KoGloss:

collaborative corpus-based construction glossaries

for foreign language learning in academia
and advanced training

Duisburg-Essen, Tartu, Vilnius, Ventspils
2012

Project group KoGloss:

University of Duisburg-Essen	University of Tartu	Vilnius University	Ventspils University College
Prof. Dr. Ulrike Haß (Project Manager) Dr. Steffi Engert Dr. Holger Gollan Dr. Anke Petschenka Derya Gür, M.A. Victoria Sirbu, M.A. Julia Steube, M.A. Julia Wrede, M.A. Eva Zitta, M.A. Beratung: Jörg Thiesing, Dipl. Phys.	Dr. Anne Arold Dr. Terje Loogus Siiri Aluoja, M.A. Kaari Antzon, M.A. Pille Rinne, M.Sc.	Dr. Daumantas Katinas Dr. Eglė Kontutytė Dr. Virginija Masiulionytė Dr. Lina Plaušinaitytė Dr. Vaiva Žeimantienė Alina Isilionis, M.A. Skaistė Volungevičienė, M.A.	Prof. Dr. Dzintra Lele-Rozentāle Sintija Blumberga, M.A. Agnese Dubova, M.A. Egita Proveja, M.A. Ivita Skripačenoka, M.A.

Funded by the European Union.

TABLE OF CONTENTS

I. WHAT IS KOGLOSS? INTRODUCTION TO THE METHOD

The area of languages for specific purposes is still strongly focused on collection, explanation and translation of separate technical terms. There are reference works on the terminology of most diverse specialities. However, what is the situation like if one rises above the word level? Where would one turn to when questions do not crop up about separate expressions but about word combinations and language patterns? For a layman, for example, it might be difficult to decide what kind of constructions are used in Business German to speak about percentages, inflation rates and labour market figures. One might be surprised that liquidity can be provided or fed or someone can be supplied with it. What does it actually mean when growth is negative? Conventional reference works offer only limited advice on these questions. The same problem arises in foreign language acquisition. In learning and teaching, the tendency is to move away from cramming isolated words and to turn to acquisition of word collocations and language constructions. Systematisation of such an approach is, however, an area with much unused potential.

KoGloss as a language teaching method provides opportunities for satisfying this need by independent learning.

What is KoGloss?

On the one hand, the name KoGloss reflects the glossary as a kind of lexicographic product; on the other hand, the element 'ko' represents the three other central aspects of the project. It stands for collaborative, corpus and constructions, and indicates that cooperation between several groups takes place here (collaborative); that it deals with linguistic units above the word level (constructions) while the analysis is based on digital text collections (corpus).

KoGloss works with a low-threshold form of digital text analysis. For that purpose, data are collected that cover the target area representatively and sufficiently. The collection can be comprised, for example, of a number of texts for specific purposes from the press, speciality publications and official publications that are accessible on-line or are available to someone professionally. Next, the material is analysed with a freely available software program in order to detect the presence and function of particular language patterns. The results are stored in a digital reference work – the glossary.

A particularity of the KoGloss method is that it can store authentic text material from any domains and specialities with regard to their linguistic peculiarities. Typical language patterns of the sphere in focus can be identified and utilized without staying at the level of separate

technical terms. A further advantage is that composition and evaluation of texts hardly require any linguistic expertise as, after short initial training, KoGloss can be used by users with most different language competences.

The potential users of the KoGloss method are all who in their job and/or further education have to pay special attention to drawing up and processing of texts: language learners, language teachers, and specialists like translators, interpreters, journalists, PR officers, secretaries, people engaged in international correspondence, and terminologists. On the one hand, KoGloss can help both native speakers and foreign language learners to achieve higher language awareness and, on the other hand, to solve quite concrete translation problems with the help of a glossary. The method makes it possible to acquire new areas in language for specific purposes in one's own or a foreign language quickly and purposefully. Hence, it can be expected that the ability to analyse native and foreign languages will clearly increase and will be of immediate use for persons in communicative professions.

This manual will present a step-by-step overview of the language didactic method KoGloss which consists of three fundamental stages – corpus building, corpus analysis and description of language constructions in a glossary.

For further information see:
http://www.uni-due.de/kogloss.eu/index.php

II. CORPUS BUILDING

As a first step to use the KoGloss method for studying particular language patterns, a digital text collection (a text corpus) should be created. In educational or professional surroundings, this can be done in two ways: the text corpus is built by teachers or project managers and presented to the learners or project participants as a ready-made product, or the text corpus is created in cooperation with learners or project participants.

Before starting to build the corpus, it is important to set certain criteria for the selection of texts and data sources. This concerns mainly the content of the texts, the access to the texts and the scope of the corpus (e.g. according to the number of texts).

1. SELECTION OF TEXTS

You can copy the texts from the Internet or scan them. Decide which criteria the texts included in the corpus should meet (e.g. the themes, text sorts, media, text length, the time span of the texts, etc.).

To ensure the thematic unity of the corpus, a list of search words can be compiled for searching the Internet. It is, however, necessary to bear in mind that no such texts should be selected for the corpus where the search word occurs only accidentally.

It is also important to avoid incorporating a great number of texts that are based on the same pattern or source and include no or only a little linguistic variation. Remember that some constructions in the corpus may have high frequency, but this need not mean that they are highly necessary.

For technical reasons it is advisable not to include such PDF files that contain 'photographed' text as their conversion is too complicated because of their layout (several columns, pictures in the text, etc.).

2. CONVERSION, ARCHIVING AND CLEANING OF TEXTS

When converting texts, keep to the following order: PDF → WORD → TXT.

The TXT files are saved and provided with uniform file names. Therefore, it is necessary to create a separate document that contains the **abbreviations** for the sources used, e.g.

Source	Abbreviation
SPIEGEL	SP
Bundesministerium der Finanzen	BMF
Financial Times	FT

The **file names** can be constructed according to the following pattern:
Source (abbreviated)_ Year_Month_Date of publication of the text_Title (abbreviated)

For example: SP_2008_05_15_Konjunktur.txt

For **cleaning** the texts included in the corpus, it is advisable to use the WORD format, as this is better suited for the purpose than the TXT format. The document should contain only cleaned text. For that, it is necessary:

a) to connect word fragments that result from division of words at line ends;

b) to delete the following text elements: page numbers, headers and footers, footnotes, marginal notes, names of authors, references to literature, forums or user comments, links, tables and figures, numeration of paragraphs, passages in foreign languages, special characters (e.g. enumeration signs) and mathematical formulas. Furthermore, text and paragraph headings, captions to pictures and enumerations have also to be deleted if they do not form full sentences.

c) Text and paragraph headings, captions to pictures and enumerations that form full sentences have to be provided with a full stop and included in the continuous text.

At the beginning of the cleaned text, information on the **source** can be given in angle brackets according to the following pattern:

<name of the saved file: full link to the file>
For example:
< SP_2008_05_15_Konjunktur: http://www.spiegel.de/wirtschaft/0,1518,553373,00.html>

After cleaning, the document can be converted into the TXT format. At that, it is necessary to follow the uniform **UTF-8** encoding.

Next, the TXT file can be entered into a separate **list of sources** according to the following pattern:

File name. Full title Full link (date of access)

For example:

SP_2008_05_15_Konjunktur. Deutsche Wirtschaft wächst überraschend stark. http://www.spiegel.de/wirtschaft/0,1518,553373,00.html (22.01.2011)

As the next step of the KoGloss method, the text corpus can be researched with the analysis program *AntConc* which will be described in the following chapter.

III.   CORPUS ANALYSIS WITH *ANTCONC*

1. WHAT IS *ANTCONC*?

In research of particular language patterns, KoGloss relies on its own digital text collections (see Chapter II, Corpus building). There are various analysis programs for digital use of such corpora. They are meant for gathering information about particular language patterns from the texts collected earlier. To do it, you can submit search requests in the form of separate words, word parts or word sequences. Such analysis software programs are of different complexity and offer various analysis tools. The freely available programs are, e.g., *Simple Concordance Program* (SCP) and *KWiCFinder*, while programs like *ParaConc* or *WordSmith* Tools have to be acquired for a fee. Depending on your needs, you can decide which software meets your requirements best. KoGloss uses the freeware program *AntConc* that is easy to use even for non-professionals in linguistics.

AntConc was created by Professor Laurence Anthony, Head of the Centre for English Language Education in Science and Engineering at Waseda University in Tokyo and is particularly suitable for smaller text collections (up to 20 million text words). It is a widely used program for analysis of digital text collections. It is simple to use, need not be installed and has detailed documentation and other resources (see http://www.antlab.sci.waseda.ac.jp/software/antconc335/AntConc_readme.pdf). The available tools allow submitting search requests for word lists, clusters, concordances, collocations and can provide a detailed view of contexts of such search results.

2. INSTALLATION AND START OF *ANTCONC*

AntConc software can be downloaded free of charge from its author's website (http://www.antlab.sci.waseda.ac.jp/software.html). After download, you can start the program (without installation) by clicking the .exe file. Its structure is based on the well-known layout of Windows. The program also uses many shortcuts that should be known from Windows.

After opening the following window will appear:
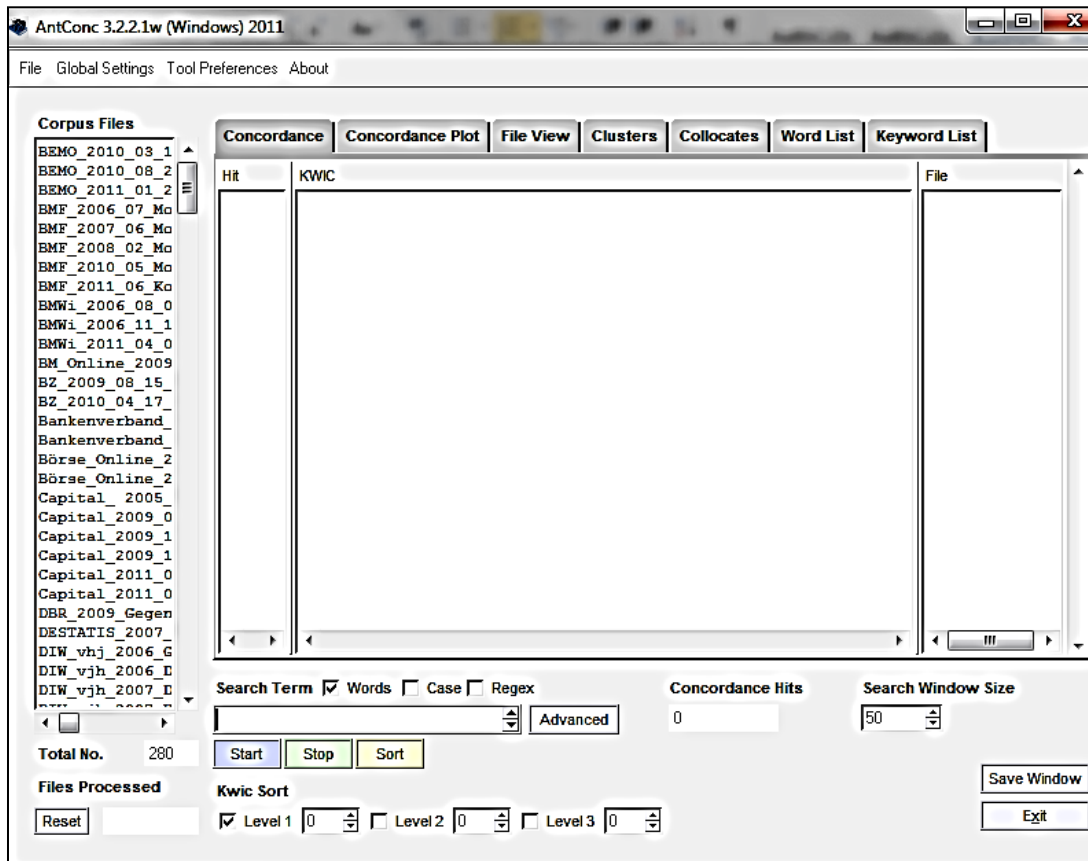
Corpus Files (on the left side): Here you can see a list of the downloaded text files. Their total number is shown at the bottom of the page under *Total No.*

Tabs Various tools appear in the form of tabs; with their help you can, for example, find concordances, identify word clusters, calculate collocates or frequencies. After submitting a search request, the results will appear in the middle window of this area.

