the collocations listed indicates <i>instant cure / solution</i>
useless	x		none of the collocations listed indicates <i>ineffective cure / solution</i>
flu		x	can be derived from <i>cold remedy</i> and <i>cough remedy</i> as <i>Mittel gegen Grippe</i> or <i>Grippemittel</i>
recuperative		x	a pleonasm for me ( <i>a remedy for remedy</i> )
rough and effective		x	can be derived from <i>drastic / kill-or-cure remedy</i> as <i>Radikalkur</i> and <i>Rosskur skur</i>
sovereign		x	can be derived from <i>patent / universal remedy</i> as <i>Allheilmittel</i> , <i>Patentlösung</i> , <i>Patentrezept</i> and <i>Spezialrezept</i> .

Table 6: Possible reasons for inclusion and exclusion

As a result, the 6:2 decision can be taken as a clear confirmation of my methodology, although other reasons might be suggested, e.g. extralinguistic and psychological ones such as reassurance.

The alert reader might have realised that my encoding collocation dictionary actually also includes semantically and predictable collocations, e.g. *adequate remedy* = *adäquates Mittel* and *homoeopathic remedy* = *homöopathisches Mittel*.

Bahns (1993) and Gabrovšek (1998a/b and 2000) suggest that the overwhelming number of collocations be best dealt with in contrastive terms, i.e. that the focus should

be only on semantically unpredictable and structurally incongruent ones. Still, in my restricted framework, in the specialised sector of adjective-noun collocations, I opted for inclusion of all the collocations my corpus-based approaches suggested since we, the lexicographers, should never leave aside the reassurance factor on the part of the dictionary user – and if the information is there, why leave it out?

It goes without saying that in a broader framework, i.e. a general-purpose bilingual collocation dictionary, this decision might be disputable, at least for the print versions (where space is constrained) as opposed to an electronic version.

So, in Chapter 6, I will deal with the question of whether it is realistic to further and expand the research work along the lines I took and if yes, in which way. Furthermore, I will recapitulate the main points of this thesis.

## 6. Further Implications and Conclusions

With this thesis I have endeavoured to demonstrate the need for a bilingual collocation dictionary. Collocations are recurrent, usage- rather than semantically based word combinations. Yet, “the notion of conventional syntagms [collocation] finds its full dimension only when it is considered contrastively, i.e. when the native language of the learner [or the translator] and his target language are taken into consideration. Real ... problems are caused by these syntagms which are comparatively speaking, lexically non-congruent, or, in other words, by those in which there is no direct translational equivalence between their corresponding elements.” (Marton 1977: 40-41).

Since collocations are passwords to native-like fluency in a foreign language (Cop 1991: 2776), and although an incorrect or unacceptable collocation may be still understood by the native speaker, it might cause laughter or even annoyance. Collocational errors should thus be minimised in translation and advanced writing in particular. This is why advanced learners and translators are the main target user groups of a collocation dictionary.

In Chapter 1 I outlined my current line of study which focuses on the lexicographical treatment of the translation of adjective-noun collocations from English into German. I tested the four major comprehensive English-German dictionaries on the market and pointed out the current drawbacks in general bilingual lexicography (cf. 1.3.). Additionally, I called attention to the fact that there are today very few monolingual and hardly any bilingual collocation dictionaries available (cf. 1.3., 1.5. and Chapter 4).

Right from the beginning I suggested what a bilingual adjective-noun collocation dictionary should look like to be of maximum use to the English learner of German or the English-to-German translator.

Before embarking on a discussion of collocations in the contrastive light, in Chapter 3, I outlined major approaches to and applications to of the concept of collocation in Chapter 2. I started with an overview of the different interpretations as seen from the historical angle and discussed the major research areas and the work done by the

practitioners in the field of collocation studies. Particularly, I dealt with corpus linguistics and its merits for collocation research, also sketching out the contributions of computational linguistics to collocational analysis. In section 2.3. I showed where the concept of collocation is to be placed within the wider framework of contemporary mainstream phraseology and other branches of linguistics.

The focus of Chapter 3 was on collocations regarded in the interlingual, English-German, light. The result of two case studies further enhanced the need for a bilingual collocation dictionary and highlighted the necessity of a better lexicographical or computational treatment, i.e. in machine translation, of collocations in the modern science of information retrieval.

In the opening sections of Chapter 4 I discussed several general and specialised English monolingual collocation dictionaries, one monolingual German dictionary with a significant collocational component and some bilingual collocation dictionaries, including two works in the English-German field. I went on to outline the principles of bilingual lexicography and assessed the collocational treatment in existing bilingual dictionaries.

Finally, Chapter 5 was dedicated to three different methodologies of compiling an English-German adjective-noun collocation dictionary. I compared two different corpus-driven methods, viz. making use of parts of speech and a purely statistical one, with my own corpus-based approach. In section 5.2., I described my technique of creating the dictionary I have in mind with an illustrative sample entry. The 16 entries I compiled for the purpose of this thesis can be found, together with the user's guide and an index of German nouns in the back matter, in Appendix D.

There yet remains the question of whether widening the scope of my research is indeed realistic in terms of both lexicographical and computational perspectives.

As can be seen from the sample entries of an English-German adjective-noun collocation dictionary, a lot of adjective-noun structures are embedded in structures where it is only the added word, e.g. a verb or a noun, which gives the combination its full contextual meaning, e.g. *scientific basis* = *wissenschaftliche Grundlage*, *to put a*

*procedure on a scientific basis* = *eine Prozedur verwissenschaftlichen* and *to approach a problem on a scientific basis* = *an ein Problem wissenschaftlich herangehen*.

This finding reveals that it would indeed make sense to widen the scope of my research and compile a general-purpose collocation dictionary, including not only adjective-noun structures but also nouns and their verb collocators, etc.

Given the fact that only about 75% of the English adjective-noun collocations tested can be reliably translated with the help of all four major English-German dictionaries on the market (cf. 3.5.1), I am convinced that the figures would look similar in verbal collocations.

A detailed investigation into this is yet to be made; however, a full-text search for *remedy* and its verbal collocators in *COLLINS-E* and *DUOX-E* resulted in five collocations: *to find a remedy*, *to apply drastic remedies*, *to take Bachs flower remedy*, *to think up a remedy* and *to devise a remedy* - two of which have found their way into my adjective-noun dictionary.

According to *OCD*, *remedy* can enter into collocations with 21 verbs:

1. (= *cure*) *to take a remedy*, *to use a remedy*, *to give somebody a remedy*, *to prescribe somebody a remedy*, *to prepare a remedy*, *to find a remedy*; *a remedy is available and a remedy works*
2. (= *way of dealing with a problem*) *to have a problem*, *to pursue a remedy*, *to seek a remedy*, *to resort to a remedy*, *to exhaust a remedy*, *to create a remedy*, *to offer a remedy*, *to afford somebody a remedy*, *to grant somebody a remedy*, *to provide somebody with a remedy*; *a remedy is available and a remedy lies in something*

Roughly one quarter of the verbal collocations of *remedy* are covered by the bilingual dictionaries. Furthermore, only one collocation, *to find a remedy*, is listed in the entry for the noun *remedy* in *COLLINS-E* (while it is in the entry for *Abhilfe* [= action to improve matters] in *DUOX-E*). The other four noun-verb collocations are listed in German entries, which would therefore take the English user a lot of effort to find in the print version.

The methodology I tested for the compilation of an English-German adjective-noun collocation dictionary can be easily extended to cover also verbal collocations; the only change that would have to be made is information retrieval from the corpora since you can no longer look for sequences of adjacent words or simply query for parts of speech. In contrast, you would have to either stick to parsed corpora or use statistical significance measures like the MI- or t-score (c.f. 2.5.).

Likewise, with the help of an English native speaker, my method can also be implemented for an encoding German into English part.

Finally, what are the chances for such a dictionary to be actually produced? As already pointed out before (cf. p. 110), first of all, the headwords to be included in the dictionary have to be carefully selected. Second, one should think of the publication of a bilingual collocation dictionary along either of two lines. The methodology can either be fully implemented in the compilation process of a general bilingual dictionary, both print and electronic version, or the dictionary is to be produced, although in co-operation with a publishing house, as a stand-alone work.

In very realistic terms I see a bilingual collocation dictionary compiled along the methodology I proposed within this thesis as an additional feature in the electronic versions of comprehensive general bilingual dictionaries. The user would be able to consult the bilingual dictionary in the usual manner and would, upon finding the entry, have the option to click on the bilingual collocation section that goes with the headword. A pop-up window would then show entries similar to the ones I compiled for this thesis. In full-text search, the collocational information would, naturally, also be included.

In the end, let me emphasize once again that collocations are passwords to native-like fluency and collocational correctness is indispensable, especially in translation and other written tasks and we, as lexicographers, should provide the users with the information they need in the best possible way. Let us, therefore, pay tribute to all those who have furthered collocation studies in all their possible applications.

## 7. References and Bibliography

**Abu-Ssaydeh, A.F.** 1995. "An Arabic-English Collocational Dictionary: Issues in Theory and Methodology". *Babel*, Vol. 41, No. 1: 12-23.

**Agnes, M.**, ed. 2002. *Webster's New World® Collegiate Dictionary*. 4<sup>th</sup> edn. Cleveland, OH : Wiley Publishing, Inc.

**Aisenstadt, E.** 1979. "Collocability Restrictions in Dictionaries". In R.K.K. Hartmann, ed. 1979, pp. 71-74.

**Alexander, R.J.** 1987. "Problems in Understanding and Teaching Idiomaticity in English". *Anglistik und Englischunterricht*, Vol. 32 (Wortschatzarbeit): 105-22.

**Allerton, D.J.** 1982. *Valency and the English Verb*. London: Academic Press.

**Allerton, D.J.** 1984. "Three (or Four) Levels of Word Cooccurrence Restriction". *Lingua*, Vol. 63, No. 1 (May): 17-40.

**Allerton, D.J., Nesselhauf, N. and Skandera, P.**, eds. 2004. *Phraseological Units: Basic Concepts and Their Application*. Basel: Schwabe Verlag.

**Asher, R.E.**, ed. 1994. *The Encyclopedia of Language and Linguistics*. 10 vols. Oxford: Pergamon Press.

**Atkins, B.T.S.** 1994. "A Corpus-based Dictionary". In M.-H. Corréard and V. Grundy, eds. 1994, pp. xix - xxxii.

**Atkins, B.T.S.**, ed. 1998. *Using Dictionaries - Studies of Dictionary Use by Language Learners and Translators*. Lexicographica: Series Maior, Vol. 88. Tübingen: Max Niemeyer Verlag.

**Atkins, B.T.S. and Varantola, K.** 1998. “Monitoring Dictionary Use”. In B.T.S. Atkins, ed. 1998, pp. 83-122.

**Bahns, J.** 1993. “Lexical Collocations: A Contrastive View”. *ELT Journal*, Vol. 47, No. 1 (January): 56-63.

**Bahns, J.** 1996. *Kollokationen als lexikographisches Problem – Eine Analyse allgemeiner und spezieller Lernerwörterbücher des Englischen*. Lexicographica: Series Maior, Vol. 74. Tübingen: Max Niemeyer Verlag.

**Bahns, J.** 1997. *Kollokationen und Wortschatzarbeit im Englischunterricht*. Tübingen: Gunter Narr Verlag.

**Baker, M.** 1992. *In Other Words – A Coursebook on Translation*. London and New York: Routledge.

**Bally, Ch.** 1909. *Traité de stylistique française*. 2 vols. Paris: Klincksieck.

**Barlow, M.** 2002. “ParaConc: Concordance Software for Multilingual Parallel Corpora”. In *Proceedings of the First International Workshop on Language Resources (LR) for Translation Work & Research*. Las Palmas. [This version taken from <http://www.ifi.unizh.ch/cl/yuste/postworkshop/repository/mbarlow.pdf>] [11/13/2004]

**Bauer, L.** 1988. “When is a Sequence of Two Nouns a Compound in English?”. *English Language and Linguistics*, Vol. 2, No. 1 (May): 65-86.

**Bazell, C.E., Catford, J.C., Halliday, M.A.K. and Robins, R.H.**, eds. 1966. *In Memory of J.R. Firth*. London: Longmans, Green & Co. Ltd.

**Beaugrande, R. de, Grosman, M. and Seidlhofer B.**, eds. 1998. *Language Policy and Language Education in Emerging Nations: Focus on Slovenia and Croatia with Contributions from Britain, Austria, Spain and Italy*. Advances in Discourse Processes, Vol. 63. Stamford, CT: Ablex Publishing.

**Béjoint, H.** 1994. *Tradition and Innovation in Modern English Dictionaries*. Oxford Studies in Lexicography and Lexicology. Oxford: Clarendon Press.

**Benson, M.** 1985. "Lexical Combinability". *Papers in Linguistics*, Vol. 18: 3-15.

**Benson, M., Benson, E. and Ilson, R.** 1986a. *Lexicographic Description of English..* Amsterdam and Philadelphia: John Benjamins Publishing Company.

**Benson, M., Benson, E. and Ilson, R.** 1986b. *The BBI Combinatory Dictionary of English*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

**Benson, M., Benson E., Ilson, R. and Young, R.** 1991. *Using the BBI – A Workbook with Exercises*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

**Benson, M. and Benson, E.** 1993. *Russian-English Dictionary of Verbal Collocations (REDVC)*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

**Benson, M., Benson, E. and Ilson, R.** 1997. *The BBI Dictionary of English Word Combinations*. Amsterdam and Philadelphia: John Benjamins Publishing Company. [Rev. edn. of: *The BBI Combinatory Dictionary of English*. 1986].

**Bloomfield, L.** 1933. *Language*. New York: Henry Holt.

**Bogatz, H.** 1997. *The Advanced Reader's Collocation Searcher (ARCS)*. Available at <http://www.geocities.com/Athens/Acropolis/7033/mainFrame2.htm> [13/09/2004]

**Bolinger, D.** 1976. "Meaning and Memory". *Forum Linguisticum*, Vol. 1: 1-14.

**Borsley, R. and Ingham, R.** 2003. "More on 'Some Applied Linguists': A Response to Stubbs". *Lingua*, Vol. 113, No.3 (March): 193-96.

**Brough, S. and Messinger, H.**, eds. 2001 *Langenscheidts Handwörterbuch Englisch. Part I: English-German. Part II: German-English*. [Part I ed. by Heinz Messinger; Part II ed. by Sonia Brough]. Berlin, etc.: Langenscheidt Publishers.

**Butler, C.S.** 1985. *Systemic Linguistics – Theory and Applications*. London: Batsford Academic and Educational.

**Choueka, Y., Klein, T. and Neuwitz, E.** 1983: "Automatic Retrieval of Frequent Idiomatic and Collocational Expressions in a Large Corpus". *Journal of Literary and Linguistic Computing*, Vol. 4: 34-38.

**Church, K., Gale, W., Hanks, P. and Hindle, D.** 1991. "Using Statistics in Lexical Analysis". In U. Zernik, ed. 1991, pp. 116-64.

**Clas, A.** 1999. Review of *The BBI Dictionary of English Word Combinations* by M. Benson, E. Benson and R. Ilson (1997). *Meta*, Vol. 44, No. 2. Available at: <http://www.erudit.org/revue/meta/1999/v44/n2/002319ar.html> [10/10/2004]

**COBUILD on CD-ROM.** 2001. [Includes: *Collins COBUILD English Dictionary for Advanced Learners* (2001). *Collins Thesaurus* (2001). *Collins COBUILD English Usage* (1992). *Collins COBUILD English Grammar* (1990). *Wordbank* (a corpus of five million words of text extracted from the "Bank of English")]. Glasgow and Brno: HarperCollins Publishers and Lingea s.r.o.

**Collins COBUILD Collocations on CD-ROM.** 1995. [ed. by J.McH. Sinclair]. Glasgow: HarperCollins Publishers.

**COLLINS e-Großwörterbuch Englisch 4.0 on CD-ROM.** 2004. Berlin etc.: Langenscheidt KG and HarperCollins Publishers Ltd.

**Cop, M.** 1990. "The Function of Collocations in Dictionaries". In T. Magay and J. Zigány, eds. 1990, pp. 34-46.

**Cop, M.** 1991. "Collocations in the Bilingual Dictionary". In F.J. Hausmann, O. Reichmann, H.E. Wiegand and L. Zgusta, eds. 1991, pp. 2775-78.

**Corréard, M.-H. and Grundy, V.**, eds. 1994. *The Oxford-Hachette French Dictionary: French-English, English-French*. Oxford: Oxford University Press and Paris: Hachette Livre.

**Cowie, A.P., Mackin, R. and McCaig, I.R.** 1983. *Oxford Dictionary of Current Idiomatic English. Vol. 2: Phrase, Clause and Sentence Idioms*. Oxford: Oxford University Press.

**Cowie, A.P.** 1985. "Collocational Dictionaries - A Comparative View". In M.G. Murphy, ed. 1986, pp. 61-69.

**Cowie, A.P. and Mackin, R.** 1993. *Oxford Dictionary of Phrasal Verbs*. [Previous edn. published as *Oxford Dictionary of Current Idiomatic English, Vol. 1: Verbs with Prepositions and Particles*, 1975]. Oxford: Oxford University Press.

**Cowie, A.P.** 1994. "Phraseology". In R.E. Asher, ed. 1994, pp. 3168-3171.

**Cowie, A.P.** 1998a. "Introduction". In A.P. Cowie, ed. 1998a, pp. 1-20.

**Cowie, A.P.**, ed. 1998b. *Phraseology – Theory, Analysis, and Applications*. Oxford: Clarendon Press.

**Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners – A History*. Oxford: Oxford University Press. Paper edn., 2002.

