

Kathrin Steyer

9 Proverbs from a Corpus Linguistic Point of View

9.1 Introduction¹³⁸

This chapter is not a general introduction to corpus linguistics. Instead, I will focus on some aspects which are particularly relevant for the empirical study of proverbs in written language.¹³⁹ My examples will be based on German written language corpora, specifically the *Deutsches Referenzkorpus* [German Reference Corpus] (DeReKo) which is located at the Institute for the German Language in Mannheim. The corpus analysis tool used for accessing the corpus data is COSMAS II (CII). However, the questions, search strategies and examples presented in this chapter should be transferable to other corpora and languages. I will discuss how a user who is interested in proverbs can *exploit* the corpus and which kind of knowledge he can gain in this way. It will become clear that no computer makes thinking obsolete – in the end it is always the human who needs to interpret the results. However, automatic methods can be very useful as they allow high quality pre-structuring of mass data. The most important skill is asking the computer questions as intelligently as possible. Mastering this skill and the methods associated with it is something everyone must learn for themselves by practical experience. As the proverbs say: *Grau is alle Theorie* [ww: Grey is all theory; ee: An ounce of experience is worth a ton of theory] and *Übung macht den Meister* [ww: Practice makes the master; ee: Practice makes perfect].¹⁴⁰

138 I thank Annelen Brunner for translating this chapter into English.

139 For more in depth information I recommend: Sinclair, 1991; Sinclair, 2004 who was a pioneer of corpus linguistics; Tognini-Bonelli, 2001 and her definitions of the empirical concepts ‘corpus-based’ and ‘corpus-driven’; McEnery & Wilson, 2001; Lüdeling & Kytö, 2008; McEnery & Hardie, 2012 (introductions in English); Mukherjee, 2009; Lemnitzer & Zinsmeister, 2010; Perkuhn & Keibel & Kupietz, 2012 and the website from Bubenhofer, 2006-2013 (introductions in German); for more about corpus-based phraseology and computer linguistic aspects of phrasemes Cowie, 1998; Heid, 2007; Moon, 1998; Moon, 2007; Rothkegel, 2007 Granger & Meunier, 2008; Ptashnyk & Hallsteinsdóttir & Bubenhofer, 2010; Sailer, 2007 and Steyer i. a. 2003; Steyer 2004; overview in Steyer, 2013. also Mieder, 2009.

140 For finding English proverb equivalents I used the OXFORD Dictionary of Proverbs (Speake, 2008) as well as dict.cc (Dict.cc).

9.2 Why Corpora?

The other day, my children said to me: *Das Leben ist kein Ponyhof* [ww: Life is not a pony farm; ee: Life isn't always a bowl of cherries]. This sentence meant nothing to me. I only know, for example, *Das Leben ist kein Wunschkonzert* [ww: Live is not a request show]. However, they insisted that the pony-farm sentence is a common German proverb, as they used it frequently in their social circle. Checking the internet and the corpora confirmed their claim and my ignorance. Test yourself and your friends: which proverbs spontaneously come to your mind? Which do you think are outdated and which modern? You will see that there are big differences between speakers. Of course there is something like a *proverb memory* of a language community, but your ability to recognize proverbs and your habits of proverb use depend heavily on your specific language biography. No speaker has mastery of all sociolects and dialects. Each speaker has a very limited, very subjective subset of knowledge about proverbs. Even proverb collections and dictionaries only help in a limited way. Without doubt they are a precious and essential part of cultural heritage. But these collections do not contain undisputable facts about the real use of a proverb in contemporary language or its constant changes and adaptations. Many proverbs will be passed on from collection to collection, from dictionary to dictionary and a collective *example memory* will be formed. This is unsatisfactory for second language learners who will be confronted with material that does not adequately reflect the current state of the language.

Long-standing empirical solutions for this problem have been surveys (asking for level of recognition and familiarity) or completion and association tests. With these methods you can gain valuable insight about the distribution of proverbs and their role in a language community.¹⁴¹

During the last decades, new possibilities for recognizing and describing language use have been opened up by the compilation of large electronic text databases (corpora). Corpora are collections of written or spoken texts. Typically the corpus data is digitalized i.e. machine readable and saved on a computer. In addition to the text data itself, corpora can also contain metadata, which describe the data, and linguistic annotations (Lemnitzer & Zinsmeister, 2010: 8).

Many corpora are collections of electronic texts which have been compiled to address a specific research question and are selected for parameters such as author, source, topic, text type, time period or medium. In our context these are special proverb corpora, e.g. searchable collections of texts or text excerpts from data bases which contain proverbs.

But there is also another type of corpora. These are not built with a specific research goal in mind but try to incorporate a representative subset of a language and

141 Overviews can be found in Grzybek, 2012 and Juska-Bacher, 2012. One goal of these studies can be to determine so-called 'paremiological minima' or 'paremiological optima' (Đurčo in this volume).

therefore strive for a broad coverage in regard to time, text type, regional variants etc. They are called general language corpora or reference corpora. However, it must be emphasized that there will never be a corpus that can truly represent a language in its entirety, though a huge corpus certainly gets closer to this goal than a small one.

One of the earliest reference corpora is the British National Corpus (BNC) which was built in the 1960s. The largest general language corpora for German which are publicly available free of charge (with registration, no download) are:

- The corpus of the *Digitales Wörterbuch des 20. Jahrhunderts* [Digital dictionary of the 20th century] (DWDS corpus)
- The *Deutsches Referenzkorpus* [German Reference corpus] (DeReKo) of the Institute for the German language in Mannheim (DeReKo 2013).

Under development is currently the C4 corpus, a combination of the DWDS corpus, the *Schweizer Text Korpus* [corpus of Switzerland], the Austrian Academic corpus and the *Korpus Südtirol* [corpus of South Tyrol] (Korpus C4).

An additional large resource for language studies these days is the World Wide Web. Internet searches gain importance in phraseology and paremiology as well (Umurova, 2005; Colson, 2007; Ptashnyk & Hallsteinsdóttir & Bubenhofer, 2010), though evidence from this empirical source should not be used without close checking and careful interpretation (Steyer, 2013).

I would like to close this introduction with a caveat – hopefully not too disheartening: Proverb use in the corpus is only a snapshot in time, one of the many facets of communication. All conclusions that you can draw are therefore only representative for this language snapshot. As many corpora are dominated by newspaper and journal texts you learn a lot about how journalists and professional writers use proverbs. This does not mean that the saleslady in the shop round the corner speaks the same way. So you can never say: I did not find this proverb in the corpus therefore it does not exist. Or: This proverb appears this many times, therefore it is one of the most frequent proverbs in my language. This would not be good scientific practice. What you can say is: This proverb appears frequently in my modern language corpus, therefore it cannot be outdated (as long as your examples are real uses of the proverb). If a proverb has variants which are not from a single source, but from different sources and different times, you can also claim that these are typical and not singular variations. So if nothing else, the typical examples of proverb use that can be found in modern corpora are definitely a good reference to decide which proverbs are most useful to learn and to use for second language learners.