Search line and search settings (below the middle). In this area you can type search requests and start. Various search criteria can be set for each tool. Particulars will be given in the section *Submit search* requests.

3. BEFORE THE RESEARCH

a) Feed in the text collection

Before you can start analysis with *AntConc*, you have to feed the texts into the program. From the menu item **File**, you can select locally saved corpus files or whole directories of corpus files and feed them into *AntConc* with *Open File(s)* and *Open Dir...*

Next, the list of downloaded files will appear on the left side of the screen.
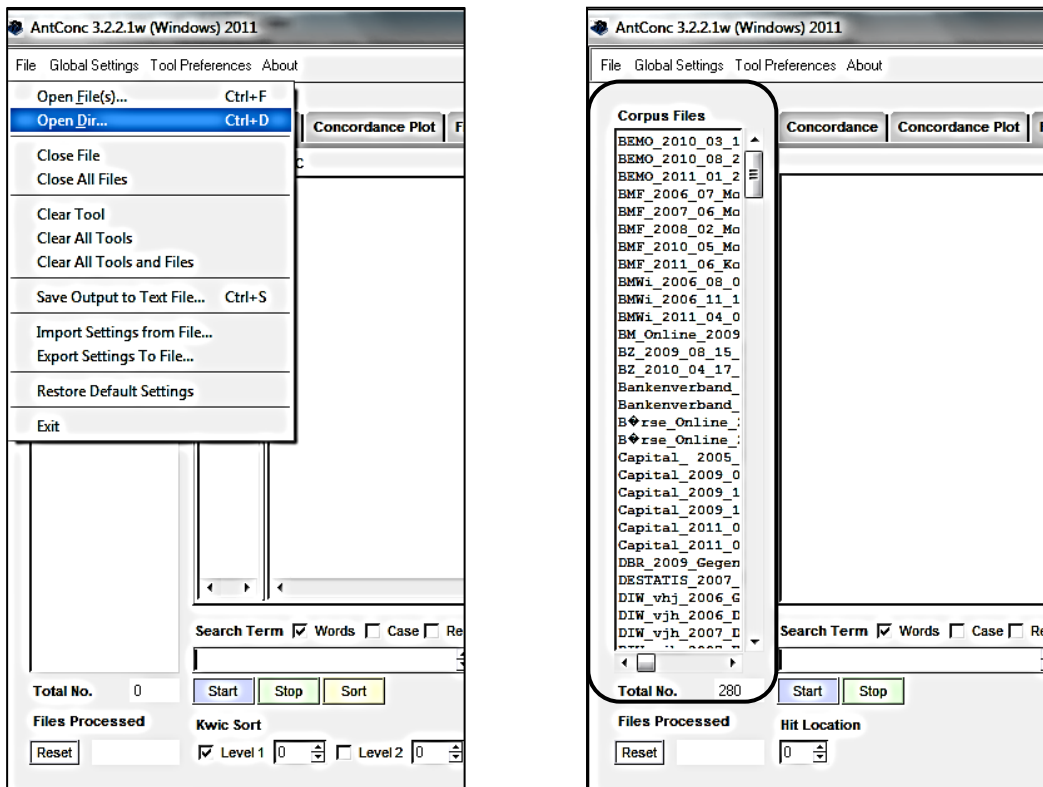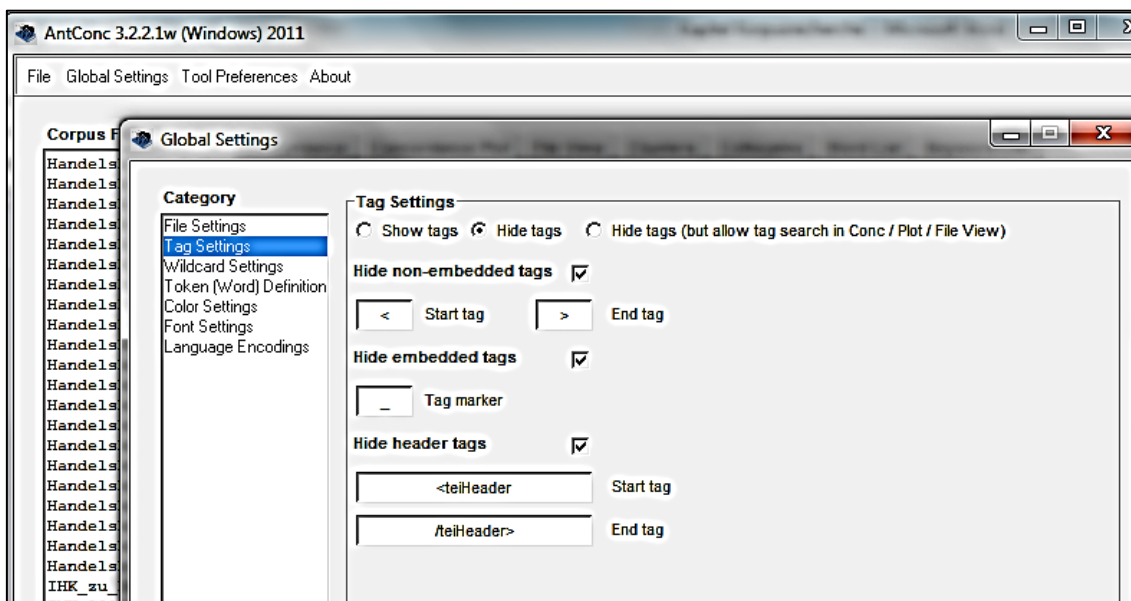

b) Make settings

To guarantee the correct analysis of data with *AntConc*, default values can be set in the menu option **Global Settings**. This concerns, for example, *Language Encoding* or *Tag Setting* (see below). All the information about default values can be found in the manual on *AntConc* use (chapter "Menu Options").
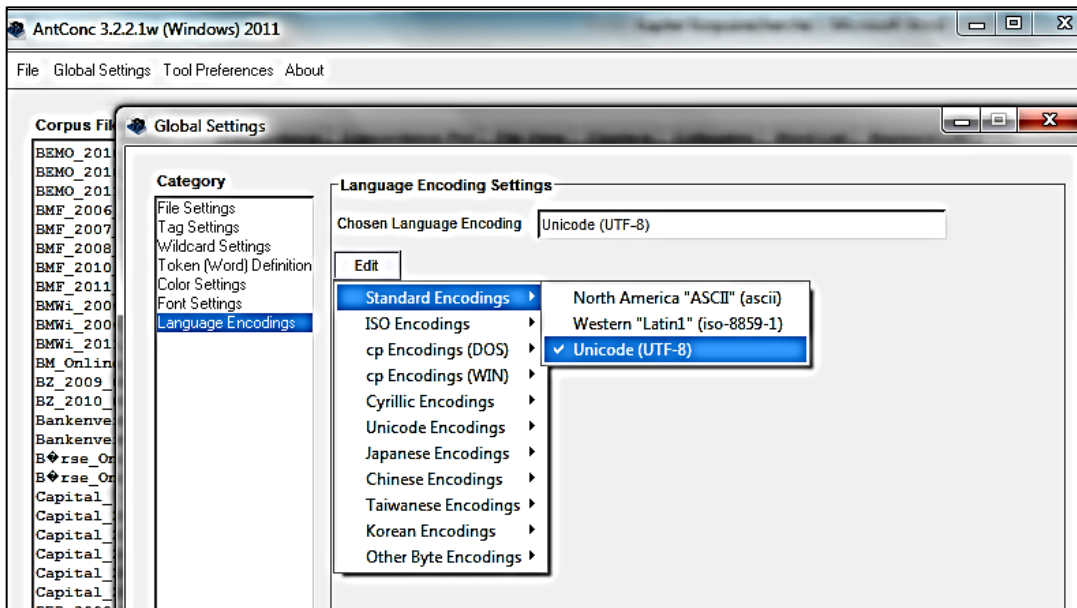
Language Encodings: To show characters correctly, select the coding UTF8.

Tag Settings: With the setting *Hide Tags*, you can designate by angle brackets which text sequences are excluded from the search, e.g. meta-data like <author's name>, <title>, etc.

c) Tools and their functions

The main functions of the KoGloss method can be used with the tools ***Word List***, ***Concordance***, ***Clusters***, ***Collocates*** and ***File View***. These can be used by clicking the corresponding tab.

With the tool **Word List**, you can create an ordered list of words occurring in the text file that is being used. The words can be sorted by frequency or alphabetically by the beginning or the end of the word; the order of the list can be reverted.

The **Concordance** tool enables to create so-called KWIC lines. It lists the lines where the **k**ey **w**ord **i**n **c**ontext (KWIC) occurs and gives information about the usage of a certain word or language pattern in the text collection.

The function **Clusters** enables to create an ordered list of word clusters that occur around a search term in the selected text files. Here, a particularly interesting subfunction is **NGrams**, as it enables to identify word clusters that often appear together without focusing on a certain word or pattern. You can select the minimum and maximum length (number of words) in each word cluster and the minimum frequency of the indicated word clusters.

With the **Collocates** tool, you can create a list of collocations that frequently occur with a certain search term or pattern. At that, the span considered at search can be varied.

The **File View** function enables at any time to call up texts that contain a certain search word. It shows then the complete view of the corresponding text file, thus showing the full linguistic environment of the selected search term.

