

Korpusbasierte Sprachreflexion mit Online-Ressourcen

Heike Zinsmeister
Institut für Maschinelle Sprachverarbeitung Universität
Stuttgart

SPONSORED BY THE



Federal Ministry
of Education
and Research

CLARIN-D

Common **L**anguage **R**esources and Technology
Infrastructure

Katalogisierung

Rechtliche und
ethische Fragen

Webservices

Schulung und
Ausbildung



CLARIN-D Zentren



<http://de.clarin.eu>

Korpusbasierte Sprachreflexion

- **Qualitative Aspekte**
 - natürlich klingende Beispiele
 - Wort-/Bedeutungsfelder (auch Polysemie)
- **Quantitative Aspekte**
 - Gängige Wortkombinationen (Kollokationen)
 - Vergleiche von alternativen Realisierungen
 - Diachroner, genreabhängiger etc. Gebrauch

Ziele des Tutorium

- Online-Korpora als Quelle für Sprachmaterial nutzen
- Grundeigenschaften von linguistischen Korpora verstehen
- Zwei Online-Korpussammlungen abfragen
- Eigenes Textmaterial online linguistisch aufbereiten und lokal sichten

Korpus-Schnittstellen und Programme

- Projekt Digitales Wörterbuch der deutschen Sprache (DWDS) der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW)



- Institut für deutsche Sprache (IDS) in Mannheim



- Webservices von CLARIN-D



- Lokale Datensichtung mit TIGERSearch





CLARIN-D

Anmeldungen

Siehe auch:
<http://ling.uni-konstanz.de/pages/home/zinsmeister/TutoriumParis/>



<http://www.dwds.de/user/register/>



COSMAS II

<http://www.ids-mannheim.de/cosmas2/projekt/registrierung/>



WEBLICHT
WEB-BASED
LINGUISTIC CHAINING TOOL

<https://weblicht.sfs.uni-tuebingen.de/> (Zugangsdaten im Kurs)

Was ist ein linguistisches Korpus?

- Eine Sammlung **gesprochener** oder **geschriebener**
- Typischerweise digitalisiert und **maschinenlesbar**
- Ebenen eines Korpus
 - Sprachdaten
 - Metadaten
 - Annotationen

Ebenen eines Korpus

Rohdaten/Primärdaten

- die "authentischen" Sprachdaten
- Video-/Audiodateien, Transkriptionen von gesprochener Sprache, Text

Metadaten

- "Daten über Daten"
- Informationen über Sprache, Datenumfang, Datenformat, Sprecher, Autoren, Erstellungszeit, Annotatoren, Annotationstagssets, ...

Annotationen

- linguistische und außerlinguistische Interpretationen
- Markierung von Wortgrenzen, Wortart, syntaktischen Phrasen und Funktionen, Zeitrelationen, Diskursrelationen, Mimik, Torstand...

Primärdaten: Stein von Rosetta





Übersetzung:

"an offering that the king gives; and Osiris, Foremost of Westerners [i.e., the Dead], the Great God, Lord of Abydos; and Wepwawet, Lord of the Sacred Land [i.e., the Necropolis]." (Allen 2000:§24.10)

Transliteration

Htp-di-nswt wsir xnty imntiw nTr aA nb AbDw
wp-wAwt nb tA Dsr (Buurman, Grimal, et al. 1988)

Terminabsprache im Verbmobilkorpus

-  also wir k"onnten <Ger"ausch> entweder uns heute noch im Laufe des Tages bis zwei Uhr treffen , <Schmatzen> <A> oder <Ger"ausch> am Dienstag oder Mittwoch <Ger"ausch> h"att' ich den ganzen Tag Zeit , und man <!2 u'ma> m"u"ste sich dann einfach einigen auf einen <!1 ein'> Termin .

(XXX014 von <http://www.phonetik.uni-muenchen.de/forschung/Verbmobil/VerbDialog.html>)

Genehmigung des Protokolls

Online
abfragbar:
[http://
opus.lingfil
.uu.se/](http://opus.lingfil.uu.se/)

Das Protokoll der gestrigen Sitzung wurde verteilt.

Rohdaten

Wenn keine Einwände, werden die redigierte Transkriptionen von Seminardebatten des Europäischen Parlaments und Übersetzungen in 11 offizielle EU-Sprachen genehmigt.

Herr Präsident, wir haben über Nacht die Meldung erhalten, daß sich Ibrahim Rugova jetzt in Rom befindet, und vielleicht könnten Sie den Parlamentspräsidenten bitten, Herrn Rugova und seiner Familie im Namen des Parlaments unsere besten Wünsche zu übermitteln.

Sie werden sich erinnern, daß wir im April-Plenum in diesem Haus eine Entschließung verabschiedeten, in der wir Herrn Rugova einluden, vor dem Ausschuß für auswärtige Angelegenheiten zu sprechen.

(ep-99-05-06.txt)

Metadaten

- Datum der Debatte
- Sprechername
- teilweise Sprache des Sprechers = Originalsprache des Beitrags ('turn')

Genehmigung des Protokolls

<SPEAKER ID=1 NAME="Der Präsident">

Das Protokoll der gestrigen Sitzung wurde verteilt.

Wenn keine Einwände vorgebracht werden, betrachte ich es als genehmigt.

<SPEAKER ID=2 LANGUAGE="EN" NAME="Spencer">

Herr Präsident, wir haben über Nacht die Meldung erhalten, daß sich Ibrahim Rugova jetzt in Rom befindet, und vielleicht könnten Sie den Parlamentspräsidenten bitten, Herrn Rugova und seiner Familie im Namen des Parlaments unsere besten Wünsche zu übermitteln.

Sie werden sich erinnern, daß wir im April-Plenum in diesem Haus eine EntschlieÙung verabschiedeten, in der wir Herrn Rugova einladen, vor dem Ausschuß für auswärtige Angelegenheiten zu sprechen.