**Cueto, N.R.** 1998. Review of *Dictionary of Selected Collocations* by J. Hill and M. Lewis, eds., 1997. *TESL-EJ*, Vol. 3, No. 3, R-11 (September): 1-3.  
Available at [http://www-writing.berkeley.edu/TESL\\_EJ/ej11/r11.html](http://www-writing.berkeley.edu/TESL_EJ/ej11/r11.html) [02/09/2004]

**Dale, R., Moisl, H. and Somers, H.**, eds. 2000. *A Handbook of Natural Language Processing*. New York and Basel: Marcel Dekker.

**Danielsson, P. and Ridings, D.** 1997. "Practical Presentation of a 'Vanilla' Aligner". In *Proceedings of the TELRI Workshop on Alignment and Exploitation of Texts*. Ljubljana, 1997. Available at <http://nl.ijs.si/telri/Vanilla/doc/ljubljana/> [13/09/2004]

**Deuter, M., Greenan, J, Noble, J. and Phillips, J.**, eds. 2002. *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.

**Ellis, N.C.** 1997. "Vocabulary Acquisition : Word Structure, Collocation, Word-Class and Meaning". In N. Schmitt and M.J. McCarthy, eds. 1997, pp. 122-39.

**Evens, M., ed.** 1988. *Relational Models of the Lexicon*. Cambridge: Cambridge University Press.

**Farghal, M and Obiedat, H.** 1995. "Collocations: A Neglected Variable in EFL". *IRAL* (International Review of Applied Linguistics), Vol. 33, No. 4 (November): 315-31.

**Farina, D.M.T.C.** 1995. "L.V. Shcherba's "Oypt": A Contribution to Theoretical Lexicography". *International Journal of Lexicography*, Vol. 8, No. 4 (Winter): 304-13.

**Firth, J.R.** 1951. "Modes of Meaning" *Essays and Studies*, n.s. 4: 118-49. Rptd. in Firth 1957, pp. 190-214.

**Firth, J.R.** 1957. *Papers in Linguistics 1934-1951*. London, etc.: Oxford University Press.

**Fletcher, W.H.** 2004. "Facilitating the Compilation and Dissemination of Ad-Hoc Web Corpora". To appear in G. Aston, S.Bernardini and D. Stewart, eds., *Papers from the Fifth International Conference on Teaching and Language Corpora*. Amsterdam: John Benjamins Publishing Company. [This version taken from [http://kwicfinder.com/Facilitating\\_Compilation\\_and\\_Dissemination\\_of\\_Ad-Hoc\\_Web\\_Corpora.pdf](http://kwicfinder.com/Facilitating_Compilation_and_Dissemination_of_Ad-Hoc_Web_Corpora.pdf)] [13/11/2004]

**Flexner, S.B.**, ed. 1987. *The Random House Dictionary of the English Language*. 2<sup>nd</sup> edn. Unabridged. New York: Random House, Inc.

**Fliedl, G., Homa, A., Maurer-Stroh, P., Weber G. and Maurer-Stroh, S.** 2004. “ANCR - The Adjective-Noun Collocation Retriever (A Tool for Generating a Bilingual Dictionary from a German-English Parallel Corpus)”. In G. Williams and S. Vessier, eds. 2004, pp. 885-90. Available at <http://www.philingua.com> [13/09/2004]

**Fontenelle, T.** 1997. *Turning a Bilingual Dictionary into a Lexical-Semantic Database*. Lexicographica: Series Maior, Vol. 79. Tübingen: Max Niemeyer Verlag.

**Friederich, W. and Canavan, J.** 1979. *Dictionary of English Words in Context*. Dortmund: Verlag Lambert Lensing.

**Gabrovšek, D.** 1998a. “Toward an Interlingual Collocational Dictionary of English and Slovene”. In R. de Beaugrande, M. Grosman and B. Seidlhofer, eds. 1998, pp. 127-42.

**Gabrovšek, D.** 1998b. “Coping with Stubborn Stains and Persistent Headaches – For What It’s Worth: Word Combinability in Action”. *Vestnik*, DTJK, Vol. 32, No. 1-2: 111-54.

**Gabrovšek, D.** 2000. “Phraseology Galore: Words and Their Combinations”. *Vestnik*, DTJK, Vol. 34, No. 1-2: 183-238.

**Gabrovšek, D.** 2003. “On the Concept of Cline in English Phraseology”. *Vestnik*, DTJK, Vol. 37, No. 1-2: 185-93.

**Gillard, P.**, ed. 2003. *Cambridge Advanced Learner’s Dictionary*. New [2<sup>nd</sup>] edn. Cambridge: Cambridge University Press. [First published 1995 as *Cambridge International Dictionary of English*. Ed. by Paul Procter]

**Gómez, R.** 2003. “Maschinelle Übersetzung”. Available at [http://user.cs.tu-berlin.de/~ramiro/downloads/Maschinelle\\_Uebersetzung.pdf](http://user.cs.tu-berlin.de/~ramiro/downloads/Maschinelle_Uebersetzung.pdf) [16/11/04]

**Halliday, M.A.K.** 1966. "Lexis as a Linguistic Level". In C.E. Bazell, J.C. Catford, M.A.K. Halliday and R.H. Robins, eds. 1966, pp. 148-62.

**Hartmann, R.K.K.**, ed. 1979. *Dictionaries and Their Users*. Exeter Linguistic Studies, Vol. 4. Exeter, Devon: University of Exeter.

**Hartmann, R.K.K.** 1987. "Wozu Wörterbücher? Die Benutzforschung in der zweisprachigen Lexikographie". *Lebende Sprachen*, Vol. 32: 154-56.

**Hausmann, F.J.** 1984. "Wortschatzlernen ist Kollokationslernen: Zum Lehren und Lernen französischer Wortverbindungen". *Praxis des neusprachlichen Unterrichts*, Vol. 31: 395-406.

**Hausmann, F.J. and Cop, M.** 1985. "Short History of English and German Lexicography". In K. Hyldgaard-Jensen and A. Zettersten, eds. 1985, pp. 183-97.

**Hausmann, F.J.** 1988. "Grundprobleme des zweisprachigen Wörterbuchs". In K. Hyldgaard-Jensen and A. Zettersten, eds. 1988, pp. 137-54.

**Hausmann, F.J.** 1989. "Le dictionnaire de collocations". In F.J. Hausmann, O. Reichmann, H.E. Wiegand and L. Zgusta, eds. 1989, pp. 1010-19.

**Hausmann, F.J., Reichmann, O., Wiegand, H.E. and Zgusta, L., eds.** 1989. *Wörterbücher / Dictionaries / Dictionnaires – Ein internationals Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie*. Vol. 1 [pp. 1-1056]. Berlin and New York: Walter de Gruyter.

**Hausmann, F.J.** 1991. "Collocations in Monolingual and Bilingual English Dictionaries". In V. Ivir and D. Kalogjera, eds. 1991, pp. 225-36.

**Hausmann, F.J., Reichmann, O., Wiegand, H.E. and Zgusta, L., eds.** 1991. *Wörterbücher / Dictionaries / Dictionnaires – Ein internationals Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie*. Vol. 2 [pp. 1-1056]. Berlin and New York: Walter de Gruyter.

*internationale de lexicographie*. Vol. 3 [pp. 2339-3167]. Berlin and New York: Walter de Gruyter.

**Hayes, M.L.** 1999. *Encouraging Words - A Source Book of Words and Phrases for Dissertation and Report Writers*. Silver Spring, MD: Beckham Publications Group, Inc.

**Hecken, A.E.** 2004. “‘Weiter im Text’ – zu den kommunikativ-pragmatischen Funktionen von Phraseologismen in Texten: Ein Forschungsüberblick”. *Ost-West-Linguistik* (OWL). Available at: <http://viadrina.euv-frankfurt-o.de/~owl/index.html> [13/09/2004]

**Heintze, A.** 1900. *Deutscher Sprachhort. Ein Stil-Wörterbuch*. Leipzig: Renger.

**Heliel, M.H.** 1990. "Lexical Collocations and Translation". In M. Thelen and B. Lewandowska-Tomaszczyk, eds. 1990, pp. 129-39.

**Hill, J. and Lewis M.**, eds. 1997. *LTP Dictionary of Selected Collocations*. Hove, East Sussex: Language Teaching Publications. [Based on the original work of Christian Douglas Kozłowska and Halina Dzierżanowska]

**Hoey, M.** 2000. “A World Beyond Collocations: New Perspectives on Vocabulary Teaching”. In M. Lewis, ed. 2000, pp. 224-43.

**Homa, A. and Sühs, W.** 2004. *ANCR (Adjective-Noun Collocation Retriever)*. Unpublished Bachelor's thesis, University of Klagenfurt, Austria. [To be made available (online) through the library of the University of Klagenfurt]

**Hornby, A.S., Gatenby, E.V. and Wakefield, H.** 1942. *Idiomatic and Syntactic English Dictionary*. Tokyo: Kaitakusha.

**Howarth, P.A.** 1998. “Phraseology and Second Language Proficiency”. *Applied Linguistics*, Vol. 19, No. 1 (March): 24-44.

**Hunston, S.** 2002. *Corpora in Applied Linguistics*. Cambridge Applied Linguistics. Cambridge: Cambridge University Press.

**Hussein, R.F.** 1990. "Collocations: The Missing Link in Vocabulary Acquisition amongst EFL Learners". *Papers and Studies in Contrastive Linguistics*, Vol. 26: 123-36.

**Hyldgaard-Jensen, K. and Zettersten, A.**, eds. 1985. *Symposium on Lexicography II. Proceedings of the Second International Symposium on Lexicography May 16-17, 1984 at the University of Copenhagen*. Lexicographica: Series Maior, Vol. 5. Tübingen: Max Niemeyer Verlag.

**Hyldgaard-Jensen, K. and Zettersten, A.**, eds. 1988. *Symposium on Lexicography III. Proceedings of the Third International Symposium on Lexicography May 14-16, 1986 at the University of Copenhagen*. Lexicographica: Series Maior, Vol. 19. Tübingen: Max Niemeyer Verlag.

**Ichikawa, S. et al.**, eds. 1995. *The Kenkyusha Dictionary of English Collocations*. Tokyo: Kenyusha. [pp. xii + 2782; as quoted by C.-Y. Tsai]

**Iordanskaja, L. and Paperno, S.** 1996. *A Russian-English Collocational Dictionary of the Human Body*. English equivalents by Lesli LaRocco and Jean MacKenzie; edited by Richard L. Leed. Columbus, OH: Slavica Publishers, Inc.

**Ivir, V. and Kalogjera, D.**, eds. 1991. *Languages in Contact and Contrast*. Berlin: Mouton de Gruyter.

**Jespersen, O.** 1924. *The Philosophy of Grammar*. London: Allen & Unwin, 1924.

**Jones, S. and Sinclair, J.McH.** 1974. "English Lexical Collocations – A Study in Computational Linguistics". *Cahiers de Lexicologie*, Vol. 24: 15-61.

**Katsumata, S.** ed. 1958. *A New Dictionary of English Collocations*. Tokyo: Kenyusha.

**Kennedy, G.** 1998. *An Introduction to Corpus Linguistics*. Studies in Language and Linguistics. London and New York: Longman.

**Kipfer, B.A.** 2003. *Roget's Descriptive Word Finder - A Dictionary / Thesaurus of Adjectives*. Cincinnati, OH: Writer's Digest Books.

**Kjellmer, G.** 1994. *A Dictionary of English Collocations: Based on the Brown Corpus. In Three Volumes*. Oxford: Clarendon Press.

**Koehn, P.** 2000. "German-English Parallel Corpus 'de-news'", Daily News 1996-2000". Available at <http://www.isi.edu/~koehn/de-news> [13/09/2004].

**Koehn, P.** 2002. "Europarl: A Multilingual Corpus for Evaluation of Machine Translation". Available at <http://www.isi.edu/~koehn/europarl> [13/09/2004].

**Kozłowska, Ch.D. and Dzierżanowska, H.** 1982/88. *Selected English Collocations*. Warsaw: Wydawnictwo Naukowe PWN.

**Kozłowska, Ch.D.** 1991. *English Adverbial Collocations*. Warsaw: Wydawnictwo Naukowe PWN.

**Kromann, H.P., Riiber, T. and Rosbach, P.** 1991. "Principles of Bilingual Lexicography". In F.J. Hausmann, O. Reichmann, H.E. Wiegand and L. Zgusta, eds. 1991, pp. 2711-28.

**Leech, G.N.** 2002. "Corpora". In K. Malmkjaer, ed. 2002, pp. 84-93.

**Leonhardi, A. and Welsh, B.W.W.** 1966 [1982]. *Grammatisches Wörterbuch Englisch*. Dortmund: Verlag Lambert Lensing.

**Lewis, M.**, ed. 2000. *Teaching Collocation*. Hove, East Sussex: Language Teaching Publications.

**Liontas, J.I.** 2002. "Exploring Second Language Learners' Notions of Idiomaticity". *System*, Vol. 30, No. 3 (September): 289-313.

**Longman Dictionary of Contemporary English on CD-ROM.** 2003. [Includes *Longman Dictionary of Contemporary English*. 4<sup>th</sup> edn. (2003). *Longman Language Activator*. 2<sup>nd</sup> edn. (2002). *Longman Etymologies Database* (2001). *Longman British English and American English Sound Database* (1999). *Longman Interactive Exercise Bank* (2001). *Longman Phrases Database* (2002). *Longman Corpus Examples Database* (2002). *Longman Verb Inflections Database* (2001)] Pearson Education Limited, Addison Wesley Longman Limited and Mozilla Public License.

**Lyons, J.** 1977. *Semantics*. 2 Vols. Cambridge: Cambridge University Press.

**MacKenzie, C.**, ed. 2003. *PONS Wörterbuch für Schule und Studium Englisch-Deutsch*. Studienausgabe. New edn. Stuttgart: Ernst Klett Sprachen GmbH.

**Malmkjær, K.**, ed. 1991/2002. *The Linguistics Encyclopedia*. London and New York: Routledge. 2<sup>nd</sup> edn., 2002. [North American Consultant Editor: J.M. Anderson]

**Manning, C.D. and Schütze, H.** 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press. [The chapter on collocations can be found online at <http://nlp.stanford.edu/fsnlp/promo/>] [13/11/04]

**Marton, W.** 1977. "Foreign Vocabulary Learning as Problem No.1 of Language Teaching at the Advanced Level". *Interlanguage Studies Bulletin*, Vol. 2: 33-57.

**Maurer-Stroh, P.** 2003. "How to Make Use of Corpora in Translation – Benefits and Limitations". *Vestnik*, DTJK, Vol. 37, No. 1-2: 289-300. Available at <http://www.philingua.com> [13/09/2004]

**Maurer-Stroh, P.** 2004a. "'House-High Favourites?' – A Contrastive Analysis of Adjective-Noun Collocations in German and English". Forthcoming in *Mostovi*.

**Maurer-Stroh, P.** 2004b. "Bilingual Dictionaries and the Translation of Adjective-Noun Collocations from German into English". Unpublished seminar paper at the University of Klagenfurt.

**Maurer-Stroh, P.** 2004c. "Translating Adjective-Noun Collocations from English into German – Do Bilingual Dictionaries Help?". Unpublished case study.

**Maurer-Stroh, P.** 2004d. "Deutsche und Englische Adjektiv-Nomen-Kombinationen: Probleme der Maschinellen Übersetzung". Unpublished case study.

**McKeown, K.R. and Radev, D.R.** 2000. "Collocations". In R. Dale, H. Moisl and H. Somers, eds. 2000, pp. 507-23. [This version taken from <http://perun.si.umich.edu/~radev/papers/handbook00.pdf>] [13/09/2004]

**Mel'čuk, I.A. and Zholkovsky, A.** 1984. *Tolkovo-kombinatornyj slovar' russkogo jazyka / Explanatory Combinatorial Dictionary of Modern Russian*. Explanatory Combinatorial Dictionary of Modern Russian. Wiener Slawistischer Almanach. Sonderband 14. Vienna. [992 pp.]

**Mel'čuk, I.A. et al.** 1984-1999. *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques*. 4 Vols. Montréal: Les Presses de l'Université de Montréal.

**Meyer, Ch.F.** 2002. *English Corpus Linguistics – An Introduction*. Cambridge: Cambridge University Press.

**Mikhail, E.H.** 1994. *Dictionary of Appropriate Adjectives*. London: Cassell. Paper edn., 1995.

**Mish, F.C., ed.** 2003. *Merriam Webster's Collegiate® Dictionary*. New [11<sup>th</sup>] edn. Springfield, MA: Merriam-Webster.

**Mitchell, T.F.** 1966. "Some English Phrasal Types". In C.E. Bazell, J.C. Catford, M.A.K. Halliday and R.H. Robins, eds. 1966, pp. 335-58.

**Mitchell, T.F.** 1975a. *Principles of Firthian Linguistics*. Longman Linguistics Library. London: Longman Group Ltd.

**Mitchell, T.F.** 1975b. "Linguistic 'Goings-On': Collocations and Other Lexical Matters Arising on the Syntagmatic Record". In T.F. Mitchell, 1975a, pp. 99-136.

**Murphy, M.G.**, ed. 1986. *Proceedings of the Fourth Joint Anglo-Soviet Seminar on English Studies. Manchester University, 21-27 September 1985*. London: The British Council.

**Nattinger J.R. and DeCarrico J.S.** 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

**Nesselhauf, N.** 2004. "What Are Collocations?". In D.J. Allerton, N. Nesselhauf and P. Skandera, eds. 2004, pp. 1-21.

**Oakes, M. P.** 1998. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

**Oxford Phrasebuilder Genie on CD-ROM.** 2003. [Includes *Oxford Advanced Learner's Dictionary*, 6<sup>th</sup> edn., ed. by S. Wehmeier and D. Lea, 2000 and *Oxford Dictionary of Collocations for Students of English*, ed. by M. Deuter et al., 2002]. Oxford University Press and TEXTware A/S.