4. USING OF ANTCONC. PROCEDURE AND FUNCTIONS

4.1 SUBMITTING SEARCH REQUESTS

a) Search requests with and without a search word

Generally, a search request starts by clicking the **Start** button. It is possible to submit search requests with or without a search word. Irrespective of a concrete search word, it is possible to create lists of words that occur frequently in the text collection (tab **Word List**). Additionally, it is possible to find frequent word clusters by starting a search request in the **N-Gram** function in the **Clusters** tab.

To find information about a concrete search word, type the word on the input line and start the search by clicking **Start**. *AntConc* enables you to use the **Collocates** function for identifying the words that often occur together with the search word or to use the **Cluster** tab to find word clusters where the search word frequently occurs. To see all the lines where the key word occurs, start the search request with the **Concordance** tab.

To learn more about the context of search words and/or word clusters, click on those and, in the **Concordance** and **File View** tabs, you can get a view of their direct linguistic environment or the whole text.

Details of all the other functions are given below.

b) Application of wildcards

For search requests with search words, you can use operators that have a placeholder function, so-called *wildcards*. With the help of wildcards, you can include various endings of the same word and the following words in the search request.

It is possible to fix more precisely how many characters the placeholder should cover and at which place the characters should occur.

Overview of automatically set placeholder functions:

+   one character or none
*   one or more characters
?   exactly one character
@   one word or none
#   exactly one word
| search word 1 OR search word 2

If you type '**+**' into a search request, all the words will be found which have an indefinite character or no additional character at that place. So, the search request stell+ yields word forms like *stell, Stelle, stelle* and *stellt*. To exclude the case that the placeholder remains empty, type '**?**'.

If you type '*****' into a search request, all the words will be found which have one or more indefinite characters at that place. This way, the list will also include words like *Stellen*, *stellte*, *Stellung*, *stellenweise*, etc.

If the placeholders have to stand for whole words or word combinations, use the operators '**@**', '**#**' und '**|**'. To find one word or no word next to the entered search word, use the placeholder '**@**', while to find exactly one word, use '**#**'.

To search for a combination of two search words, use the operator '**|**' which has the meaning of 'or'. It will find all the places where one or the other search word occurs, while the cases where both words occur together will not be specially marked. So, in the case of the search

request *Frage|Antwort* you will get all the search results that include *Frage* and all the search results that include *Antwort*.

c) Expanding your search to variant forms of words

To set possibly well-aimed and effective search requests, you should bear in mind certain peculiarities of linguistic expressions.

This concerns, for example, the inflection of words, primarily of verbs, nouns and adjectives. To receive all the possible variants of the principal form of a word, all the possible endings of the word should be covered by placeholders. In detail, this could look as follows:

Example of a noun:

Input:	*System++*
word forms found:	*System*, *Systeme*, *Systems*, *Systemen*
Examples[1]:	[…] *koexistierten die* **Systeme** *friedlich* […]
	Das regulative **System** *der Versicherungswirtschaft* […]
	Die Hauptkomponenten des **Systems** *bilden* […]

Thus, the search results include hits where the word occurs in the singular or in the plural and in various grammatical cases. To find new word formations with the search word as a stem, choose a placeholder that covers any number of characters. Thus the input *System****** yields results like *systematisch*, *Systematisierung*, *systemerelevant*, etc.

Example of an adjective:

In the case of adjectives, the placeholder '+' has often to be used repeatedly, as the adjective endings of degrees of comparison tend to be longer.

Input:	*klein+++++*
word forms found:	*klein, kleines, kleinen, kleiner, kleineren, kleinsten,* etc.
Examples:	*Wenn nur ein* **kleiner** *Teil der 2,3 Millionen* […]
	Dies dürfte vor allem **kleinere** *Börsen in Europa belasten* […]
	[…] *sei die Abweichung* **kleiner** *als 0,5 % gewesen* […]

---

[1] All the examples have been taken from the KoGloss corpus of Business German. In general, when working with AntConc, attention should always be paid to the specific features of a particular language (languages with diacritics, gradation, etc.).

<u>Example of a verb:</u>

Input: *++such+++*

word forms found: *suchen, sucht, suchte, suchten, gesucht*

Examples: *Jede Regierung **sucht** nun nach der geeigneten Strategie* […]

*Deshalb wird 2009 nach einem Maß **gesucht*** […]

[…] *wer einen **sucht**, rechnet sich mehr Chancen aus* […]

Here, the search results include hits in which the word occurs in the singular or in the plural and in various tenses and persons. At verbs, it is particularly important to cover the possible past forms by placeholders (e.g. place double '+' in front of the word to find the past participle *gesucht*).

To find new word formations with the verb as a stem, use a placeholder that covers any number of characters. Thus, the input **such** yields results like *untersuchen*, *Untersuchung*, *Arbeitssuchende*, *heimgesucht*, etc. By placing the placeholder '*' before the stem, you can also find words with prefixes.

<u>Special case – inflection</u>

In general, pay attention to cases when the word stem is changeable (e.g. *Plan*: *Pläne*). Then not all the inflected forms can be can be found this way, and separate search requests have to be submitted, like, for example, in the case of *gehen*, as the forms *gegangen* or *ging* cannot be found when searching for the verb stem. The German language also uses separable verbs (like *einbrechen* or *losfahren*) which in certain contexts occur at two different places in the sentence, for example, in the past tense: „*Die Erzeugung im Bauhauptgewerbe **brach** im März um 10,8 % **ein**.*" (Handelsblatt online, 09.05.2005)

In other particular languages, further specific forms should be taken into consideration.

AntConc also enables to feed in so-called lemma lists that cover all the possible forms of a word beforehand and make further search requests easier.

4.2   WORD LIST

To create a list that includes all the word forms occurring in a text collection, open the tab **Word List** and click the *Start* button to start the search. At that, no search word need be typed on the search line. You can use the following search settings for sorting the results:

- **Sort by**: The words can be sorted alphabetically or by frequency. The sorting order can be inverted (*Invert Order*).

- **Treat all Data as Lowercase**: If this setting is activated, no difference is made between higher-case and lower-case letters.

The result gives the following information:
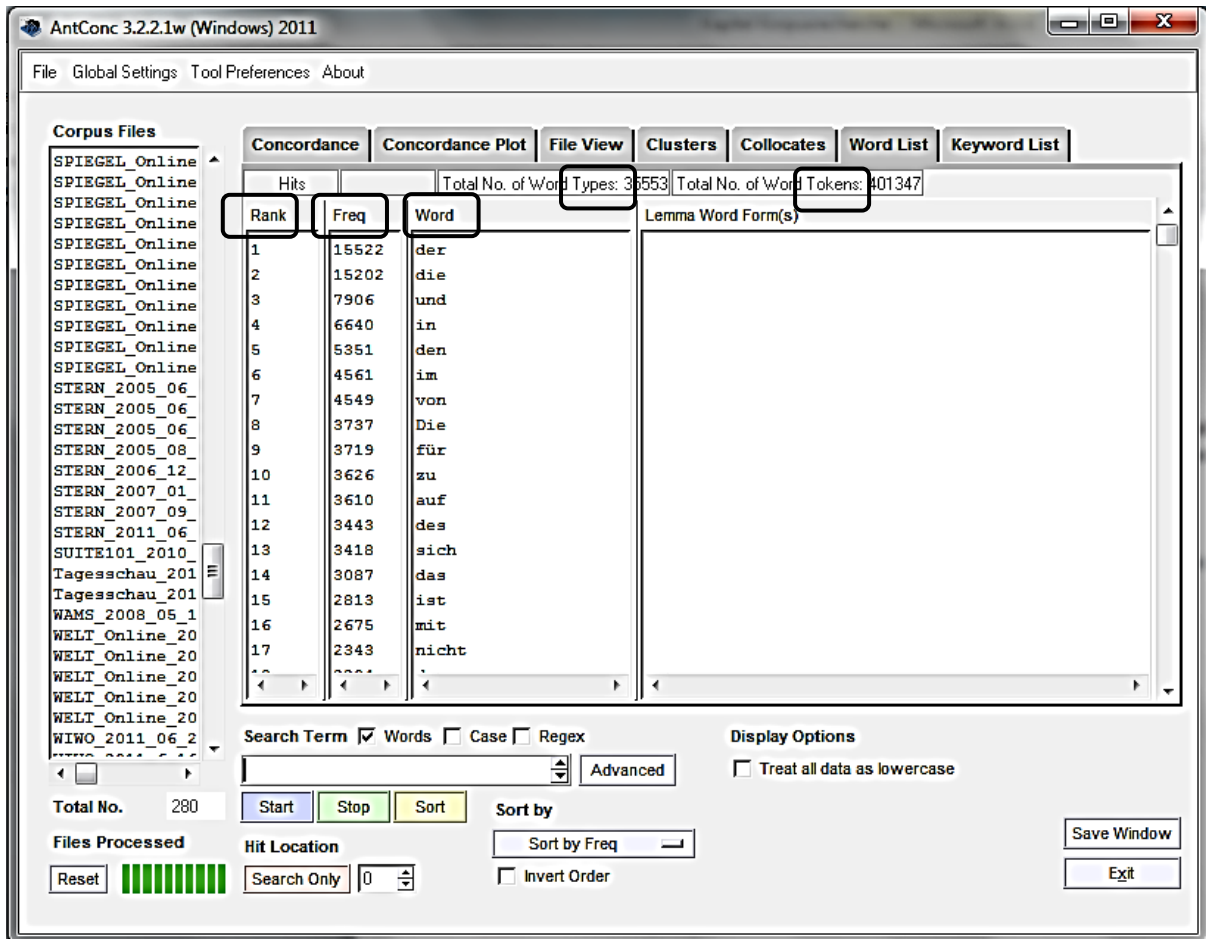
The line **Word** shows all the word forms in the text.

Rank shows the ranking of each word form in the text collection according to their frequency.

**Freq** shows how often the searched word form occurs in the text collection.

Types: total number of all the word forms occurring in the texts.

Tokens: total number of all the text words occurring in the texts.