(ep-99-05-06.txt)

Beispiel: Europarl-Korpus



Terminologie

- Token
- Wortart 'part of speech'
- Wortartenlabel 'pos tag'
- Definiertes Inventar 'tag set'
- Quasi-Standard zur Beschreibung der Wortarten des Deutschen: STTS (Stuttgart-Tübingen-TagSet)

```
<words>
<word id="word_1" pos="NN">Genehmigung</word>
<word id="word_2" pos="ART">des</word>
<word id="word_3" pos="NN">Protokolls</word>
<word id="word_4" pos="ART">Das</word>
<word id="word_5" pos="NN">Protokoll</word>
<word id="word_6" pos="ART">der</word>
<word id="word_7" pos="ADJA">gestrigen</word>
```

Annotationen

- Satzsegmentierung
- Wortsegmentierung
- Wortart

```
ord>
/word>
/word>
<word id="word_9" pos="$.>.</word>
```

(ep-99-05-06.txt)

D W D S

heike.zinsmeister abmelden | Profil Zur alten DWDS-Webseite

D W D S Hilfe

DWDS Suche **Ressourcen** Erschließung Projekt Aktuelles

Ägide Suche (in meinem Korpus) Mein Korpus Suchhistorie Hilfe anzeigen

Aktuelle Sicht: **DWDS Standardsicht**

DWDS-Wörterbuch Detailansicht

Ägide
 fem.; -, ohne Plural
 Gebrauch: **gehoben**
 Herkunft: Griechisch
 unter jmds. Ä. ♦ unter jmds. Leitung, Obhut ↓
 dwdswb-0.3.9 Eintrag | Zusammensetzungen | Beleg | Beispiele

OpenThesaurus OpenThesaurus Webseite

Synonymgruppen für **Ägide**
 Leitung, Obhut, Präsidium, Schirmherrschaft, **Ägide**
 OpenThesaurus 2011-09-09

DWDS-Wortprofil 2010 Relationenfilter

Statistisches Wortprofil für **Ägide**
Nationen Präsidenten neuen stell
 Wortart: Substantiv Zeige Tags | Tabellenansicht

Etymologisches Wörterbuch des Deutschen (nach Pfeiler) Kein Eintrag vorhanden

DWDS-Kernkorpus Treffer: 55

1 ...ypertrophierung unter der **Ägide** Honecker erfahren hat «, ...
 2 ...taatsdirigismus unter der **Ägide** ihrer eigenen Apparate, w...
 3 ... deutsche Kino" unter der **Ägide** der sozialliberalen Koali...
 4 ...usgegeben. - Unter der **Ägide** des so verdammten Vaters,...
 5 ...n Opernbühne unter seiner **Ägide** zu einer der besten in De...
 6 ...kvermögen. Unter ihrer **Ägide** bekam die Partei so etwas...
 7 ... Umgestaltung « unter der **Ägide** der SED. Während sich ...
 8 ...den. ZEIT: Unter Ihrer **Ägide** als Mercedes-Chef wurden ...
 9 ...unter der durchgreifenden **Ägide** Meisners nacheinander der...
 10 ...ge Reformen, unter dessen **Ägide** die meisten Preise freige...
 55 von 63 Treffern anzeigbar DDC-Query | Darstellung | Suchfilter

Die ZEIT & ZEIT Online Treffer: 1045

Im Panel-Menü unten rechts:

DWDS-Kernkorpus

Darstellungsoptionen:
 Änderungen an Darstellungsoptionen können sich auf die Anzahl und Reihenfolge der Treffer auswirken.

Sortierung nach: Datum absteigend

nur einen Treffer pro Dokument anzeigen:

Volltextanzeige statt Keywords in context (KWIC):