9.3 Corpus Linguistic Approaches to Proverb Study

The corpus-empirical approach has only recently started to impact on paremiology. Pioneering work has been done by Ďurčo, among others 2005), Čermák, among others 2006 and by the EU-Project SprichWort (2008-2010, 143376-LLP-1-2008-1-SI-KA2-KA2MP) (Steyer, 2012 a and b; SWP).¹⁴²

Generally, two corpus linguistic approaches to the study of current proverb use can be distinguished:

- Knowing a proverb, because it is codified in a dictionary or as an entry in the mental lexicon of speakers and then searching this proverb in the corpus (corpus based) (3.1).
- Detecting a proverb in a corpus (corpus driven). One prominent approach for this is the statistical collocation analysis of paremiological keywords (e.g. cultural symbols like numbers, colors, animals or body parts) or introductory formulas like *proverb*, *saying*, *slogan* or *says an old wisdom* (3.2).¹⁴³

Regardless of the methodological approach, all contemporary corpus analysis tools provide multiple ways of visualizing the search results. Apart from a list of sources which can be sorted in different ways, two presentation formats are standard today:

- concordance lines (keyword in context: KWIC)
- full text results.

The KWIC format is helpful for recognizing base forms of proverbs, fixedness and variance and pragmatic markers (3.1.1 – 3.1.3). Full text results give information about the meaning of the proverb and its usage context (3.1.4).

9.3.1 Corpus-based Questions About Proverbs

9.3.1.1 Proverb – Yes or No?

According to the proverb definition of Röhrich & Mieder, 1977 the following criteria (amongst others) must be considered when identifying a proverb:

- 1) Do the components appear in the form of a sentence or a non-finite clause – like *Ohne Fleiß kein Preis* [ee: No pain, no gain] – that is equivalent to a sentence?

¹⁴² Another corpus-based project was EPHRAS (EPHRAS).

¹⁴³ I will not elaborate on the method of n-gram analysis for automatically finding frequent word clusters, as it needs some expert knowledge and special tools and is also not yet commonly applied in proverb studies. An overview over applications of this method gives Bürki, 2012.

2) Are they not – or at least not only – used as quotations with a cited source but have become part of everyday language as an expression of wordly wisdom?

3) Do they express generalized experiences or value judgements?

While 3) can only be determined by a qualitative interpretation, corpus analysis gives valuable indications in regard to 1) and 2).

Searching for proverbs in a general language corpus is no trivial matter. There are few or no assumptions about the occurrence and behaviour of a proverb that can be made in advance. Therefore each proverb candidate must be examined individually in an iterative alternation of automatic analysis and the formulation of hypotheses.¹⁴⁴ During this process, it is crucial to have as few preconceived notions about the form of the proverb as possible because again and again corpus evidence proves our intuition wrong. If I search for a fixed sentence in the corpus, I will only find this sentence. All possible variations, extensions and reductions will not be covered by this search. Therefore it is a good strategy to start with a wide search which is then gradually restricted.

The first step is to check whether the lexical components of the proverb candidate appear in the same sentence at all. Consider the proverb candidate *Alte Ochsen machen gerade Furchen* [ww: Old oxes make straight furrows]. In this case *Ochse* [ox], *Furche* [furrow] and *alt* [old] never appear together in the corpus. This is evidence that this proverb is probably outdated.

If the search for the lexical components in the same sentence was successful, this can indicate a proverb. KWIC concordance lines help to quickly check an important proverb criterion (Lüger, 1999): whether the form is that of a sentence or a non-finite clause equivalent to a sentence. For example searching for *Speck* [bacon] and *Maus* [mouse] in the same sentence already gives a clear picture of the proverb *Mit Speck fängt man Mäuse* [ww: With bacon one catches mice; ee: Good bait catches fine fish].

(1)

K00 *Mit Speck fängt man die Mäuse – und mit Dollars Leichtathleten.*
 RHZ06 *Mit Speck fängt man Mäuse, aber keine Stimmen.*
 RHZ00 *Mit Speck fängt man Mäuse, weiß der Volksmund.*
 RHZ06 *Mit Speck fängt man Mäuse und mit Käse die Narren*
 DPA09 *Mit Speck fängt man Mäuse, aber keine Wähler.*
 Z06 *Mit Speck fängt man Mäuse – und macht man auch Mäuse*

144 The following strategies for validating proverbs in a corpus were developed from experiences in the EU project „SprichWort“ (SWP) where for the first time a comprehensive corpus validation was conducted for 2000 German proverbs (Steyer, 2012 b). All examples for searches and their result numbers are based on DeReKo. The exact search queries are documented in Steyer, 2013; Umurova, 2005; Ďurčo, 2006; Hrisztova-Gotthardt, 2010 and Hrisztova-Gotthardt & Gotthardt, 2012 use specialized search strategies as well. A comprehensive corpus -based study of English proverbs from a diachronic perspective was conducted by Aurich, 2009. also Charteris-Black, 1999.

NUN91 *Mit Speck fängt man Mäuse, mit Kultur gewinnt man Kunden.*

For the proverb candidate *Niemand ist ohne Fehl und Tadel* [ww: Nobody is without faults and blames; ee: Nobody is perfect] however, searching for the components *Fehl* [fault] and *Tadel* [blame] together gives a high number of results, but the complete proverb sentence was rarely found. In this case, only the propositional phrase *ohne Fehl und Tadel* [ww: without fault and blame] is fixed but the contexts vary:

(2)

F95 *Wer ohne Fehl und Tadel ist, der werfe den ersten Stein.*
 N92 *löste diese Aufgabe ohne Fehl und Tadel.* Mehr noch:
 O94 *die Musik ist ohne Fehl und Tadel;*
 M98 *selbst Heilige sind nicht frei von Fehl und Tadel.*

For other proverb candidates the wide search must be restricted gradually to capture the real instances of proverb use. An example is the proverb *Zeit ist Geld* [ee: Time is money]. Even if you allow only one word between the components *Zeit* [time] and *Geld* [money], you get a lot of hits that have nothing to do with the proverb, but capture the binomial pair *Zeit und Geld* [time and money] – the same happens, by the way, when searching for *time* and *money* in the English corpus BNC. Only when integrating the verb form *ist* [is] into the search, you will find the proverb. The following KWIC examples are from the BNC:

(3)

A3C 74 *Time is money* and, after money, time is what the masses
 crave most.
 ABK 149 *TIME is money.*
 ANY 1167 That costs time, and *time is money.*
 ASF 1442 It was soon realized by many of the middle class that
 ‘*time is money*’ and consequently must be carefully regulated
 and used economically.