4.3 CONCORDANCE

To get a list with KWIC lines, open the tab **Concordance**. Here the search request can start with the input of a keyword. *Concordance Hits* shows the number of concordances found

The search settings can be adjusted as follows:
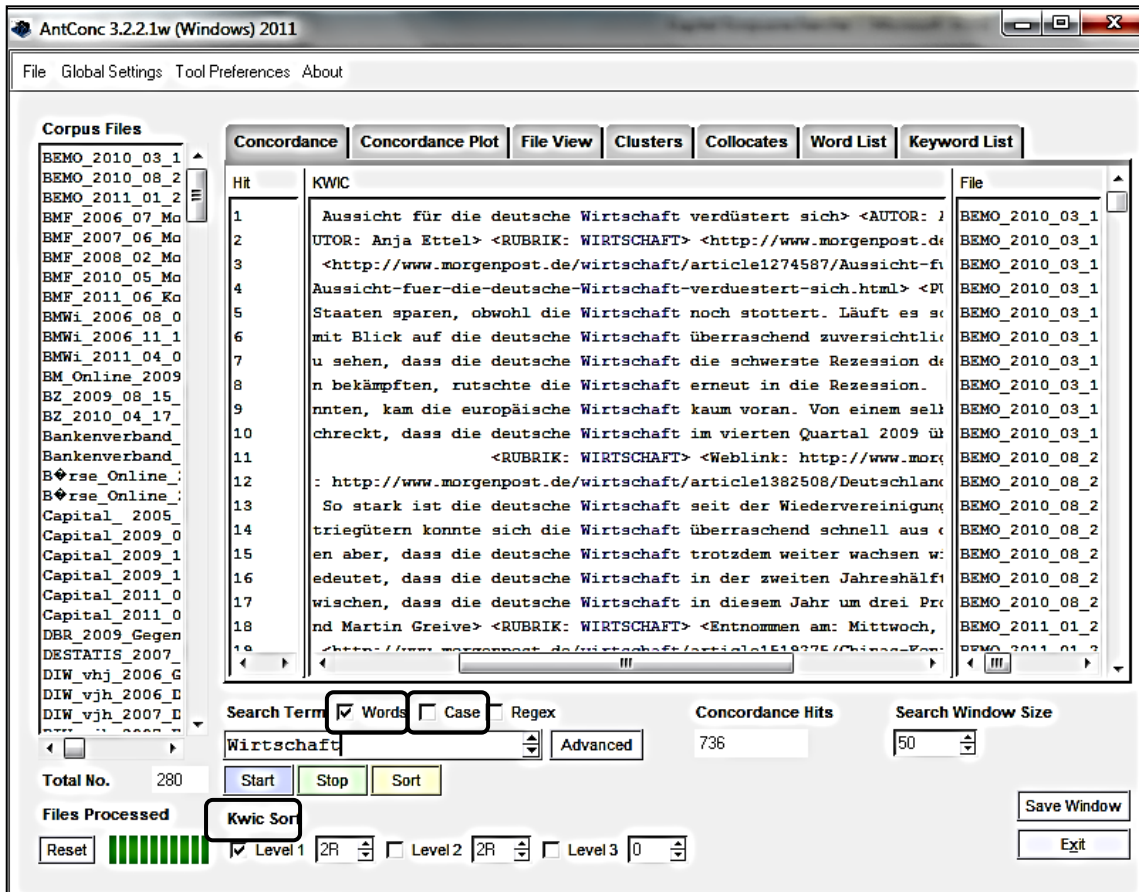
Case: If this setting is selected, a difference is made between higher-case and lower-case letters in the search word, otherwise it is ignored.

Words: The search term is one full word. If this option is not selected, word fragments can also be found.

Kwic Sort: An option to determine how many words on the left/right side of the search word will be shown.

Alternatively, the results of an earlier search request in the Cluster, Word List or Collocates function can be clicked, and automatically the Concordance tab with the corresponding KWIC lines will appear.

4.4 FILE-VIEW

To see the environment of a search word hit in the original text, click on the corresponding word or pattern and the view will turn into the tab **File View**. The window shows now the whole text file, in which the corresponding hit was found.
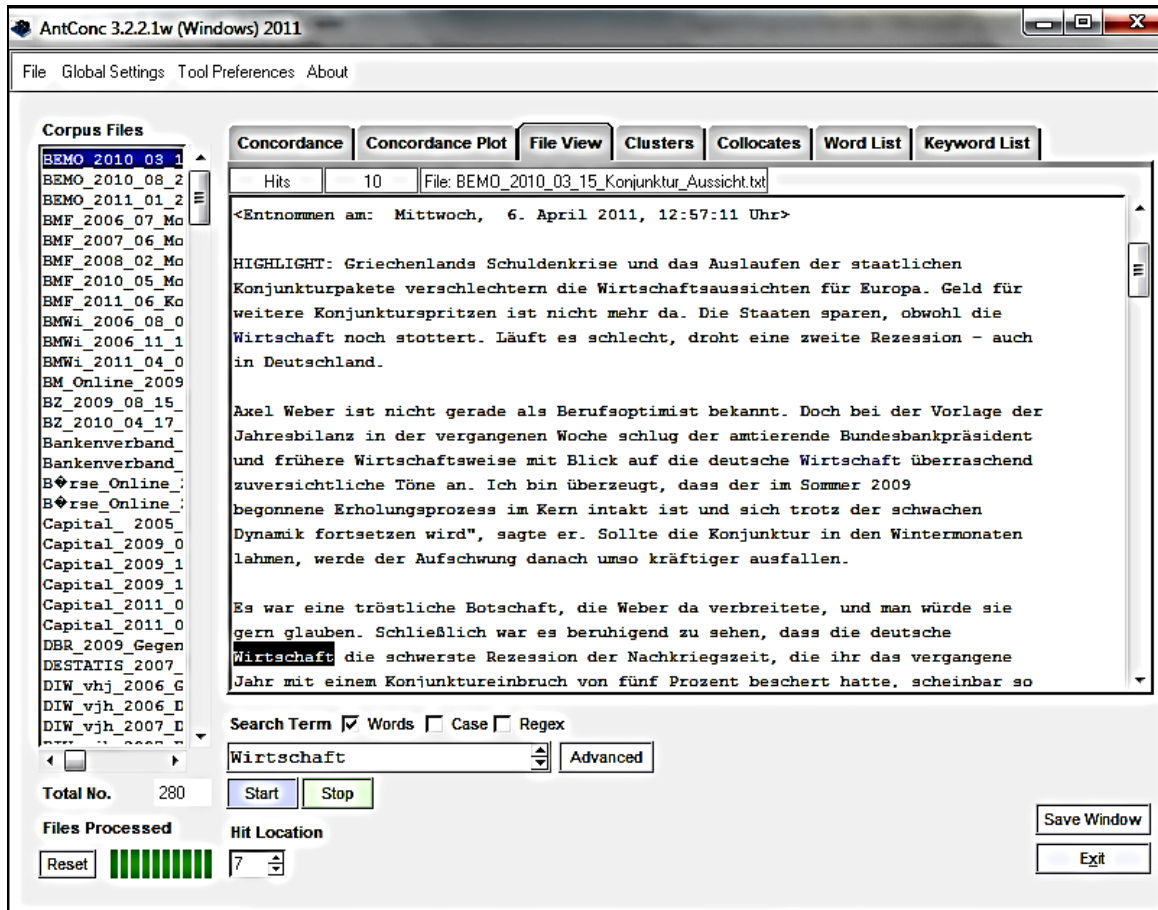
4.5   COLLOCATES

The **Collocates** tab enables to create lists of neighbouring words that particularly often occur together with a certain search word.

At that, you can apply the following search and sorting settings.
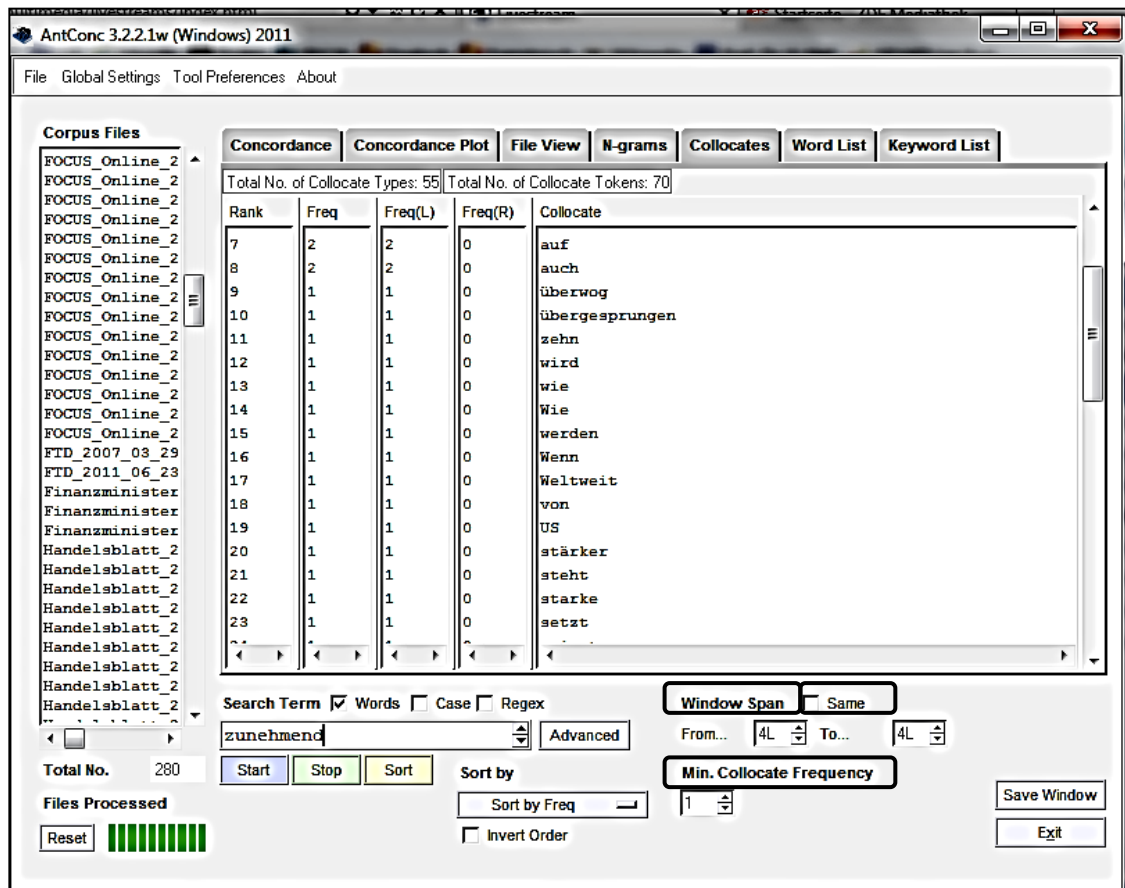
With **Window Span** you can choose how many words on the left/right of the search word will be shown (here a four-word span).

To keep the minimum and maximum span the same, activate the check box **Same**.

With **Min. Collocate Frequency** you can set the minimum frequency of the co-occurrence partners.

