DiaCollo: Kollokationsanalyse in diachroner Perspektive

Ein Use Case des CLARIN-D-Zentrums an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW)

Beteiligte Projekte: Deutsches Textarchiv (www.deutschestextarchiv.de)
Digitales Wörterbuch der Deutschen Sprache (www.dwds.de)
CLARIN-D-Zentrum der BBAW (http://clarin.bbaw.de)

Was ist DiaCollo?

Von dem berühmten Sprachphilosophen Ludwig Wittgenstein ist der Satz überliefert: „Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache." (Philosophische Untersuchungen, S. 43) Anders ausgedrückt wird also über den Kontext, in welchem ein Wort benutzt wird, dessen Bedeutung sichtbar. Ein mehrdeutiges Wort wie „Bank" kann somit zum Beispiel anhand seines Kontextes disambiguiert werden: Die Bank, auf der man sitzt, wird eher mit Begriffen wie „setzen, ausruhen, Garten" gemeinsam auftreten; die Bank als Finanzinstitut eher mit Begriffen wie „Geld, Zahlung, zur … gehen".

Ebenso lässt sich Bedeutungswandel von Wörtern oftmals an typischen Wortverbindungen erkennen (also an den Wörtern, mit welchen ein Wort typischerweise zusammensteht, den Kollokationen). Schließlich können auch politische, kulturelle oder gesellschaftliche Veränderungen im Zusammenhang mit einem zentralen Begriff an dessen typischen Wortverbindungen erkennbar und nachvollziehbar werden (siehe das Beispiel „Revolution" unten).

Für das Auffinden solcher typischen Wortverbindungen zu einem Stichwort steht das Werkzeug DiaCollo bereit, dessen Funktionalität im Folgenden gezeigt werden soll. In DiaCollo werden typische Wortverbindungen auf Grundlage verschiedener Korpora in einem bestimmten Zeitraum und vergleichend über verschiedene Zeiträume hinweg ermittelt. Die Ergebnisse werden dann in unterschiedlicher Weise visuell aufbereitet ausgegeben.

Schritt-für-Schritt-Anleitung

In der folgenden Schritt-für Schritt-Anleitung erfahren Sie, wie man Suchen mit DiaCollo vornimmt und wie Sucheinstellungen vorgenommen werden können. Beispiel: Wandel von Diskursen am Beispiel *Revolution:*

- Gehen Sie zu http://kaskade.dwds.de/dstar/dta/diacollo/ - DiaCollo-Abfragen auf der Basis der Daten des „Deutschen Textarchiv".

- Geben Sie im Feld QUERY das Wort *Revolution* ein.
- Wählen im FORMAT-Feld den Wert „Cloud“. Die Werte in allen anderen Feldern können so bleiben, wie sie bereits vorgegeben sind. Klicken Sie dann „submit“ (neben dem QUERY-Feld).
- Unterhalb der Abfragemaske erscheint nun ein Kasten, in dem Wörter, die mit *Revolution* typischerweise vorkommen, eingetragen sind. Das Fenster zeigt den Stand von 1610. Die Darstellung ist eine Wortwolke, d.h. die Wörter werden in verschiedenen Größen, Farben und Farbintensitäten dargestellt
- Oberhalb dieses Fensters befindet sich eine Zeitleiste, die in diesem Fall von 1610 – 1910 reicht. Es sind Abstände von jeweils 10 Jahren darauf markiert.
- Am linken Rand des Fensters finden Sie eine Skala der (relativen) Bindungsstärke, die Ihnen die Einordnung und Interpretation der angezeigten Ergebnisse erleichtert.
- Sie können nun den Schieberegler auf einen bestimmten Wert in dieser Leiste positionieren (z.B. 1790) und die in dieser Dekade typischerweise mit Revolution vorkommenden Wörtrer betrachten und durch Anklicken der Wörter zu den Belegen mit dieser Wortverbindung gelangen.
- Alternativ können Sie auf die „Start“-Taste neben der Zeitleiste klicken. Die Veränderungen der typischen Wortverbindungen über die Zeit werden nun als Film abgespielt. Sie können die Abspielgeschwindigkeit mit dem Regler neben der Start-Taste verändern.

Sie können diese Basiseinstellungen in vielerlei Hinsicht verändern. So können Sie z.B. den abgefragten Zeitraum (YEAR) und die Zeitabstände auf der Zeitleiste verändern (SLICE). Sie können die Höchstzahl der angezeigten Wortverbindungen ändern (KBEST) und auch die Darstellung der Ergebnisse (FORMAT).

Es stehen weitere Korpora und auch weitere Arten der Anwendung zur Verfügung. Sie können z. B. mit DiaCollo zwei oder mehr Wörter in ihrem Gebrauch in einem bestimmten Zeitraum anhand von Gemeinsamkeiten und Unterschieden ihrer typischen Wortverbindungen miteinander verglichen werden, ebenso ein Wort in zwei weiter auseinanderliegenden Zeiträumen.

Weitere Details werden unten ausführlicher und anhand von Beispielen dargestellt. Darüber hinaus gibt es eine ausführliche Bedienungshilfe (http://kaskade.dwds.de/dstar/dta/diacollo/help.perl, in Englisch).