For the proverb candidate *Viel Lärm um nichts* [ww: Much noise about nothing, ee: Much ado about nothing] the search must be heavily restricted, with many components and a close focus. Searching for *Lärm* [noise] and *nichts* [nothing] in the same sentence in DeReKo gives more than 3000 results, however these include many instances that have nothing to do with the proverb, e.g. *Die Fahrer hören nichts vom Lärm* [The drivers hear **nothing** of the **noise**]. After including *viel* [much] and *um* [about] into the search, about 77% of the hits capture the sentence *Viel Lärm um nichts*. However, these results must be further examined, as many of them are citations, referring to the comedy of William Shakespeare. We are interested in how the sentence is used as proverb therefore we try to exclude as many words as possible that

indicate a Shakespeare context in any form. The final, very complex search query is: Search for *viel* and *Lärm* and *nichts* in the same sentence, but the sentence must not include *Shakespeare* or *Komödie* [comedy] or *Uhr* [clock]¹⁴⁵ or *Kino* [movie theatre] or *Film* [movie] or *Regie* [stage direction] or *Branagh*¹⁴⁶ or *Branaghs* or *Schauspieler* [actor] or *Hollywood* or *Hollywoods* or any compound words with *Theater* [theatre]. This search yields still over 1500 hits for *Viel Lärm um nichts* and you can now assume that these reflect its usage as a real proverb.

This double life as a quotation and as a proverb is very frequent phenomenon in the corpus, as you can see from the fact that references to the real or supposed origin of a proverb are very common. For example, the biblical or Latin roots of the proverb are mentioned or the person who is credited with its creation. These markers should not be treated as true or false statements about authorship, but as indicators for origin contexts which are still present in the minds of the speakers. For example, in the context of the proverb candidate *Zeit ist Geld* [ee: Time is money] you often find references to Benjamin Franklin:

(4)

Handel und Wandel nahmen neue Formen an, sodass der amerikanische Staatsmann Benjamin Franklin den stehenden Begriff «Zeit ist Geld» prägte. Ein Geist der Unrast begann um sich zu greifen. (St. Galler Tagblatt, 28.10.1999)

There is a transition zone for proverb candidates which are already frequent in the corpus, but cannot (yet) fulfil the proverb criterion that they are applicable in many contexts of communication. For example, the proverb candidate *Es gibt kein schlechtes Wetter, nur schlechte Kleidung* [ww: There's no bad weather, but only bad clothes] is only used in weather contexts and only with the meaning that any weather can be endured wearing the right clothes and that you should therefore not cancel your outdoor activities because of the weather.

(5)

Frei nach dem schlaunen Spruch „Es gibt kein schlechtes Wetter, nur schlechte Kleidung“ halten sich die Jungen und Mädchen in der Betreuungszeit ausschließlich im Wald auf, auch bei Regen und Schnee. Da leistet die mitwachsende Kinderlederhose gute Dienste (Rhein-Zeitung, 29.10.2008)

¹⁴⁵ The appearance of *Uhr* indicates a time table, as in (movie) theatre programs.

¹⁴⁶ Kenneth Branagh is the director of one of the most successful movie adaptations of the Shakespeare play from the year 1993 (*Much Ado About Nothing*).

However, it is easy to imagine this proverb in a sense: “If you prepare appropriately, you can master or enjoy any situation”. The following usage example from the corpus already hints at an evolution towards such a status as a saying of worldly wisdom:

(6)

Wenn man glaubt, das Glück geschenkt zu bekommen, liegt man falsch. Wenn mich einer fragt: Wie war im Urlaub das Wetter?, antworte ich stets: *Es gibt kein schlechtes Wetter, nur schlechte Kleidung*. So ist es nämlich auch mit dem Glücklichsein. (die tageszeitung, 01.07.1997, S. 22)

An important feature – e.g. in CII – is the possibility to sort the search results chronologically. Such lists give indications whether a proverb is more likely archaic or newly coined. You can also examine aspects like the earliest occurrence or a decrease in use frequency. For some proverbs, corpus analysis even allows the reconstruction of the evolution from an original quotation to a proverb, for example for the German proverb *Wer zu spät kommt, den bestraft das Leben* [ww: He who comes late is punished by life]. This proverb is a transformation of the original sentence from Mikhail Gorbachev *Gefahren warten nur auf jene, die nicht auf das Leben reagieren* [ww: Dangers await only those who do not respond to life] during a state visit to the former East Germany in October 1989, shortly before the fall of the Berlin Wall. (Mieder, 2004; Mieder, 2010; Steyer, in print). Figure 9.1 shows the chronological list of hits (899)¹⁴⁷.

You can see, that this proverb is documented for the first time in DeReKo indeed in 1989. More than 30 years later, you find many examples for the use as a true proverb in the corpus, like the following:

„Limburgerhof als Stadtteil von Ludwigshafen? Die Stadt hätte sicherlich nichts dagegen. Die Frage ist nur: Ist dies das strategische Interesse von Limburgerhof? Ein Sprichwort sagt: *Wer zu spät kommt, den bestraft das Leben*. Für Limburgerhof heißt das: Entweder jetzt die Chance mit Neuhofen zu ergreifen oder aber später zum Spielball fremder Interessen zu werden“ (Die Rheinpfalz, 20.08.2012, S. 23).

And many young people do not know the context anymore, as a small survey amongst students showed me. The sentence is truly established as a proverb.

¹⁴⁷ Query: &wer /s0 (&kommen oder &spätkommen) /s0 &Leben /s0 &bestrafen.

Hits	Texts	Year
11	10	1989
36	34	1990
11	10	1991
18	18	1992
20	18	1993
9	8	1994
12	12	1995
32	32	1996
54	54	1997
52	51	1998
94	89	1999
55	53	2000
39	39	2001
38	36	2002
39	37	2003
43	42	2004
32	32	2005
44	42	2006
35	33	2007
39	38	2008
63	60	2009
36	34	2010
61	52	2011
26	26	2012
899 all in all		

Figure 9.1: Chronological list of *Wer zu spät kommt, den bestraft das Leben*

9.3.1.2 Fixedness and Variance

One of the central questions in proverb studies and proverb lexicography is about fixedness and variance of proverbs. On the one hand there are variations which happen at the syntactic and morphological level, as in this example of the proverb *Der Fisch stinkt vom Kopf* [ee: The fish always stinks from the head downwards]:

(7)

- T05 *Der Fisch stinkt vom Kopf her*, sagt der Volksmund.
 NUZ09 Chinesen haben ein unappetitliches, aber wahres Sprichwort: *Der Fisch stinkt vom Kopfe her*.
 NUN05 »Am *Kopf* fängt *der Fisch* das Stinken an« ist eine alte Weisheit.