**Pätzold, K.-M.** 1991. "Bilingual Lexicography: English-German, German-English". In F.J. Hausmann, O. Reichmann, H.E. Wiegand and L. Zgusta, eds. 1991, pp. 2961-69.

**Palmer, H.E.** 1933. *Second Interim Report on English Collocations*. 4<sup>th</sup> impr 1966. Tokyo: Kaitakusha.

**Palmer, H.E.** 1938. *A Grammar of English Words*. London, etc.: Longmans, Greens & Co.

**Patterson, A.** 2003. Review of *Oxford Collocations Dictionary for Students of English* by M. Deuter et al. (eds. 2002). *TLT-Book Review*, Vol. 27, No. 6 (June). Available at [http://www.jalt-publications.org/tlt/reviews/show\\_br.php?id=41](http://www.jalt-publications.org/tlt/reviews/show_br.php?id=41) [13/09/2004]

**PC-Bibliothek on CD-ROM.** 2.0 (1997) and 3.0. Rev. 6 (2001). [Includes: *Duden-Deutsches Universalwörterbuch*. 3<sup>rd</sup> edn. (1997). *Duden-Deutsches Universalwörterbuch*. 4<sup>th</sup> edn. (2001). *Duden-Oxford Großwörterbuch Englisch* (1990). *Duden-Oxford Großwörterbuch Englisch*. 2<sup>nd</sup> edn (1999). *Duden - Die deutsche Rechtschreibung*. 22<sup>nd</sup> edn. (2000)] Mannheim: Bibliographisches Institut and F.A. Brockhaus AG.

**Porzig, W.** 1934. "Wesenhafte Bedeutungsbeziehungen". *Beiträge zur Geschichte der deutschen Sprache und Literatur*, Vol. 58: 70-97.

**Reum A.** 1931. *A Dictionary of English Style*. Leipzig: Weber.

**Rodale, J.I.** 1947. *The Word Finder*. Emmaus, PA: Rodale Press.

**Roos, E.** 1976. "Contrastive Collocational Analysis". *Papers and Studies in Contrastive Linguistics*, Vol. 5: 65-75.

**Rundell, M.**, ed. 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education.

**Saussure, F.** 1916 *Cours de linguistique générale*. Edited by Ch. Bally and A. Sechehaye. Paris and Lausanne: Éditions Payot.

**Schmitt, N. and McCarthy M.J.**, eds. 1997. *Vocabulary – Description, Acquisition and Pedagogy*. Cambridge Language Teaching Library. Cambridge: Cambridge University Press.

**Shcherba, L.V.** 1995. "Towards a General Theory of Lexicography". Trans. by D.M.T.C. Farina. *International Journal of Lexicography*, Vol. 8, No. 4 (Winter): 314-350. [Originally published in Russian as "Oypt obshchei teorii leksikografii" in the

Journal of the Soviet Academy of Sciences, *Izvestiia Akademii Nauk SSSR, Otdelenie literatury i iazyka*, 1940, No. 3: 89-117]

**Siepmann, D.** 1997. *The Advanced Learner's Trilingual Lexicon. Dictionnaire thématique anglais-français-allemand. Politique-Economie-Expression du temps et de la quantification*. Paris: Ellipses. [Sample pages available from [http://www.dirksiepmann.de/Publications/Dico\\_Excerpt/Introduction.html](http://www.dirksiepmann.de/Publications/Dico_Excerpt/Introduction.html)] [13/09/2004]

**Sinclair, J.McH.** 1966. "Beginning the Study of Lexis". In C.E. Bazell, J.C. Catford, M.A.K. Halliday and R.H. Robins, eds. 1966, pp. 288-302.

**Sinclair, J.McH., ed.** 1987a. *Collins COBUILD English Language Dictionary*. London: HarperCollins Publishers.

**Sinclair, J.McH., ed.** 1987b. *Looking Up – An Account of the COBUILD Project in lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London and Glasgow: Collins ELT.

**Sinclair, J.McH.** 1991. *Corpus, Concordance, Collocation. Describing English Language*. Oxford: Oxford University Press.

**Sinclair, J.McH.** 1997. "Corpus Evidence in Language Description". In A. Wichmann et al. eds. 1997, pp. 27-39.

**Sinclair, J.McH.** 2000. "Lexical Grammar". *Darbai ir Dienos*, Vol. 24 (Naujoji Metodologija): 191-203. Available at <http://donelaitis.vdu.lt/publikacijos/sinclair.pdf> [13/09/2004]

**Sinclair, J.McH.** 2004. *Trust the Text: Language, Corpus and Discourse*. London and New York: Routledge. [edited with Ronald Carter]

**Skandera, P.** 2004. "What Are Idioms?". In D.J. Allerton, N. Nesselhauf and P. Skandera, eds. 2004, pp. 23-35.

**Smadja, F.** 1993. "Retrieving Collocations from Text: Xtract". *Computational Linguistics*, Vol. 19, No. 1 (March): 143-77.

**Smadja, F., McKeown, K.R. and Hatzivassiloglou, V.** 1996. "Translating Collocations for Bilingual Lexicons: A Statistical Approach". *Computational Linguistics*, Vol. 22, No. 1 (March): 1-38.

**Smolik, W.** 1969. "Aktives' Wörterbuch Deutsch-Russisch". *Nachrichten für Sprachmittler*, Vol. 3: 11-13.

**Soanes, C. and Stevenson, A.**, eds. 2003. *Oxford Dictionary of English*. 2<sup>nd</sup> edn. Oxford: Oxford University Press. [Originally ed. by Judy Pearsall and Patrick Hanks (1998)]

**Spencer, A.** 1975. *Noun-Verb Expressions in Legal English. A Handbook for Speakers of Arabic*. Khartoum: Khartoum University Press.

**Stein, G.** 1985. "English-German / German-English Lexicography: Its Early Beginnings". *Lexicographica* 1: 134-64.

**Stubbs, M.** 1996. *Text and Corpus Analysis – Computer-Assisted Studies of Language and Culture*. Language in Society. Oxford and Cambridge, MA: Blackwell Publishers.

**Stubbs, M.** 2001. *Words and Phrases – Corpus Studies of Lexical Semantics*. Oxford and Malden, MA: Blackwell Publishers.

**Summers, D.**, ed. 2003. *Longman Dictionary of Contemporary English*. New [4<sup>th</sup>] edn. Harlow, Essex: Pearson Education Limited / Longman.

**Svensén, B.** 1993. *Practical Lexicography – Principles and Methods of Dictionary-Making*. Translated from the Swedish by John Sykes and Kerstin Schofield. Oxford and New York: Oxford University Press.

**Swan, M.** 1997. "The Influence of the Mother Tongue on Second Language Vocabulary Acquisition and Use". In N. Schmitt and M.J. McCarthy, eds. 1997, pp. 156-80.

**Sweet, H.** 1899. *The Practical Study of Languages: A Guide for Teachers and Learners*. London: Dent.

**ten Hacken, P.** 2004. "What are Compounds?". In D.J. Allerton, N. Nesselhauf and P. Skandera, eds. 2004, pp. 53-66.

**Ter-Minasova, S.** 1992. "The Freedom of Word-Combinations and the Compilation of Learners' Combinatory Dictionaries". In H. Tommola, K. Varantola, T. Salmi-Tolonen and J. Schopp, eds. 1992, pp. 533-39.

**Terrell, P., Schnorr, V., Morris, W.V.A. and Breitsprecher, R.**, eds. 1999. *Collins German-English, English-German Dictionary Unabridged*. 4<sup>th</sup> edn. Glasgow and New York: HarperCollins Publishers.

**The New Oxford American Dictionary on CD-ROM.** 2001. [Ed. by Elizabeth J. Jewell and Frank Abate]. Oxford University Press and i-Finger.

**Thelen, M. and Lewandowska-Tomaszczyk, B.**, eds. 1990. *Translation and Meaning, Part 1. Proceedings of the Maastricht Session of the 1990 Maastricht-Lódz Duo Colloquium on "Translation and Meaning", Maastricht, The Netherlands, 4-6 January 1990*. Maastricht: Euroterm Maastricht.

**Thyen, O.; Scholze-Stubenrecht, W.; Seubel, M.; Clark, M.; Mohan, B.; Sawers, R.; Prowe, G.; Morris, N. and Morris, R.**, eds. 1999. *Duden Oxford Großwörterbuch Englisch: Deutsch-Englisch, Englisch-Deutsch*. 2<sup>nd</sup> rev. edn. Mannheim etc.: Dudenverlag and Oxford: Oxford University Press.

**Tognini-Bonelli, E.** 2001. *Corpus Linguistics at Work. Studies in Corpus Linguistics*, Vol. 6. Amsterdam and Philadelphia: John Benjamins Publishing Company.

**Tommola, H., Varantola, K., Salmi-Tolonen, T. and Schopp, J.**, eds. 1992. *EURALEX '92 Proceedings I-II*. Studia Translatologica, ser. A, Vol. 2. Tampere: University of Tampere, Department of Translation Studies.

**Toury, G.** 1995. *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

**Tsai, C.-Y.** 2000. Review of *The Kenkyusha Dictionary of English Collocations* by S. Ichikawa et al. (eds. 1995). *International Journal of Lexicography*, Vol. 13, No. 3 (September): 205-9.

**Wehmeier, S. and Lea D.**, eds. 2000. *Oxford Advanced Learner's Dictionary of Current English*. 6<sup>th</sup> edn. Oxford: Oxford University Press. [Originally ed. by A.S. Hornby (1942/1948)]

**Weng-chang, W.**, ed. 1990. *A Dictionary of English Collocations*. Taipei: Chung-yung Publications, Ltd. [pp. viii + 1928]

**Wermke, M., Klosa, A., Kunkel-Razum, K. and Scholze-Stubenrecht, W.**, eds. 2001. *Duden - Das Stilwörterbuch*. Der Duden in 12 Bänden: Das Standardwerk zur Deutschen Sprache, Vol. 2. rev. [8<sup>th</sup>] edn. Mannheim, etc.: Dudenverlag.

**Wichmann, A., Fligelstone, S., McEnery, T. and Knowles G.**, eds. 1997. *Corpus Evidence in Language Description*. Harlow, Essex: Longman.

**Williams, E.B.** 1959. "The Problem of Bilingual Lexicography Particularly As Applied to Spanish and English". *Hispanic Review*, Vol. 27: 246-53.

**Williams, G. and Vessier, S.**, eds. 2004. *Proceedings of the Eleventh EURALEX International Congress*. Lorient, France, 6-10 July, 2004. 3 vols. Lorient, Cedex: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

**Wray, A. and Perkins, M.R.** 2000. "The Functions of Formulaic Language: An Integrated Model". *Language and Communication*, Vol. 20, No. 1 (January): 1-28.

**Zernik, U.** ed. 1991. *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*. New Jersey: Lawrence Erlbaum Associates.

**Zgusta, L. et al.** 1971. *Manual of Lexicography*. Paris: The Hague. [as quoted by Kromann et al. 1991]

## APPENDIX A – Corpus Linguistics

The following is a collection of my bookmarks relating to corpus linguistics. For a detailed summary of available corpora, their description and availability, cf. Meyer (2002: 142-150).

### (1) General (as of 17 November 2004)

#### Bookmarks for Corpus-based Linguists (by David Lee)

**Contents:** (1) Corpora, Collections, Data Archives (mainly for English) (2) Non-English Corpora (3) Courses FAQs, Info, E-Lists, Standards (4) Software, Tools, Frequency Lists, etc. (5) References, Papers, Journals (6) Teaching & Misc Links (7) People, Places & Conferences Index. At: <http://devoted.to/corpora>

#### Concordancing (by Peter Ruthven-Stuart)

**Contents:** The pages on this site contain useful information for anyone contemplating using a concordancer. Included is a Bibliography of concordance-, collocation-, corpus- and vocabulary-related books. At: <http://www.nsknet.or.jp/~peterr-s/index.html>

#### Corpora and Corpus-based Computational Linguistics (Manuel Barbera's Web Resources Reference Guide)

**Contents:** (1) - Corpora and Corpus Linguistics. (2) - Multilingual and Parallel Corpora. (3) - Electronic Literary Text Archives. (4) - References, Standards & Educational Resources. (5) - Tools. At: [http://www.bmanuel.org/clr2\\_tt.html](http://www.bmanuel.org/clr2_tt.html)

### **Corpus Linguistics (by Tony McEnery and Andrew Wilson)**

**Contents:** (1) Early Corpus Linguistics and the Chomskyan Revolution. (2) What Is a Corpus and What Is in It? (3) Quantitative Data. (4) The Use of Corpora in Language Studies. At: <http://bowland-files.lancs.ac.uk/monkey/ihe/linguistics/contents.htm>

### **Corpus Linguistics**

Contents: (1) Definition of a corpus (2) Definition of corpus linguistics (3) Examples of English language corpora (4) Examples of corpus linguistics studies (5) Test the hypothesis of co-selection (6) Form-meaning distinction (7) Schematic knowledge (8) Exploring texts through the concordancer (9) Some fundamental precepts in corpus-based research (10) Some implications of corpus-based study (11) Why use a corpus (12) Different corpora (13) Constructing a corpus (14) Finding and cleaning up text (15) Useful sites and homepages on corpora and concordancers. At: <http://www.engl.polyu.edu.hk/corpuslinguist/corpus.htm>

### **Corpus Linguistics**

**Contents:** (1) What Is a Corpus? (2) What Is Corpus Linguistics? (3) Corpus Links (4) Archives & Resources Links (5) Corpus Research Centre Links (6) Conferences & Journals (7) Corpus Linguists (8) Courses on Corpus Linguistics. At: <http://plaza.snu.ac.kr/~hskwon/corpus.html>

### **Corpus Linguistics, Translation, and Language Learning (by Federico Zanettin)**

**Contents:** (1) CORPORA mailing list and CORPORA list archive (2) Text analysis software (3) Online textbooks and tutorials on corpus linguistics (4) Online concordancing (5) Corpora: associations and distribution (6) Text sources on the Web (7) Translation-driven, bilingual and multilingual corpora (8) Corpus annotation (9) Conferences (10) Other corpus linguistics pages. At: <http://www.federicozanettin.net/sslmit/cl.htm>

### **Gateway to Corpus Linguistics on the Net**

**Contents:** (1) Corpus Linguistics (2) Tutorials (3) Corpora (4) Software (5) CL in Applied Linguistics (6) Data-driven learning (7) Virtual Resources (8) Bibliography. At: <http://www.corpus-linguistics.de/sitemap.html>

### **UCREL (=University Centre for Computer Corpus Research on Language) Home Page**

**Contents:** (1) Web-based course in corpus linguistics (2) Corpora (3) Corpus search tools (4) Corpus Annotation (5) British National Corpus (6) English word frequency lists (6) DEMO CLAWS English part-of-speech tagger (7) NEW USAS English semantic tagger (8) Log-likelihood calculator (9) Relevant web links (10) Bookmarks for Corpus-based Linguists. At: <http://www.comp.lancs.ac.uk/computing/research/ucrel/>

### **W3-Corpora Project at the University of Essex**

Contents: (1) Learn about corpus linguistics (2) Corpus linguistics tutorial (3) Search engine to access publicly available corpora and search for any word/phrase you like. At: <http://clwww.essex.ac.uk/w3c/>

## **(2) Corpus Software – Download and Online Use (as of 17 November 2004)**

### **Alex Catalogue of Electronic Texts**

**Contents:** The Alex Catalogue of Electronic Texts is a collection of public domain documents from American and English literature as well as Western philosophy. At: <http://www.infomotions.com/alex/>

### **CA: Linguistic Demos - Language-Specific Tools**

**Contents:** Online demos for tokenization, analysis, disambiguation etc. for various languages.

At: <http://www.xrce.xerox.com/competencies/content-analysis/toolhome.en.html>

### **CLAWS Part-of-Speech Tagger for English**

**Contents:** Free online trial service of the CLAWS tagger. At: <http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/>

### **Collins Corpus Concordance and Collocation Sampler**

**Contents:** Search the Collins *WordbanksOnline* English corpus which is composed of 56 million words of contemporary written and spoken text. At: <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>

### **Cooccurrence Database**

**Contents:** Online collocation tool using the (German) COSMAS corpora. At: <http://corpora.ids-mannheim.de/ccdb/>

### **ConcApp Concordancing Programs**

**Contents:** Free windows concordancer (download).

At: <http://www.edict.com.hk/pub/concapp/>

### **Demonstration der statistischen Kookkurrenzanalyse (Statistical Co-occurrence Analysis)**

**Contents:** Online collocation tool using the (German) COSMAS corpora. At: <http://corpora.ids-mannheim.de/~cosmas/>

**Digitales Wörterbuch der Deutschen Sprache (Digital Dictionary of the German Language)**

**Contents:** Dictionary and online corpus research. At: <http://www.dwds.de/>

**J-BAT KWIC - A Java Concordancer**

**Contents:** Free download of the concordancer and online concordance of various (literary) texts. At: <http://www.hokuriku-u.ac.jp/p-ruthven/javaconc/>

**KWIC Concordance for Windows**

**Contents:** Free download of the concordancer. At: [http://www.chs.nihon-u.ac.jp/eng\\_dpt/tukamoto/kwic\\_e.html](http://www.chs.nihon-u.ac.jp/eng_dpt/tukamoto/kwic_e.html)