DiaCollo im Detail

Die Textbasis

Um die Veränderung eines Wortes und seiner Verwendung oder die Veränderung einer Gruppe von Wörtern über einen längeren Zeitraum verfolgen zu können, bedarf es einer großen Menge von digitalisierten Texten (eines großen Textkorpus), deren jeweilige Entstehungs- oder Publikationszeit bekannt und in den jeweiligen Metadaten angegeben ist. Dadurch können die typischen Wortverbindungen einer bestimmten „Zeitscheibe“ (z.B. einer bestimmten Dekade) zugeordnet werden. Das Zentrum Sprache der BBAW bietet hier verschiedene Textsammlungen an, deren Entstehung sich über einen jeweils längeren Zeitraum erstreckt, unter anderen:

- *dta*: *Das Deutsche Textarchiv (1650-1900) – Startpunkt:* http://kaskade.dwds.de/dstar/dta/diacollo/
- *zeit*: Die Wochenzeitung *Die ZEIT* (1946-2015) – Startpunkt: http://kaskade.dwds.de/dstar/zeit/diacollo/
- nd: Die Tageszeitung „Neues Deutschland“ (1946-1990) – Startpunkt: http://kaskade.dwds.de/dstar/nd/diacollo/

Es gilt: Je größer die Textbasis ist, umso interessantere und verlässlichere Ergebnisse liefert das Werkzeug. Eine weitere wünschenswerte Eigenschaft des Korpus ist seine Ausbalanciertheit, d.h. eine möglichst breite Streuung und ausgewogene Verteilung der Texte über verschiedene Zeiträume und Textsorten.

Im Folgenden stellen wir einige Anwendungsbeispiele (Use Cases) für die sprachhistorische Analyse mit DiaCollo vor.

Kollokationen eines Stichworts in verschiedenen Zeiträumen

Wir betrachten im Folgenden mit Hilfe von DiaCollo den Wandel der typischen Wortverbindungen einiger Stichwörter. Beispielhaft soll gezeigt werden, wie sich anhand der zeitlichen Veränderungen der mit einem Stichwort verbundenen Kollokationen der Wandel von Diskursen entdecken (Beispiel 1) und die Emergenz neuer Lesarten und Verwendungsweisen nachweisen lassen (Beispiel 2). Der Analysemodus (in DiaCollo *Profile* genannt) ist *collocations*. Weitere Analysemodi finden Sie im ausführlichen DiaCollo-Manual (http://kaskade.dwds.de/dstar/zeit/diacollo/help.perl).

Beispiel 1: Der Wandel von Diskursen am Beispiel Krise

Zunächst wollen wir, gestützt auf das digitale Archiv der Wochenzeitung *Die ZEIT*, das Stichwort „Krise“ in den Blick nehmen, s.

http://kaskade.dwds.de/dstar/zeit/diacollo/?query=Krise&format=cloud.

Als Ausgabeformat (*FORMAT*) haben wir die Wortwolke (*Cloud*) gewählt. Die mit „Krise“ gemeinsam vorkommenden Wörter werden lemmatisiert, d.h. auf die jeweilige Grundform abgebildet. Beispiel:

Auch wenn im Text „ernste Krise“ vorkommt, wird als (häufig) vorkommendes Wort die Grundform „ernst“ angezeigt (Abb. 1).

Im Beispiel in Abb. 1 werden für jede Dekade (also einen Zeitabschnitt, der *SLICE* genannt wird und im Beispiel auf *‚10* (Jahre)’ eingestellt ist) die 10 stärksten Kollokationen angezeigt (Wert *10* für *KBEST*). Die Skala an der linken Seite zeigt die Stärke der angezeigten Kollokationen mit Krise: Ein dunkles Rot steht für eine starke Kollokation (z.B. „schwer“ in den 1940er Jahren); je mehr sich die Beschriftung in Richtung gelb, grün oder blau verschiebt, desto schwächer ist die Bindung (z.B. „wirtschaftlich“ in den 1940er-Jahren).