4.6 CLUSTERS/N-GRAMS

a) Search request with a search word

The function **Clusters** enables to create an ordered list of word clusters that occur around a search term in the selected text files. To do this, type the corresponding word on the input line and start the search by clicking **Start**.
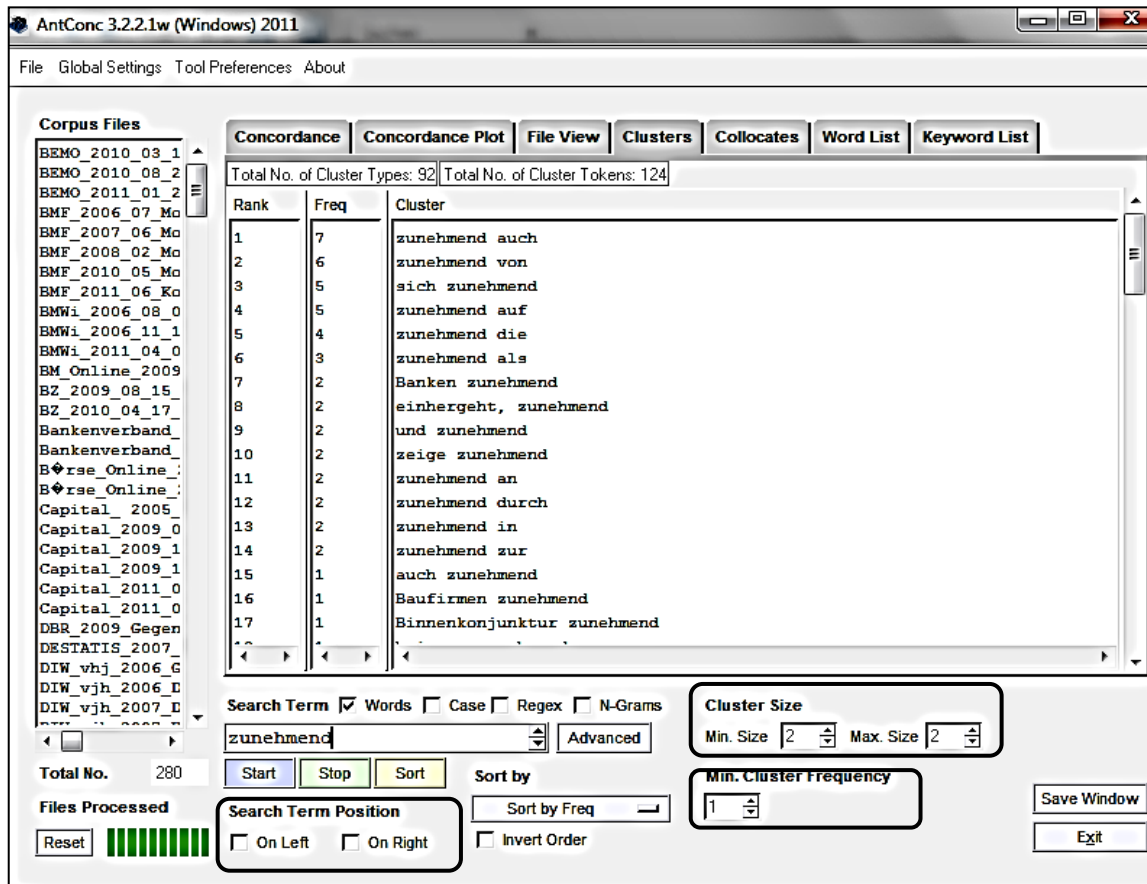
The following search settings can be applied:

**Cluster Size**: here the number of words in the word cluster (in this example two words) can be set.

Min. Cluster Frequency: here the minimum frequency can be set from which upwards a word cluster will be shown (in the example, starting from one occurrence)[2].

Search Term Position: here you can set if the search word should always occur on the left or the right side of the word cluster.

---

[2] As a rule, the more frequent hits are of greater interest. This, however, depends on the aim of the search.

b) Search request without a search word

By ticking the field **N-Grams** in the **Cluster** tab, it is possible to start a search request without a search word. The result is a list of frequently occurring word clusters in the text collection, while the size of the cluster can be set.
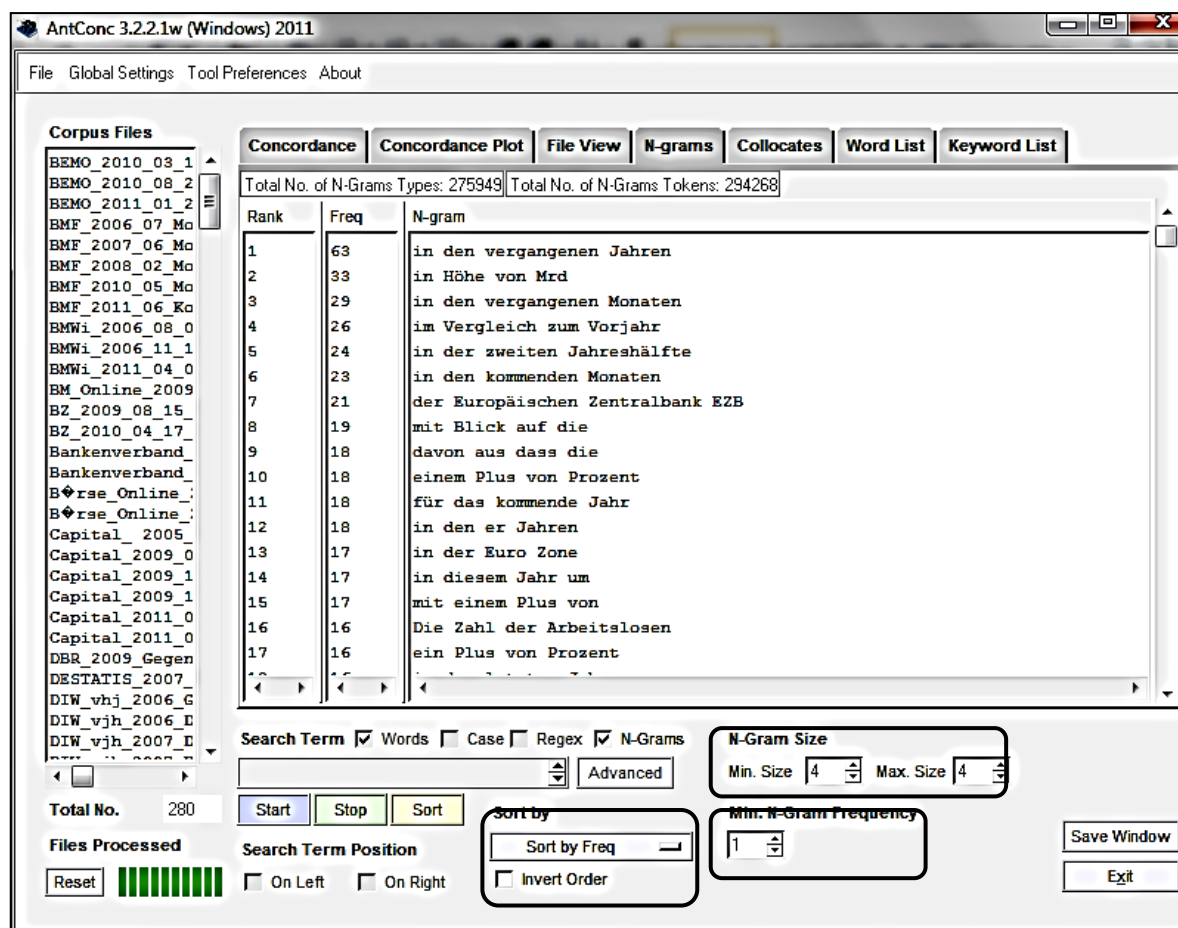
The size of the N-Gram can be set with **Min./Max. N-Gram Size**.

With **Min. N-Grams Frequency** you can set the minimum frequency of N-Grams[3].

The N-Grams can be sorted either by frequency (**Sort by Freq**) or by the beginning or the end of the word (**Sort by Word / Word End**).

---

[3] Here, it is also advisable to consider only the hits with greater frequency.

5. EXPORT AND SAVING OF RESULTS

After separate research stages, the search results can be exported and saved locally, after which they can be used directly or as a basis for a glossary.
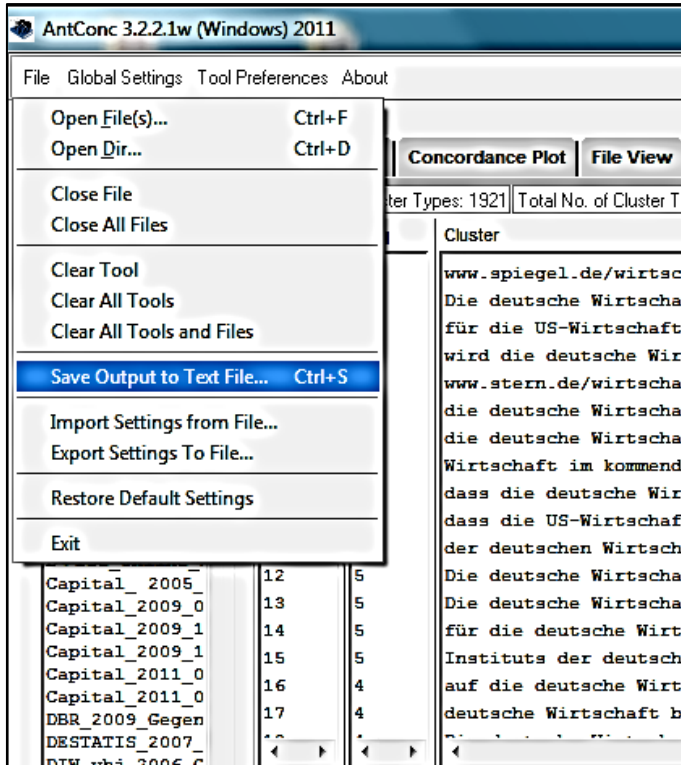
In the **File** menu, the analysis results can be saved with **Save Output to Text File**. The analysis results are released as a txt file and can be saved locally on the same computer or on some other data storage device.
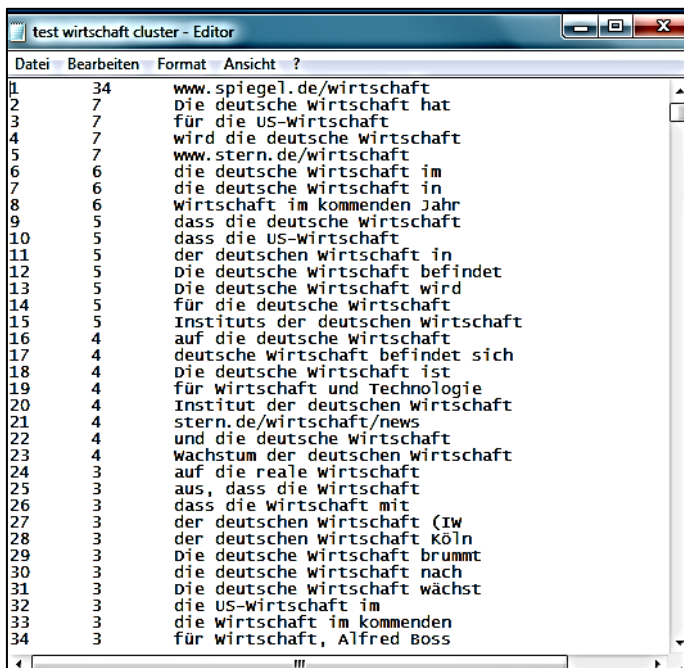
The text file contains the same information as the search window in *AntConc,* depending on the particular search request.

The result here shows the word clusters with the search word *Wirtschaft.*

After search requests, the next step is the application of the actual glossary entry that can include various information as required. The analysis software offers numerous options for use, and the aspects of analysis may vary according to the level of knowledge and purpose. The KoGloss method generally aims at fulfilling the categories presented in Chapter IV.