**LEXA – Corpus Processing Software (ICAME)**

**Contents:** Free download of the software. At: <http://helmer.aksis.uib.no/lexainf.html>

**Mike Scott's Web**

**Contents:** Free download of a demo version of WordSmith Tools (a lexical analysis software for the PC, published by Oxford University Press since 1996). At: <http://www.lexically.net/wordsmit/index.html>

**Online Tagger für das Deutsche (German Online Tagger)**

**Contents:** Online tagging of German sentences (based on Brill's Tagger). At: <http://www.ifi.unizh.ch/CL/tagger/>

**"Phrases in English"**

**Contents:** Exploring words and phrases from the British National Corpus (online). At: <http://pie.usna.edu>

### **Project Gutenberg**

**Contents:** Online resource of free electronic texts. At <http://gutenberg.net/> (English) and <http://gutenberg.spiegel.de/> (German)

### **Reference Tools & Search Engines**

**Contents:** Inter alia, resources for corpus linguistics (online concordancers). At: <http://www.englishstudydirect.com/OSAC/dicmenu.htm>

### **Simple Concordance Program 4.0.6.**

**Contents:** Free download of the concordance program. At: <http://www.textworld.com/scp/index.html>

### **Simple Search of BNC-World**

**Contents:** Search the BNC online. At: <http://thetis.bl.uk/lookup.html>

### **Text Analysis Info - Information Retrieval Software**

**Contents:** Download of text analysis software (freeware and demos). At: <http://www.textanalysis.info/inforet.htm>

### **TextSTAT - Simple Text Analyse Tool (Simple Text Analysis Tool)**

**Contents:** Free download of the tool. At: <http://www.tekom-nuernberg-aktuell.de/download/textstat-doku-de.html>

### **The Compleat Lexical Tutor (v.4) - For data-driven language learning on the web**

**Contents:** Inter alia, online concordance and phrase extraction in English (BNC, Brown, etc.) and French (Le Monde 1998, Maupassant, etc.). At: <http://www.lextutor.ca/>

### **The Uplug Sentence Aligner**

**Contents:** Online parallel sentence aligner (English, German, Swedish). At: <http://stp.ling.uu.se/~joerg/uplug/SentAlign.html>

### **TnT – Statistical Part-of-Speech Tagging (by Thorsten Brants)**

**Contents:** Online tagging of German and English sentences. At: <http://www.coli.uni-sb.de/~thorsten/tnt/>

### **TreeTagger - A Language-Independent Part-of-Speech Tagger**

**Contents:** The TreeTagger has been successfully used to tag German, English, French, Italian, Greek and Old French texts and is easily adaptable to other languages if a lexicon and a manually tagged training corpus are available. Download for Sun-Solaris, PC-Linux, Mac OS-X, and Windows (demo version). At: <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html>

### **Web Concordancer (English)**

**Contents:** Online concordances of various corpora (including Brown and Lob). At: <http://www.edict.com.hk/concordance/WWWConcappE.htm>

### **Webcorp**

**Contents:** Use the WWW as a corpus. Online concordances. At: <http://www.webcorp.org.uk/>

### **WMTrans Lemmatizer**

**Contents:** Lemmatizer for English, German and Italian. Free download (registration required). At: <http://www.canoo.com/wmtrans/products/lemmatizer.html>

**Wortschatz Lexikon**

**Contents:** Search for collocations in German, English, French, Dutch, etc. At:  
<http://wortschatz.uni-leipzig.de/>

## **APPENDIX B – Online Machine Translation Tools**

Here are the links to the MT tools I tested in Chapter 3.

**Free2ProfessionalTranslation by SDL International (PROF)**

At: <http://freetranslation.com/>

**Promt-Online by Smart Link Corporation (PROM)**

At: <http://translation2.paralink.com/>

**Softissimo by Reverso (SOFT)**

At [http://www.reverso.com/text\\_translation.asp](http://www.reverso.com/text_translation.asp)

**Systran (SYST)**

At: <http://www.systransoft.com>

## APPENDIX C – Parallel Test Corpus

### (1) Text Version of the Corpus

**GERMAN:** Der Mann ist im Garten, obwohl es in Strömen gießt. Es regnet jetzt schon seit Donnerstag ohne Unterbrechung. Es ist kein Regenschauer, sondern Dauerregen. Eine dicke Regenwolke hängt über dem Dorf. Sogar die Katzen flüchten vor dem wolkenbruchartigen Schauer. Ein eisiger Wind bläst von Norden her. Ein schlechtes Zeichen, denn der Wetterbericht sagt schwere Stürme voraus. Manche Menschen mögen es, wenn draußen ein Sturm tobt. Eigentlich ziehen sie sogar einen tobenden Sturm einer angenehmen Brise vor. Sogar stürmischer Wind ist nicht genug. Starker Regen muss es sein, nicht stellenweise Regen. Strömender Regen und böiger Wind. Das Kind wird einen Regenschirm kaufen, obwohl es nicht in Strömen gießt. Es regnet nicht schon seit Wochen ohne Unterbrechung. Kein Regenschauer, kein Dauerregen. Nur stellenweise Regen. Eine dicke Regenwolke ist eine Seltenheit. Es gibt keine wolkenbruchartigen Schauer, keinen strömenden Regen. Eisiger Wind bläst nur im Herbst. Manchmal gibt es auch schwere Stürme. Wenn draußen ein Sturm tobt, bleiben die Leute zu Hause und beobachten den stürmischen Wind, den böigen Wind und den starken Regen. Eigentlich mögen sie beides, einen tobenden Sturm und eine angenehme Brise. Die Frau war am Strand, obwohl es in Strömen goss. Es regnete schon seit Mittwoch ohne Unterbrechung. Es war kein Regenschauer, sondern Dauerregen. Eine dicke Regenwolke hing über der Stadt. Sogar die Hunde flüchteten vor dem wolkenbruchartigen Schauer. Ein eisiger Wind blies von Westen her. Ein schlechtes Zeichen, denn der Wetterbericht sagte schwere Stürme in dieser Gegend voraus. Manche Menschen mögen es nicht, wenn draußen ein Sturm tobt, sie ziehen eine angenehme Brise einem tobenden Sturm vor. Sie hassen stürmischen Wind und starken Regen. Stellenweise Regen ist in Ordnung, aber nicht strömender Regen und böiger Wind.

**ENGLISH:** The man is in the garden although it is pouring with rain. It's been raining solidly since Thursday now. It's not a rain shower, but steady rain. A thick rain cloud is hanging over the village. Even the cats flee from the torrential rain. An icy wind blows

from the north. A bad sign. The weather forecast is for severe gales. Some people like it when it is blowing a gale outside, they actually prefer a fierce gale to a pleasant breeze. Even blustery wind is not enough. It has to be heavy rain not patchy rain. Pouring rain and gusty wind. The child will buy an umbrella although it isn't pouring with rain. It has not been raining solidly for weeks now. No rain shower, no steady rain. Only patchy rain. A thick rain cloud is a rarity. There is no torrential rain, no pouring rain. Icy wind only blows in autumn. Sometimes there are even severe gales. When it is blowing a gale outside, people stay at home and watch the blustery wind, the gusty wind and the heavy rain. They actually like both, a fierce gale and a pleasant breeze. The woman was on the beach although it was pouring with rain. It had been raining solidly since Wednesday. It wasn't a rain shower, but steady rain. A thick rain cloud hung over the town. Even the dogs fled from the torrential rain. An icy wind blew from the west. A bad sign. The weather forecast was for severe gales in this region. Some people don't like it when it is blowing a gale outside, they prefer a pleasant breeze to a fierce gale. They hate blustery wind and heavy rain. Patchy rain is okay, but not pouring rain and gusty wind.

## (2) Aligned and Annotated Parallel Corpus

ADJ	<i>adjective</i>	PERPRO	<i>personal pronoun</i>
ADV	<i>adverb</i>	PROIND	<i>indefinite pronoun</i>
ARTDEF	<i>definite article</i>	PRP	<i>preposition</i>
ARTIND	<i>indefinite article</i>	SUB	<i>noun</i>
DEMPRO	<i>demonstrative pronoun</i>	SZE	<i>sentence end</i>
INF	<i>infinitive</i>	SZK	<i>punctuation = comma</i>
KONNEB	<i>co-ordinating conjunction</i>	VER	<i>verb</i>
KONUNT	<i>subordinating conjunction</i>	VERAUX	<i>auxiliary verb</i>

1. der\_ARTDEF Mann\_SUB sein\_VERAUX im\_PRP Garten\_SUB ,\_SZK obwohl\_KONUNT ich\_PERPRO in\_PRP Strömen\_SUB gießen\_VER .\_SZE
2. ich\_PERPRO regnen\_VER jetzt\_ADV schon\_ADV seit\_PRP Donnerstag\_SUB ohne\_PRP Unterbrechung\_SUB .\_SZE
3. ich\_PERPRO sein\_VERAUX kein\_PROIND Regenschauer\_SUB ,\_SZK sondern\_KONNEB Dauerregen\_SUB .\_SZE

4. ein\_ARTIND dick\_ADJ Regenwolke\_SUB hängen\_VER über\_PRP der\_ARTDEF Dorf\_SUB .\_SZE
5. sogar\_ADV der\_ARTDEF Katze\_SUB flüchten\_VER vor\_PRP der\_ARTDEF wolkenbruchartiger\_ADJ Schauer\_SUB .\_SZE
6. ein\_ARTIND eisig\_ADJ Wind\_SUB blasen\_VER von\_PRP Norden\_SUB her\_ADV .\_SZE
7. ein\_ARTIND schlecht\_ADJ Zeichen\_SUB ,\_SZK denn\_KONNEB der\_ARTDEF Wetterbericht\_SUB sagen\_VER schwer\_ADJ Sturm\_SUB voraus\_ADV .\_SZE
8. manch\_PROIND Mensch\_SUB mögen\_VERMOD ich\_PERPRO ,\_SZK wenn\_KONUNT draußen\_ADV ein\_ARTIND Sturm\_SUB toben\_VER .\_SZE eigentlich\_ADV ziehen\_VER ich\_PERPRO sogar\_ADV ein\_ARTIND tobend\_ADJ Sturm\_SUB ein\_ARTIND angenehm\_ADJ Brise\_SUB vor\_ADV .\_SZE
9. sogar\_ADV stürmisch\_ADJ Wind\_SUB sein\_VERAUX nicht\_ADV genug\_ADV .\_SZE
10. stark\_ADJ Regen\_SUB muss\_VERMOD ich\_PERPRO sein\_VERINF ,\_SZK nicht\_ADV stellenweise\_ADV Regen\_SUB .\_SZE
11. strömend\_ADJ Regen\_SUB und\_KONNEB böig\_ADJ Wind\_SUB .\_SZE
12. der\_ARTDEF Kind\_SUB werden\_VERAUX ein\_ARTIND Regenschirm\_SUB kaufen\_VERINF ,\_SZK obwohl\_KONUNT ich\_PERPRO nicht\_ADV in PRP Strömen\_SUB gießen\_VER .\_SZE
13. ich\_PERPRO regnen\_VER nicht\_ADV schon\_ADV seit\_PRP Woche\_SUB ohne\_PRP Unterbrechung\_SUB .\_SZE
14. kein\_PROIND Regenschauer\_SUB ,\_SZK kein\_PROIND Dauerregen\_SUB .\_SZE
15. nur\_ADV stellenweise\_ADV Regen\_SUB .\_SZE
16. ein\_ARTIND dick\_ADJ Regenwolke\_SUB sein\_VERAUX ein\_ARTIND Seltenheit\_SUB .\_SZE
17. ich\_PERPRO geben\_VER kein\_PROIND wolkenbruchartiger\_ADJ Schauer\_SUB ,\_SZK kein\_PROIND strömend\_ADJ Regen\_SUB .\_SZE
18. eisig\_ADJ Wind\_SUB blasen\_VER nur\_ADV im\_PRP Herbst\_SUB .\_SZE
19. manchmal\_ADV geben\_VER ich\_PERPRO auch\_ADV schwer\_ADJ Sturm\_SUB .\_SZE
20. wenn\_KONUNT draußen\_ADV ein\_ARTIND Sturm\_SUB toben\_VER ,\_SZK bleiben\_VER der\_ARTDEF Leute\_SUB zu\_PRP Haus\_SUB und\_KONNEB

beobachten\_VER der\_ARTDEF stürmisch\_ADJ Wind\_SUB ,\_SZK der\_ARTDEF böig\_ADJ Wind\_SUB und\_KONNEB der\_ARTDEF stark\_ADJ Regen\_SUB .\_SZE

21. eigentlich\_ADV mögen\_VERMOD ich\_PERPRO beide\_DEMPRO ,\_SZK ein\_ARTIND tobend\_ADJ Sturm\_SUB und\_KONNEB ein\_ARTIND angenehm\_ADJ Brise\_SUB .\_SZE

22. der\_ARTDEF Frau\_SUB sein\_VERAUX am\_PRP Strand\_SUB ,\_SZK obwohl\_KONUNT ich\_PERPRO in\_PRP Strömen\_SUB gießen\_VER .\_SZE

23. ich\_PERPRO regnen\_VER schon\_ADV seit\_PRP Mittwoch\_SUB ohne\_PRP Unterbrechung\_SUB .\_SZE

24. ich\_PERPRO sein\_VERAUX kein\_PROIND Regenschauer\_SUB ,\_SZK sondern\_KONNEB Dauerregen\_SUB .\_SZE

25. ein\_ARTIND dick\_ADJ Regenwolke\_SUB hängen\_VER über\_PRP der\_ARTDEF Stadt\_SUB .\_SZE

26. sogar\_ADV der\_ARTDEF Hund\_SUB flüchten\_VER vor\_PRP der\_ARTDEF wolkenbruchartiger\_ADJ Schauer\_SUB .\_SZE

27. ein\_ARTIND eisig\_ADJ Wind\_SUB blasen\_VER von\_PRP Westen\_SUB her\_ADV .\_SZE

28. ein\_ARTIND schlecht\_ADJ Zeichen\_SUB ,\_SZK denn\_KONNEB der\_ARTDEF Wetterbericht\_SUB sagen\_VER schwer\_ADJ Sturm\_SUB in\_PRP dieser\_DEMPRO Gegend\_SUB voraus\_ADV .\_SZE

29. manch\_PROIND Mensch\_SUB mögen\_VERMOD ich\_PERPRO nicht\_ADV ,\_SZK wenn\_KONUNT draußen\_ADV ein\_ARTIND Sturm\_SUB toben\_VER ,\_SZK ich\_PERPRO ziehen\_VER ein\_ARTIND angenehm\_ADJ Brise\_SUB ein\_ARTIND tobend\_ADJ Sturm\_SUB vor\_ADV .\_SZE

30. ich\_PERPRO hassen\_VER stürmisch\_ADJ Wind\_SUB und\_KONNEB stark\_ADJ Regen\_SUB .\_SZE

31. Stellenweise\_ADV Regen\_SUB sein\_VERAUX in\_PRP Ordnung\_SUB ,\_SZK aber\_KONNEB nicht\_ADV strömend\_ADJ Regen\_SUB und\_KONNEB böig\_ADJ Wind\_SUB .\_SZE

1. the\_ARTDEF man\_SUB be\_VERAUX in\_PRP the\_ARTDEF garden\_SUB although\_KONUNT i\_PERPRO pour\_VER with\_PRP rain\_SUB .\_SZE

2. i\_PERPRO rain\_VER solidly\_ADV since\_PRP Thursday\_SUB now\_ADV .\_SZE

3. i\_PERPRO be\_VERAUX not\_ADV a\_ARTIND rain\_SUB shower\_SUB ,SZK but\_KONNEB steady\_ADJ rain\_SUB .S\_SZE
4. a\_ARTIND thick\_ADJ rain\_SUB cloud\_SUB hang\_VER over\_PRP the\_ARTDEF village\_SUB .SZE
5. even\_ADV the\_ARTDEF cat\_SUB flee\_VER from\_PRP the\_ARTDEF torrential\_ADJ rain\_SUB .SZE
6. an\_ARTIND icy\_ADJ wind\_SUB blow\_VER from\_PRP the\_ARTDEF north\_SUB .SZE
7. a\_ARTIND bad\_ADJ sign\_SUB .SZE The\_ARTDEF weather\_SUB forecast\_SUB be\_VERAUX for\_PRP severe\_ADJ gale\_SUB .SZE
8. some\_PROIND people\_SUB like\_VER i\_PERPRO when\_KONUNT i\_PERPRO blow\_VER a\_ARTIND gale\_SUB outside\_ADV ,SZK i\_PERPRO actually\_ADV prefer\_VER a\_ARTIND fierce\_ADJ gale\_SUB to\_PRP a\_ARTIND pleasant\_ADJ breeze\_SUB .SZE
9. even\_ADV blustery\_ADJ wind\_SUB be\_VERAUX not\_ADV enough\_ADV .SZE
10. i\_PERPRO have\_VERMOD to\_INF be\_VERAUX heavy\_ADJ rain\_SUB not\_ADV patchy\_ADJ rain\_SUB .SZE
11. pouring\_ADJ rain\_SUB and\_KONNEB gusty\_ADJ wind\_SUB .SZE
12. the\_ARTDEF child\_SUB will\_VERAUX buy\_VER an\_ARTIND umbrella\_SUB although\_KONUNT i\_PERPRO not\_ADV pour\_VER with\_PRP rain\_SUB .SZE
13. i\_PERPRO not\_ADV rain\_VER solidly\_ADV for\_PRP week\_SUB now\_ADV .SZE
14. no\_PROIND rain\_SUB shower\_SUB ,SZK no\_PROIND steady\_ADJ rain\_SUB .SZE
15. only\_ADV patchy\_ADJ rain\_SUB .SZE
16. a\_ARTIND thick\_ADJ rain\_SUB cloud\_SUB be\_VERAUX a\_ARTIND rarity\_SUB .SZE
17. there\_ADV be\_VER no\_PROIND torrential\_ADJ rain\_SUB ,SZK no\_PROIND pouring\_ADJ rain\_SUB .SZE
18. icy\_ADJ wind\_SUB only\_ADV blow\_VER in\_PRP autumn\_SUB .SZE
19. sometimes\_ADV there\_ADV be\_VERAUX even\_ADV severe\_ADJ gale\_SUB .SZE
20. when\_KONUNT i\_PERPRO blow\_VER a\_ARTIND gale\_SUB outside\_ADV ,SZE people\_SUB stay\_VER at\_PRP home\_SUB and\_KONNEB watch\_VER

the\_ARTDEF blustery\_ADJ wind\_SUB ,\_SZK the\_ARTDEF gusty\_ADJ  
wind\_SUB and\_KONNEB the\_ARTDEF heavy\_ADJ rain\_SUB .\_SZE