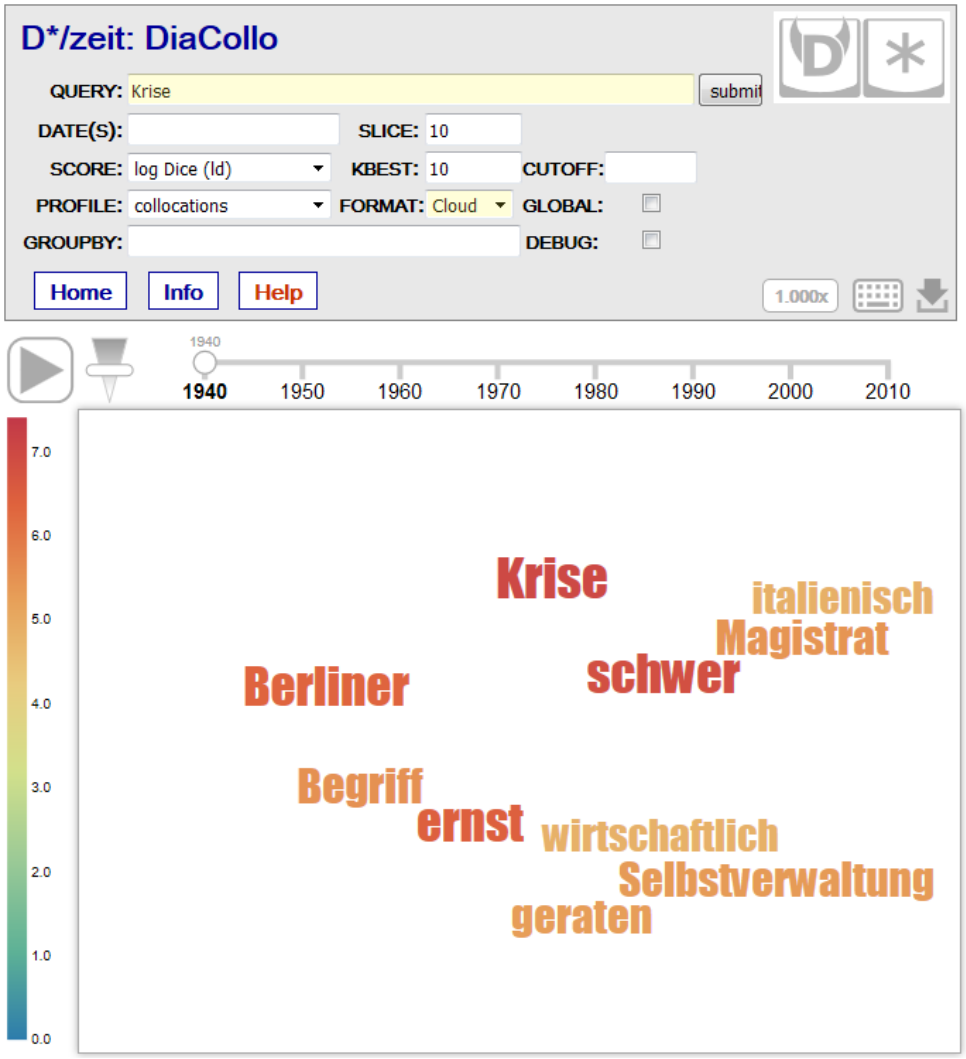


Abbildung 1

Die Zeitleiste am oberen Rand der Begriffswolke ermöglicht die Navigation zwischen den Zeitabschnitten. Durch Klicken auf eine der angezeigten Jahreszahlen kann gezielt ein bestimmter Zeitabschnitt angewählt werden. Klicken wir etwa auf das Jahr 1990, sehen wir eine veränderte Wortwolke, in der sich „schwer“ leicht abgeschwächt hat, „italienisch“ sowie einige weitere Kollokationen verschwunden sind und „Golf“ bzw. „Asien“ neu hinzugekommen sind (Abb. 2).

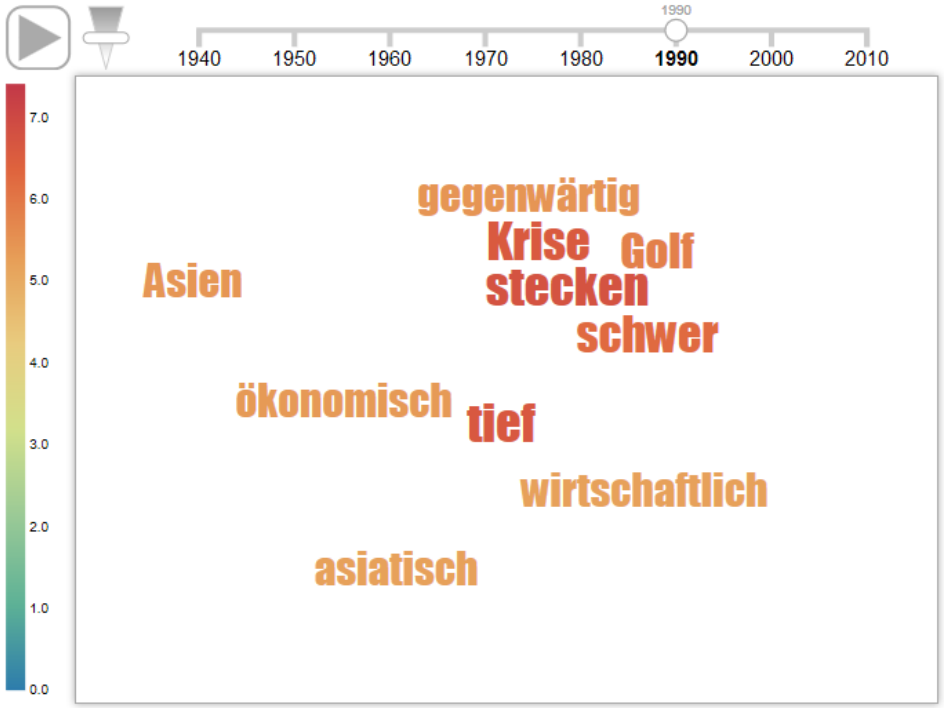


Abbildung 2

Alternativ wird nach einem Klick auf die Abspiel-Taste oben links eine Animation über den gesamten Zeitraum gezeigt. Die Geschwindigkeit der Animation kann mittels der an einen Lautstärkeregler erinnernden Schaltfläche direkt daneben angepasst werden.