On the other hand, there is the substitution of certain components with other lexical elements. These variations can be found by using restrictive search queries as explained above. For example to find variations of the proverb *Übung macht den Meister* [ww: Practice makes the master] you can search for *Übung macht den/die/*

das [Practice makes the] WITHOUT *Meister* [master] or *macht den Meister* [makes the master] WITHOUT *Übung* [practice]:

(8)

Übung macht den X-Meister [ww: Practice makes the X-master]

M99 *Übung macht den Handball-Meister*

M99 *Übung macht auch hier den Zaubermeister*

These are variants which use compounds with *Meister*, e.g.: *Practice makes the master of handball/magic.*

(9)

Übung macht den X [ww: Practice makes the X]

A97 *Übung macht den Radioprediger*

A00 *Übung macht den Schützen*

A09 *Übung macht den Feuerwehrmann*

F99 *Übung macht den Gourmet*

In these variants *Meister* is substituted, e.g.; *Practice makes the radio preacher/marksman/ fire fighter/gourmet.*

(10)

X macht den Meister [ww: X makes the master]

B06 *Technik macht den Meister*

B07 *Energie macht den Meister*

B07 *Vorsicht macht den Meister*

E99 *Doch erst Playoff macht den Meister*

Here, *Übung* [practice] is substituted, e.g.; *Technique/Energy/Caution/Playoff makes the master.*

It is also possible to find variations with more abstract search queries,¹⁴⁸ e.g. *Wer A sagt, muss auch B sagen* [ee: Who says A must say B].¹⁴⁹

¹⁴⁸ So-called proverb construction plans or schemas are already mentioned in Röhrich & Mieder, 1977. See also Burger, 2010.

¹⁴⁹ In this case, the exact search query in COSMAS II syntax would be: (\$wer /+w2:2 &sagen /+w1:1 &müssen /+w5 &sagen) %s0 (\$a und \$b).

(11)

R99 *Wer „Puppe“ [doll] sagt, muss auch „Barbie“ sagen*
 RHZ06 *Wer Argentinien [Argentina] sagt, muss auch Tango sagen*
 T10 *Wer Beatles sagt, muss aber auch Rolling Stones sagen*

In this case, the proverb variants play with the abstract meaning ‘one thing necessarily calls for another’ of the pattern *Who says X must say Y*, e.g.: *doll* → *Barbie*; *Argentina* → *Tango*; *Beatles* → *Rolling Stones*.

9.3.1.3 Proverb Frequency

Calculating proverb frequency is a complex problem which has no standard solution. There will be different results depending on the corpus and the search query that was used. Consider the proverb *Not macht erfinderisch* [ww: Necessity makes ingenious; ee: Necessity is the mother of invention]. If you use a restrictive search query like “*Not* [necessity] directly followed by *macht* [makes] directly followed by *erfinderisch* [ingenious]” you get more than 900 hits for the proverb in exactly this form. If you widen the search and allow all morphological forms of the verb *machen* [make], the number of hits already increases by approx 40. If in addition to that you allow wider gaps between the components so that you also capture occurrences with an introductory formula (*Not macht bekanntlich/sagt man* [ww: as is generally known/as they say] *erfinderisch*) or minimal lexical additions (e.g. particles like *Not macht eben immer/nun mal/schließlich* [ww: always/after all/as you know] *erfinderisch*) the number of hits for this example increases by approx. 600!

As you can see, there can be no absolute proverb frequencies (this is also true for multi-word expressions in general). Statements about frequency are only meaningful if you make transparent on which corpus basis and with which search queries the numbers have been obtained (Appendix 2). It is also recommended to refer to proportional frequencies or frequency trends rather than absolute numbers.

9.3.1.4 Meaning and Usage¹⁵⁰

Corpus analysis gives us an empirically sound way to find out what aspects are always parts of the meaning of a proverb and what aspects are typically connected to the use of the proverb and form its wider context. Pragmatic aspects of meaning in particular are often perceived from a very subjective perspective. Corpus analysis helps to get a more objective view as it identifies usage properties which repeat themselves and are therefore typical. Thus, meaning can often be captured in a more accurate and

¹⁵⁰ In addition to Steyer, 2013, this section is based mainly on Steyer & Hein, 2010; Steyer, 2012b; Hein, 2012; Steyer & Ďurčo, in print.

nanced way than by pure introspection. For example, the proverb *Die Ratten verlassen das sinkende Schiff* [ww: The rats are leaving the sinking ship] on first glance refers to someone who notices a danger in time and flees.¹⁵¹ However, corpus evidence clearly points to the following meaning-in-usage (cited from the article on the SprichWort platform, Steyer & Hein, 2010):

(12)

Bedeutung / Besonderheiten im Gebrauch [meanig / current usage]

„Sagt man, wenn sich Verantwortliche bei von ihnen mit verschuldeten negativen Entwicklungen oder Schwierigkeiten zurückziehen und sich damit aus der Verantwortung stehlen“ (Steyer & Hein, 2010)

[English translation of this paraphrase: “Is used when someone draws back from negative consequences or problems he has caused and by doing so shirks responsibility.”]

Korpusbeleg [corpus citation]:

Die Ratten haben *das sinkende Schiff* verlassen. Der irakische Oppositionspolitiker Hassan Halboos ist fest davon überzeugt, dass Tausende von Helfern Saddam Husseins ihre Uniformen ausgezogen und sich unters Volk gemischt haben. (RHZ03/APR.09770 Rhein-Zeitung, 12.04.2003; “Saddams Helfer tarnen sich”)

Sometimes corpus evidence shows that there is a pragmatic aspect that is always tied to the use of a proverb, such as a negative connotation *Die dümsten Bauern haben die dicksten Kartoffeln* [ww: The dumbest peasants have the biggest potatoes; ee: Fortune favors fools]. In most cases pragmatic aspects like ‘devaluation’ are not present for all, but still for many of the usage instances from the corpus. The following usage aspects can typically be found with the help of corpus analysis:

- Context of usage; e.g. typically in sport reports: *Man ist so alt, wie man sich fühlt* [ww: You are as old as you feel]; *Rache ist süß* [ww: Vengeance is sweet]; *Knapp daneben ist auch vorbei* [ww: Narrowly missed is still missed];
- Function in communication, e.g. typically as ‘admonition’: *Erst denken, dann handeln* [ww: First thinking, then acting]; *Hochmut kommt vor dem Fall* [ee: Pride goes before a fall]; *Die Konkurrenz schläft nicht* [ww: The competition does not sleep];

¹⁵¹ The basis for this proverb is an old belief among sailors the rats rather flee towards the open sea than drown with the ship (Duden, 11, 2013: 593).

- Text type, e.g. typically in horoscopes: *Der Klügere gibt nach* [ee: The cleverer give in]; *Eile mit Weile* [ee: Haste makes waste]; *Wer rastet, der rostet* [ee: You snooze, you loose].

