

Korpushermeneutische Analyse politischer Reden mittels *CorpusExplorer*

script

3-min-talk

Jan Oliver Rüdiger

Sprecher:
Carsten Bänsch

Die Idee hinter dem Projekt – einen schnellen und einfachen Einstieg in die Analyse großer Korpusdaten mittels *CorpusExplorer* geben. Diese frei verfügbare Software bietet aktuell über 45 Analysen/Visualisierungen für vielfältige korpuslinguistische Zwecke und ist durch ihre Nutzerfreundlichkeit auch für den Einsatz in der universitären Lehre geeignet. Als Beispiel dient das EuroParl-Korpus, man kann aber auch eigenes Textmaterial (z. B. Textdateien, eBooks, Xml, Twitter, Blogs, etc.) mit dem *CorpusExplorer* annotieren, analysieren und visualisieren. Die Videos zeigen Schritt-für-Schritt die einzelnen Funktionen.

Überspannt werden die Videos von einer kleinen zweistufigen Aufgabe: Zuerst sollten ein paar Fragen/Thesen/Annahmen überlegt werden, die sich mit den Plenarprotokollen des EuroParl auswerten lassen – einige Videos geben auch explizite Anregungen oder man nutzt die Inspiration der anderen Beiträge im Issue #3. Die einfachsten Fragen/Thesen lassen sich bereits mit den hier vorgestellten Videos beantworten. Sobald es komplexer wird, betritt man den zweiten – reflexiven Teil der überspannenden Aufgabe: Es ist zu überlegen, wie durch (mehrfache) Kombination der

einzelnen Video-/ Wissensbausteine das Ziel erreicht werden kann (ein Beispiel, s.u.). Im Zweifelsfall stehen außerdem ein Handbuch und ein E-Mail Support zur Verfügung.

Ein Beispiel (Idee):

Über die vorgestellten Analysen lassen sich in mehreren Schritten politische Positionen (durch die Analyse des Sprachgebrauchs) isolieren. Man könnte hier wie folgt vorgehen: Zuerst identifiziert man typische Redewendungen/ Phrasen/ N-Gramme/ Syntagmen/ Sprachgebrauchsmuster - dann betrachtet man Abweichungen sowie den Verwendungskontext.

Ein Beispiel (Durchführung):

Es gibt sehr viele Parlamentarier, die in ihren Redebeiträgen das Syntagma „Raum der Freiheit, der Sicherheit und des Rechts“ verwenden (siehe Video 11). Greift man "Freiheit" aus dem Syntagma heraus (z. B. mit der Frequenzanalyse – Video 07), würde man vermuten, dass alle Parlamentarier parteiübergreifend das Gleiche meinen. Die Detailanalyse des Zeichens zeigt aber, dass "Freiheit" jeweils unterschiedlich befüllt wird - entweder mit "wirtschaftlicher Freiheit" (EVP) oder "sozialer Gerechtigkeit/ Chancengleichheit" (S&D).

Hier bieten sich zwei Möglichkeiten an, entweder man gibt eine Gruppe vor (z. B. weil eine bestimmte Fraktion/Partition gewünscht ist) oder man bestimmt diese über den Sprachgebrauch (z. B. alle, die ein bestimmtes Syntagma nutzen). In beiden Fällen werden die Gruppen mittels Schnappschuss isoliert (Video 09). Eine weitere Unterteilung kann sinnvoll sein z. B. in jeweils eine Gruppe, die Freiheit wirtschaftlich oder sozial befüllen. Gerade durch die immer detaillierter werdenden Schnappschüsse werden selbst große Textmengen wie das EuroParl-Korpus schnell handhabbar (übrig bleiben meist nur eine Handvoll Beiträge). So kann auf diesem Weg erkennbar werden, dass einzelne Sprachgebrauchsmuster selbst von den Fraktionen mit eigenen Bedeutungen aufgeladen werden. Dies geht über Fahnenworte/ Schlagworte hinaus. Bedeutungsmodifikation von Syntagmen durch die bewusste Bedeutungsbesetzung einzelner Wörter im sie umgebenden Text aufgeladen werden. Allein am Beispiel-Syntagma könnte man einzelne Kämpfe um die Deutungs-/ Auslegungshoheit zeigen (vergl. auch Video 10). Nicht nur die Befüllung des Zeichen(-komplexes) ist hier von Interesse, sondern auch wie der konkrete Einsatz innerhalb von Argumenta-

tionsstrukturen erfolgt – so taucht „Freiheit“ in Sozialen/ Bildungs-/ Gerechtigkeits-Debatten (S&D) ebenso auf wie in Debatten um Handel/Außenpolitik/Geopolitik (EVP) – Auch diese Debatten lassen sich mit Video 09 in einen Schnappschuss abtrennen. Wenn ein Redner seine Position unterstreichen will, dann appelliert er an "Grundwerte", indem er eines von vielen Sprachgebrauchsmustern verwendet, um den Frame "Grundwerte" aufzurufen.