21. i\_PERPRO actually\_ADV like\_VER both\_DEMPRO ,\_SZK a\_ARTIND  
fierce\_ADJ gale\_SUB and\_KONNEB a\_ARTIND pleasant\_ADJ breeze\_SUB  
.SZE

22. the\_ARTDEF woman\_SUB be\_VERAUX on\_PRP the\_ARTDEF beach\_SUB  
although\_KONUNT i\_PERPRO pour\_VER with\_PRP rain\_SUB .SZE

23. i\_PERPRO rain\_VER solidly\_ADV since\_PRP Wednesday\_SUB .SZE

24. i\_PERPRO be\_VERAUX not\_ADV a\_ARTIND rain\_SUB shower\_SUB ,\_SZK  
but\_KONUNT steady\_ADJ rain\_SUB .SZE

25. a\_ARTIND thick\_ADJ rain\_SUB cloud\_SUB hang\_VER over\_PRP the\_ARTDEF  
town\_SUB .SZE

26. even\_ADV the\_ARTDEF dog\_SUB flee\_VER from\_PRP the\_ARTDEF  
torrential\_ADJ rain\_SUB .SZE

27. an\_ARTIND icy\_ADJ wind\_SUB blow\_VER from\_PRP the\_ARTDEF west\_SUB  
.SZE

28. a\_ARTIND bad\_ADJ sign\_SUB .SZE The\_ARTDEF weather\_SUB forecast\_SUB  
be\_VERAUX for\_PRP severe\_ADJ gale\_SUB in\_PRP this\_DEMPRO region\_SUB  
.SZE

29. Some\_PROIND people\_SUB do\_VERAUX not\_ADV like\_VER i\_PERPRO  
when\_KONUNT i\_PERPRO blow\_VER a\_ARTIND gale\_SUB outside\_ADV  
.SZK i\_PERPRO prefer\_VER a\_ARTIND pleasant\_ADJ breeze\_SUB to\_PRP  
a\_ARTIND fierce\_ADJ gale\_SUB .SZE

30. i\_PERPRO hate\_VER blustery\_ADJ wind\_SUB and\_KONNEB heavy\_ADJ  
rain\_SUB .SZE

31. patchy\_ADJ rain\_SUB be\_VERAUX okay\_ADJ ,\_SZK but\_KONNEB not\_ADV  
pouring\_ADJ rain\_SUB and\_KONNEB gusty\_ADJ wind\_SUB .SZE

## APPENDIX D – A Sample of an English-German Adjective-Noun Collocation Dictionary

### Introduction

Collocations are frequently co-occurring words which act very much like single choices and are so-called passwords to native-like fluency and typicality.

This model dictionary covers 16 English nouns and their adjective collocators. The nouns are a random sample taken from the intersection of Michael West's *General Service List* (1953), the *Longman Defining Vocabulary* (1987) and M.L. Hayes's (1999) list of nouns frequently used in academic writing: *basis, collection, comparison, condition, content, discovery, discussion, examination, framework, mechanism, population, principle, proposal, remedy, sample and tendency*.

The collocators of the headwords are a selection from existing collocation dictionaries (*BBI, LTP* and *OCD*) and the four major English monolingual learner's dictionaries (*COBUILD, LDOCE, MAC* and *OALD*). All collocations were checked against the *British National Corpus*.

The German translations are taken from two of the most comprehensive English-German dictionaries on the market (*COLLINS-E* and *DUOX-E*), online dictionaries like *LEO* and *DICT.CC* and are backed up by corpus data from *Digitales Wörterbuch der Deutschen Sprache*.

The lexicographer is a native speaker of German.

### Purpose of the dictionary

This dictionary is primarily intended for English advanced learners of German or English-to-German translators. With the help of a collocation dictionary like this not only common collocational errors can be avoided but also native-like fluency in speech and writing can be made easier.

### User's Guide

Adjective collocators are listed in their base noun entries. The headwords are printed in bold face and are bigger than the rest of the entry.

Each entry is divided into a maximum of four sections which are marked by square bullet points:

(1) English adjective-noun collocations with a German adjective-noun representation: On the English side only the adjectives are listed (they are printed in bold), while on the German side the collocation as a whole is represented. To help the user German adjectives are inflected according to the gender of the following noun. The English adjectives are listed alphabetically. If, on the German side, there are several equivalent collocations, these are ordered alphabetically first by the noun then by the adjective.

(2) English adjective-noun collocations that are rendered into German with a different structure are listed in this section of the entry (they are listed alphabetically according to the adjective collocator). Often a verb or other word class items have to be added to give the collocation its correct contextual meaning in both languages. Such English 'phrases' are printed in bold italics, while the German translation is set in normal type.

(3) English adjective-noun collocations forming a part of a pattern, e.g. prepositional phrase: The full pattern is listed (in alphabetical order of the collocators) together with its German translation equivalents. Again, English 'phrases' are printed in bold italics.

(4) In this section the user finds English adjective-noun collocations with are rendered into German as 'authentic compounds', that is, word combinations that are spelled as single words, as well as

German simplex words. English combinations are printed in bold italics and are put in alphabetical order according to the collocators. German translations are followed by their grammatical gender.

German collocations phrases and compounds are normally unmarked, so labels (preceding the German translation) indicate a special register or style. When the German number differs from the English one, this is also marked by a preceding label. These labels are listed in the abbreviations section.

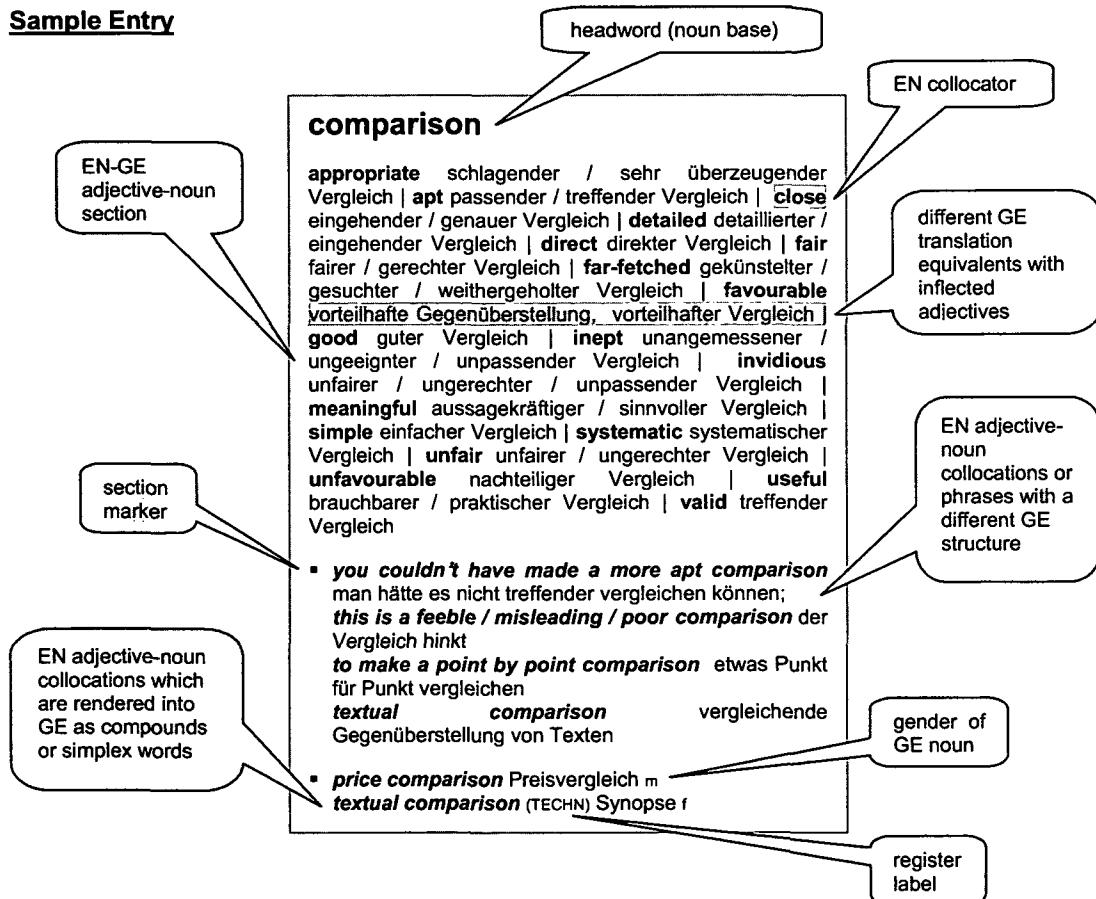
Entries are normally not subdivided into different senses when the German translations reflect the English uses

(although an English headword might have different German translation equivalents). In longer entries, where it is possible to divide the German translations into different senses, this is done for the benefit of the user.

When needed for a particular translation, sense discriminators are added in italics in round brackets.

In the back matter of the dictionary there is a list of German nouns (included are also nouns from sections 2-4).

### Sample Entry



### Abbreviations

a.	<i>and</i>	inf.	<i>informal</i>	PL	<i>plural</i>
BIOL	<i>biology</i>	JUR	<i>jurisprudence</i>	POL	<i>politics</i>
BUS	<i>business</i>	LIBR	<i>library science</i>	PSYCH	<i>psychology</i>
CHEM	<i>chemistry</i>	LING	<i>linguistics</i>	SG	<i>singular</i>
COMPUT	<i>computer science</i>	lit.	<i>literal</i>	SOCIOL	<i>sociology</i>
CONSTR	<i>construction</i>	LOGIC	<i>logic</i>	SPORT	<i>sports</i>
derog.	<i>derogatory</i>	MED	<i>medicine</i>	TECHN	<i>technology</i>
fig.	<i>figurative</i>	MIL	<i>military</i>	UNIV	<i>university</i>
form.	<i>formal</i>	PHYS	<i>physics</i>	usu.	<i>usually</i>

## basis

**broad** breite Basis | **firm** feste Basis / Grundlage, stabile Basis | **scientific** wissenschaftliche Grundlage | **sound** solide / gesunde Basis | **theoretical** theoretische Grundlage

- **to work on a casual basis** Gelegenheitsarbeiter sein
- **to have a factual basis** auf Tatsachen beruhen
- **to have no firm / sound basis** (fig.) auf wackeligen Füßen stehen
- **to work on a commission basis** auf Provisions-/ Prozentbasis arbeiten, gegen Provision arbeiten
- **to put something on a different basis** etwas auf eine neue Grundlage stellen
- **to be engaged on a part-time basis to teach French** als Teilzeitlehrer/-in für Französisch angestellt sein /arbeiten
- **to be employed on a permanent / regular basis** fest angestellt sein
- **to do something on a professional basis** etwas [haupt]beruflich machen / betreiben
- **to do something on a rota basis** sich ablösen, etwas abwechselnd machen
- **to approach a problem on a scientific basis** an ein Problem wissenschaftlich herangehen
- **to put a procedure on a scientific basis** ein Verfahren verwissenschaftlichen
- **to meet with someone on a social basis** mit jemandem außerdienstlich verkehren
- **to be on a sound basis (business)** auf festen Füßen stehen, (theory) auf einer soliden Basis ruhen
- **to help out on a temporary basis** als Aushilfe arbeiten
- **to be employed on a trial basis** auf Probe, probehalber/-weise beschäftigt sein
- **How does it average out on a weekly basis?** Wie viel ist das im Schnitt pro Woche?
- **on an annual basis** ein Mal im Jahr, jährlich
- **on a commercial basis** gewerblich
- **on a competitive basis** nach Leistung
- **on a daily / day-to-day basis** tageweise
- **on a daily basis** als Tagelöhner
- **on an equal basis** auf gleicher Basis, (bus) als [gleichgestellte(r)] Partner
- **on an experimental basis** versuchsweise
- **on a first-come first-served basis** (fig.) nach dem Prinzip "Wer zuerst kommt, mahlt zuerst"
- **on a full-time basis** ganztags, hauptamtlich, hauptberuflich
- **on a hand-to-mouth basis** (fig.) von der Hand in den Mund [leben]
- **on an hourly basis** stundenweise
- **on an interim basis** (form.) interimistisch, vorübergehend
- **on a long-term basis** auf lange Sicht, langfristig
- **on a monthly basis** ein Mal im Monat, monatlich
- **on a part-time basis** halbtags
- **on a pro-rata basis** anteilmäßig
- **on a random basis** stichprobenweise
- **on a regular basis** regelmäßig
- **on a rota basis** abwechselnd, umschichtig
- **on a sale-or-return basis** auf Kommission[sbasis]
- **on a selective basis** selektiv
- **on a systematic basis** systematisch
- **on a temporary basis** aushilfsweise
- **on a trial basis** probe- / versuchsweise
- **on a voluntary basis** auf freiwilliger Basis, freiwillig
- **on a weekly basis** ein Mal in der Woche, wöchentlich
- **legal basis** Rechtsgrundlage f

## collection

### 1. GROUP OF COLLECTED THINGS

**amazing** sensationelle Sammlung | **bizzare** bizarre Sammlung | **celebrated** berühmte Sammlung | **classic(al)** klassische Sammlung | **complete** komplett / vollständige Sammlung | **comprehensive** umfangreiche / umfassende Sammlung | **distinguished** hervorragende Sammlung | **diverse** bunte / vielfältige Sammlung | **extensive** umfangreiche Sammlung | **extraordinary** außergewöhnliche Sammlung | **famous** berühmte Sammlung | **fascinating** faszinierende Sammlung | **fine** ausgezeichnete / herrliche Sammlung | **free** freie Sammlung | **good [recht]** ansehnliche Sammlung | **great** große Sammlung | **ever-growing** ständig wachsende Sammlung | **heterogeneous** heterogene / vielschichtige Sammlung | **historical** geschichtliche Sammlung | **huge** riesige / gewaltige Sammlung | **important** bedeutende / wichtige Sammlung | **impressive** stattliche Sammlung | **incomplete** unvollkommene / unvollständige Sammlung | **interesting** interessante Sammlung | **international** internationale Sammlung | **large** große Sammlung | **magnificent** ausgezeichnete / hervorragende / kostbare Sammlung | **major** bedeutende Sammlung | **miscellaneous** bunte / bunt gemischte Sammlung | **national** staatliche Sammlung | **outstanding** hervorragende Sammlung | **permanent** permanente Sammlung | **private** private Sammlung | **public** öffentliche Sammlung | **rare** seltene Sammlung | **remarkable** beachtliche / bemerkenswerte Sammlung | **rich** reiche / reichhaltige / vielfältige Sammlung | **small** kleine Sammlung | **splendid** ausgezeichnete / hervorragende / kostbare Sammlung | **strange** seltsame Sammlung | **stunning** atemberaubende / fantastische Sammlung | **substantial** umfangreiche / umfassende Sammlung | **superb** großartige Sammlung | **unique** einzigartige Sammlung | **useful** brauchbare, nützliche Sammlung | **valuable** wertvolle Sammlung | **varied** vielseitige / [sehr] gemischte Sammlung | **vast** riesige / gewaltige / überwältigende Sammlung

- **finest collections** Sammlungen vom Feinsten
- **loans from private collections** Leihgaben aus / von privater Hand

▪ **art collection** Kunstsammlung f, (belonging to a person) Kunstbesitz m, (of paintings only) Gemäldesammlung f

▪ **coin collection** Münzsammlung f

▪ **loan collection** (for exhibition) PL Leihgaben f

▪ **photographic collection** Fotothek f

▪ **private collection** Privatsammlung f

▪ **record collection** [Schall]Plattensammlung f

▪ **reference collection** (LIBR) Handapparat m

▪ **reverse collection** (UNIV, LIBR) Semesterapparat m

▪ **special collection** Spezialsammlung f

▪ **stamp collection** Briefmarkensammlung f

### 2. RANGE OF NEW CLOTHES

**new** neue Kollektion

- **autumn collection** Herbstkollektion f
- **ready-to-wear collection** Prêt-à-porter-Kollektion f
- **spring collection** Frühjahrskollektion f
- **summer collection** Sommerkollektion f
- **winter collection** Winterkollektion f