Examining corpus evidence also gives a good idea of the textual integration of proverbs. For example, by doing an alphabetical sort of the KWIC lines, you can easily see frequent elements in front of the proverb. In the case of the proverb *Der Schein trügt* these are conjunctions like *aber*, *denn* or *doch* [ww: **But/Because/However** appearances are deceitful; ee: **But/Because/However** you can't tell a book by its cover]:

(13)

WPD11 **Aber** *der Schein trügt*: In einer Bar wandern
 RHZ11 **Aber** *der Schein trügt*. Auch diese Felder
 HMP09 **Aber** *der Schein trügt*: Während der 27-Jährige sorglos
 seine PIN eingibt,
 F01 **Denn** *der Schein trügt* nie in Maastricht:
 F95 **Denn** *der Schein trügt*, lernen wir,
 BRZ10 **Denn** *der Schein trügt*: Trotz der Minusgrade
 T87 **Doch** *der Schein trügt*: Unser Backwerk entsteht aus dem
 Mehl nur weniger
 T86 **Doch** *der Schein trügt*

Other typical textual context elements can be: frequent modification with adverbs or particles like *Der Ton macht **eben/halt/nun mal** die Musik* [ww: The sound makes PARTICLE the music (the particles mean roughly 'after all'); ee: It's not what you say, but how you say it]; frequent sentence mode, e.g. question: *Ende gut, alles gut?* [ee: All's well that ends well?], frequent negation: *Die Zeit heilt **nicht** alle Wunden* [ww: Time does **not** heal all wounds]; frequent use as an opening phrase for a topic: *Der Schein trügt*: [ww: *Appearances are deceitful*: ee: **But** You can't tell a book by its cover:] frequent use as a parenthesis: *Aber andererseits, **sicher ist sicher**, lud er zum Empfang [...] ein* [ww: But on the other hand, **sure is sure**; ee: just to be sure, he invited to the reception]; frequent reduction to only a part of the proverb: *Reden ist Silber* [ONLY: Speech is silver] or *Schweigen ist Gold* [ONLY: Silence is golden], frequent syntactical transformation: *süße Rache* [sweet revenge from Revenge is sweet].

9.3.2 Proverbs – Corpus Driven

The second corpus analytical approach is the automatic extraction of multi-word units – including proverbs – from the corpus. A useful method that is available in many corpus analysis tools is collocation analysis which calculates the significant partner words of a target word. A sophisticated implementation of this method can be used

via COSMAS II (Belica, 1995).¹⁵² Collocation analysis allows us to find common multi-word units that cluster around proverb key words like *Geld* [money]; *Welt* [world]; *Liebe* [love] or *Mensch* [human], but also around introductory formulas.¹⁵³ Figure 2 shows a clipping from the collocation profile of the introductory formula *bekanntlich* [as is generally known]:¹⁵⁴

collocation	syntagmatic pattern
erfinderisch	Not macht bekanntlich [...] erfinderisch
Totgesagte	Totgesagte [leben] bekanntlich länger
Magen	Liebe geht bekanntlich durch den Magen geht
Kleinvieh	Kleinvieh [macht] bekanntlich auch Mist
Hochmut	Hochmut [kommt] bekanntlich vor dem
mahlen	der Mühlen mahlen [...] bekanntlich langsam
heiligt	Der der Zweck heiligt [...] bekanntlich die Mittel
belebt	Konkurrenz belebt [...] bekanntlich das Geschäft
geduldig	Papier ist bekanntlich [...] geduldig
Derbys	Derbys [haben] bekanntlich ihre eigenen
stinkt	Der Geld Fisch stinkt [...] bekanntlich nicht vom
Tüchtige	Glück hat bekanntlich nur der Tüchtige

Figure 9.2: Clipping from the collocation profile of *bekanntlich*

You can see that the partner words point towards several common proverbs.

In addition to that, collocaton analysis gives valuable hints in regard to typical usage situations as attested by many context-based approaches (Sinclair, 2004; Stubbs, 2001 and many other). The clippings in Figure 9.3 are from the collocation profile of *Die Ratten verlassen das sinkende Schiff* [ww: The rats are leaving the sinking ship], (section 3.1.4).

¹⁵² The following examples have been calculated with the collocation analysis tool developed by Belica, which is available via COSMAS II. In contrast to other tools, this tool also outputs common syntagmatic patterns. Other tools for collocation analysis or collocation resources are e.g. available at DWDS. Another tool for various languages is Sketch Engine (SkE; Ďurčo, 2010).

¹⁵³ About the corpus-based analysis of proverb markers or introducers Čermák, 2004; Ďurčo, 2005.

¹⁵⁴ The clippings from the collocation profiles are cited without quantitative information, as they just serve as an illustration of the general principle. Documenting and explaining the parameters and numbers would be beyond the scope of this contribution. Further information can be found in the collocation analysis tutorial (Perkuhn & Belica, 2004).

collocation	syntagmatic pattern
1)	
Spruchwort	Spruchwort [Die] Ratten verlassen das
Devise	Devise ... Ratten
bekanntlich	Ratten ... bekanntlich
Redewendung	die Redewendung daß die ... Ratten das ...
Motto	nach dem Motto Die ... Ratten verlassen das
2)	
Exodus	Ratten ... das verlassen sinkende Schiff ... die Kabila-Regierung ... Exodus
Titanic	Ratten ... Titanic
Untergang	verlassen ... Untergang
Panik	Panik ... verlassen
fliehen	fliehen ... Ratten
Kapitän	Ratten ... verlassen das sinkende Schiff die der ... Kapitän
Seeleute	Ratten ... Seeleute
Bord	Ratten verlassen das sinkende Schiff ... der ... Bord
Ufer	Ratten ... Ufer
Unglück	Ratte ... Unglück
Flucht	Ratten das sinkende verlassen Schiff ... Flucht
Welle	Welle ... Ratten
Chaos	Chaos ... Es ist die ... Ratten
Brand	verlassen ... Brand
3)	
Multis	Multis ... Multis ... Ratten
Diktator	Ratte ... Diktator
Regierungen	Ratten verlassen das sinkende Schiff ... Regierungen
Geschäftsführerin	Geschäftsführerin ... Ratte
Milosevic	Milosevic ... verlassen
Rücktritt	Rücktritt ... Ratte

Figure 9.3: Clippings from the collocation profile of *Ratten–verlassen–Schiff*

Significant partner words in the wider context of this proverb indicate

- 1) that this sentence is commonly used as a saying of worldly wisdom (e.g. *proverb/maxim/as is generally known/saying/motto*),
- 2) that the nautical origin metaphor is often referenced, (e.g. *Titanic/sinking/panic/captain/escape/to disembark* etc.),
- 3) that the behaviour described by the proverb is typically attributed to persons or social groups that hold leadership positions, (e.g. *dictator/governments/manager/Milosevic* etc.).

Finally, collocation analysis can help to analyse proverb patterns and schemas. Figure 9.4 shows a clipping from the collocation profile of the proverb pattern *Wer X, der Y* [ww: He who X Y], an even more abstract pattern of *Who says X must say Y* (discussed in section 3.1.2).

collocation	syntagmatic pattern
rastet	Wer [...] rastet der rostet
rostet	Wer [rastet der] rostet
suchet	Wer suchet der findet
findet	Wer [suchet sucht der] findet
bestellt	Wer [...] bestellt der bezahlt
anschafft	Wer anschafft der zahlt
dopt	Wer dopt der fliegt
sündigt	Wer schläft der sündigt nicht -
schreibt bleibt	Wer schreibt der bleibt
schlägt	Wer [...] schlägt der geht
lügt	Wer [schreit der] lügt der ...
gewinnt wagt	Wer wagt der gewinnt
austeilt	Wer austeilt der muss muß
erntet	Wer sät der erntet
aufmuckt	Wer aufmuckt der ...
schläft	Wer [...] schläft der sündigt

Figure 9.4: Clipping from the collocation profile of the pattern *Wer* GAP OF ONE WORD, *der*

Typically realisations of the pattern *Wer X, der Y* are: *Wer rastet, der rostet* [ww: Who rests rusts; ee: You snooze, you lose]; *Wer wagt, gewinnt* [ee: Who dares wins].