Um einen bestimmten Zeitabschnitt wie die Finanz- und Eurokrise seit 2008 genauer zu untersuchen, können wir die Zeitintervalle verkleinern (z.B. *SLICE=2* für Abschnitte von jeweils zwei Jahren), den betrachteten Zeitraum einschränken (z.B. *DATE(S)=2008:2015*) und die Zahl der angezeigten Kollokatoren erhöhen (*KBEST: 20*):

http://kaskade.dwds.de/dstar/zeit/diacollo/?query=Krise&date=2008%3A2015&slice=2&kbest=20&format=cloud (Abb. 3).

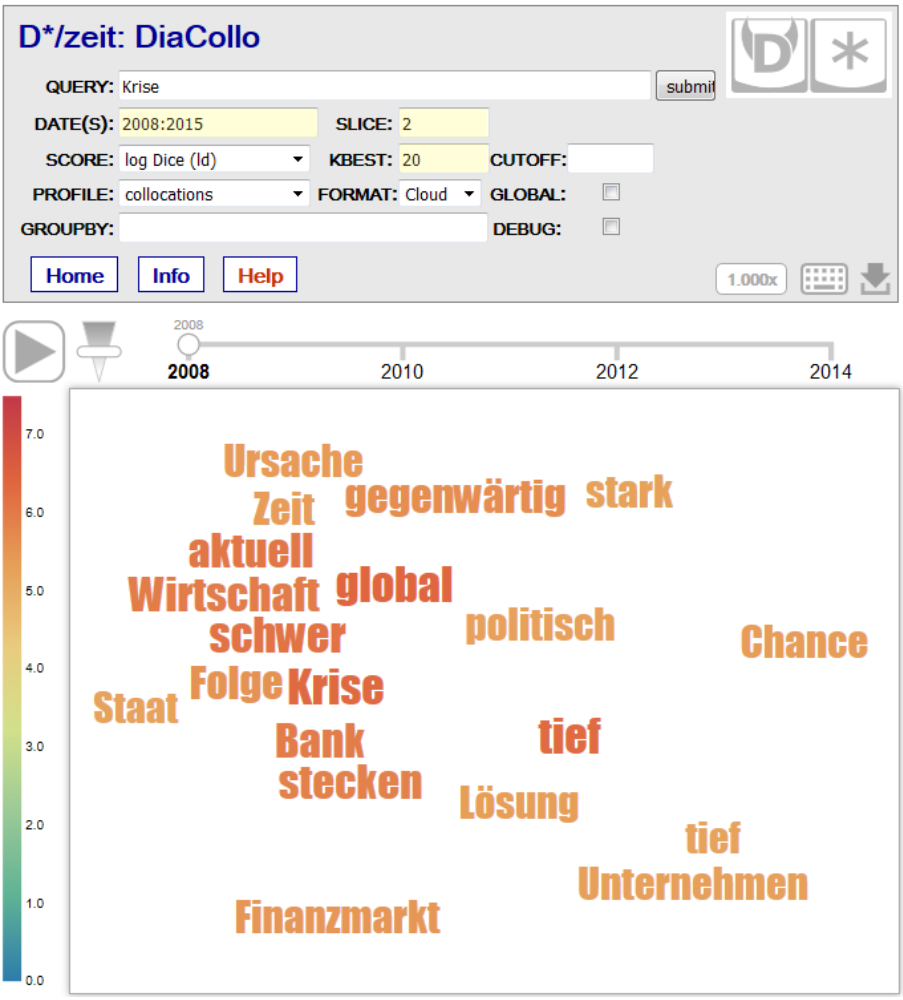


Abbildung 3

Wenn wir die Animation über die Zeit ablaufen lassen, sehen wir, dass im Zeitabschnitt 2010/2011 „Finanzmarkt“ und „Wirtschaft“ eng mit „Krise“ verbunden sind, 2012/2013 „Eurozone“ auftaucht und im letzten Zeitabschnitt „Ukraine“ und „Russland“ hinzukommen. In den beiden letzten Abschnitten finden wir außerdem das Wort „sportlich“, was vielleicht überrascht. Mit DiaCollo lassen sich Kookkurrenzen direkt zur Fundstelle in der Zeitung zurückverfolgen (wodurch eine punktgenaue Analyse ermöglicht wird). Um herauszufinden welche sportlichen Krisen Eingang in die *ZEIT* fanden, können wir mit der Maus auf „sportlich“ klicken und erhalten ein kleines Fenster mit Detailinformationen (Abb. 4).