Darstellung mit manuell definierten Optionen

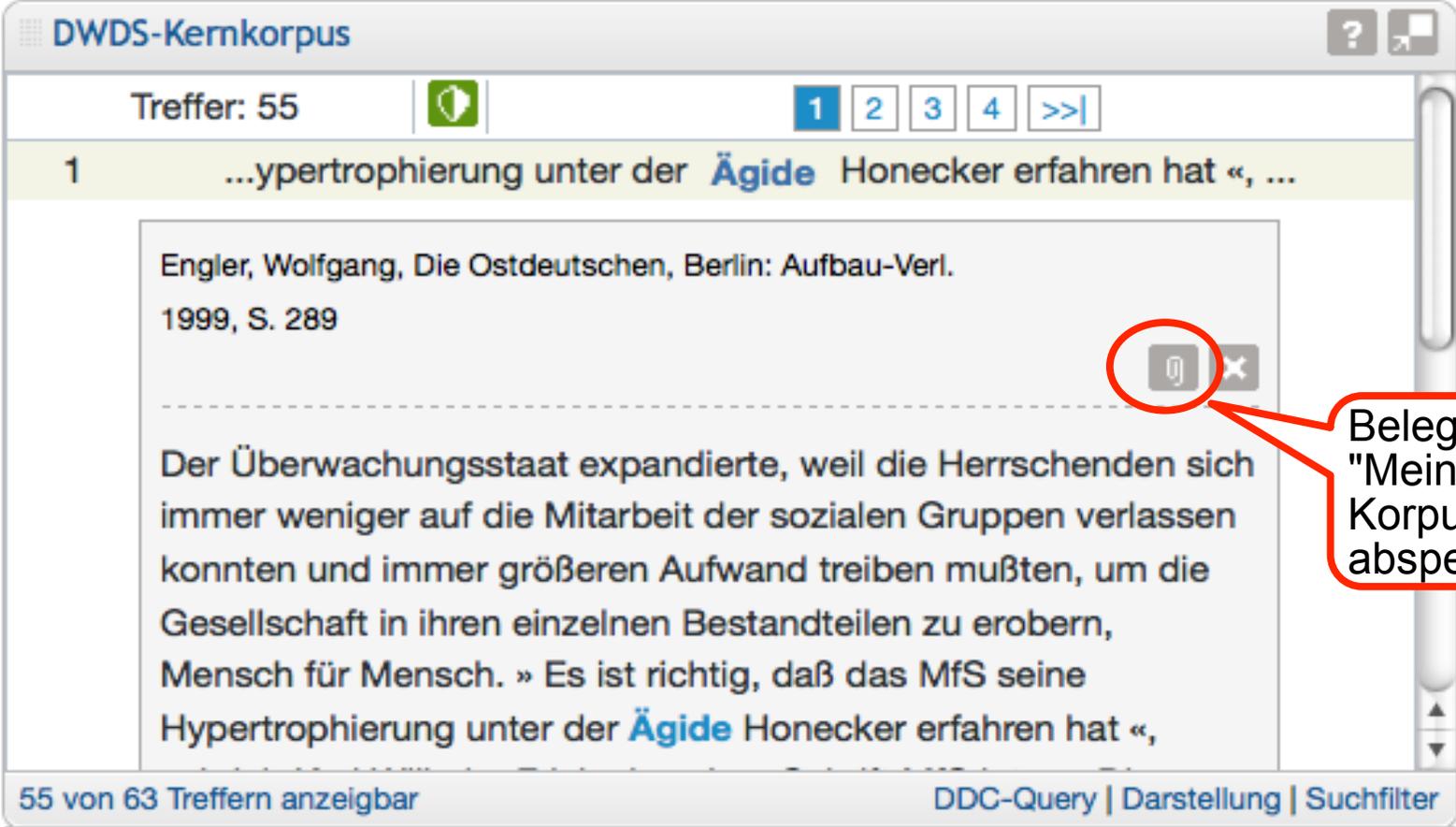
DWDS-Kernkorpus

Treffer: 5 1 2 3 4 >>|

1	...ybertrophierung unter der Ägide Honecker erfahren hat «, ...
2	...taatsdirigismus unter der Ägide ihrer eigenen Apparate, w...
3	... deutsche Kino" unter der Ägide der sozialliberalen Koali...
4	...usgegeben. - Unter der Ägide des so verdammten Vaters,...
5	...n Opernbühne unter seiner Ägide zu einer der besten in De...
6	...kvermögen. Unter ihrer Ägide bekam die Partei so etwas...
7	... Umgestaltung « unter der Ägide der SED. Während sich ...
8	...den. ZEIT: Unter Ihrer Ägide als Mercedes-Chef wurden ...
9	...unter der durchgreifenden Ägide Meisners nacheinander der...
10	...ge Reformier, unter dessen Ägide die meisten Preise freige...

55 von 63 Treffern anzeigbar DDC-Query | Darstellung | Suchfilter

Klick auf Belegwort im KWIC: Kontextanzeige



DWDS-Kernkorpus

Treffer: 55

1 2 3 4 >>|

1 ...ypertrophierung unter der **Ägide** Honecker erfahren hat «, ...

Engler, Wolfgang, Die Ostdeutschen, Berlin: Aufbau-Verl.
1999, S. 289

Der Überwachungsstaat expandierte, weil die Herrschenden sich immer weniger auf die Mitarbeit der sozialen Gruppen verlassen konnten und immer größeren Aufwand treiben mußten, um die Gesellschaft in ihren einzelnen Bestandteilen zu erobern, Mensch für Mensch. » Es ist richtig, daß das MfS seine Hypertrophierung unter der **Ägide** Honecker erfahren hat «,

55 von 63 Treffern anzeigbar

DDC-Query | Darstellung | Suchfilter

Beleg in
"Meinem
Korpus"
abspeichern

DWDS-Wortprofil 2010

Statistisches Wortprofil für **Ägide**

Nationen Präsidenten

neuen steht_unter

Wortart: **Substantiv** Zeige Tags **Tabellenansicht**

Relationenfilter

DWDS-Wortprofil 2010

Relationenfilter

Stammform	Wortart	syntaktische Relation	Saliency	Frequenz
Präsident	Substantiv	Präsidenten ist Genitivattribut von Ägide	3.68	5
Nation	Substantiv	Nationen ist Genitivattribut von Ägide	3.35	5
neu	Adjektiv	neuen ist Attribut von Ägide	1.19	7
stehen_unter	PP	steht hat Präpositionalphrase unter_Ägide	0.01	4

Wortart: **Substantiv** Wolkenansicht



Zusätzliche Panels aktivieren:

The screenshot shows a web browser window with the URL www.dwds.de/?qu=Ägide&view=1. The page displays the DWDS interface for the word 'Ägide'. A dialog box titled 'Panel hinzufügen' is open, providing instructions: 'Fügen Sie ein Panel ihrer View hinzu. Klicken Sie hierfür auf eine Kategorie aus der das Panel stammt. Sie können eine Kategorie wieder schließen, in dem Sie wieder auf den Kategoriebalken klicken.' Below the instructions are three buttons: 'Korpus hinzufügen', 'Wörterbücher hinzufügen', and 'Statistik hinzufügen'. In the left sidebar, the 'Panel' button is circled in red.

Zusätzliches Panel: Korpusfrequenzen

Der aktuellen Sicht ein neues Panel hinzufügen.

Panel hinzufügen

Fügen Sie ein Panel Ihrer View hinzu. Klicken Sie hierfür auf eine Kategorie aus der das Panel stammt. Sie können eine Kategorie wieder schließen, indem Sie wieder auf den Kategoriebalken klicken.