Collocation analysis not only generates lists of significant partner words and syntagmatic patterns as shown in the clippings above, but also sorts KWIC concordance lines and full text results according to their context patterns. These should always be included in the interpretation. Below you can study some KWIC lines (12) and full text results (13) for the proverb *Wer austeilt, der muss auch einstecken können* [ww: He who dishes it out, must also be able to take it].

(14)

K97 Nein, mich kann man nicht beleidigen. *Wer austeilt, der muß auch einstecken können.*

PNI haben wir uns schöne Auseinandersetzungen, auch verbal, geliefert. *Wer austeilt, der muss auch einstecken können, okay.*

PRP Man muß auch nicht zimperlich sein. *Wer austeilt, der muß auch einstecken können.* Herr Kollege Dr. Mertes,

(15)

Nein, mich kann man nicht beleidigen. *Wer austeilt, der muß auch einstecken können.*

Was hätte Matthias Schuh werden sollen, wenn es nach seinen Eltern gegangen wäre? (Kleine Zeitung, 17.05.1997, Ressort: Menschen; „Mich kann doch niemand beleidigen!“)

In dem Bereich, den ich früher hier vertreten durfte, haben wir uns schöne Auseinandersetzungen, auch verbal, geliefert.

Wer austeilt, der muss auch einstecken können, okay. Aber Sie werden bei allen Auseinandersetzungen verbaler Art mit der früheren niedersächsischen Umweltministerin oder auch mit dem jetzigen Umweltminister von mir nie Ausdrücke wie „mangelnder Sachverstand“, „keine Ahnung“ oder so etwas gehört haben. (Protokoll der Sitzung des Parlaments Landtag Niedersachsen am 30.07.1999. 30. Sitzung der 14. Wahlperiode 1998-2003. Plenarprotokoll, Hannover, 1999 [S. 2734])

Man muß auch nicht zimperlich sein. *Wer austeilt, der muß auch einstecken können.*

Herr Kollege Dr. Mertes, eines geht nicht. (PRP/W13.00035 Protokoll der Sitzung des Parlaments Landtag Rheinland-Pfalz am 17.07.1997. 35. Sitzung der 13. Wahlperiode 1996-2001. Plenarprotokoll, Mainz am Rhein, 1997)

It must again be stressed that collocation analysis only gives pointers and can never replaces human interpretation.

9.4 Summary and Outlook

This chapter gave a first overview of how proverb research can benefit from corpus linguistic approaches. My main goal was to illustrate the general principles. Corpus linguistics revolutionizes especially application-related areas like *proverb lexicography and didactics* which are already focussed on language usage. An example for a strictly corpus-based approach is the above-mentioned EU-project SprichWort. Future research should aim towards comparing the results of corpus linguistic methods with those of proverb surveys and of the traditional historical paremiology. A vision would be the creation of new proverb collections or modern proverb information systems which combine all this knowledge in a meaningful way.

However, corpus linguistic proverb research also opens new perspectives in regard to linguistic theory, e.g. a connection with construction grammar and with a new, pattern-based phraseology. As paremiologists who use corpus linguistic methods generally work with a large data basis, they view regularities in many similar cases of usage. A new quality of research is not only achieved by using more data, but also by discovering structures which have been hidden from human perception before. There is a chance to find unusual cross-connections and unexpected relations. Because of this, you can now study proverbs as part of the larger picture of language and vocabulary.

Empirical corpus research once again raises the general issue of the special status of proverbs on the one hand and their role in the network of multi-word units on

the other hand: Corpus empirical evidence shows that speakers consciously recall fixed phrases and use them for their special communicative function of condensing complex messages. At the same time, the limitless number and variance of introductory formulas in the corpus is evidence that speakers do not make a fine distinction between *proverb*, *saying*, *slogan* or *motto*, but that only the concept *sentence as message* is important.

The corpus linguistic perspective also shows that proverbs themselves can be realisations of more general patterns and schemas (example *Wer X der Y*) and share attributes and characteristics with non-proverb multi-word units. Pilot studies suggest that proverbs often represent the prototypical realisations of those schemas and, as lexically fixed expressions, are more salient in the mind of the speakers, while non-proverb units of the same schema tend to be subject to creative ad-hoc variations. This raises the question why some proverbs have hardly any variants while others have many. As you can see, strictly corpus-based proverb studies can create a fresh impetus for a pattern-based theory of the lexicon.

References

- Aurich, C. (2009). *Proverb Structure in the History of English: Stability and Change: A Corpus-Based Study*. (Phraseologie und Parömiologie 26). Baltmannsweiler: Schneider Hohengehren.
- Belica, C. (1995). *Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analyse-methode*. Mannheim: Institut für Deutsche Sprache.
<http://www1.ids-mannheim.de/kl/projekte/methoden/ur.html> (accessed November 12, 2013).
- BNC: British National Corpus. <http://www.natcorp.ox.ac.uk/> (accessed November 12, 2013).
- Bubenhofner, N. (2006-2013). *Einführung in die Korpuslinguistik: Praktische Grundlagen und Werkzeuge. Elektronische Ressource*. <http://www.bubenhofner.com/korpuslinguistik/>. (accessed November 12, 2013).
- Burger, H. (2010). *Phraseologie: Eine Einführung am Beispiel des Deutschen*. 4., neu bearb. Aufl. Berlin: Schmidt.
- Bürki, A. (2012). Korpusgeleitete Extraktion von Mehrwortsequenzen aus (diachronen) Korpora: Vorgehenswege für deutschsprachige Daten. In N. Filatkina, A. Kleine-Engel, M. Dräger & H. Burger (Eds.), *Aspekte der historischen Phraseologie und Phraseographie* (pp. 263-292). Heidelberg: Winter.
- Čermák, F. (2004). Text Introducers of Proverbs and Other Idioms. In C. Földes & J. Wirrer (Eds.), *Phraseologismen als Gegenstand sprach- und kulturwissenschaftlicher Forschung* (Phraseologie und Parömiologie 15) (pp. 27-46). Baltmannsweiler: Schneider Hohengehren.
- Čermák, F. (2006). Statistical Methods for Searching Idioms in Text Corpora. In: A. Häcki Buhofer, & H. Burger (Eds.), *Phraseology in Motion I. Methoden und Kritik* (Phraseologie und Parömiologie 19) (pp. 33-42). Baltmannsweiler: Schneider Hohengehren.
- Charteris-Black, J. (1999). The Survival of English Proverbs: A Corpus Based Account. *DeProverbio*, 5(2). <http://www.deproverbio.com/display.php?a=3&r=96> (accessed November 11, 2013).
- CL: COSMAS II: Korpusrecherche- und analysesystem. Internet:
<http://www.ids-mannheim.de/cosmas2/> (accessed November 12, 2013).