IV. PROCESSING OF CONSTRUCTIONS IN THE GLOSSARY

The constructions selected by *AntConc* will be recorded as glossary entries. In a glossary entry, the information can be edited and presented according to the following basic pattern:

Headword or construction: ...

Morphology ...

Information about the syntactic structure of the construction ...

Meaning(s) ...

Pragmatics (Usage): ...

Usage examples of the construction from the corpus: ...

Other information: ...

Teachers, students or professional users can complement, shorten or modify this basic pattern. The basic pattern contains information about morphology, syntax, semantics (meaning) and usage of the construction (style, specifics of text sorts, frequency of usage, etc.). The entry also contains usage examples from the corpus. The compilers of entries can also include necessary information from corpus analysis or their own knowledge in the glossary entry, regardless whether this concerns grammatical, semantic, pragmatic or purely factual aspects. The end of the entry can include the name of the compiler and the date.

The glossary entries can be submitted to the virtual learning platform Moodle and saved there as described in the chapter about Moodle. A completed glossary entry in Moodle might look as follows:

1. ELEMENTS OF GLOSSARY ENTRIES

a) Construction

The glossary entry above presents a construction detected with *AntConc* software that will be described in the accessory entry, e.g.:

Headword or construction: *im Vergleich zu X*

For collecting certain linguistic patterns and constructions characteristic of the language for specific purposes or a foreign language, *AntConc* can be used in various ways. On the one hand, the *Cluster* function (see Chapter III) can be applied to create lists of common collocations; on the other hand, *Collocates* search (see Chapter III) can be used for targeted search for collocates to a certain expression. In both cases, the first search results should be considered provisional. Targeted research should be used to check whether the word clusters found this way really form a unit of meaning, and if the word patterns found really are specific to the area under observation, e.g. business language, scientific language, the language of advertising design, etc. While making such decisions, it is helpful to view in detail the contexts of the words and word clusters found by means of KWIC lines (*Concordance* function) and *File View* function.

It should be taken into account that there in some cases variants exist inside word clusters, and at certain places a different word can occur in a construction. Such patterns can be included in the glossary using the placeholder 'X' (see example above).

b) Information on morphology

The information on morphology (parts of speech and inflection) is useful when the method is applied to language teaching. Then, the **word classes** of the constituent parts of a construction, e.g. *Bruttosozialprodukt* (noun), and information on the **inflectional paradigm** can be given (plural die *Bruttosozialprodukte*, genitive des *Bruttosozialprodukt(e)s*, etc.). Details of **word formation** could also be given here. To make the composition of a sample word clear, the word could be divided into its constituent parts: *Brutto + sozial + produkt*.

The information on morphology of constructions can also be expanded, e.g.
im Vergleich zu X

Morphologie:

$$\text{Präposition/Artikel}_{im} + \text{Substantiv}_{Sg.,Dat.} + \text{Präposition}_{zu/m} + X_{Dat.}$$

$$X = \text{Nominalphrase, z.B. Artikel + Substantiv, Artikel + Adjektiv + Substantiv}$$

Here, the user can learn, on the one hand, which form the components of the construction have (the first lines) and, on the other hand, what can be used instead of variable elements (=X) (the second line). Variant forms of words and word clusters can be found by using placeholders (wildcards) (see Chapter III).

c) Information on the syntactic structure of the construction

Here, it is shown which elements the construction consists of and which function it typically fulfils in the texts analysed. At that, the degree of formalisation can depend on the level of the target group.

For example, the construction *im Vergleich zu X* can be classified according to its composition as a prepositional phrase and according to its function in the sentence as an adverbial modifier.

Angabe zur syntaktischen Struktur der Konstruktion:

- **Aufbau:** Präpositionalphrase mit eingebetteter Präpositionalphrase - [im Vergleich [zu/m X]$_{PP}$]$_{PP}$
- **Einbettung im Satz:** als Adverbialbestimmung

The necessary information can be found when you examine the KWIC lines (*Concordance* function) of the word patterns and find and categorise their possible usages.

For professional users, the form of the entry can be as simple as possible, i.e. contain only the minimum of grammatical information, e.g.

> **im Vergleich zu X:**
>
> **Grammatik:** Vergleich, der; im Vergleich zu + Dativ

d) Information on meaning

The information on meaning (semantics) is equally important for language learners and professional users. Under "meaning(s)" a circumlocution (paraphrase) of the content of a particular construction is presented. The meaning is deduced from the corpus. For example, the meaning of the construction *im Vergleich zu X* can be paraphrased as in *direkter Gegenüberstellung zu einer zweiten Größe*. In description, words of the same or similar meaning (synonyms, in the example above *gegenüber*, *im Gegensatz zu*, etc.), and words of opposite meaning (antonyms) are also helpful and can be listed here. For example, the meaning of the construction *im Vergleich zu X* can be described as follows:

> **Bedeutung(en):**
>
> - 'in direkter Gegenüberstellung zu einer zweiten Größe'
> - **Synonyme:** *gegenüber, im Gegensatz zu, verglichen mit, in Relation zu*

Various dimensions of meaning of words and word clusters can be deduced and generalised with the help of research results from the text collection. Here, too, you can use the KWIC lines and the full text view (*File View*).

e) Information on usage

The information on usage (pragmatics) contains notes on style, usage frequency of constructions, etc. At that, the typical text sorts or limitation to a certain topic or domain can play a role.

For example, the description of the construction *im Vergleich zu X*, can explain in which function the construction occurs in texts and in which contexts it can mainly be found.

Gebrauch: wird genutzt als Bindeglied zum Aufzeigen einer Gegenüberstellung
oder Verhältnismäßigkeit zweier Größen, vor allem von Zahlenwerten
in ihrer zeitlichen Veränderung oder von unterschiedlichen Markt- und Finanzbereichen;
kontrastive Funktion;

Finally, it can be mentioned here if the usage of a word/construction expresses a certain attitude on part of the speaker (a negative or a positive evaluation, deliberate distancing of oneself from the situation, etc.). For example, the construction *Negativwachstum* is used for linguistic enhancement of an actually negative development.

The information on usage is important for both language learners and professional users.

f) Usage examples of constructions from the corpus

The usage examples from the text collection show, on the one hand, in which context a certain construction occurs, also which syntactic form it has, and in which function it can be found. On the other hand, the meaning and usage of the construction can better be explained by means of authentic examples. The usage examples are taken from the corpus by means of *AntConc* and included in the glossary entry. For that, approximately five full sentences are chosen, which show as many different usages as possible, including occasional meaning variants. To classify particular examples, the source can be given in brackets.[4]

---

[4]From the viewpoint of intellectual property rights, examples of sentence length are allowed. However, if you want to include a link directly in a text of your text collection, be sure that you comply with the national intellectual property rights, e.g. by limiting access to certain users or getting a permit from the proprietor of rights.

Verwendungsbeispiele aus dem Korpus:

*"Die Zinsausgaben sind abhängig von der Steigerung der Bundesschuld durch die jährliche Nettokreditaufnahme, vom Zinsniveau und von den Konditionen der Anschlussfinanzierungen **im Vergleich zu den Tilgungsleistungen**."* (Bundesministerium der Finanzen, Monatsbericht 05/2010)

*"Im Schnitt erwarten die Analysten 2010 **im Vergleich zu 2009** einen Gewinnanstieg von 40 Prozent und bis 2011 sogar eine Gewinnverdopplung."* (FOCUS Money, 29/12/2009: "Holprige Strecke")

*"**Im Vergleich zu deutschen Banken** wachsen die spanischen Großbanken vor allem durch Zukäufe im Ausland."* (Handelsblatt, 30/12/2005: "Spanische Banken trumpfen auf")

*"Nach einem Zuwachs von 5,6 Prozent bei den realen Auftragseingängen in den Betrieben mit 20 und mehr Beschäftigten im Jahr 2007 konnte der Bereich Wirtschaftsbau im Januar **im Vergleich zu dem entsprechenden Vorjahresmonat** ein Plus von gut 20 Prozent realisieren."* (Institut der deutschen Wirtschaft Köln, 08/07/2008: "Schwächung ohne Absturz: IW-Konjunkturprognose Frühjahr 2008")

*"Zwar ist die Summe der Nettolöhne **im Vergleich zum Vorjahr** im ersten halben Jahr gestiegen, allerdings brachte das dritte Quartal 2010 – **im Vergleich zu den drei Monaten** zuvor – bereits wieder ein Minus von 0,1 Prozent."* (ZEIT online, 02/12/2010: "Glaube statt Fakten")

The information on the usage of constructions is important in both education of linguists and professional practice.

g) Other information

In this section, the compiler of the entry can write notices on various other aspects of the construction which did not fit other sections. For example, the following information can be added to the construction *(ein/der) Zuwachs von X Prozent*:

Sonstige Informationen: Alternativ zum Wort "Prozent" wird in der Konstruktion häufig das Zeichen "%" benutzt.

The following chapter describes the treatment of glossary entries on the virtual learning platform Moodle.

V. MOODLE

1. WHAT IS MOODLE?

Moodle is a learning management system. It facilitates the administration and performance of courses and online trainings. At that, it is possible to add further content to teaching materials. For example, additional materials, tasks, tests and the entire communication can be mediated through the system.

As Moodle is an online application, it has to be installed centrally and made available online by your institution (school, university, company). For further information on Moodle, see http://moodle.org.

If Moodle is already available for you, your Moodle administration should create an empty course for you, in which you have the rights of a teacher (see also "Local allocation of roles in the glossary" in the section "Administration of the glossary").

Moodle can be used in various ways. In a newly created course, various learning activities can be installed, including the learning activity "Glossary". As a learning activity in Moodle, the glossary enables to set up, edit and expand word lists. As soon as you have registered as a teacher, you can create and administer the activity "Glossary" in your course. In the following, you will obtain fundamental information for the creation and use of the learning activity "Glossary".

2. CREATING THE GLOSSARY

When you first enter the course space in Moodle, you will find it "empty" (Fig. 1). All the course activities and resources must first be created.

The Moodle course space consists of **blocks** (the narrow column on the left) and **course segments** (the wide column in the middle). The **blocks** "Navigation" and "Settings" make it possible to navigate in the course.