Korpus hinzufügen

Wörterbücher hinzufügen

Statistik hinzufügen

Statistikpanel hinzufügen

Kernkorpus - Wortverlaufsstatistik

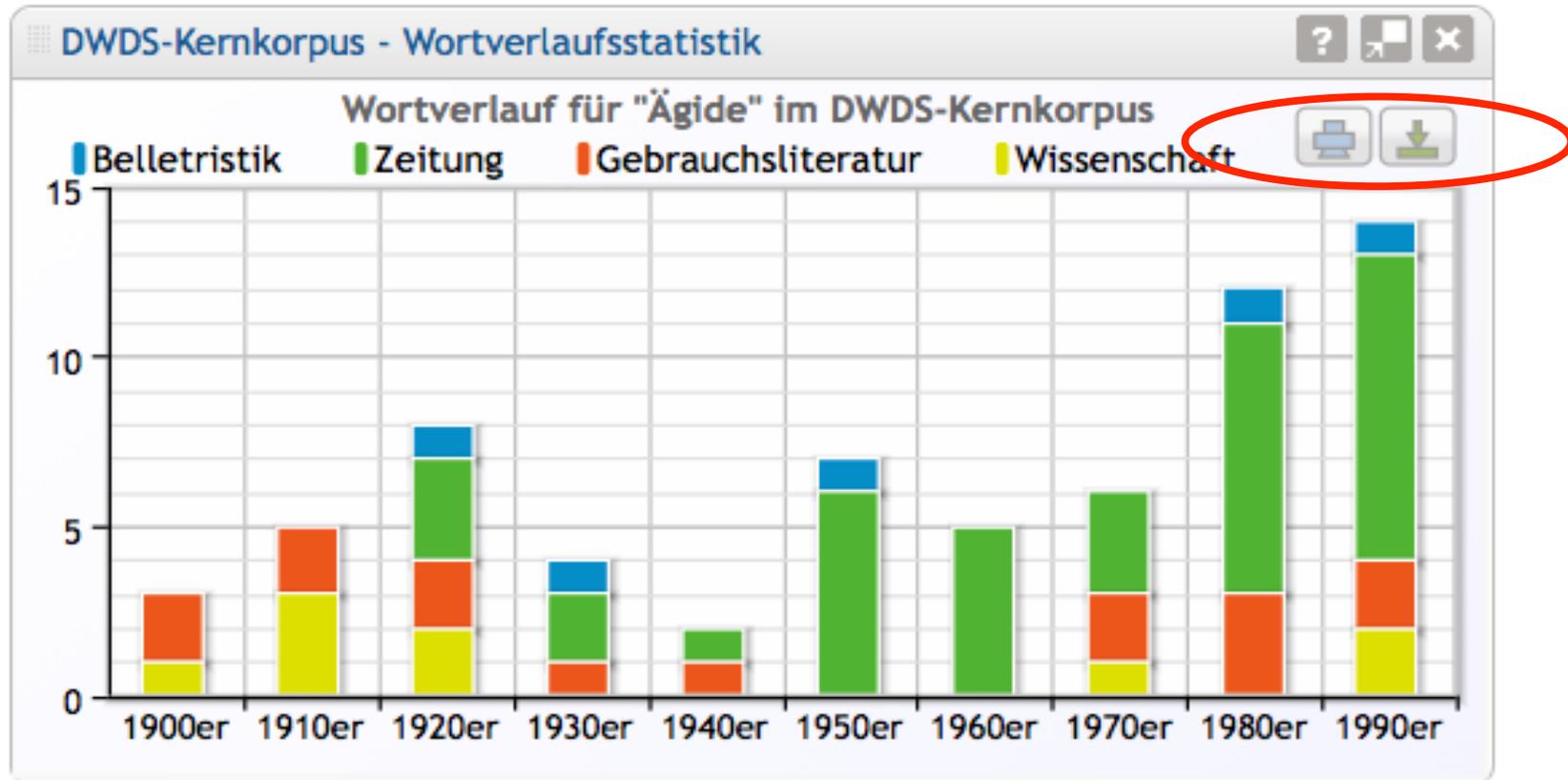
DWDS-Wortprofil 2010

Korpusfrequenzen

Vorschau des Panel (wird automatisch geladen, wenn man mit der Maus über den Button fährt)

Korpus	Hits	Hits [ppm]	Korpusgröße [Mill. Token]
Berliner Tagesspiegel	221217	1339.1865	165.19
Berliner Zeitung	278639	1099.7457	253.37
C4-Korpus	31752	396.9000	80.00
Compact Memory Corpus	20133	767.0477	26.25
DDR-Korpus	5502	635.5810	8.66
Die ZEIT & ZEIT Online	579626	1260.0565	460.00
DWDS-Kernkorpus	98543	979.5430	100.60
DWDS-Korpus21	2579	1380.1526	1.87
Juillard-Korpus	513	1026.0000	0.50
Potsdamer NN	19993	1266.0766	15.79

Zusätzliches Panel: Wortverlaufstatistik



"Mouse-over"-Effekte



Exportmöglichkeiten

- in der aktuellen Version
 - nur Copy & Paste der Einzelbelege
- in der alten Version (Link oben rechts)
 - Export in Textformat
- Ausblick
 - Exportfunktion wird in neuem Release wieder angeboten



COSMAS II



**COSMAS II - Übersicht****COSMAS II_{web}**

Online-Hilfe

Entwicklung

COSMAS II_{web} starten

Browserunterstützung

[Seite drucken](#)[Thema drucken](#)[Sitemap](#)

Die Applikation COSMAS II_{web}

Übersicht

COSMAS II_{web} ist eine betriebssystemunabhängige WWW-Applikation, die die Ko-
WINDOWS-Betriebssystem gebunden zu sein oder COSMAS II_{win} installieren zu

COSMAS II_{web} verfügt - mit Ausnahme der grafischen Suchanfragekomponente -

Aktuelle Version: 1.7

 Applikation **starten**.

Detaillierte Informationen

- Freigabeinformationen: **aktuelle Version**, ältere Versionen
- Liste der **unterstützten Browser**



Einstieg **Anmeldung** 1. Recherche Optionen

Portal COSMAS II Liebe Benutzer des COSMAS II-Webdienstes,

Online-Hilfe Einstieg Anmeldung Recherche Optionen

- Forum Login 2.
- Kontakt Logout
- Browserunterstützung Registrieren
- Passwort vergessen ?

Benutzungshinweise

Die Benutzung von COSMAS II ist kostenlos und nur für nicht-kommerzielle Zwecke erlaubt. Beacht "Registrieren" die **Nutzungsbeschränkungen**.

Einstieg Abmeldung **Recherche** 3. Optionen Hilfe

- Archiv
- Suchanfrage
- Wortformliste
- Ergebnisse
- Kookkurrenzanalyse
- Gesamt-KWIC
- Gesamt-Volltext
- Export

Recherche

Das normale Vorgehensmodell bei einer Recherche ähnelt einem Wasserfallmodell. Die einzelnen Schritte sind unten dargestellt. Selbstverständlich können sie im Modell zurückspringen. Änderungen in voran gestellten Aktionen wirken sich aber auf nachstehende aus. So löscht die Auswahl eines neuen Archivs beispielsweise die frühere Auswahl eines Korpus, sodass sie ein neues Korpus wählen müssen.