### 3. MONEY

**charity** karitative Sammlung

- **make a house-to-house collection** von Haus zu Haus gehen und sammeln

<ul style="list-style-type: none"> <li>▪ <b>charity collection</b> Wohltätigkeitssammlung f  <b>door-to-door collection</b> Haussammlung f  <b>house-to-house collection</b> Haussammlung f  <b>street collection</b> Straßensammlung f</li> </ul> <p>4. ACT OF BRINGING TOGETHER</p> <ul style="list-style-type: none"> <li>▪ <b>data collection</b> Datenerfassung f</li> </ul> <p>5. ACT OF TAKING AWAY</p> <p><b>weekly</b> wöchentliche Leerung / Abholung</p> <ul style="list-style-type: none"> <li>▪ <b>garbage / refuse / rubbish collection</b> Müllabfuhr f  <b>waste paper collection</b> Altpapersammlung f</li> </ul> <p>6. GROUP OF PEOPLE</p> <p><b>motley</b> (inf.) SG buntzusammengewürfelter Haufen   <b>odd</b> (inf.) SG seltsamer Verein</p> <p>7. GROUP OF OBJECTS AT DISPLAY</p> <p><b>random</b> willkürliche Auswahl   <b>small</b> kleine Auswahl   <b>strange</b> seltsame Auswahl</p>	<p>2. CIRCUMSTANCES (PL)</p> <p>adverse ungünstige Bedingungen, widrige Umstände   <b>appalling</b> unmenschliche Bedingungen, himmelschreiende Verhältnisse   <b>artic</b> arktische Verhältnisse   <b>bad</b> arge Verhältnisse   <b>controlled</b> kontrollierte Bedingungen   <b>crowded</b> beengte Verhältnisse   <b>difficult</b> missliche Umstände / Verhältnisse   <b>economic</b> wirtschaftliche Gegebenheiten   <b>excellent</b> erstklassige Bedingungen   <b>extreme</b> extreme Bedingungen   <b>favourable</b> günstige / optimale Bedingungen   <b>financial</b> finanzielle Verhältnisse / (form.) Umstände   <b>foggy</b> nebeliges Wetter   <b>harsh</b> harte Bedingungen   <b>impoverished</b> ärmliche Verhältnisse   <b>intolerable</b> unerträgliche / unhaltbare Zustände   <b>lamentable</b> sg bejammernswerte Lage   <b>local</b> hiesige Verhältnisse   <b>unhuman</b> (lit. a. fig) menschenunwürdige / unmenschliche Bedingungen   <b>optimum</b> optimale Bedingungen   <b>physical</b> räumliche Verhältnisse   <b>pitiable</b> beklagenswerte Zustände, sg bejammernswerte Lage   <b>pitiful</b> beklagenswerte Zustände, traurige Verhältnisse   <b>poor</b> ärmliche Verhältnisse   <b>prevailing</b> gegenwärtige / derzeitige Umstände, aktuelle Verhältnisse   <b>roadworthy</b> verkehrssichere Verhältnisse   <b>severe</b> schwierige Bedingungen / Verhältnisse   <b>social</b> gesellschaftliche Verhältnisse, soziale Gegebenheiten   <b>unfavourable</b> ungünstige Bedingungen   <b>varying</b> (usu.: <i>weather</i>) veränderliche Verhältnisse   <b>wretched</b> erbärmliche / miserable Zustände, traurige Verhältnisse  </p> <ul style="list-style-type: none"> <li>▪ <b>arctic conditions</b> sibirische Kälte  <b>this is caused by climatic conditions</b> das ist klimatisch bedingt  <b>bad road conditions</b> mangelhafte / schlechte Straßenverhältnisse  <b>to live in cramped conditions</b> [räumlich] beengt / beschränkt wohnen  <b>to live in crowded conditions</b> beengt wohnen  <b>freezing conditions</b> Temperaturen unter null  <b>the icy conditions on the roads</b> das Glatteis auf den Straßen  <b>to live in miserable conditions</b> ein elendes Leben führen  <b>keep animals in natural conditions</b> Tiere naturnah halten  <b>the poor conditions</b> die Ärmlichkeit der Verhältnisse  <b>to live in poor conditions</b> ärmlich leben  <b>the usual conditions</b> die allgemein gültigen Bedingungen  <b>clothing appropriate to the weather conditions</b> wettermäßige Kleidung  <b>variations in weather conditions</b> unbeständiges Wetter  <b>to live in wretched conditions</b> ein elendes Leben führen</li> </ul>
<h3>comparison</h3> <p><b>appropriate</b> schlagender / sehr überzeugender Vergleich   <b>apt</b> passender / treffender Vergleich   <b>close</b> eingehender / genauer Vergleich   <b>detailed</b> detaillierter / eingehender Vergleich   <b>direct</b> direkter Vergleich   <b>fair</b> fairer / gerechter Vergleich   <b>far-fetched</b> gekünstelter / gesuchter / weithergeholt Vergleich   <b>favourable</b> vorteilhafte Gegenüberstellung, vorteilhafter Vergleich   <b>good</b> guter Vergleich   <b>inert</b> unangemessener / ungeeigneter / unpassender Vergleich   <b>invidious</b> unfairer / ungerechter / unpassender Vergleich   <b>meaningful</b> aussagekräftiger / sinnvoller Vergleich   <b>simple</b> einfacher Vergleich   <b>systematic</b> systematischer Vergleich   <b>unfair</b> unfairer / ungerechter Vergleich   <b>unfavourable</b> nachteiliger Vergleich   <b>useful</b> brauchbarer / praktischer Vergleich   <b>valid</b> treffender Vergleich</p> <ul style="list-style-type: none"> <li>▪ <b>you couldn't have made a more apt comparison</b> man hätte es nicht treffender vergleichen können; <b>this is a feeble / misleading / poor comparison</b> der Vergleich hinkt  <b>to make a point by point comparison</b> etwas Punkt für Punkt vergleichen  <b>textual comparison</b> vergleichende Gegenüberstellung von Texten</li> <li>▪ <b>price comparison</b> Preisvergleich m  <b>textual comparison</b> (TECHN) Synopse f</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>arctic conditions</b> sibirische Kälte  <b>this is caused by climatic conditions</b> das ist klimatisch bedingt  <b>bad road conditions</b> mangelhafte / schlechte Straßenverhältnisse  <b>to live in cramped conditions</b> [räumlich] beengt / beschränkt wohnen  <b>to live in crowded conditions</b> beengt wohnen  <b>freezing conditions</b> Temperaturen unter null  <b>the icy conditions on the roads</b> das Glatteis auf den Straßen  <b>to live in miserable conditions</b> ein elendes Leben führen  <b>keep animals in natural conditions</b> Tiere naturnah halten  <b>the poor conditions</b> die Ärmlichkeit der Verhältnisse  <b>to live in poor conditions</b> ärmlich leben  <b>the usual conditions</b> die allgemein gültigen Bedingungen  <b>clothing appropriate to the weather conditions</b> wettermäßige Kleidung  <b>variations in weather conditions</b> unbeständiges Wetter  <b>to live in wretched conditions</b> ein elendes Leben führen</li> </ul>
<h3>condition</h3> <p>1. PREREQUISITE</p> <p><b>fundamental</b> unabdingbare Voraussetzung, (form.) Postulat   <b>necessary</b> (LOGIC) notwendige Voraussetzung   <b>physical</b> (slightly form.) physische Voraussetzung   <b>structural</b> (CONSTR) bauliche Voraussetzung   <b>sufficient</b> hinreichende Voraussetzung   <b>varying</b> PL unterschiedliche Voraussetzungen</p> <ul style="list-style-type: none"> <li>▪ <b>externally imposed conditions</b> von außen auferlegte Bedingungen</li> <li>▪ <b>basic condition</b> Grund- / Rahmenbedingung f  <b>prevailing condition</b> Rahmenbedingung f  <b>structural condition</b> (sociol) Rahmenbedingung f</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>in / under / subject to certain conditions</b> unter bestimmten Bedingungen  <b>under different conditions</b> (fig.) mit veränderten Vorzeichen  <b>in / under difficult conditions</b> unter harten Bedingungen  <b>in good conditions</b> unter guten Bedingungen  <b>in icy conditions</b> bei Eis  <b>in foggy conditions</b> bei Nebel  <b>in / under present conditions</b> unter den gegenwärtigen Bedingungen / Umständen, bei den gegenwärtigen / derzeitigen Verhältnissen  <b>in wet and slippery conditions</b> bei Wasserglätte</li> <li>▪ <b>driving conditions</b> PL Straßenverhältnisse n  <b>controlled conditions</b> (TECHN) Regelgröße f  <b>cramped conditions</b> sg Beengtheit f</li> </ul>

<p><b>ground conditions</b> Bodenverhältnisse n, (SPORT) Platzverhältnisse n</p> <p><b>light conditions</b> Lichtverhältnisse n</p> <p><b>living conditions</b> Existenz- / Lebens- / Unterkunftbedingungen f, (family) Wohnverhältnisse n, Umgebung f</p> <p><b>local conditions</b> Standortbedingungen f</p> <p><b>market conditions</b> Marktlage f, Konjunktur f</p> <p><b>road conditions</b> Straßenverhältnisse n, SG Straßenzustand m</p> <p><b>snow conditions</b> Schneeverhältnisse n</p> <p><b>weather conditions</b> Wetterverhältnisse n, Witterungsverhältnisse n, SG Wetterlage f, SG Wetter n</p> <p><b>working conditions</b> Arbeitsbedingungen f</p> <p><b>3. STATE (SG)</b></p> <p><b>lamentable</b> beklagenswerter Zustand   <b>mental</b> seelisches Befinden   <b>miserable</b> jämmerlicher Zustand   <b>original</b> ursprünglicher Zustand   <b>pitiful</b> jämmerlicher / trauriger Zustand   <b>serious</b> (MED) bedenklicher / ernster Zustand   <b>stable</b> (MED) stabiler Zustand   <b>wretched</b> erbärmlicher / kläglicher / miserabler / trauriger Zustand</p> <ul style="list-style-type: none"> <li>▪ <b>the building is in bad condition</b> das Gebäude ist in einem baulich schlechten Zustand</li> <li>▪ <b>he is in a critical condition</b> (MED) sein Zustand ist kritisch, er schwiebt in Lebensgefahr</li> <li>▪ <b>the dreadful conditions in this country</b> die Furchtbarkeit der Zustände in diesem Land</li> <li>▪ <b>not be in good condition</b> (person) mitgenommen sein</li> <li>▪ <b>to get something into good condition</b> (thing) etwas instand setzen</li> <li>▪ <b>to keep a machine in good condition</b> eine Maschine warten / instand halten</li> <li>▪ <b>the human condition</b> die Situation der / des Menschen, das Menschsein</li> <li>▪ <b>she is in an interesting condition</b> (inf.) PL Sie ist in anderen Umständen</li> <li>▪ <b>this stamp is in mint condition</b> diese Briefmarke ist postfrisch</li> <li>▪ <b>something is in mint condition</b> etwas sieht fabrikneu aus</li> <li>▪ <b>to keep the garden in a proper condition</b> (coll.) den Garten in Schuss halten</li> <li>▪ <b>the building is in ruinous condition</b> das Gebäude ist verfallen / baufällig</li> <li>▪ <b>he is in a stable condition</b> (MED) der Patient ist stabil, der Zustand des Patienten ist stabil</li> <li>▪ <b>the patient is in an unstable condition</b> (MED) der Zustand des Patienten ist labil / nicht stabil</li> <li>▪ <b>this machine is not in working condition</b> diese Maschine funktioniert nicht</li> <li>▪ <b>in bad condition</b> (person) in schlechter Verfassung</li> <li>▪ <b>in bad / poor condition</b> (thing) in schlechtem Zustand, schlecht instand</li> <li>▪ <b>in mint / pristine condition</b> (thing) in tadellosem Zustand</li> <li>▪ <b>in perfect condition</b> (thing) in tadellosem Zustand, in Bestzustand, (goods) fehlerfrei</li> <li>▪ <b>in prime condition</b> (food) von hervorragender Qualität, (fruit) voll ausgereift</li> <li>▪ <b>in proper</b> (thing) in ordnungsgemäßem Zustand</li> <li>▪ <b>in a satisfactory condition</b> (MED) in einem zufriedenstellenden Zustand</li> <li>▪ <b>in tip-top condition</b> (thing) in einem Topzustand</li> <li>▪ <b>in working condition</b> (machine) betriebsfähig</li> <li>▪ <b>general condition</b> Allgemeinzustand m, (MED) Allgemeinbefinden n</li> <li>▪ <b>initial condition</b> Ausgangszustand m</li> </ul>	<p><b>mental condition</b> Geisteszustand m</p> <p><b>miserable condition</b> Armseligkeit f</p> <p><b>natural condition</b> Rohzustand m</p> <p><b>original condition</b> Ausgangszustand m</p> <p><b>weak condition</b> Schwächezustand m</p> <p><b>working condition</b> (TECH) Betriebszustand m</p> <p><b>4. ILLNESS</b></p> <p><b>feverish</b> fiebige Erkrankung</p> <ul style="list-style-type: none"> <li>▪ <b>he can't do that with his heart condition</b> er mit seinem Herzleiden kann das nicht</li> <li>▪ <b>she is suffering from / has a heart condition</b> sie hat ein Herzleiden, sie ist herzleidend, (inf.) sie hat am Herzen</li> <li>▪ <b>to have a serious heart condition</b> schwer herzkrank sein</li> <li>▪ <b>with a heart condition</b> herzleidend</li> </ul> <ul style="list-style-type: none"> <li>▪ <b>heart condition</b> Herzleiden n, Herzkrankheit f</li> <li>▪ <b>lung condition</b> Lungenleiden n</li> <li>▪ <b>medical condition</b> Erkrankung f</li> <li>▪ <b>nervous condition</b> Nervenleiden n</li> <li>▪ <b>skin condition</b> Hautkrankheit f</li> </ul> <p><b>5. FITNESS (SG)</b></p> <p><b>good</b> gute Kondition   <b>bad</b> schlechte Kondition   <b>physical</b> körperliche Verfassung</p> <ul style="list-style-type: none"> <li>▪ <b>a body in peak condition</b> ein gut durchtrainierter Körper</li> <li>▪ <b>poor physical condition</b> schlechte körperliche Verfassung</li> <li>▪ <b>in bad condition</b> in schlechter Verfassung</li> <li>▪ <b>in good condition</b> (person) [gut] in Form, in guter Verfassung, (coll.) auf der Höhe</li> <li>▪ <b>in peak condition</b> (athlete) in Best- / Höchstform</li> <li>▪ <b>in prime condition</b> (sportsmen, animals) in bester Verfassung</li> <li>▪ <b>in superb condition</b> gut durchtrainiert</li> <li>▪ <b>in tip-top condition</b> (person) in Topform</li> <li>▪ <b>physical condition</b> Form f</li> </ul> <p><b>content</b></p> <p><b>1. AMOUNT OF A SUBSTANCE CONTAINED IN SOMETHING (SG)</b></p> <ul style="list-style-type: none"> <li>▪ <b>to have a high [alcohol / acid / fat etc.] content</b> [alkohol- / säure- / fett]reich / stark [alkohol- / säure- / fett]haltig sein</li> <li>▪ <b>to have a low [alcohol etc.] content</b> schwach / wenig [alkohol etc.]haltig / [alkohol etc.]arm sein</li> <li>▪ <b>to have a high carbon monoxide / gold / lead content</b> einen hohen Gehalt an Kohlenmonoxid / Gold / Blei haben</li> <li>▪ <b>with a high alcohol / fat content</b> mit hohem Alkohol- / Fettgehalt</li> <li>▪ <b>with a high fibre content</b> mit hohem Anteil an Ballaststoffen</li> <li>▪ <b>with a very low fat / iron content</b> stark entfettet / enteisen sein</li> <li>▪ <b>with a high salt content</b> sehr salzhaltig sein</li> <li>▪ <b>with a low salt content</b> salzarm</li> <li>▪ <b>acid content</b> Säuregehalt m</li> <li>▪ <b>alcohol content</b> Alkoholgehalt m</li> </ul>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>blood-alcohol content</b> Blutalkoholspiegel m  <b>calorie content</b> Kaloriengehalt m  <b>fat content</b> Fettgehalt m  <b>fibre content</b> Ballaststoffgehalt m  <b>gold content</b> Goldgehalt m  <b>iron content</b> Eisengehalt m  <b>lead content</b> Bleigehalt m  <b>meat content</b> Fleischeinwaage m  <b>moisture content</b> Feuchtigkeitsgehalt m  <b>nicotine content</b> Nikotingehalt m  <b>ore content</b> Erzgehalt m  <b>protein content</b> Eiweißgehalt m  <b>silver content</b> Silbergehalt m  <b>sugar content</b> Zuckergehalt m  <b>sulphur content</b> Schwefelgehalt m  <b>vitamin content</b> Vitamingehalt m  <b>water content</b> Wassergehalt m</p> <p>2. CAPACITY</p> <ul style="list-style-type: none"> <li>▪ <b>cubic content</b> Rauminhalt m, Kubikinhalt m</li> </ul> <p>3. THINGS THAT ARE INSIDE SOMETHING (PL)</p> <p>(the) <b>entire</b> sg der gesamte Inhalt</p> <ul style="list-style-type: none"> <li>▪ <b>home contents</b> sg Haustrat m</li> </ul> <p>4. INFORMATION / MEANING CONTAINED IN SOMETHING (SG)</p> <p><b>intellectual</b> der geistige Gehalt   <b>political</b> der politische Gehalt   <b>religious</b> der religiöse Gehalt   <b>technical</b> der fachliche Gehalt</p> <ul style="list-style-type: none"> <li>▪ <b>facutal content</b> Sachgehalt m  <b>information content</b> (COMPUT) Information f, Informationsgehalt m  <b>informational content</b> Informationsgehalt m  <b>semantic content</b> (LING) Bedeutungsgehalt m</li> </ul>	<p><b>Discussion</b>   <b>detailed</b> ausführliche / eingehende Diskussion, gründliche Erörterung   <b>earnest</b> ernsthafte Diskussion   <b>endless</b> endlose / uferlose Diskussion   <b>extensive</b> ausführliche Diskussion   <b>formal</b> formale / offizielle Diskussion   <b>frank</b> ehrliche / freimütige / offene Diskussion   <b>full</b> ausführliche / inhaltsreiche Diskussion   <b>general</b> discussion allgemeine Diskussion   <b>heated</b> erregte / hitzige Diskussion   <b>inconclusive</b> ergebnislose Diskussion   <b>informal</b> informelle / offene / ungezwungene / zwanglose Diskussion   <b>initial</b> anfängliche Diskussionen   <b>intense</b> intensive / heftige Diskussion, heftige Debatte   <b>interesting</b> interessante Diskussion   <b>lengthy</b> endlose / langwierige Diskussion   <b>lively</b> angeregte / lebhafte / rege Diskussion   <b>long</b> lange Diskussion   <b>open</b> offene Diskussion   <b>peaceful</b> friedliche Diskussion   <b>political</b> politische Diskussion   <b>previous</b> vorangegangene / vorherige Diskussion   <b>prolonged</b> langwierige Diskussion   <b>protracted</b> langwierige Diskussion   <b>public</b> öffentliche Diskussion   <b>quiet</b> friedliche Diskussion   <b>rational</b> sinnvolle / vernünftige Diskussion   <b>reasoned</b> vernünftige Diskussion   <b>sensible</b> vernünftige Diskussion   <b>serious</b> ernste / ernsthafte Diskussion   <b>stormy</b> hitzige / stürmische Diskussion   <b>technical</b> sachliche Diskussion   <b>unfruitful</b> fruchtlose/ unproduktive Diskussion   <b>useful</b> fruchtbare Diskussion   <b>wider</b> umfassendere Diskussion   <b>wide-ranging</b> ausführliche / weitreichende Diskussion  </p>
<p><b>discovery</b></p> <p><b>amazing</b> erstaunliche Entdeckung   <b>archaeological</b> archäologische Entdeckung, archäologischer Fund   <b>great</b> große Entdeckung   <b>grim</b> grauenvolle / grausige Entdeckung   <b>important</b> wichtige Entdeckung / wichtige Erkenntnis   <b>incredible</b> erstaunliche /, (coll.) unglaubliche Entdeckung   <b>major</b> wichtige Entdeckung   <b>new</b> (PL) neue Entdeckungen   <b>recent</b> die jüngste / neueste Entdeckung   <b>remarkable</b> bemerkenswerte / verblüffende Entdeckung   <b>revolutionary</b> revolutionierende Entdeckung   <b>scientific</b> wissenschaftliche Entdeckung   <b>significant</b> bedeutsame / bedeutungsvolle Entdeckung   <b>startling</b> aufregende / sensationelle Entdeckung   <b>surprising</b> überraschende Entdeckung   <b>terrible</b> (coll.) furchterliche / schreckliche Entdeckung   <b>unexpected</b> unerwartete Entdeckung</p> <ul style="list-style-type: none"> <li>▪ <b>an inspired discovery</b> (fig.) das Ei des Kolumbus  <b>a recent discovery</b> eine Entdeckung neueren Datums</li> <li>▪ <b>new discovery</b> (lit. a. fig.) Neuentdeckung f</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>they had an animated / lively discussion</b> sie diskutierten angeregt / eifrig / lebhaft  <b>we had a fairly extensive discussion</b> wir haben es ziemlich ausführlich diskutiert  <b>open to further discussion</b> ungeklärt  <b>something invites further discussion</b> etwas fordert zu weiteren Diskussionen auf  <b>a general discussion developed</b> die Diskussion darüber wurde allgemein  <b>you can't have a good discussion with him</b> mit ihm kann man schlecht debattieren  <b>to have a very heated discussion</b> mit großer Härte diskutieren  <b>a high-level discussion</b> eine Diskussion auf höchster Ebene  <b>an inconclusive discussion</b> eine Diskussion ohne [schlüssiges] Ergebnis  <b>to have a lengthy / prolonged discussion</b> about something langwierig über etwas beraten  <b>why do we need all this long discussion?</b> Was gibt es da noch groß zu diskutieren?  <b>there was a great deal of political discussion</b> es wurde viel politisiert  <b>a private discussion</b> ein Gespräch unter vier Augen  <b>a round-table discussion</b> eine Diskussion am runden Tisch  <b>to have a sensible discussion about something</b> über etwas vernünftig diskutieren</li> </ul> <ul style="list-style-type: none"> <li>▪ <b>exploratory discussion</b> Sondierungsgespräch n  <b>forum discussion</b> Forumsdiskussion f, Forumsgespräch n  <b>group discussion</b> Gruppendiskussion f</li> </ul>
<p><b>discussion</b></p> <p><b>acrimonious</b> erbitterte Diskussion   <b>amicable</b> friedliche Diskussion   <b>animated</b> angeregte / lebhafte / rege Diskussion   <b>brief</b> kurze Diskussion   <b>candid</b> ehrliche / offene Diskussion   <b>confidential</b> vertrauliche Diskussion   <b>considerable</b> eingehende</p>	<p><b>examination</b></p> <p>1. DETAILED INSPECTION</p> <p><b>brief</b> kurze Untersuchung   <b>careful</b> gründliche / sorgfältige Untersuchung, sorgfältige Überprüfung   <b>close</b> eingehende / genaue Untersuchung   <b>critical</b></p>