The **course segments** will later include the concrete study materials of the course. Here resources and activities can be set up. The resources can include various electronic content: files in any format, links, lists, etc. Activities are the learning activities the learners have to perform, e.g. to compile a glossary.
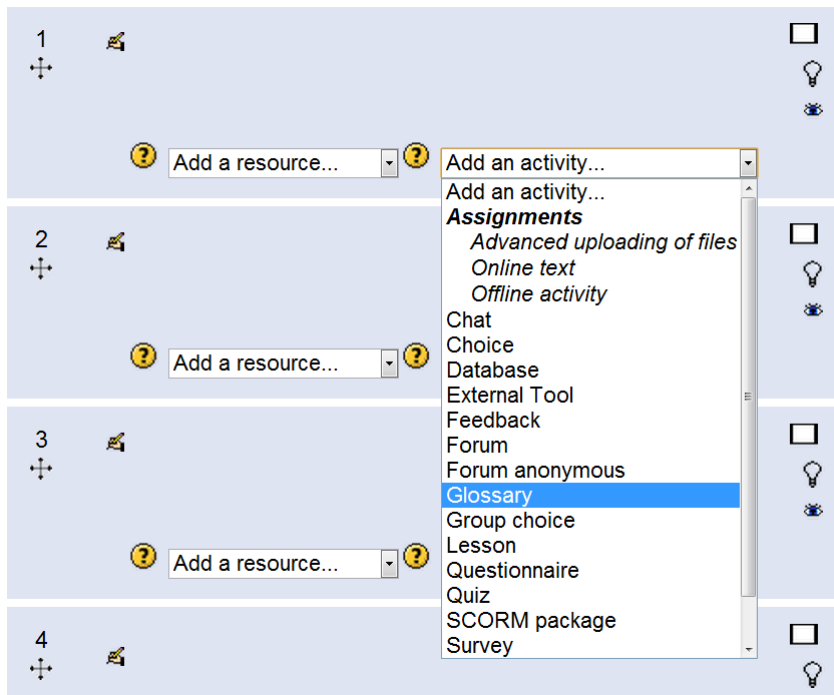
Figure 1: Course space design

To create a glossary in Moodle, take the following steps:
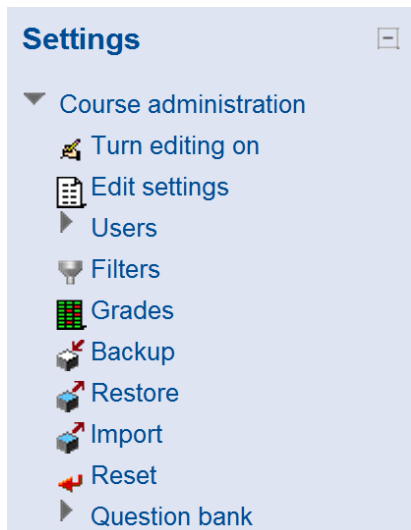
Figure 2

a) In the "Settings" area, find the item "Course administration". Here you can switch to the "Turn editing on" mode (Fig. 2).
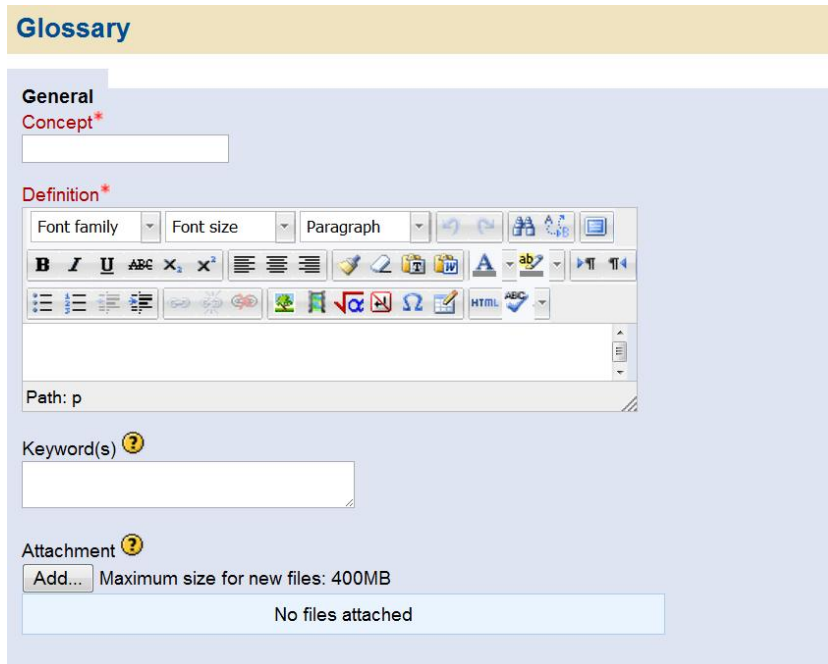
Figure 3: Creating the glossary

b) The symbols that appear (e.g. ✍, ❓, ✕ etc.) lead you directly to the places in the course that can be edited.

c) On the right, in the course segments, two drop-down menus appear: "Add a resource" and "Add an activity". From "Add an activity", select "Glossary" (Fig. 3).

d) As the next step, you can modify the settings of your glossary according to your own preferences, e.g., "Allow comments on entries" in order to write comments on the entries or "Automatically link glossary entries" in order to link one entry to another. By clicking the question marks on the yellow background (❓), you can get additional information on how to use each setting. Give your glossary a name and describe its purpose. By finally saving all your settings, you will get back to your course.

2.1   CREATING A GLOSSARY ENTRY

After you have created a glossary, you can create a glossary entry on the view page of the glossary. Click the button "Add a new entry", and a new window will open, into which you can write your entry (Fig. 4).
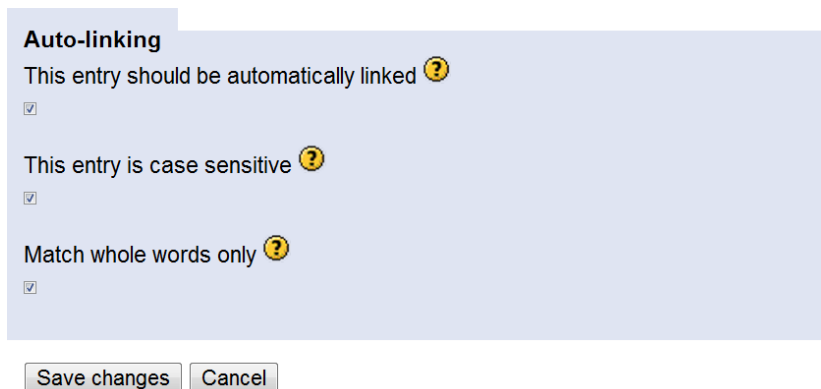

Figure 4

a) Type the desired expression (a word or a multi-word construction) in the "Concept" field.

b) In the text field of the editor, you can fill in the structure of the entry item by item. The information filled in the glossary entry must follow a schema that has preliminarily been agreed on (for details on entry structure, see Chapter IV). This can later be applied as a glossary template. In the editor, several additional text formatting functions are available for you.

c) You can add "Keywords" (synonyms) to each glossary entry. In the case of extensive glossaries, this is a good option for searching for them.

d) To link entries with one another, you should activate the auto-linking function by clicking on "This entry should be automatically linked" (Fig. 5).

Figure 5

e) For each glossary entry, the options "This entry is case sensitive" and "Match whole words only" can be activated under auto-linking.

f) **This entry is case sensitive.** This setting specifies whether exact matching of upper- and lower-case letters is necessary when auto-linking to an entry.

g) **Match whole words only.** If this option is activated, an entry will be linked only when a whole word in the text coincides with the glossary entry. For example, no link is generated between the glossary entry "Konstrukt" and the word "Konstruktivität" in the text.

h) Finally, save the filled entry.

2.2 EDITING THE COMPILED ENTRY

a) The participants may always edit their entries if you have set the "Edit always" option in glossary entries on "Yes".

b) To edit a glossary entry, click on the hand symbol in it.

c) To delete a glossary entry, click on the symbol of the cross.

2.3   WRITING OF COMMENTS

Under the function "Comments", the course participants and teachers can express their opinions about the created entries, make proposals for improvements or exchange experience as can be seen in an example from the German KoGloss glossary (here a professor has wanted to motivate the students and request for corrections):

Samstag, 7. Januar 2012, 19:07

Fantastisch. Nur den Sinn der Angabe unter Syntax/Paradigmen verstehe ich nicht. Können Sie das expliziter formulieren?

Comments encourage cooperation between the participants and thus, add to the quality of the glossary, particularly if the participants exchange comments on one another's work.

To write comments,

a)   click on the link "Comments"

b)   type your text in the text field of the editor and save it.

Earlier comments are visible in the glossary entry with the name of the author and the creation date.

3. USING EXISTING GLOSSARIES

3.1.   VIEWING THE GLOSSARY

An existing glossary can be viewed and sorted alphabetically, by date or by author (Fig. 6).
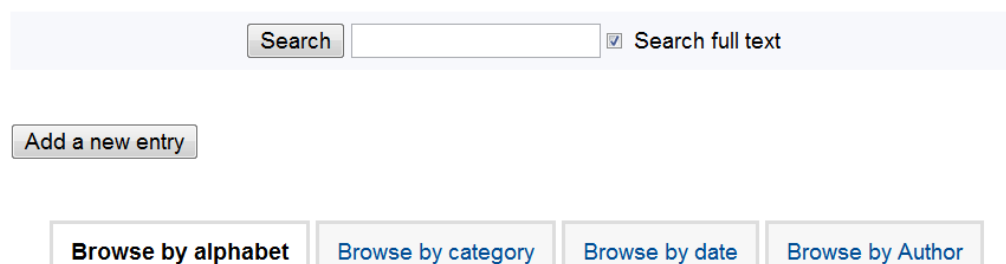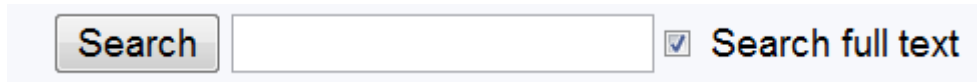
Figure 6

3.2. SEARCHING THE GLOSSARY

To search the glossary, use the field "Search". If you want to search only from among concepts, untick "Search full text" to the right of the search field.

If the "Search full text" option is activated, the desired word will be searched not only from among the concepts but also from the whole text of the glossary.

3.3. IMPORTING/EXPORTING THE GLOSSARY

Moodle enables you to use the compiled glossaries again. You can import glossaries from other courses to your own course or export the glossary of your course.

Exporting a glossary:
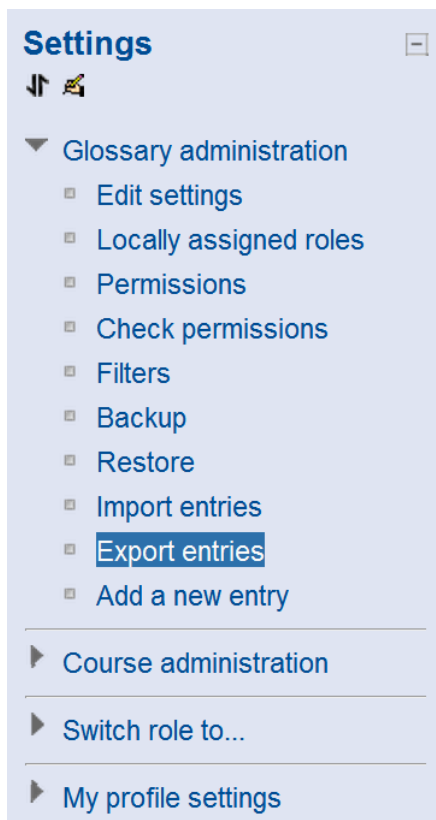
a) open the glossary that you want to export. In the Settings area under "Glossary administration" click on "Export entries" (Fig. 7).