1. Archiv wählen
2. Korpus wählen
3. Anfrage formulieren
4. Wortformliste bearbeiten
5. Ergebnisanzeige
6. KWIC
7. Export von Ergebnissen





Einstieg

Abmeldung

Recherche

Optionen

Hilfe

Suche

Lemmatisierung

Korpuspräsentation

Ergebnispräsentation

KWIC/Volltext

Newsletter

Die Einstellungen, die Sie in COSMAS II vornehmen können, sind in unterschiedliche Kategorien gegliedert:

- **Suche**
Damit können Sie darauf Einfluss nehmen, ob die Groß-/Kleinschreibung berücksichtigt werden soll, ob Expansionslisten (Wortformlisten) beim Einsatz von Platzhalter- und Lemmatisierungsoperatoren erzeugt und wie sie sortiert werden sollen; ob die Treffer mit einem Zufallsgenerator reduziert werden sollen.
- **Lemmatisierung**
Damit können Sie Einfluss nehmen auf die Arbeitsweise des Lemmatisierungsoperators &. Die Flexionsformen bleiben immer eingeschaltet.
- **Korpuspräsentation**
Mit diesen Einstellungen können Sie Einfluss auf die Sortierung der Korpuspräsentation nehmen.
- **Ergebnispräsentation**
Mit diesen Einstellungen können Sie Einfluss auf die Sortierung und die Darstellung der Ergebnispräsentation nehmen.
- **KWIC/Volltext**
Mit den KWIC- und Volltextoptionen können Sie Einfluss auf die Anzeige und die alphabetische Sortierung des KWICs und Volltexts nehmen.
- **Newsletter**
Wenn Sie den COSMAS II-Newsletter abonnieren, erhalten Sie eine Benachrichtigung, wenn eine neue Version von COSMAS II erschienen ist, die IDS-Korpora ergänzt wurden, etc.

Ihre Einstellungen werden beim Abmelden vom COSMAS II-Service **sitzungsübergreifend gespeichert.**



Einstieg

Abmeldung

Recherche

Optionen

Archiv

Korpus

Suchanfrage

Wortformliste

Ergebnisse

Kookkurrenzanalyse

Gesamt-KWIC

Gesamt-Volltext

Export

**Aktuelles
Archiv**

Aktuelles Korpus

Suchanfrage

Archivauswahl

Öffnen Sie bitte eines der folgenden Archive:

W - Archiv der geschriebenen Sprache

Alte Version von "W - Archiv der geschriebenen Sprache" (verfügbar bis 13.05.2012)

W-ÜBRIG - Archiv der aussortierten geschriebenen Korpora

TAGGED-C - Archiv morphosyntakt. annotierter Korpora (CONNEXOR)

TAGGED-T - Archiv morphosyntakt. annotierter Korpora (TreeTagger)

TAGGED-M - Archiv morphosyntakt. annotierter Korpora (MECOLB; ehem. TAGGED)

HIST - Archiv der historischen Korpora

GFDS - Kartei der Gesellschaft für deutsche Sprache

WK-PH - Archiv der phasengegliederten Wendekorpora



Informationen zu den Archiven finden Sie unter [Organisation des Textmaterials](#).



Einstieg

Abmeldung

Recherche

Optionen

Archiv

Korpus

→ vordefiniert

benutzerdefiniert

Korpusverwaltung

Suchanfrage

Wortformliste

Ergebnisse

Kookkurrenzanalyse

Gesamt-KWIC

Gesamt-Volltext

Export

**Aktuelles
Archiv**

TAGGED-T – Archiv morphosyntakt. annotierter Korpora (T

Aktuelles Korpus

-

Suchanfrage

-

Korpusauswahl

Laden Sie bitte eines der folgenden virtuellen Korpora:

Vordefinierte Korpora

TAGGED-T-öffentlich - alle öffentlichen Korpora des Archivs TAGGED-T

brz - Braunschweiger Zeitung, September 2005 - Dezember 2009

bvz - Burgenländische Volkszeitung, Januar 2007 - Dezember 2009

haz - Hannoversche Allgemeine, August 2007 - Dezember 2009

hmp - Hamburger Morgenpost, Januar 2006 - Dezember 2009

mm - Mannheimer Morgen, Januar 2006 - Dezember 2009

non - Niederösterreichische Nachrichten, Januar 2007 - Dezember 2009





Einstieg **Abmeldung** **Recherche** **Optionen** **Hilfe**

- Archiv
- Korpus
- Suchanfrage**
- Wortformliste
- Ergebnisse
- Kookurrenzanalyse
- Gesamt-KWIC
- Gesamt-Volltext
- Export

Aktuelles Archiv TAGGED-T – Archiv morphosyntakt. annotiert **Aktuelles Korpus** hmp – Hamburger Morgenpost, Januar 🔍

Suchanfrage &licht

Suche formulieren **Ergebnisse:**

Eingabe:
&licht

Weggelassener Verknüpfungoperator bedeutet:

Wortabstand /+w1 logisches "ODER"

Beispiele:
direkt nacheinander [Hilfe zu den Suchanfragebeispielen](#)

MORPH-Assistent **Suchen** **Suchanfragesyntax**



Alle aktivieren Alle deaktivieren Liste zu: &licht, 12 Wortformen, alfab. aufsteigend Öffnen

Ergebnisse

Wortformenliste: Deaktivierung von irrelevanten Formen

Alle aktivieren	Alle deaktivieren	Liste zu: &licht, 12 Wortformen, alfab. aufsteigend	Schließen
<input type="checkbox"/>	<input type="checkbox"/>	LICHT	16
<input type="checkbox"/>	<input type="checkbox"/>	Licht	2215
<input checked="" type="checkbox"/>	<input type="checkbox"/>	licht	7
<input type="checkbox"/>	<input type="checkbox"/>	Lichte	28
<input checked="" type="checkbox"/>	<input type="checkbox"/>	lichte	13
<input type="checkbox"/>	<input type="checkbox"/>	Lichten	1
<input checked="" type="checkbox"/>	<input type="checkbox"/>	lichten	27
<input type="checkbox"/>	<input type="checkbox"/>	LICHTER	7
<input type="checkbox"/>	<input type="checkbox"/>	Lichter	268
<input checked="" type="checkbox"/>	<input type="checkbox"/>	lichter	5
<input type="checkbox"/>	<input type="checkbox"/>	Lichtes	6
<input checked="" type="checkbox"/>	<input type="checkbox"/>	lichtes	3

Ergebnisse



Einstieg Abmeldung Recherche Optionen Hilfe

Archiv

Korpus

Suchanfrage

Wortformliste

Ergebnisse

Quellenansicht
→ **Korpusansicht**
Dokumentansicht
Ansicht vor/seit
Zeitpunkt
Jahrzehntansicht
Jahresansicht
Monatsansicht
Tagesansicht
Länderansicht
Stat. KWIC-Ausw.