kritische Untersuchung / Überprüfung | **cursor** flüchtige / oberflächliche Untersuchung | **detailed** eingehende / genaue Untersuchung | **forensic** forensische / gerichtsmedizinische Untersuchung | **full** eingehende / genaue Untersuchung | **important** wichtige Untersuchung | **initial** anfängliche Untersuchung / Überprüfung | **lengthy** ausführliche Untersuchung | **medical** ärztliche / medizinische Untersuchung | **physical** ärztliche Untersuchung | **scientific** wissenschaftliche Prüfung | **superficial** oberflächliche Überprüfung | **thorough** gründliche Untersuchung

- **on closer examination** bei genauer Prüfung, bei genauerer / näherer Untersuchung, bei näherem Hinsehen
- **to give somebody a follow-up examination** (MED) jemanden nachuntersuchen
- **to have a gynaecological examination** sich gynäkologisch untersuchen lassen
- **to have a medical examination** sich ärztlich untersuchen lassen
- **to give somebody a psychiatric examination** jemanden auf seinen Geisteszustand untersuchen, jemanden psychiatrisch untersuchen
- **give somebody a thorough examination** (MED) jemanden gründlich untersuchen
- **make a thorough examination of something** etwas gründlich durchforschen
- **customs examination** Zollkontrolle f
- **follow-up examination** Nachuntersuchung f
- **further examination** Nachuntersuchung f
- **medical examination** (MIL) Musterung f
- **post-mortem examination** (MED) Autopsie f, Leichenschau f, Obduktion f, Sektion f
- **preliminary examination** (MED) Voruntersuchung f
- **routine examination** (MED) Routineuntersuchung f
- **x-ray examination** (MED) Röntgenuntersuchung f

## 2. TEST OF KNOWLEDGE

difficult schwierige Prüfung | **easy** leichte Prüfung | **external** externe Prüfung | **formal** formelle / offizielle Prüfung | **important** bedeutende / wichtige Prüfung | **mock** simulierte Prüfung | **oral** mündliche Prüfung | **practical** praktische Prüfung | **public** öffentliche Prüfung | **rigorous** strenge Prüfung | **stiff** schwere / schwierige Prüfung | **written** schriftliche Prüfung

- **to take an external examination to get one's diploma** die Diplomprüfung extern ablegen
- **to take an internal examination** an der Universität, an der man immatrikuliert ist, eine Prüfung ablegen
- **multiple-choice examination** Prüfung nach dem Multiplechoiceverfahren
- **to submit someone to an oral examination** jemanden mündlich prüfen
- **public examination** eine der Öffentlichkeit zugängliche Prüfung
- **to have a written examination** schriftlich geprüft werden
- **competitive examination** Auswahlprüfung f
- **end-of-term examination** Semesterprüfung f
- **entrance examination** Aufnahmeprüfung f
- **final examination** Abschlussprüfung f, Examen n
- **intermediate examination** Zwischenprüfung f
- **matriculation examination** Hochschulreife f
- **mock examination** Übungstest m
- **preliminary examination** Vorprüfung f
- **professional examination** Fachprüfung f
- **qualifying examination** Auswahlprüfung f
- **selection examination** Ausleseprüfung f, Auswahlprüfung f
- **state examination** (usu.: JUR) Staatsexamen n

## framework

administrative administrative Struktur | **analytical** analytischer Rahmen, analytische Struktur | **basic** grundlegende Struktur | **broad** breiter / grober Rahmen | **chronological** chronologischer Rahmen, chronologische Struktur | **coherent** einheitlicher Rahmen, kohärente Struktur, zusammenhängende Abfolge | **comprehensive** umfassender Rahmen | **conceptual** konzeptueller Rahmen | **constitutional** gesetzlicher / konstitutioneller Rahmen | **convenient** günstiger / passender Rahmen | **democratic** demokratische Struktur | **economic** wirtschaftliche Struktur, wirtschaftlicher Rahmen | **existing** bestehende Struktur | **financial** finanzieller Rahmen, finanzielle Struktur | **historical** historischer Rahmen | **legal** gesetzlicher / rechtlicher Rahmen | **political** politischer Rahmen, politische Struktur | **theoretical** theoretisches Gerüst | **wider** größerer Rahmen

- **to give a ceremony a musical framework** (fig.) eine Feier musikalisch umrahmen
- **basic framework** PL Rahmenbedingungen f, Grundstruktur f
- **conceptual framework** SG Regelwerk n, (ACCOUNT) PL Buchungsrichtlinien, (fig.) Raster n
- **economic framework** Wirtschaftsgefüge n
- **existing framework** PL Rahmenbedingungen f, Grundstruktur f
- **general framework** PL Rahmenbedingungen f, Grundstruktur f
- **legal framework** Rechtsstruktur f
- **regulatory framework** Rechtsrahmen m

## mechanism

**cellular** (BIOL, CHEM) zellulärer Mechanismus | **clumsy** primitiver Mechanismus | **complex** komplexer Mechanismus | **fail-safe** störungssicherer Mechanismus | **genetic** (BIOL, MED) genetischer Mechanismus | **immunological** (MED) immunologischer Mechanismus | **molecular** (BIOL, CHEM) molekularer Mechanismus | **neural** (MED) neuraler Mechanismus | **physiological** (BIOL, MED) physiologischer Mechanismus | **precise** genauer Mechanismus | **psychological** (MED) psychologischer Mechanismus | **social** gesellschaftlicher Mechanismus | **special** spezieller Mechanismus | **underlying** zugrundeliegender Mechanismus | **useful** nützlicher Mechanismus

- **adaptation / adaptive mechanism** (SOCIOL) Anpassungsmechanismus m
- **automatic mechanism** (lit. a. fig.) Automatik f, (TECHN) Automatismus m
- **brake mechanism** Bremsvorrichtung f
- **clock[work] mechanism** Uhrwerk n
- **control mechanism** Kontrollmechanismus m
- **defence / escape mechanism** (PSYCH) Abwehrmechanismus m
- **immunological mechanism** (MED) Immunmechanismus m
- **locking mechanism** (TECHN) Verriegelung f, Verriegelungsmechanismus m, Gesperrte n, Sperrgetriebe n
- **steering mechanism** (TECHN) Steuerung f, Steuerungsvorrichtung f
- **striking mechanism** Schlagwerk n
- **survival mechanism** (PSYCH) Überlebensmechanismus m
- **trigger mechanism** (PSYCH) Auslöser m

## population

active aktive Bevölkerung | **ageing** älter werdende Bevölkerung | **American** amerikanische Bevölkerung | **black** schwarze Bevölkerung | **city** städtische Bevölkerung | **dense** dichte Besiedelung | **elderly** ältere Bevölkerung | **expanding** wachsende Bevölkerung | **female** weibliche Bevölkerung | **floating** (fig.) wandernde Bevölkerung | **growing** population wachsende Bevölkerung | **heterogeneous** heterogene / vielschichtige Bevölkerung | **huge** riesige Bevölkerung, sehr große Bevölkerungsdichte | **increasing** zunehmende Bevölkerung / Bevölkerungsdichte / Bevölkerungszahl | **indigenous** einheimische Bevölkerung | **jewish** jüdische Bevölkerung | **large** große Bevölkerung / Bevölkerungsdichte | **male** männliche Bevölkerung | **mixed** gemischte Bevölkerung | **native** eingeborene / heimische Bevölkerung | **original** ursprüngliche Bevölkerung / Bevölkerungszahl | **resident** [orts]ansässige Bevölkerung | **rising** wachsende Bevölkerung | **scattered** verstreut lebende / weit verstreute Bevölkerung | **shrinking** abnehmende Bevölkerung, sinkende Bevölkerungszahl | **small** geringe Population, (people) geringe / kleine Bevölkerung, (animals) geringer / kleiner Bestand | **sparse** schwache Besiedelung | **stable** stabile Bevölkerungszahl | **thin** dünne Besiedelung | **white** weiße Bevölkerung | **working** arbeitende / erwerbstätige Bevölkerung | **young** junge Bevölkerung | **youthful** jugendliche Bevölkerung

- **the growing black population in London** die wachsende Zahl von Schwarzen in London
- **the entire population** die Gesamtheit der Bevölkerung
- **the growing immigrant population** der wachsende Einwandereranteil
- **the prison population** die Zahl der in Haft befindlichen Personen
- **the entire student population** die Gesamtheit der Studenten
- **city population** Großstadtbevölkerung f
- **civil / civilian population** Zivilbevölkerung f
- **elephant population** Elefantenbestand m, Elefantepopulation f
- **fish population** Fischbestand m
- **general population** Allgemeinbevölkerung f
- **global population** Gesamtbevölkerung f, Weltbevölkerung f
- **immigrant population** PL Einwanderer m, Einwanderinnen f
- **local population** Bewohner m/f, Einwohner m/f
- **mixed population** Mischbevölkerung f
- **native population** PL Eingeborene m/f, Einheimische m/f
- **overall population** (people) Gesamtbevölkerung f, (animals) Gesamtbestand m
- **resident population** Wohnbevölkerung f
- **rural population** Landbevölkerung f
- **surplus population** Bevölkerungsüberschuss m
- **total population** Gesamtbevölkerung f
- **transient population** Nichtansässige m/f
- **urban population** Stadtbevölkerung f
- **working population** Erwerbsbevölkerung f
- **world population** Gesamtbevölkerung f, Weltbevölkerung f

## principle

**abstract** abstraktes Prinzip | **Archimedes' principle** (PHYS) archimedisches Prinzip | **cardinal** oberster Grundsatz | **central** zentrales Prinzip | **legal**

rechtliche Grundlage | **common** allgemeine Grundlage, allgemeines Prinzip | **democratic** demokratische Grundlage / Grundsatz / Prinzip | **elementary** grundlegendes Prinzip | **essential** wichtiger Grundsatz / wichtiges Prinzip | **female** weibliches Prinzip | **general** durchgängiges Prinzip | **high** PL hohe Prinzipien | **inflexible** (usu.: PL) starre Prinzipien | **main** oberster Grundsatz | **male** männliches Prinzip | **moral** moralischer Grundsatz, moralisches Prinzip | **overriding** vorrangiges Prinzip | **political** politische Grundlage, politisches Prinzip, (of an individual) politische Einstellung | **rigid** strenges Prinzip, PL starre Vorschriften | **scientific** wissenschaftliche Grundlage | **simple** einfaches Prinzip | **socialist** sozialistisches Prinzip | **sound** vernünftiger Grundsatz, vernünftiges Prinzip | **strict** strenges Prinzip, PL starre Vorschriften | **theoretical** PL theoretische Grundlagen | **underlying** zugrundeliegendes Prinzip | **universal** durchgängiges / universelles Prinzip | **vital** wichtigstes Prinzip

- **to stick to the broad principle** die große Linie wahren
- **to go back to first principles** zu den Grundlagen zurückkehren
- **the governing principle** das Prinzip, von dem sich jemand leiten lässt
- **a man of high principle** ein Mann mit hohen Prinzipien
- **following liberal principles** liberal ausgerichtet sein
- **achievement principle** Leistungsprinzip n
- **active principle** Wirkprinzip n
- **basic principle** PL Grundlagen f, Grundprinzip m, Grundregel f, Leitsatz m
- **cardinal / central principle** Grundprinzip n, Hauptprinzip n
- **elementary principle** Grundprinzip n, Grundregel f
- **first principle** Hauptsatz m
- **formal principle** Gestaltungsprinzip n
- **fundamental principle** PL Grundlagen f, Grundprinzip n
- **general principle** Grundsatz m
- **guiding principle** Leitgedanke m, Leitsatz m, (fig.) Richtschnur f
- **hereditary principle** Erbprinzip n
- **legal principle** Rechtsgrundsatz m
- **clear legal principles** Rechtsklarheit f
- **key principle** Hauptprinzip n, Schlüsselprinzip n
- **moral principles** PL Moraigesetze n, Prinzipien n
- **pleasure principle** (PSYCH) Lustprinzip n
- **political principles** sg Politik f
- **structural principle** Aufbauprinzip n
- **ultimate principle** Grundprinzip n
- **uncertainty principle** (PHYS) Unschärferelation f
- **underlying principle** Grundprinzip n

## proposal

alternative alternativer Vorschlag | **concrete** konkreter Vorschlag | **controversial** umstrittener Vorschlag | **detailed** ausführlicher / detaillierter Vorschlag | **economic** wirtschaftlicher Vorschlag | **far-reaching** weit greifender / weit reichender Vorschlag | **firm** verbindliches Angebot | **general** allgemeiner Vorschlag | **grotesque** unsinniger Vorschlag | **inadequate** inakzeptables Angebot, inakzeptabler Vorschlag | **major** bedeutender Vorschlag | **modest** bescheidenes Angebot, bescheidener Vorschlag | **new** neuer Vorschlag | **original** ursprüngliches Angebot, ursprünglicher Vorschlag