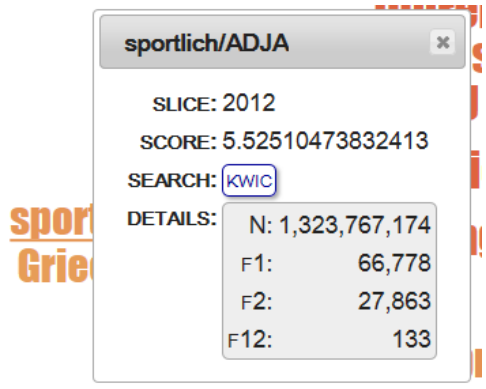
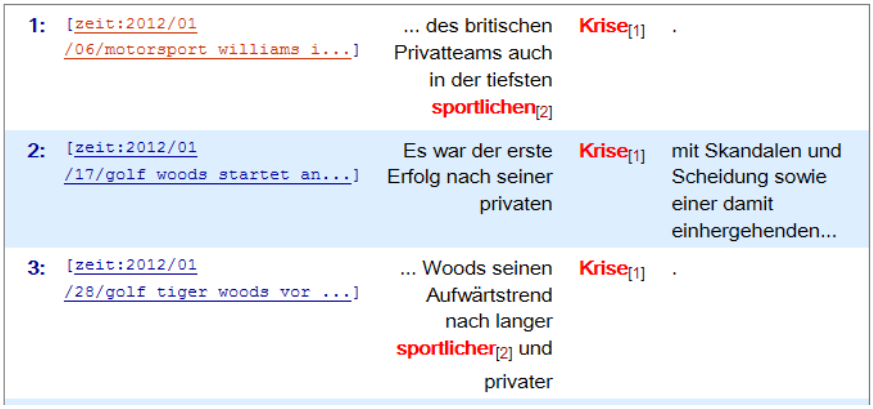


Abbildung 4

/ADJA hinter *sportlich* gibt an, dass es sich dabei um ein (attributives) Adjektiv handelt.

Ein Klick auf *KWIC* (kurz für *Keyword in context*, also eine Anzeige des Kontextes um das Stichwort herum) zeigt die entsprechenden Belegstellen: Das Formel-1 Team von Williams, der Golfspieler Tiger Woods wie auch eine ganze Reihe von Fußballspielern befanden sich 2012 oder 2013 in einer sportlichen Krise (Abb. 5).



1:	[zeit:2012/01 /06/motorsport williams i...]	... des britischen Privatteams auch in der tiefsten	**Krise**[1]	.
2:	[zeit:2012/01 /17/golf woods startet an...]	Es war der erste Erfolg nach seiner privaten	**Krise**[1]	mit Skandalen und Scheidung sowie einer damit einhergehenden...
3:	[zeit:2012/01 /28/golf tiger woods vor ...]	... Woods seinen Aufwärtstrend nach langer	**sportlicher**[2] und privater	**Krise**[1] .

Abbildung 5

Ein Klick auf einen Eintrag in der zweiten Spalte, also die Angabe zur Fundstelle des Belegs ([zeit:2012/01/06/motorsport_williams_i...] usw.) führt zum vollständigen Artikel, in diesem Fall im Archiv von http://zeit.de.

Beispiel 2: Emergenz neuer Lesarten und Verwendungsweisen anhand des Begriffs 'Revolution'

Wortuntersuchungen, die in der Sprachgeschichte weiter zurück gehen, lassen sich anhand der Korpora des Deutschen Textarchivs (ca. 1600-1900) durchführen, die auch mit DiaCollo verwendet werden können (Kürzel: *dta*). Anhand des Terminus „Revolution" soll nun gezeigt werden, wie sich Bedeutungsverschiebungen von Stichwörtern über die Zeit anhand ihrer typischen Wortverbindungen beobachten lassen:

http://kaskade.dwds.de/dstar/dta/diacollo/?query=Revolution&format=cloud.

Im Etymologischen Wörterbuch von Wolfgang Pfeifer findet sich noch die Bedeutung: „Umlaufbewegung der Gestirne, bes. der Planeten um die Sonne [...]" bzw. „Umlauf eines Himmelskörpers um ein Hauptgestirn, (tägliche) Umdrehung (der Erde um die eigene Achse) [...]." (Cf. http://zwei.dwds.de/wb/Revolution#et-1)

Diese Bedeutung zeigt sich auch in DiaCollo. Der Zeitschnitt 1670 zeigt „Tag" und „Stunde" als typische Kollokationen für „Revolution" (Abb. 6).