Video 1 Installation

- Öffnen Sie einen Webbrowser und rufen Sie die Seite <http://www.CorpusExplorer.de> auf.
- Auf dieser Seite finden Sie neben dem Programmdownload auch einen Link zum Handbuch sowie frei verfügbare Korpora und Programmerweiterungen.
- Klicken Sie auf „Download“ und dann auf „Installieren“.
- Speichern Sie die Datei setup.exe auf ihrer Festplatte.

- Nachdem die Datei erfolgreich heruntergeladen wurde, führen Sie diese aus.
- Klicken Sie auf „Installieren“ und bestätigen Sie etwaige Sicherheits-Abfragen. Die Abfragen sind abhängig von der eingesetzten Windows-Version. Im Fall von Windows 10 klicken Sie auf „Weitere Informationen“ und dann auf „Trotzdem ausführen“.
- Akzeptieren Sie die Lizenzbedingungen und klicken Sie auf „Installieren“.
- Die Installation dauert wenige Minuten. Im Hintergrund wird ein kompletter Toolchain aus verschiedenen linguistischen Werkzeugen installiert.
- Nach erfolgreicher Installation startet der CorpusExplorer – die Software ist nun einsatzbereit.

Video 2 Orientierung

- Nach der erfolgreichen Installation des CorpusExplorer finden Sie in ihren „Eigenen Dokumenten“ einen Ordner

mit dem Titel: CorpusExplorer. Dieser Ordner hat vier Unterordner: Meine Datenquellen, ... Erweiterungen, ... Korpora, und ... Projekte. Im Ordner für „Datenquellen“ werden durch den CorpusExplorer Skripte für die Onlineakquise von Korpora hinterlegt.

- Der Ordner „Erweiterungen“ dient zur Ablage von CorpusExplorer Erweiterungen/Add-ons. Frei verfügbare Erweiterungen finden Sie auf folgender Webseite: www.corpusexplorer.de.
- Im „Korpora“-Ordner legt der CorpusExplorer alle Korpora ab.
- Jedes Mal, wenn Sie den CorpusExplorer starten, wird ein neues Projekt angelegt. Ein Projekt kann beliebig viele Korpora beinhalten. Außerdem werden in einem Projekt Einstellungen und Schnappschüsse gespeichert. Alle angelegten Projekte finden Sie im Ordner: Meine Projekte.
- Starten Sie den CorpusExplorer. Auf der Startseite des CorpusExplorers

finden Sie eine Checkliste für die gelungene Analyse sowie eine weiterführende Hilfe.

- Im oberen Fensterbereich finden Sie das Hauptmenü. Klicken Sie auf das CorpusExplorer Symbol. Mit Hilfe dieses Menüs können Sie neue Projekte erstellen, bestehende Projekte laden und speichern sowie die Projekteinstellungen ändern.
- Bei jedem Programmstart erstellt der CorpusExplorer automatisch ein neues Projekt.
- Gehen Sie zur Checkliste und klicken Sie auf Projektnamen vergeben. Geben Sie einen Projektnamen ein. Bestätigen Sie Ihre Eingabe mit OK. Ein grüner Haken zeigt an, dass das Projekt jetzt benannt ist.
- Klicken Sie im Hauptmenü auf Projekt speichern, wird das Projekt unter dem jeweiligen Namen gespeichert (Eigene Dokumente > CorpusExplorer > Meine Projekte).

- Neben dem Hauptmenü sehen Sie weitere Symbole. Um auf die gerade angezeigte Startseite zurückzukommen, können Sie jederzeit auf das Haus-symbol klicken. Die drei Symbole neben dem Haus finden Sie auch in der Checkliste. Erfahren Sie in den folgenden Videos dieses Webcasts, wie Sie Korpora laden, Schnappschüsse erstellen und Analysen durchführen.