Aktuelles Archiv TAGGED-T - Archiv morphosyntakt. annotiert **Aktuelles Korpus** hmp - Hamburger Morgenpost, Januar

Suchanfrage &licht

Korpusansicht **Ergebnisse: 55**

Gesamt-KWIC Gesamt-Volltext Export Kookkurrenzanalyse

Treffer **Texte** **von** **bis** **Korpus** ▲

+	17	17	2006	2006	HMP06 Hamburger Morgenpost 2006
+	19	19	2007	2007	HMP07 Hamburger Morgenpost 2007
+	7	7	2008	2008	HMP08 Hamburger Morgenpost 2008
+	12	11	2009	2009	HMP09 Hamburger Morgenpost 2009
	55	54	2006	2009	4 Korpora

Gesamt-KWIC **Gesamt-volltext** **Export** **Kookkurrenzanalyse**

Auswahl aktiv. Alle aktiv. Alle deaktiv. |< << Seite 1 von 1 >> >| Volltext

1	<input type="checkbox"/>	HMP09/FEB.01071	"Die Augenpartie wird mit einem lichten Zitruston in Kombination mit K
2	<input type="checkbox"/>	HMP09/FEB.01071	Sie einen Lidschatten in einem lichten Sommerton (Nr. 222).
3	<input type="checkbox"/>	HMP09/FEB.01325	bei Numerus-Clausus Fächern zu lichten .
4	<input type="checkbox"/>	HMP09/MAR.03275	Das Angebot ist groß: Unter dem lichten Glasdach der ehemaligen Maschi
5	<input type="checkbox"/>	HMP09/APR.00258	uchtungskonzepte kann unter www.licht.de
6	<input type="checkbox"/>	HMP09/JUN.01711	ist dank üppiger Verglasung so licht und luftig wie in kau
7	<input type="checkbox"/>	HMP09/JUL.02943	als Rezitatoren, möchte er ein lichtes , poetisches Gesam

Gesamt-KWIC
→ **unsortiert**
alphabetisch
chronologisch
zufällige Sort.
nach Kook.

Gesamt-volltext

Export





WARNUNG:

Die statistische Kookkurrenzanalyse einer geringen Anzahl von Belegen (hier: 55) führt nur selten zu wirklich aussagekräftigen Ergebnissen.

Kookkurrenzanalyse

Einstellungen

Kontext: Wörter links höchstens 1 Satz
 Wörter rechts Autofokus

Granularität:

Zuverlässigkeit:

Clusterzuordnung:

Lemmatisierung verwenden:

Funktionswörter ignorieren:

LLR-Wert anzeigen:

Nummerierung des Hauptkookkurrenzators:

Zurücksetzen Starten Hilfe

#	LLR	kumul.	Häufig	Kookkurrenzen	syntagmatische Muster
+	1	46	3	3 Schilderwald	100% den Schilderwald [...] lichten
+	2	34	6	3 Anker	100% Anker lichten
+	3	23	9	3 Szenen	100% lichte [...] Szenen
+	0	55	46	statistisch unspezifisch	



Einstieg Abmeldung Recherche Optionen Hilfe

Archiv
Korpus
Suchanfrage
Wortformliste
Ergebnisse
Kookkurrenzanalyse
Gesamt-KWIC
Gesamt-Volltext
Export

Aktuelles Archiv: TAGGED-T – Archiv morphosyntakt. annotier Aktuelles Korpus: hmp – Hamburger Morgenpost, Januar

Suchanfrage: &licht

Exportkonfiguration Ergebnisse: 55

Gesamt-KWIC Gesamt-Volltext Export Kookkurrenzanalyse

Allgemeines

Dateiname: Ergebnisse_licht Schriftgröße: 8

Bitte keine Umlaute, Sonderzeichen oder Leerzeichen im Dateinamen verwenden.

Dateiformat: ASCII RTF

Orientierung: Querformat Hochformat

Ergebnisübersicht Korpusansicht

KWIC

vorher 0 Sätze

nachher 0

KWIC / Volltext

Sortierung: unsortiert

Nur aktive Treffer





Karin <<

type: text/plain
lang: de

Preview:
Karin fliegt nach New York. Sie will dort Urlaub machen.

Chain 1 × +

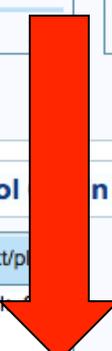
Show tools with status: dev production **Run Tools** **Clear Chain** **Clear Results**

Next Choices (Double-click on an icon to add it to the chain)

SfS To TCF Converter version: 0.3 type: text/tcf+xml lang: de text	BBAW Plaintext Converter type: text/tcf+xml version: 0.3 lang: de text
---	---

Current Tool Chain (These tools are called when the Run button is clicked)

Input text/plain Karin fliegt nach New York. will dort Urlaub machen.
--



WebLicht ▾ File ▾ View ▾ Help ▾ Feedback

Karin

type: text/p
lang: de

tools with status: dev production Run Tools Clear Chain Clear Results

Next Choices (Double-click on an icon to add it to the chain)

Sfs To TCF Converter version: 0.3 type: text/tcf+xml lang: de text	BBAW Plaintext Converter type: text/tcf+xml version: 0.3 lang: de text
---	---

Current Tool Chain (These tools are called when the Run button is clicked)

Input text/plain Karin fliegt nach New York. Sie will dort Urlaub machen.

Note: In the original image, the 'File' menu and its sub-items 'New', 'Open Recent', 'Open Sample', and 'Upload' are circled in red. The 'New' sub-menu is also open, showing 'text/plain', 'text/query-cqp', and 'text/query-ddc'.