- Colson, J. P. (2007). The World Wide Web as a corpus for set phrases. In H. Burger, D. Dobrovol'skij, P. Kühn & N. R. Norrick (Eds.), *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (HSK 28.2) (pp. 1071-1077). Berlin/New York: de Gruyter.
- Cowie, A. P. (1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press.
- DEREKO (2013). Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2013-I (Release vom 19.03.2013). Mannheim: Institut für Deutsche Sprache.
<http://www1.ids-mannheim.de/kl/projekte/korpora.html> (accessed November 12, 2013).
- Dict.cc: Dict.cc. Deutsch-Englisches Wörterbuch. <http://www.dict.cc/> (accessed November 12, 2013).
- Duden 11 (2013). *Duden Redewendungen. Wörterbuch der deutschen Idiomatik*. 4., neu bearb. und aktualisierte Aufl. Herausgegeben von der Dudenredaktion. Red. Bearb. W. Scholze-Stubenrecht & A. Haller-Wolf. Berlin/Mannheim/Zürich: Dudenverlag.
- Đurčo, P. (2005). *Sprichwörter in der Gegenwartssprache*. Trnava: Univerzita Sv. Cyrila a Metoda. Filozofická fakulta.
- Đurčo, P. (2006). Methoden der Sprichwortanalysen oder Auf dem Weg zum Sprichwörter-Optimum. In A. Häcki Buhofer & H. Burger (Eds.), *Phraseology in Motion I. Methoden und Kritik. Akten der Internationalen Tagung zur Phraseologie (Basel, 2004)*. (Phraseologie und Parömiologie 19) (pp. 3-20). Baltmannsweiler: Schneider Hohengehren.
- Đurčo, P. (2010). Einsatz von Sketch Engine im Korpus –Vorteile und Mängel. In S. Ptashnyk, E. Hallsteinsdóttir, & N. Bubenhofer (Eds.) (2010). *Korpora, Web und Datenbanken. Computer-gestützte Methoden in der modernen Phraseologie und Lexikographie* (Phraseologie und Parömiologie 25) (pp. 119-131). Baltmannsweiler: Schneider Hohengehren.
- DWDS-Corpus: Korpora des Digitalen Wörterbuchs der deutschen Sprache.
<http://www.dwds.de/ressourcen/korpora/> (accessed November 12, 2013).
- EPHRAS: EPHRAS. *Ein mehrsprachiges phraseologisches Lernmaterial auf CD-ROM*.
<http://www.ephras.org/> (accessed November 12, 2013).
- Granger, S. & Meunier, F. (Eds.) (2008). *Phraseology: An interdisciplinary Perspective*. Amsterdam; Philadelphia: John Benjamins.
- Grzybek, P. (2012). Facetten des parömiologischen Rubik-Würfels. Kenntnis \equiv Bekanntheit [\Leftrightarrow Verwendung \approx Frequenz] ?!?. In K. Steyer (Ed.), *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie* (Studien zur Deutschen Sprache 60) (pp. 99-138). Tübingen: Narr.
- Heid, U. (2007). Computational linguistic aspects of phraseology II. In H. Burger, D. Dobrovol'skij, P. Kühn & N. R. Norrick (Eds.), *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (HSK 28.2) (pp. 1036-1044). Berlin/New York: de Gruyter.
- Hein, K. (2012). Zugang zu Sprichwortbedeutung und –gebrauch mit Hilfe von Korpora. In K. Steyer (Ed.), *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie* (Studien zur Deutschen Sprache 60) (pp. 315-340). Tübingen: Gunter Narr.
- Hrisztova-Gotthardt, H. (2010). *Vom gedruckten Sprichwörterbuch zur interaktiven Sprichwortdatenbank: Überlegungen zum linguistischen und lexikographischen Konzept mehrsprachiger Sprichwortdatenbanken* (Sprichwörterforschung 26). Bern: Peter Lang.
- Hrisztova-Gotthardt, H. & Gotthardt, Z. (2012). Sprichwörter im bulgarischen nationalen Korpus. *Proverbium* 29, 25-36.
- Juska-Bacher, B. (2012). Empirische Methoden in der kontrastiven Sprichwortforschung: Möglichkeiten der Informantenbefragung. In K. Steyer (Ed.), *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie* (Studien zur Deutschen Sprache 60) (pp. 139-164). Tübingen: Gunter Narr.
- Korpus C4: Gemeinschaftskorpus des DWDS, des AAC, des Korpus Südtirol und des CHTK.
<http://www.korpus-c4.org/> (accessed November 12, 2013).

- Lemnitzer, L. & Zinsmeister, H. (2010). *Korpuslinguistik: Eine Einführung*. 2. durchges. und aktualisierte Aufl. (Narr Studienbücher). Tübingen: Gunter Narr.
- Lüdeling, A. & Kytö, M. (2008). *Corpus linguistics: An International Handbook* (Vol 1). (HSK 29.1). Berlin/New York: de Gruyter.
- Lüger, H. H. (1999). *Satzwertige Phraseologismen: Eine pragmlinguistische Untersuchung*. Wien: Praesens.
- McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice. (Cambridge textbooks in linguistics)*. Cambridge et al.: Cambridge University Press.
- McEnery, T. & Wilson, A. (2001). *Corpus linguistics*. (2nd edition). (Edinburgh Textbooks in Empirical Linguistics) Edinburgh: Edinburgh University Press.
- Mieder, W. (2004). „Andere Zeiten, andere Lehren“ – Sprach- und kulturgeschichtliche Betrachtungen zum Sprichwort. In K. Steyer (Ed.), *Wortverbindungen – mehr oder weniger fest* (Jahrbuch des Instituts für Deutsche Sprache 2003) (pp. 415-438). Berlin/New York: de Gruyter.
- Mieder, W. (2009). *International Bibliography of Paremiology and Phraseology* (2 Volumes). Berlin/New York: de Gruyter.
- Mieder, W. (2010). „Spruchschlösser (ab)bauen“: Sprichwörter, Antisprichwörter und Lehnsprichwörter in Literatur und Medien. Wien: Praesens.
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Moon, R. (2007). Corpus linguistic approaches with English Corpora. In H. Burger, D. Dobrovol'skij, P. Kühn & N. R. Norrick (Eds.), *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (HSK 28.2) (pp. 1045-1059). Berlin/New York: de Gruyter.
- Mukherjee, J. (2009). *Anglistische Korpuslinguistik: Eine Einführung*. (Grundlagen der Anglistik und Amerikanistik 33). Berlin: Schmidt.
- Perkuhn, R. & Belica, C. (2004). *Eine kurze Einführung in die Kookkurrenzanalyse und syntagmatische Muster*. Mannheim: Institut für Deutsche Sprache.
<http://www.ids-mannheim.de/kl/misc/tutorial.html> (accessed November 12, 2013).
- Perkuhn, R., Keibel, H. & Kupietz, M. (2012). *Korpuslinguistik*. (UTB 3433). Paderborn: Fink.
- Ptashnyk, S., Hallsteinsdóttir, E. & Bubenhofer, N. (Eds.) (2010). *Korpora, Web und Datenbanken: Computergestützte Methoden in der modernen Phraseologie und Lexikographie* (Phraseologie und Parömiologie 25). Baltmannsweiler: Schneider Hohengehren.
- Röhrich, L. & Mieder, W. (1977). *Sprichwort*. Stuttgart: Metzler.
- Rothkegel, A. (2007). Computerlinguistische Aspekte der Phraseme I. In H. Burger, D. Dobrovol'skij, P. Kühn & N. R. Norrick (Eds.), *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (28.2) (pp. 1027-1035). Berlin/New York: de Gruyter.
- Sailer, M. (2007). Corpus linguistic approaches with German corpora. In H. Burger, D. Dobrovol'skij, P. Kühn & N. R. Norrick (Eds.), *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (HSK 28.2) (pp. 1060-1071). Berlin/New York: de Gruyter.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.
- SkE: Sketch Engine: Corpus Query System. <http://www.sketchengine.co.uk/> (accessed November 12, 2013).
- Speake, Jennifer (Ed.) (2008). *OXFORD Dictionary of Proverbs*. Fifth Edition. Oxford: Oxford University Press.
- Steyer, K. & Āurčo, P. (2013). Ein korpusbasiertes Beschreibungsmodell für die elektronische Sprichwortlexikografie. In J.-M. Benayoun, N. Kübler & J.-P. Zouogbo (Eds.), *Parémiologie. Proverbes et formes voisines. Volume 3* (pp. 219-250). Sainte Gemme: Presses Universitaires.
- Steyer, K. & Hein, K. (2010): Deutsche Sprichwortartikel. In: Datenbank des EU-Projekts SprichWort. www.sprichwort-plattform.org/sp/Sprichwort (accessed November 12, 2013).