Video 3 Das EuroParl-Korpus

- Rufen Sie folgende Webseite in ihrem Browser auf: <http://www.statmt.org/europarl/> Hier finden Sie einige Informationen zum EuroParl Korpus. Aktuell ist die Version 7.
- Klicken Sie auf Downloads: „Source Releases“.
- Entpacken Sie die heruntergeladene Datei.
- Im Unterordner TXT finden Sie die verschiedenen Sprachversionen des EuroParl-Korpus.

- Im gesamten weiteren Verlauf dieses Webcast werden wir ausschließlich die Dateien aus dem Ordner DE verwenden. Sie können aber auch jede andere Sprachversion nutzen.
- Öffnen Sie im Ordner eine der EP-Dateien. Diese werden Ihnen als Plaintext angezeigt. D. h. es gibt keine Textformatierungen. Innerhalb des Fließtexts sind verschiedene Marker hinterlegt. Zum Beispiel die ChapterID oder der Name des jeweiligen Speakers. Diese Marker werden später vom CorpusExplorer dazu verwendet, um Metadaten zu erzeugen. Ein Beispiel: Im CorpusExplorer können Sie später abfragen, wer?... wann zu welchem Thema einen Redebeitrag im europäischen Parlament gehalten hat.
- Sie können nun das Dokument schließen und den CorpusExplorer starten.

Im nächsten Video erfahren Sie, wie Sie das Korpus in den CorpusExplorer einlesen und annotieren.

Video 4 Korpus annotieren

- Nach dem Start des CorpusExplorers sollten Sie in aller Regel zuerst einen Projektnamen vergeben.
- Als zweites müssen Sie mindestens ein Korpus laden. Klicken Sie dazu in der Checkliste auf „Mindestens ein Korpus laden“.
- Hier haben Sie vier Möglichkeiten. Die erste Möglichkeit „Existierendes Korpus laden“ steht nur dann zur Verfügung, wenn bereits Korpora im Ordner „Dokumente > CorpusExplorer > Meine Korpora“ vorhanden sind.
- Die zweite Möglichkeit „Dokumente annotieren“ erlaubt es, aus sehr unterschiedlichen Textquellen ein Korpus zu erstellen. Extraktion von Text und Metadaten, sowie Bereinigung und Annotation laufen dabei vollautomatisch ab.
- Die Optionen „Korpus importieren“ eignet sich, um bereits annotiertes

Korpusmaterial, zum Beispiel aus anderen Programmen, zu importieren.

- Die vierte und letzte Option „Online Korpus akquirieren“ erlaubt es, Webseiten nach bestimmten Begriffen zu durchsuchen und die Suchergebnisse entsprechend als Korpusmaterial zu speichern.
- Klicken Sie auf „Dokumente annotieren“ und wechseln Sie zu dem Ordner, in dem Sie die EuroParl Dateien abgelegt haben. Wie Sie sehen, tauchen keinerlei Dateien auf, das liegt daran, dass Sie zunächst unten rechts das korrekte Dateiformat wählen müssen.
- Wie Sie in der Liste erkennen können, gibt es für verschiedene Dateiendungen mehrere Einträge. Zum Beispiel existieren für TXT-Dateien mehrere Auswahlmöglichkeiten. Der wesentliche Unterschied liegt hierbei in der Erschließung von Metadaten. Für die in diesem Video verwendeten EuroParl-Dateien gibt es einen speziellen Prozess...