TreeTagger

Desc: Italian,English,French,German part-of-speech tagger
lemmatiser

Creator: IMS: University of Stuttgart

Contact: Helmut.Schmid@ims.uni-stuttgart.de

URL: http://gelbaugenpinguin.ims.uni-stuttgart.de/cgi-bin/tree-tagger4.perl

ID: 227

I/O:	Input Features	Output Features
	type: text/tcf+xml tokens lang: de version: 0.4	lemmas postags.tagset: stts

Metadaten

Chain 1 + <

Show tools with status: dev prod

Next Choices (Double-click to add)

<p>IMS TreeTagger</p> <p>lemmas postags.tagset: stts</p> <p style="text-align: right;">i</p>	<p>BBAW ...</p> <p style="text-align: right;">i</p>
--	---

IMS SMOR lemmatizer IMS SMOR Sfs POS Tagger - OpenNLP

Current Tool Chain (These tools are called when the Run button is clicked)

<p>Input text/plain</p> <p>Karin fliegt nach New York. Sie will dort Urlaub machen.</p> <p style="text-align: right;">i</p>	<p>Sfs To TCF Converter</p> <p>version: 0.4</p> <p>type: text/tcf+xml lang: de text</p> <p style="text-align: right;">i X</p>	<p>IMS Tokenizer</p> <p>tokens sentences</p> <p style="text-align: right;">i X</p>
---	--	---

Für die Aufgabe relevante Programme:

(siehe auch die Kursseite)

Show tools with status: dev production

Run Tools

Clear Chain

Clear Results

Next Choices (Double-click on an icon to add it to the chain)

Current Tool Chain (These tools are called when the Run button is clicked)

<p>Input text/plain</p> <p>Karin fliegt nach New York. Sie will dort Urlaub machen.</p>	<p>SfS To TCF Converter</p> <p>version <input type="text" value="0.3"/></p> <p>type: text/tcf+xml lang: de text</p>	<p>IMS Tokenizer</p> <p>sentences.tokref: true tokens</p>	<p>IMS TreeTagger</p> <p>lemmas postags.tagset: stts</p>	<p>IMS Constituent Parser</p> <p>parsing.tagset: tigertb</p>	<p>SfS Tcf2Negra</p> <p>type: text/negra</p>
--	--	--	---	---	---



Korpusimport mit TigerRegistry

- Klick auf Tiger-Icon oder im "Terminal":

```
cd ~/Desktop/TIGERSearchTools/  
./runTRegistry.sh  
(oder: cd /Applications/TIGERSearchTools/)
```



Siehe die Kursseite!



TIGERRegistry

Corpus Folder Options Help



CorporaDir

TIGERRegistry

developed by:

IMS, University of Stuttgart (TIGER Project)

<http://www.tigersearch.de>



TIGERRegistry

Corpus Folder Options Help

CorporaDir

CLARIN-D

Screen 1

TIGER dev

IMS, University of

Corpus Properties

Meta Types Bookmarks Templates

Corpus ID test

Corpus Name test

Author

Date

Corpus is in

TIGER-XML Format Other Format

Corpus ID test

Import file

/Users/heike/Desktop/Paris_120616/tsearch/test.plain Choose

Import filter

Negra format filter

Convert corpus graphs all

TIGER-XML file

/Users/heike/test.xml.gz Choose

Default Name for XML file

GZip XML file Delete XML file after indexing

External Header Choose

Extended indexing (increased query processing efficiency)

Start Cancel

Screen 2

VP NP AP

Tiger XML



Screen 3



TigerSearch



TIGERSearch - (no corpus loaded)

Corpus Query Help

(no corpus loaded)

- CorporaRoot
 - DemoCorpora
 - Tiger
 - TromsoCorpora
 - Tueba
 - WSJ
 - pcc
 - test

test (with extended index)

(19 corpus graphs)

1. General information

Name: test
ID: test
Format: Negra Format, Version 4

2. Corpus details

Features (T): word, lemma, pos, morph

Textual mode Graphical mode

Open Bookmarks

Welcome to TIGERSearch! Please choose a corpus.

Search

TigerSearch: Suchanfragen



TIGERSearch - Corpus: test

Corpus Query Help

test

- test
 - Documentation
 - Summary view
 - Detailed view
 - Edge labels
 - edges
 - secondary edges
 - Nonterminal features
 - cat
 - Terminal features
 - word
 - lemma
 - pos
 - morph

Textual mode Graphical mode

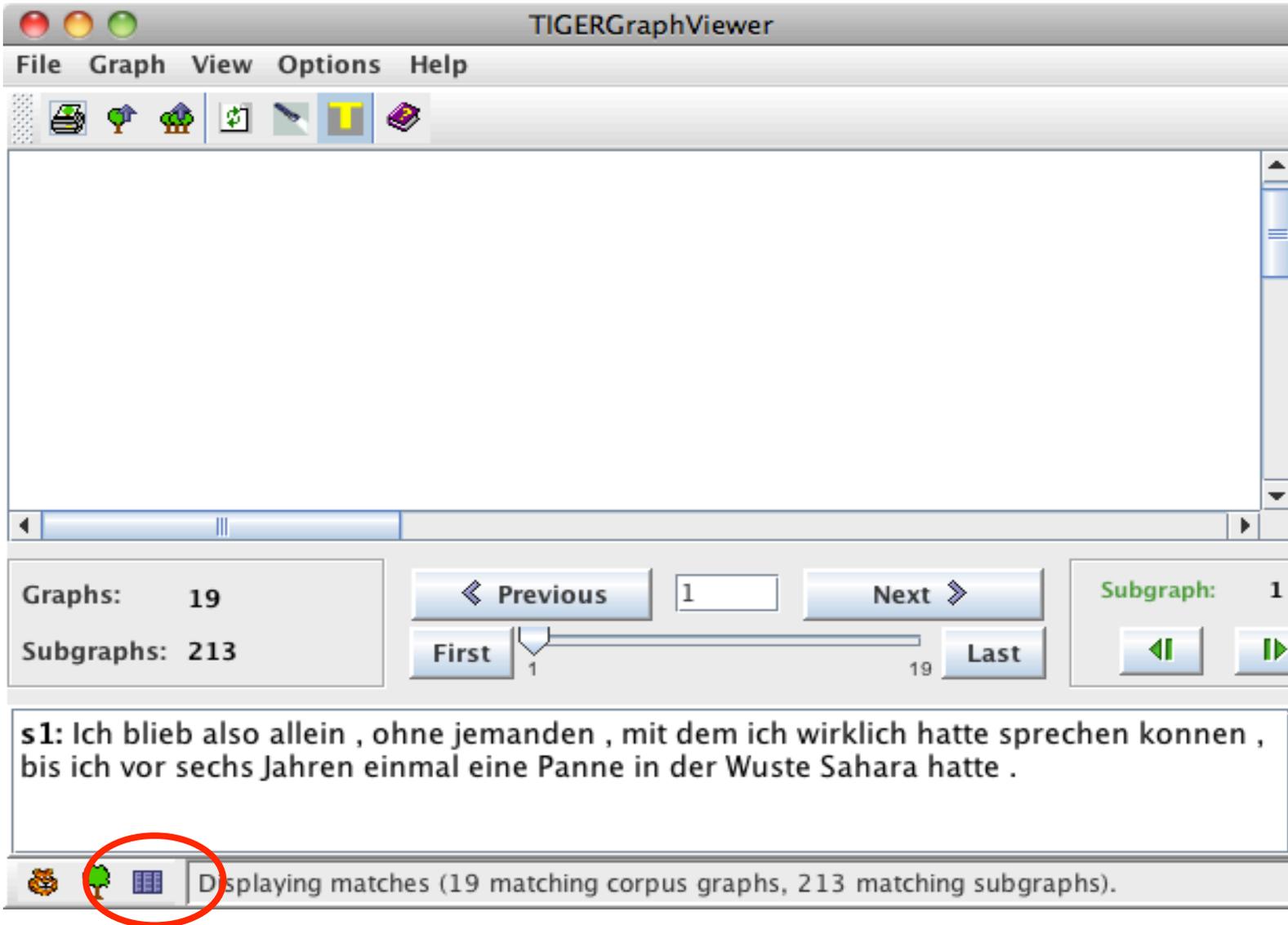
```
//Vokabular des Korpus:  
#1:[pos=/.+//]  
  
// Bi-Gramme:  
// #1:[pos=/.+//] . #2:[pos=/.+//]  
  
// Aktivieren Sie eine der Suchen,  
// indem Sie die Schrägstriche am  
// Zeilenanfang löschen.
```