- Steyer, K. (2003). Korpus, Statistik, Kookkurrenz: Lässt sich Idiomatisches „berechnen“? In H. Burger, A. Häcki Buhofer & G. Gréciano (Eds.): *Flut von Texten – Vielfalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifk der Phraseologie* (Phraseologie und Parömiologie 14) (pp. 33-46). Baltmannsweiler: Schneider Hohengehren.
- Steyer, K. (2004). Kookkurrenz: Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In K. Steyer (Ed.), *Wortverbindungen – mehr oder weniger fest* (Jahrbuch des Instituts für Deutsche Sprache 2003) (pp. 87-116). Berlin/New York: de Gruyter.
- Steyer, K. (2012b). Sprichwortstatus, Frequenz, Musterbildung: Parömiologische Fragen im Lichte korpusmethodischer Empirie. In K. Steyer (Ed.), *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie* (Studien zur Deutschen Sprache 60). (pp. 287-314). Tübingen: Narr.
- Steyer, K. (2013). *Usuelle Wortverbindungen: Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht* (Studien zur Deutschen Sprache 65). Tübingen: Narr.
- Steyer, K. (Ed.) (2012a). *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie* (Studien zur Deutschen Sprache 60). Tübingen: Narr.
- Steyer, K. (in print). Zitierkarrieren. Vom Vergehen und Überleben markanter Sätze. In C. Grandl & K. McKenna (Eds.). „Bis dat, qui cito dat“ – Gegengabe in Paremiology, Folklore, Language and Literature. *Honoring Wolfgang Mieder on His Seventieth Birthday*. Frankfurt am Main: Peter Lang.
- Stubbs, M. (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- SWP: EU-Sprichwortplattform. Internet: <http://www.sprichwort-plattform.org> (accessed November 12, 2013).
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work* (Studies in corpus linguistics 6). Amsterdam/Philadelphia: John Benjamins.
- Umurova, G. (2005). *Was der Volksmund in einem Sprichwort verpackt: Moderne Aspekte des Sprichwortgebrauchs anhand von Beispielen aus dem Internet* (Sprichwörterforschung 24). Bern: Peter Lang.

Appendix 1

High frequent proverbs of the Sprichwort Platform (SWP) with Numbers of hits and queries (DeReKo 2013) (17.10.2014):

1. <i>Weniger ist mehr.</i>	3.579 \$weniger /+w1 ist /+w3 \$mehr
2. <i>Der Schein trügt.</i>	2.738 Schein /+w3 trügt
3. <i>Ende gut, alles gut.</i>	2.638 Ende /+w3 gut /+w3 alles
4. <i>Aller guten Dinge sind drei.</i>	2.511 \$aller /+w3 guten /+w3 Dinge /+w5 (\$drei ODER 3)
5. <i>Die Hoffnung stirbt zuletzt.</i>	2.364 Hoffnung /+w3 stirbt /+w5 zuletzt
6. <i>Was lange währt, wird endlich gut.</i>	2.216 \$was /+w3 lange /+w3 währt /+w5 gut
7. <i>Aller Anfang ist schwer.</i>	1.894 \$aller /+w3 Anfang /+w5 schwer
8. <i>Der Weg ist das Ziel.</i>	1.833 Weg /+w1 ist /+w3 Ziel
9. <i>Viel Lärm um nichts.</i>	1.725 (&viel /s0 &Lärm /s0 nichts) %s2 (Shakespeare ODER Shakespeares ODER Shakespeare's ODER &Uhr ODER &Kino ODER &Theater ODER Branagh)
10. <i>Zeit ist Geld.</i>	1.605 Zeit /+w1 ist /+w5 Geld
11. <i>Totgesagte leben länger.</i>	1.560 Totgesagte /+w3 leben /+w5 länger
12. <i>Aufgeschoben ist nicht aufgehoben.</i>	1.540 \$aufgeschoben /+w3 „nicht“ /+w5 aufgehoben
13. <i>Alles hat seine Zeit.</i>	1.514 \$alles /+w3 hat /+w5 seine /+w5 Zeit
14. <i>Not macht erfinderisch.</i>	1.366 „Not“ /+w3 macht /+w5 erfinderisch
15. <i>Ehre wem Ehre gebührt.</i>	1.330 (Ehre /+w2:2 Ehre) /+w5 gebührt
16. <i>Die Konkurrenz schläft nicht.</i>	1.317 Konkurrenz /+w3 schläft /+w5 „nicht“
17. <i>Ausnahmen bestätigen die Regel.</i>	1.289 &Ausnahme /+w5 bestätigen /+w5 &Regel
18. <i>Kleider machen Leute.</i>	1.270 Kleider /+w3 machen /+w5 Leute

19. *Vertrauen ist gut, Kontrolle ist besser.* 1.132 Vertrauen /+w3 gut /+w5 Kontrolle
/+w5 besser

20. *Wer rastet, der rostet.* 984 \$wer /+w3 rastet /+w5 rostet