- **compromise proposal** Kompromissvorschlag m
- **counter proposal** Gegenvorschlag m

<p><b>marriage proposal</b> Heiratsantrag m <b>peace proposal</b> Friedensvorschlag m</p>	<p><b>free sample</b> Gratismuster n, Gratisprobe f <b>random sample</b> Stichprobe f <b>rock sample</b> Gesteinsprobe f <b>tissue sample</b> (MED) Gewebeprobe f <b>urine sample</b> Harnprobe f, Urinprobe f</p>
<p><b>remedy</b></p> <p>adequate adäquates Mittel   alternative alternatives Mittel   <b>appropriate</b> geeignetes Mittel   <b>common</b> allgemein bekannte / verbreitetes Mittel   <b>desperate</b> äußerstes Mittel   <b>drastic</b> PL drastische Mittel   <b>effective</b> wirksames Mittel   <b>good</b> gutes Mittel   <b>homoeopathic</b> homöopathisches Mittel   <b>palliative</b> schmerzstillendes Mittel   <b>proposed</b> vorgeschlagene Lösung   <b>simple</b> einfaches Mittel   <b>sure</b> sicheres / zuverlässiges Mittel   <b>traditional</b> traditionelles / herkömmliches Mittel   <b>unfailing</b> unfehlbares Mittel   <b>universal</b> universales / universelles Mittel   <b>usual</b> übliches Mittel</p> <ul style="list-style-type: none"> <li>▪ <b>to take Bach's flower remedy</b> (MED) eine Bachblütentherapie machen</li> <li>▪ <b>the best remedy for</b> das beste Mittel gegen / für <b>a cold remedy</b> ein Mittel gegen Erkältung</li> <li>▪ <b>a desperate disease requires a dangerous / drastic remedy</b> (fig.) extreme Situationen erfordern extreme Maßnahmen</li> <li>▪ <b>to apply drastic remedies</b> (lit.) drastische Mittel ergreifen, (fig.) mit eisemem Besen kehren</li> <li>▪ <b>civil remedy</b> (JUR) Rechtsbehelf n, Rechtsmittel m <b>cough remedy</b> Hustenmittel n <b>contractual remedy</b> Vertragsentschädigung f <b>folk remedy</b> Volksmittel n <b>legal remedy</b> (JUR) Rechtsbehelf n, Rechtsmittel m <b>drastic / kill-or-cure remedy</b> Radikalkur f, Rosskur f <b>herbal remedies</b> PL Kräuterheilmittel n <b>home / household remedy</b> Hausmittel n <b>judicial</b> (JUR) Rechtsbehelf n, Rechtsmittel m <b>natural remedy</b> (medication) Naturheilmittel n, (therapy) Naturheilverfahren n <b>patent remedy</b> (lit. a. fig.) Spezialrezept n, Patentrezept n, Patentlösung f <b>popular remedy</b> Hausmittel n <b>quack remedy</b> (slightly derog.) Mittelchen n <b>universal remedy</b> (lit. a. fig.) Allheilmittel n, Universalmittel n</li> </ul>	<p><b>tendency</b></p> <p>alarming beunruhigende Tendenz   <b>artistic</b> PL künstlerische Neigungen   <b>centrifugal</b> PL zentrifugale Tendenzen   <b>clear</b> eindeutige / klare Neigung / Tendenz   <b>common</b> allgemeine Tendenz, allgemeiner Trend   <b>contradictory</b> PL widersprüchliche Tendenzen   <b>general</b> allgemeiner Trend   <b>greater</b> größere Tendenz   <b>growing</b> steigende Tendenz   <b>homosexual</b> homosexuelle Veranlagung   <b>inbuilt</b> angeborene / natürliche Tendenz   <b>increasing</b> steigende / zunehmende Tendenz   <b>innate</b> angeborene / natürliche Tendenz   <b>lesbian</b> lesbische / sapphische Veranlagung   <b>marked</b> ausgeprägte Neigung, ausgeprägter Hang   <b>natural</b> angeborene / natürliche Tendenz   <b>pronounced</b> ausgeprägte Neigung, ausgeprägter Hang   <b>slight</b> leichte / schwache Tendenz   <b>strong</b> starke Tendenz   <b>underlying</b> grundlegende / unterschwellige Tendenz   <b>unfortunate</b> unglückliche Neigung / Tendenz   <b>universal</b> allgemeine Tendenz, allgemeiner Trend</p> <ul style="list-style-type: none"> <li>▪ <b>to have aggressive tendencies</b> aggressiv veranlagt sein</li> <li>▪ <b>he had an annoying tendency to forget things</b> er hatte die ärgerliche Angewohnheit, alles zu vergessen</li> <li>▪ <b>to have criminal tendencies</b> eine Veranlagung zur Kriminalität haben, kriminell veranlagt sein</li> <li>▪ <b>a homicidal tendency</b> ein Drang zum Töten</li> <li>▪ <b>he has homicidal tendencies</b> er könnte zum Mörder werden</li> <li>▪ <b>to have homosexual tendencies</b> homosexuelle veranlagt sein</li> <li>▪ <b>to have an inherent tendency to something</b> etwas ist anlagebedingt</li> <li>▪ <b>to show a left-wing / right-wing tendency</b> links / rechts angehaucht sein</li> <li>▪ <b>to have lesbian tendencies</b> lesbisch veranlagt sein</li> <li>▪ <b>he has a marked tendency towards criminality</b> er hat einen deutlichen / starken Hang zur Kriminalität</li> <li>▪ <b>to have a natural tendency to do something</b> naturgemäß dazu neigen, etwas zu tun</li> <li>▪ <b>to show a stronger tendency</b> fester / stärker tendieren</li> <li>▪ <b>to have suicidal tendencies</b> eine Neigung zum Selbstmord haben, selbstmordgefährdet sein</li> <li>▪ <b>general tendency</b> Grundrichtung f <b>left-wing / right wing tendency</b> (POL) (sl.) Links- / Rechtsdrall m</li> </ul>
<p><b>sample</b></p> <p>biased verzerrte Stichprobe / Auswahl   <b>clinical</b> klinische Stichprobe   <b>free</b> unverkäufliches Muster   <b>large</b> große Auswahl   <b>representative</b> repräsentative Auswahl, repräsentatives Beispiel   <b>small</b> kleine Auswahl</p> <ul style="list-style-type: none"> <li>▪ <b>to have a blood sample taken</b> Blut abgenommen bekommen</li> <li>▪ <b>blood sample</b> Blutprobe f <b>(the) taking of a blood sample</b> Blutentnahme f <b>commercial sample</b> [Waren]muster n <b>faecal sample</b> Kotprobe f <b>free sample</b> Gratismuster n, Gratisprobe f <b>random sample</b> Stichprobe f <b>rock sample</b> Gesteinsprobe f <b>tissue sample</b> (MED) Gewebeprobe f <b>urine sample</b> Harnprobe f, Urinprobe f</li> </ul>	

A	Abfolge	framework	Eingeborene	population
	Abholung	collection	Einheimische	population
	Abschlussprüfung	examination	Einwanderer, Einwanderinnen	population
	Abwehrmechanismus	mechanism	Einwandereranteil	population
	Alkoholgehalt	content	Einwohner	population
	Allgemeinbefinden	condition	Eis	condition
	Allgemeinebevölkerung	population	Eisengehalt	content
	Allgemeinzustand	condition	Eiweißgehalt	content
	Allheilmittel	remedy	Elefantenbestand	population
	Altpapiersammlung	collection	Elefantenpopulation	population
	Angebot	proposal	Entdeckung	discovery
	Angewohnheit	tendency	Erbprinzip	principle
	Anpassungsmechanismus	mechanism	Ergebnis	discussion
	Anteil	content	Erkältung	remedy
	Arbeitsbedingungen	condition	Erkrankung	condition
	Ärmlichkeit	condition	Erwerbsbevölkerung	population
	Armseligkeit	condition	Erzgehalt	content
	Aufbauprinzip	principle	Examen	examination
	Aufnahmeprüfung	examination	Existenzbedingungen	condition
	Augen	discussion		
	Ausgangszustand	condition		
	Aushilfe	basis		
	Ausleseprüfung	examination		
	Auslöser	mechanism		
	Auswahl	collection		
	Auswahl	sample		
	Auswahlprüfung	examination		
	Automatik	mechanism		
	Automatismus	mechanism		
	Autopsie	examination		
B	Bachblütentherapie	remedy		
	Ballaststoffe	content		
	Ballaststoffgehalt	content		
	Basis	basis		
	Bedeutungsgehalt	content		
	Bedingung, -en	condition		
	Beengtheit	condition		
	Beispiel	sample		
	Besen	remedy		
	Besiedelung	population		
	Bestform	condition		
	Bestzustand	condition		
	Betriebszustand	condition		
	Bevölkerung	population		
	Bevölkerungsdichte	population		
	Bevölkerungsüberschuss	population		
	Bevölkerungszahl	population		
	Bewohner	population		
	Blei	content		
	Bleigehalt	content		
	Blut	sample		
	Blutalkoholspiegel	content		
	Blutentnahme	sample		
	Blutprobe	sample		
	Bodenverhältnisse	condition		
	Bremsvorrichtung	mechanism		
	Briefmarke	condition		
	Briefmarkensammlung	collection		
	Buchungsrichtlinien	framework		
D	Datenerfassung	collection		
	Datum	discovery		
	Debatte	discussion		
	Diplomprüfung	examination		
	Diskussion	discussion		
	Drang	tendency		
E	Ebene	discussion		
	Ei	discovery		

<b>H</b>			
Haft	<i>population</i>	Lösung	<i>remedy</i>
Hand	<i>basis</i>	Lungenleiden	<i>condition</i>
Handapparat	<i>collection</i>	Lustprinzip	<i>principle</i>
Hang	<i>tendency</i>		
Harnprobe	<i>sample</i>		
Härte	<i>discussion</i>		
Haufen	<i>collection</i>		
Hauptprinzip	<i>principle</i>		
Hauptsatz	<i>principle</i>		
Haus	<i>collection</i>		
Hausmittel	<i>remedy</i>		
Hausrat	<i>content</i>		
Haussammlung	<i>collection</i>		
Hautkrankheit	<i>condition</i>		
Heiratsantrag	<i>proposal</i>		
Herbstkollektion	<i>collection</i>		
Herzen	<i>condition</i>		
Herzkrankheit	<i>condition</i>		
Herzleiden	<i>condition</i>		
Hinsehen	<i>examination</i>		
Hochschulreife	<i>examination</i>		
Höchstform	<i>condition</i>		
Höhe	<i>condition</i>		
Hustenmittel	<i>remedy</i>		
<b>I</b>			
Immunmechanismus	<i>mechanism</i>	Nachuntersuchung	<i>examination</i>
Information	<i>content</i>	Naturheilmittel	<i>remedy</i>
Informationsgehalt	<i>content</i>	Naturheilverfahren	<i>remedy</i>
Inhalt	<i>content</i>	Nebel	<i>condition</i>
<b>J</b>			
Jahr	<i>basis</i>	Neigung, -en	<i>tendency</i>
<b>K</b>			
Kaloriengehalt	<i>content</i>	Nervenleiden	<i>condition</i>
Kälte	<i>condition</i>	Neuentdeckung	<i>discovery</i>
Kleidung	<i>condition</i>	Nichtansässige	<i>population</i>
Kohlenmonoxid	<i>content</i>	Nikotingehalt	<i>content</i>
Kollektion	<i>collection</i>		
Kolumbus	<i>discovery</i>		
Kommission	<i>basis</i>		
Kommissionsbasis	<i>basis</i>		
Kompromissvorschlag	<i>proposal</i>		
Kondition	<i>condition</i>		
Konjunktur	<i>condition</i>		
Kontrollmechanismus	<i>mechanism</i>		
Körper	<i>condition</i>		
Kotprobe	<i>sample</i>		
Kräuterheilmittel	<i>remedy</i>		
Kriminalität	<i>tendency</i>		
Kubikinhalt	<i>content</i>		
Kunstbesitz	<i>collection</i>		
<b>L</b>			
Lage	<i>condition</i>		
Land	<i>condition</i>		
Landbevölkerung	<i>population</i>		
Leben	<i>condition</i>		
Lebensbedingungen	<i>condition</i>		
Lebensgefahr	<i>condition</i>		
Leerung	<i>collection</i>		
Leichenschau	<i>examination</i>		
Leihgabe, -n	<i>collection</i>		
Leihgaben	<i>collection</i>		
Leistung	<i>basis</i>		
Leistungsprinzip	<i>principle</i>		
Leitgedanke	<i>principle</i>		
Leitsatz	<i>principle</i>		
Lichtverhältnisse	<i>condition</i>		
Linie	<i>principle</i>		
Linksdrall	<i>tendency</i>		
<b>M</b>			
Mal			
Marktlage			
Maschine			
Maßnahmen			
Mechanismus			
Mensch, -en			
Menschsein			
Mischbevölkerung			
Mittel			
Mittelchen			
Monat			
Moralgesetze			
Mörder werden			
Müllabfuhr			
Multiplechoiceverfahren			
Münzsammlung			
Muster			
Musterung			
<b>N</b>			
Nachuntersuchung			
Naturheilmittel			
Naturheilverfahren			
Nebel			
Neigung, -en			
Nervenleiden			
Neuentdeckung			
Nichtansässige			
Nikotingehalt			
<b>O</b>			
Obduktion			
Öffentlichkeit			
<b>P</b>			
Partner			
Patentlösung			
Patentrezept			
Patient			
Personen			
Plattensammlung			
Platzverhältnisse			
Politik			
Postulat			
Preisvergleich			
Prêt-à-porter-Kollektion			
Prinzip			
Prinzip, -ien			
Privatsammlung			
Probe			
Problem			
Provision			
Provisionsbasis			
Prozentbasis			
Prüfung			
Punkt			
<b>Q</b>			
Qualität			
<b>R</b>			
Radikalkur			
Rahmen			
Rahmenbedingung			
Rahmenbedingungen			
Raster			
Rauminhalt			
Rechtsbelehr			

Rechtsdrall	<i>tendency</i>	Universalmittel	<i>remedy</i>
Rechtsgrundlage	<i>basis</i>	Universität	<i>examination</i>
Rechtsgrundsatz	<i>principle</i>	Umschäferelation	<i>principle</i>
Rechtsklarheit	<i>principle</i>	Unterkunftbedingungen	<i>condition</i>
Rechtsmittel	<i>remedy</i>	Untersuchung	<i>examination</i>
Rechtsrahmen	<i>framework</i>	Urinprobe	<i>sample</i>
Rechtsstruktur	<i>framework</i>		
Regelgröße	<i>condition</i>		
Regelwerk	<i>framework</i>		
Richtschnur	<i>principle</i>		
Rohzustand	<i>condition</i>		
Röntgenuntersuchung	<i>examination</i>		
Rosskur	<i>remedy</i>		
Routineuntersuchung	<i>examination</i>		
<b>S</b>			
Sachgehalt	<i>content</i>	Vertriebsentschädigung	<i>remedy</i>
Sammlung	<i>collection</i>	Vitamingehalt	<i>content</i>
Säuregehalt	<i>content</i>	Volksmittel	<i>remedy</i>
Schallplattensammlung	<i>collection</i>	Voraussetzung,-en	<i>condition</i>
Schlagwerk	<i>mechanism</i>	Vorprüfung	<i>examination</i>
Schlüsselprinzip	<i>principle</i>	Vorschlag	<i>proposal</i>
Schneeverhältnisse	<i>condition</i>	Vorschriften	<i>principle</i>
Schnitt	<i>basis</i>	Voruntersuchung	<i>examination</i>
Schuss	<i>condition</i>	Vorzeichen	<i>condition</i>
Schwächezustand	<i>condition</i>		
Schwarzen	<i>population</i>		
Schwefelgehalt	<i>content</i>		
Sektion	<i>examination</i>		
Selbstmord	<i>tendency</i>		
Semesterapparat	<i>collection</i>		
Semesterprüfung	<i>examination</i>		
Sicht	<i>basis</i>		
Silbergehalt	<i>content</i>		
Situation	<i>condition</i>		
Situationen	<i>remedy</i>		
Sommerkollektion	<i>collection</i>		
Sondierungsgespräch	<i>discussion</i>		
Sperrgetriebe	<i>mechanism</i>		
Spezialrezept	<i>remedy</i>		
Spezialsammlung	<i>collection</i>		
Staatsexamen	<i>examination</i>		
Stadtbevölkerung	<i>population</i>		
Standortbedingungen	<i>condition</i>		
Steuerung	<i>mechanism</i>		
Steuerungsvorrichtung	<i>mechanism</i>		
Stichprobe	<i>sample</i>		
Straßensammlung	<i>collection</i>		
Straßenverhältnisse	<i>condition</i>		
Straßenzustand	<i>condition</i>		
Struktur	<i>framework</i>		
Studenten	<i>population</i>		
Synopse	<i>comparison</i>		
<b>T</b>			
Tagelöhner	<i>basis</i>		
Tatsache, -n	<i>basis</i>		
Teilzeitlehrer/-in	<i>basis</i>		
Temperaturen	<i>condition</i>		
Tendenz, -en	<i>tendency</i>		
Tiere	<i>condition</i>		
Tisch	<i>discussion</i>		
Topzustand	<i>condition</i>		
Töten	<i>tendency</i>		
Trend	<i>tendency</i>		
<b>U</b>			
Überlebensmechanismus	<i>mechanism</i>		
Überprüfung	<i>examination</i>		
Übungstest	<i>examination</i>		
Uhrwerk	<i>mechanism</i>		
Umgebung	<i>condition</i>		
Umstände	<i>condition</i>		