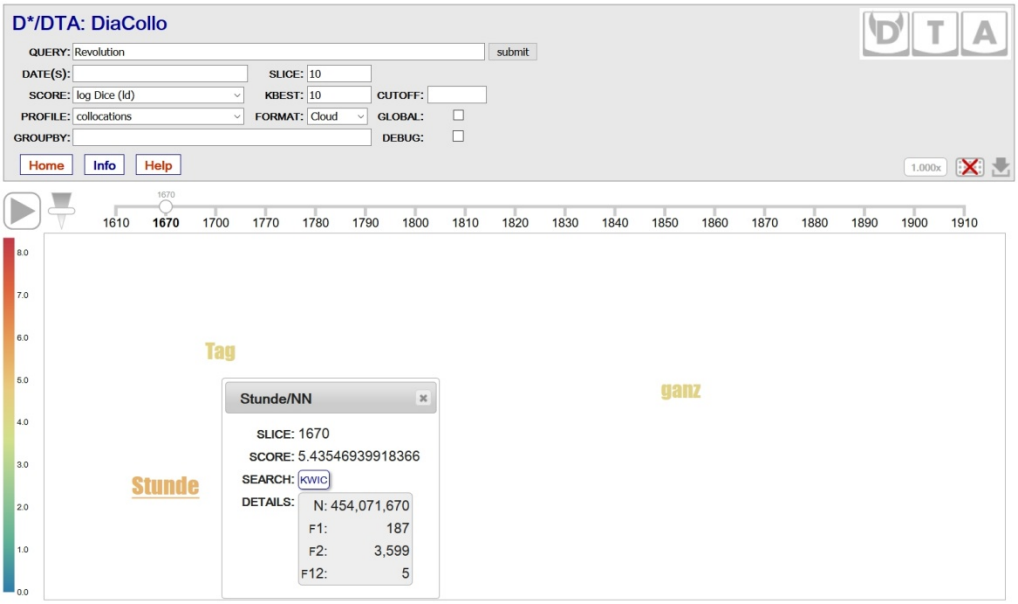


Abbildung 6

Die zugehörigen Treffer im Korpus des DTA zeigen, dass hier „Revolution“ in der o.g. Bedeutung benutzt worden ist (Abb. 7 & 8).

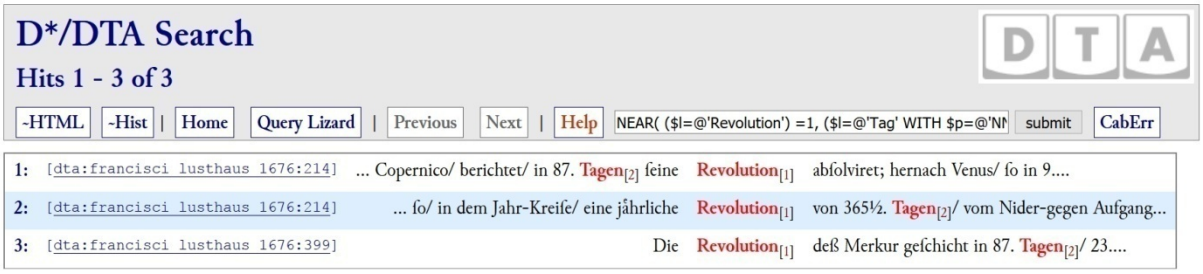


Abbildung 7

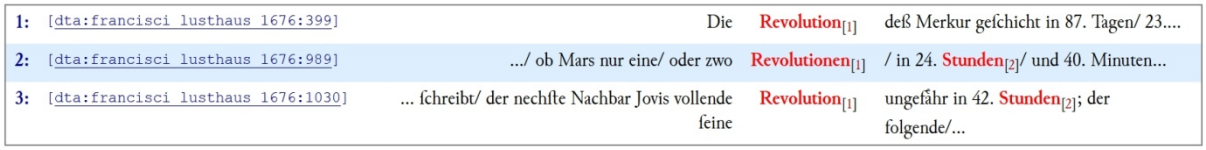


Abbildung 8

Eine weitere Bedeutung von Revolution zeigt sich im Zeitschnitt 1770, wo „Revolution“ mit den Termini „Geschlecht“ und „menschlich“ zusammen steht (Abb. 9).

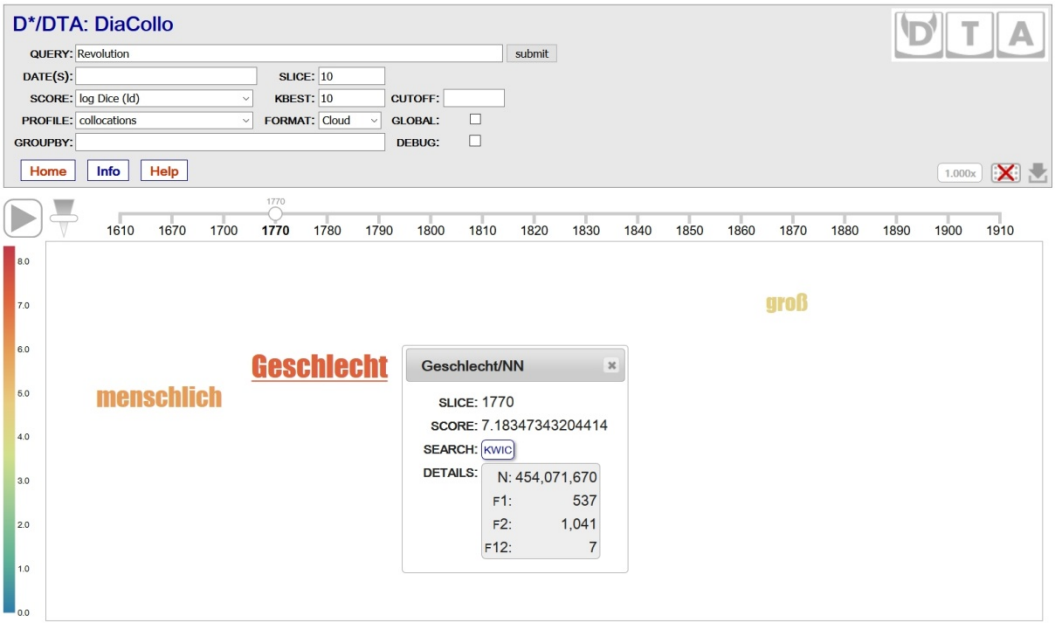


Abbildung 9

Aufschluss über diese Bedeutung gibt zum Beispiel einer der Treffer im DTA, wo es heißt:

“Jn unwohnbaren Sůmpfen, an den Ausflůſſen des Rheins, hat ſich die Freiheit und Jnduſtrie, nicht bloß Wohnungen, ſondern Thronen, gebaut. Und Spanier und Britten haben dieſe Europåiſchen Kůnſte, den Erdkreis zu verſchőnern, ůber Meer in neue Welten getragen: [...] Solche groſſe Weltbegebenheiten, ſolche **Revolutionen des menſchlichen Geſchlechts** und des Erdbodens, machen die Materie der Weltgeſchichte aus.“
(Cf. http://www.deutschestextarchiv.de/book/view/schloezer_universalhistorie01_1772?p=25).

Hier ist von „Revolutionen“ im allgemeinen Sinne eines Wandels und Umbruchs, der jedoch nicht gewaltsam erzielt werden muss, die Rede (vgl. die bei Pfeifer, ebd., angegebene Bedeutung „‘Wandel, Veränderung’ (der äußeren Umstände wie der inneren Haltung)“).

Erwartungsgemäß gesellt sich in den Zeiten der französischen Revolution die heute gängigste Bedeutung hinzu, nämlich der „gewaltsame[...] Umsturz der bestehenden Staatsform und der bestehenden Machtverhältnisse“ (Pfeifer ebd.). Sie wird deutlich an der Kollokation „gewaltsam“, der historische Kontext der französischen Revolution zeigt sich an Kollokationen wie „französisch, Volk, Frankreich“ (Abb. 10).

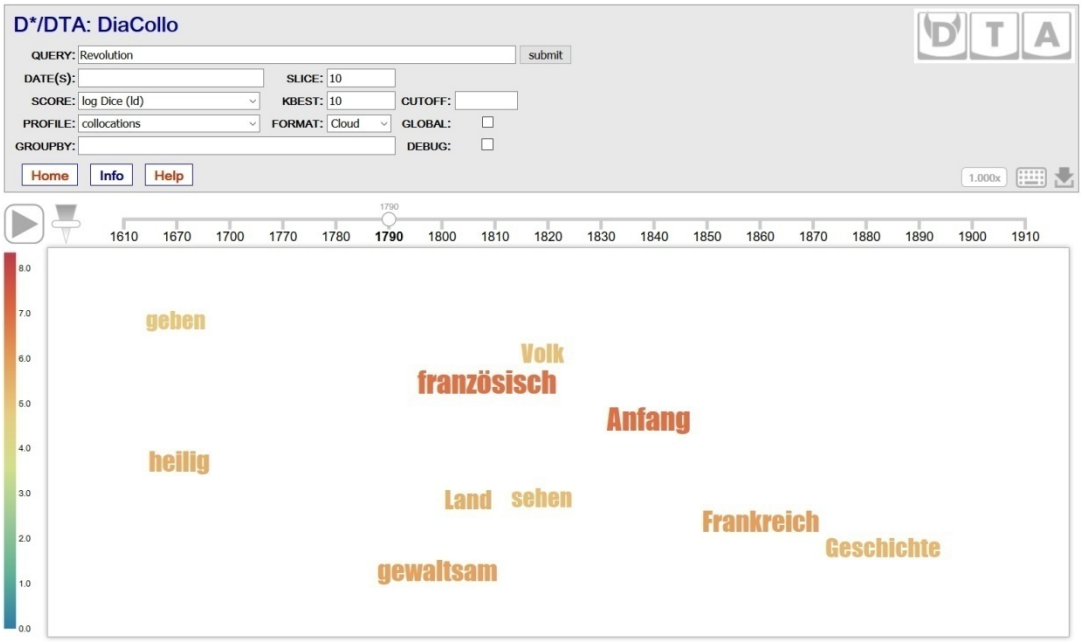


Abbildung 10

Abgesehen von diesen Bedeutungsverschiebungen werden zudem historische gesellschaftliche Veränderungen anhand der typischen Wortverbindungen deutlich. So lassen sich nach 1790 andere, der französischen Revolution folgende Bewegungen an den typischen Kollokationen ablesen, etwa der napoleonische Einfluss in Europa (Abb. 11) oder die industrielle Revolution (Abb. 12).