Figure 7

b) click on "Export glossary to file" (Fig. 8). A pop-up window will open. Select "Save" (and, as circumstances require, a place where you save it) and "OK". So the glossary will be saved locally as an XML file. You can always upload this file again using Moodle's import function.

Export entries to XML file

Export entries to file

Figure 8

Importing a glossary

a) In a theme block of your choice, create a new glossary by selecting the option "Glossary" in the "Resources" menu.

b) open the new glossary. In the "Settings" area under "Glossary administration" click on "Import entries".

Import entries from XML file

File to import ❓

Choose a file...

No files attached

Destination of imported entries ❓

Current glossary ▾

Import categories

☐

Submit

Figure 9

c) click on "Select file" (Fig. 9). A pop-up window will open (Fig. 10).

d) In the left-side area of the pop-up window, select "Upload a file" and, by clicking on "Browse", search your computer for the initially saved file (see section "Exporting a glossary", item 2).
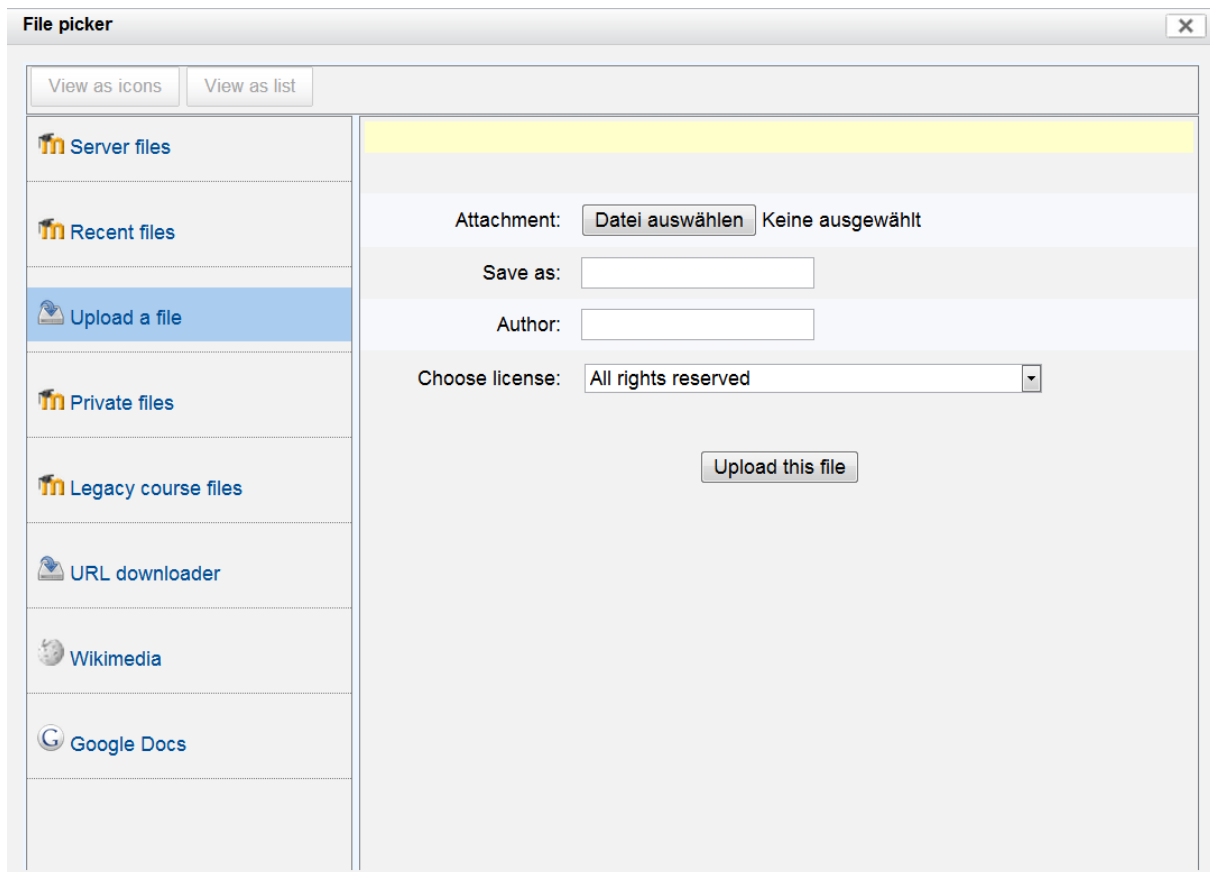
Figure 10

    e) Thereafter select "Upload this file".

    f) Finally, save the uploaded file.

3.4. ADMINISTRATION OF THE GLOSSARY. LOCAL ALLOCATION OF ROLES

How a user can work in Moodle and administer the course depends on the user's role. It is the Moodle administration that determines the roles. If necessary, these can be broadened. In a Moodle course, the following main roles can be assigned.

Teacher

Teachers administer the course and can organise it according to their needs. Teachers can also assume other roles (e.g. of the participants, in order to see the course content from their perspective). Teachers also have the right to assign roles to other participants. For that they have to be logged into the course.

You can assign a role in the following way:

    a) In the settings of the glossary click on "Locally assigned roles" and select a role for allocation, e.g. Teacher (Fig. 11).

Figure 11

  b) Thereafter, two lists will appear: "Existing users" who already have the role of a teacher and "Potential users" to whom the role of the teacher can now be assigned.

  c) from the right-hand list select the potential users. Click the button "Add" By doing so you have assigned the persons selected by you the role of the teacher.

Participant (Student)

Participants can use the course and the activities, resources and blocks it contains, but they are not allowed to make any changes in the settings of the course.

Guest

Guests can see the content of the course, but they are not allowed to use or change it. Guest access must first be activated in the course settings.

VI.   APPLICATION IN TEACHING AND FURTHER EDUCATION

The KoGloss method consists of the following steps:

Corpus building	Searching for construction	Analysing and describing constructions	Commenting
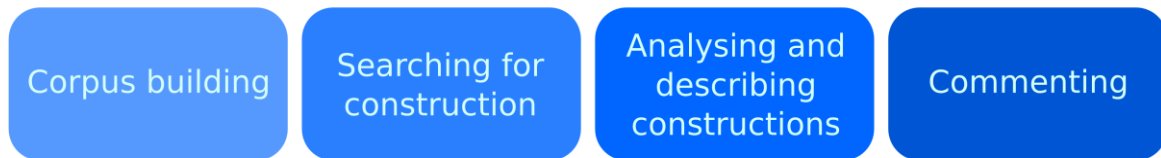
Dependent on the target group, either the whole method (a) or its separate steps (b, c, d, and e) can be applied.

Applicability of the KoGloss method

	Corpus building	Searching for construction	Analysing and describing constructions	Commenting
a)	Corpus building	Searching for construction	Analysing and describing constructions	Commenting
b)	Corpus building	Searching for construction	Analysing and describing constructions	Commenting
c)	Corpus building	Searching for construction	Analysing and describing constructions	Commenting
d)	Corpus building	Searching for construction	Analysing and describing constructions	Commenting
e)	Corpus building	Searching for construction	Analysing and describing constructions	Commenting

For each target group, a combination of steps can be chosen that best meets your needs and facilitates the acquisition of the desired skills and proficiencies.

Applicability	Potential target groups	Competencies achieved
a	students with good previous linguistic knowledge (e.g. students of translation and other linguistic specialities) language	foreign language and/or mother tongue competence language for specific purposes competence general research (methodological) competence text linguistic competence lexicographical competence language didactic competence professional linguistic communication
b	students with previous linguistic knowledge language teachers professional users (translators, interpreters, terminologists, journalists, etc.)	foreign language and/or mother tongue competence language for specific purposes competence text linguistic competence lexicographical competence
c	students with previous linguistic knowledge language learners professional users (journalists, PR officers, secretaries, etc.)	foreign language and/or mother tongue competence language for specific purposes competence professional linguistic communication
d	language learners professional users	foreign language and/or mother tongue competence language for specific purposes competence
e	students with previous linguistic knowledge language learners	foreign language and/or mother tongue competence language for specific purposes competence professional linguistic communication

The principles of corpus building can be explained to all target groups. Corpus building can be particularly recommended to students of translation in order to practise information and text search. Certainly, the text corpus can also be presented as a ready-made product, primarily in the case of short courses or projects.

The possible constructions can also be presented by teachers, which means that the learners work only on their analysis/description and comment on ready-made entries (e).

The KoGloss method can also be successfully used individually, e.g. for independent research by translators/interpreters and for creation of special information sources that can be used later. For that, one could build his/her own corpus (b) or use an existing one (d). If the method is used in groups, unified formatting of glossary entries should be agreed upon at the beginning so that the learners could work on them independently. If needed, the necessary linguistic terminology can be taught/revised.

The final product, i.e. the glossaries compiled, can be used in first and second language acquisition, for text production in the mother tongue or in a foreign language, in translation exercises, teaching of terminology, in courses for development of research competence, etc. The constructions in one or several languages can be studied in homework assignments, term and bachelor's papers addressing the themes of translation.

In professional practice, the constructions glossaries can be used according to professional necessity and linguistic competence. First of all, they can be used at production of specialised texts. According to the speciality, the common constructions in specialist discourse can be acquired in advance and used for production of text sorts characteristic of the speciality. Such constructions glossaries in language for specific purposes can be particularly helpful for users at the beginning of their careers for whom written communication is inevitable in their everyday work (secretaries, specialists). Here, glossaries in both the mother tongue and foreign languages (for example, for communication with foreign business partners) are relevant.

Another target group who would need to use constructions glossaries in their professional praxis are translators and interpreters. The common lexis of a particular speciality can be collected into a glossary in the form of full constructions as they are used in the discourse of the speciality. Through such an approach, primarily interpreters can acquire and automatise not only separate technical terms but also the commonly used phrases of a particular speciality. In this case, too, the constructions glossaries can be helpful in both the mother tongue and foreign languages.

Glossar

Korpusaufbau

language learning

Übersetzen

Textkorpus

Konkordanzanalyse

acquisition

glossary

technical translation

linguistic analysis

KoGloss

translation

concordance analysis

concordance

Konkordanz

translation didactics

corpus building

Korpusanalyse

corpus analysis

language

linguistische Analyse

lexicography e-learning

AntConc

kollaborativ

collaborative

Konstruktionen

E-learning

Sprachlernmethode

Glossarerstellung

Übersetzungsdidaktik

Konkordanz

language learning method

moodle

Lexikografie

Sprachen lernen

LSP

text corpus

glossary creation

Fachsprache

Fachübersetzung

constructions

Spracherwerb

Sprache

language for special purposes