Open Bookmarks Info

Corpus test loaded.

Search

TigerSearch: Aktivierung der Statistik



TIGERGraphViewer

File Graph View Options Help

Graphs: 19
Subgraphs: 213

Subgraph: 1

s1: Ich blieb also allein , ohne jemanden , mit dem ich wirklich hatte sprechen können , bis ich vor sechs Jahren einmal eine Panne in der Wüste Sahara hatte .

Displaying matches (19 matching corpus graphs, 213 matching subgraphs).



Statistics export - current

Build Default Add Remove Clear **Corpus** Frequency

GraphID	Submatch	Feature 1
		#1
		#1

1.

Build Default Add Remove Clear **Corpus** Frequency

GraphID	Submatch	Feature 1
		#1
		word
		lemma

2.

3.

Build Default Add Remove Clear **Corpus** Frequency

GraphID	Submatch	Feature 1
		#1
		word



Build Default Add Remove Clear **Corpus** Frequency Copy Export Close

GraphID	Submatch	Feature 1
		#1
		word
s1	1	Ich
s1	2	blieb
s1	3	also
s1	4	allein
s1	5	,
s1	6	ohne
s1	7	jemanden
s1	8	,
s1	9	mit
s1	10	dem
s1	11	ich
s1	12	wirklich
s1	13	hatte
s1	14	sprechen
s1	15	können
s1	16	,
s1	17	bis
s1	18	ich
s1	19	vor
s1	20	sechs
s1	21	Jahren

Build Default Add Remove Clear **Corpus** **Frequency** Copy Export

Freq	Feature 1
	#1
	word
11	.
10	,
7	ich
6	"
6	?
5	Ich
5	mich
4	ein
4	hatte
4	mir
3	auf
3	als
3	war
3	eine
3	das
3	die
2	...
2	bei
2	und
2	Schaf
2	bin

Tables have been updated.

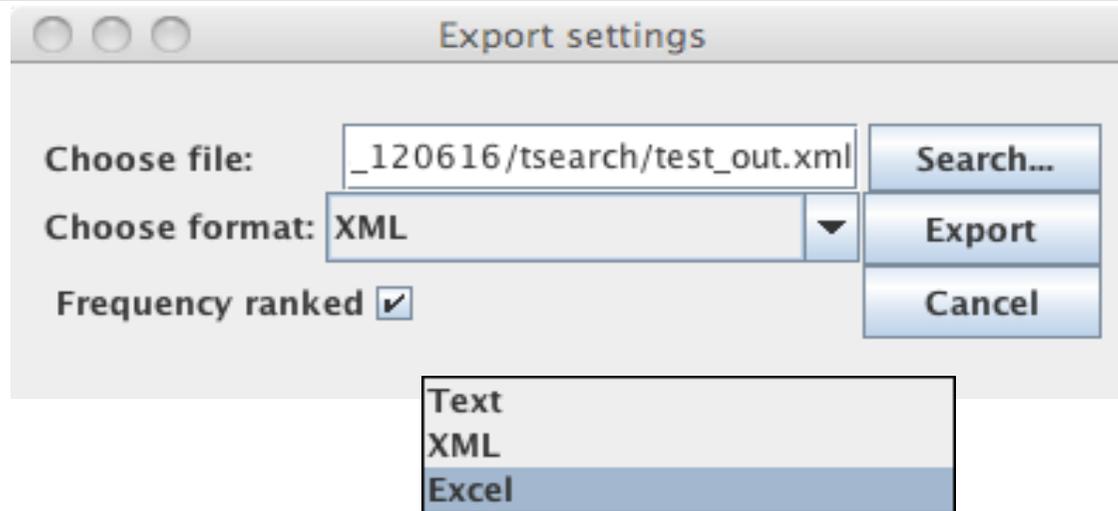
Frequency table is currently shown.

TigerSearch: Komplexere Frequenzstatistik

Build	Default	Add	Remove	Clear	Corpus	Frequency	Copy	Export	Close
Freq ▲		Feature 1		Feature 2					
	#1								
	word					pos			
11	.					\$.			
10	,					\$,			
7	ich					PPER			
6	?					\$.			
6	"					\$(
5	mich					PRF			
5	Ich					PPER			
4	mir					PPER			
4	ein					ART			
4	hatte					VAFIN			
3	eine					ART			
3	war					VAFIN			
3	die					ART			
3	auf					APPR			
2	als					KOKOM			
2	bei					APPR			
2	in					APPR			
2	bin					VAFIN			
2	der					ART			
2	Schaf					NN			
2	von					APPR			

Tables have been updated.





- Exportformat: Wählen Sie: 'XML piped through XSLT'
- Ablageformat: text file
- Klicken Sie auf SEARCH, wählen Sie einen Namen für die neue Datei.



CLARIN-D M12 Workshop

Leipzig, 27.-28. Juni 2012

<http://clarin.informatik.uni-leipzig.de/>

Vielen Dank!