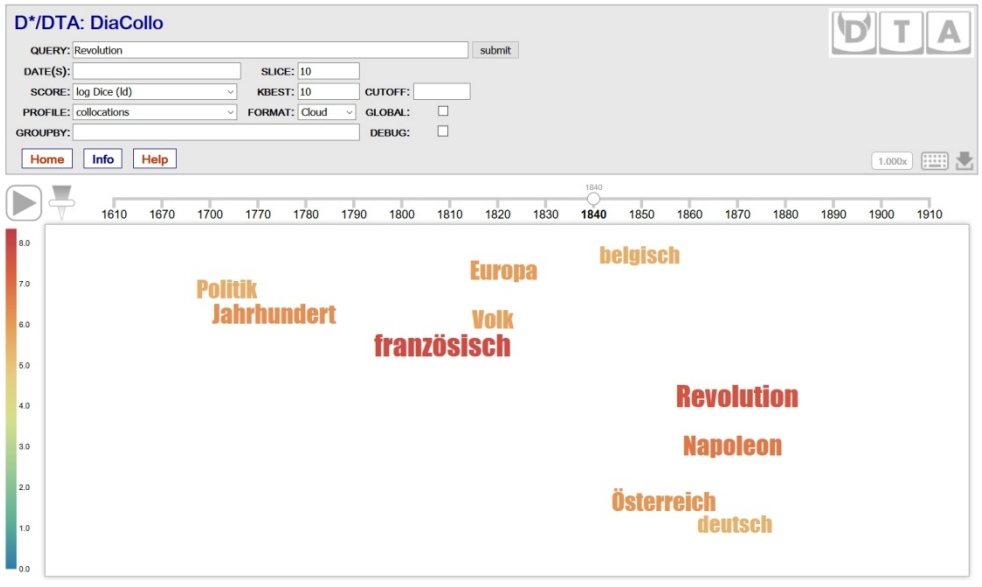


Abbildung 11

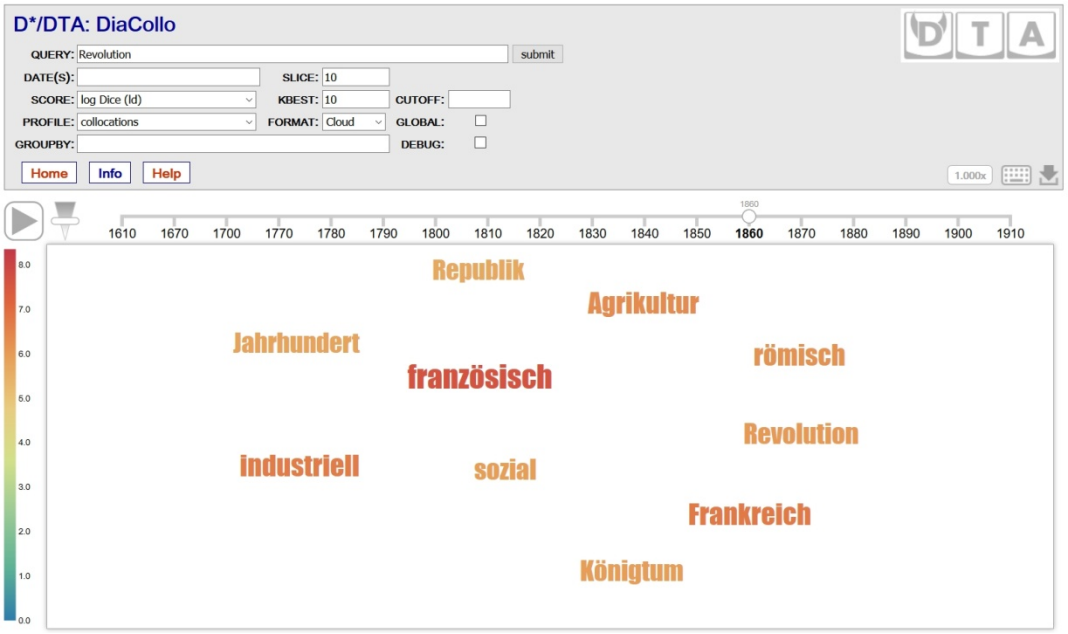


Abbildung 12

Ausblick

Über die hier beschriebene Anwendung von DiaCollo bei der Analyse des Wandels der Verwendung einzelner Wörter hinaus kann DiaCollo zur Beantwortung verschiedener weiterer Forschungsfragen benutzt werden; zum Beispiel kann man:

- Bedeutungsunterschiede zwischen zwei Wörtern herausarbeiten und veranschaulichen, indem man die Wortverbindungen, die für jeweils eines der beiden Wörter besonders typisch sind, kontrastierend gegenüberstellt. Besonders interessant ist dieses Verfahren beim Vergleich bedeutungsverwandter Wörter (ähnliche, aber auch gegensätzliche Bedeutung, z.B. „Häftling“ vs. „Gefangene“ oder „Tochter“ vs. „Sohn“)
- die ‚Karrieren‘, d.h. die Zunahme und Abnahme der Verwendungshäufigkeit ähnlicher Wörter miteinander vergleichen. Besonders interessant ist der Vergleich von Wörtern mit einem gemeinsamen Element (z.B. die häufigsten Wörter, die mit „Drogen-“ beginnen oder auf „-droge“ enden) oder mit einer gemeinsamen Bedeutung (z.B. die häufigsten Wörter zur Bezeichnung von Flüssigkeiten, wie „Wasser“, „Wein“ etc.).

Diese Anwendungen setzen die passende Wahl eines *Profils* und eines geeigneten Darstellungsformats voraus, u.U. müssen weitere Parameter angepasst werden. Details zu diesen Anwendungsmöglichkeiten werden im DiaCollo-Manual beschrieben (http://kaskade.dwds.de/dstar/zeit/diacollo/help.perl, Englisch). Wir möchten Sie dazu ermutigen, Ihre eigenen Tests zu machen und die Möglichkeiten dieses Werkzeugs mit Ihren eigenen Forschungsfragen zu entdecken.