- ... der sich im Vergleich zum Standard-TXT dadurch unterscheidet, dass nicht nur der reine Text eingelesen wird, sondern auch die Metadaten – und anhand der Metadaten auch eine Auftrennung des EuroParl-Plenarprotokolls in einzelne Redebeiträge erfolgt.
 - Wählen Sie „EuroParl (*.txt)“ aus – danach werden alle verfügbaren Dateien mit dieser Endung angezeigt.
 - Wählen Sie einige davon exemplarisch aus und klicken Sie auf Öffnen. Die Dateien werden jetzt eingelesen, aufgetrennt und bereinigt. Alles was Sie tun müssen, ist abwarten.
 - Nachdem dieser Prozessschritt abgeschlossen wurde, erscheint ein Dialog, der alle erkannten Dokumente enthält. Rechts neben dem Text sehen Sie zudem die Metadaten: Das wären zum einen der Sprecher, die Originalsprache sowie verschiedene andere Informationen die im Protokoll vorhanden sind.
 - Klicken Sie auf „OK“ – Es erscheint ein Dialog, in dem Sie den Tagger auswählen und konfigurieren können. Die Standardeinstellung ist der TreeTagger mit deutschem Sprachmodell.
 - In dieser Liste können Sie den Tagger auswählen. Standardmäßig bringt der CorpusExplorer neben dem TreeTagger auch den TNT-Tagger mit sich sowie einen sprachunabhängigen Tagger für puren Textimport.
 - Abhängig vom gewählten Tagger sind die erweiterten Einstellungen. Hier können Sie das Sprachmodell wählen sowie einige detaillierte Konfigurationen vornehmen. Wenn Sie mit den Einstellungen zufrieden sind, klicken Sie auf „OK“.
 - Abschließend müssen Sie nur noch einen Korpusnamen vergeben; dieser wird Ihnen später im CorpusExplorer angezeigt und das Korpus wird unter diesem Namen in ihren eigenen Dokumenten im Ordner CorpusExplorer > Meine Korpora gespeichert. Nachdem sie den Namen vergeben haben, klicken Sie auf „OK“ und der Annotationsprozess startet.
 - Dieser Vorgang dauert je nach Korpusumfang einige Minuten. Nachdem der Annotationsprozess abgeschlossen wurde, aktualisiert sich die Ansicht des Corpus Explorer.
- Video 5 EuroParl Korpuserweiterung**
-
- Sie werden feststellen, dass der Import einer ganzen Sprachversion des EuroParl-Korpus mehrere Minuten oder sogar Stunden in Anspruch nehmen kann. Daher gibt es für diese Webcast-Serie ein bereits annotiertes, d. h. nutzungsfähiges Korpus. Um dieses Korpus zu nutzen, beenden Sie zunächst den CorpusExplorer und rufen Sie die Webseite <http://CorpusExplorer.de> auf.
 - Klicken Sie auf „Infos/Download“ und dann auf „Korpora“.
 - Auf dieser Seite finden Sie verschiedene vorgefertigte Korpora, unter anderem auch das EuroParl.

- Klicken Sie auf Download und speichern Sie die Datei im Ordner „Dokumente“ > „CorpusExplorer“ > „Meine Erweiterungen“. Der Grund, warum Sie die Datei in diesem Ordner speichern müssen und nicht unter „Meine Korpora“ – bei dieser Datei handelt es sich um Korpus-Add-on. Korpus- und Programm-Add-ons erweitern die Funktionalität des CorpusExplorers und werden über dieselbe Infrastruktur verteilt und auch kontinuierlich aktualisiert.
- Starten Sie dann den CorpusExplorer. Sie erhalten eine Meldung, dass ein Update für Sie bereitsteht. Bestätigen Sie diese Meldung mit „Ja“. Das Korpus wird automatisch heruntergeladen und steht danach im Ordner „Meine Korpora“ zur Verfügung.

Video 6 Konzepte

- Starten Sie den CorpusExplorer und laden Sie das EuroParl-Korpus.
- Wie Sie auf den ersten Blick sehen, enthält dieses Korpus über 200'000

Texte in drei Layern mit über 54 Millionen Token.

- Aktuell befinden Sie sich auf der Korpus-Übersichts-Seite. Klicken Sie im Programm-Menü auf das Korpus-Symbol. Sie sehen, dass die Funktionen, die Sie auf der aktuellen Seite sehen, auch über das Programm-Menu verfügbar sind.
- Klicken Sie auf das Haus-Symbol, um zum Startbildschirm zurück zu gelangen. Sobald ein Korpus geladen wurde, sind alle Kriterien der Checkliste für eine erfolgreiche Analyse erfüllt.
- Sie sehen aber auch, dass zwischen den Korpora und den Analysen die Schnappschüsse stehen. Wenn Sie den CorpusExplorer erstmalig ausprobieren, brauchen Sie sich um Schnappschüsse keine Gedanken machen. Sobald ein Korpus geladen wurde, erstellt der CorpusExplorer automatisch einen Schnappschuss, der das gesamte Korpus umfasst. Die wesentlichen drei Funktionen von Schnappschüssen sind Reduktion, Kombination und Isolation. Reduktion: Nutzen Sie Schnappschüsse,

um Korpora zu filtern. Rekombination: Schnappschüsse können untereinander kombiniert werden. Isolation: Schnappschüsse isolieren Analysen und Korpora voneinander – dadurch werden Analysen reproduzierbar, selbst wenn Sie neues Korpusmaterial hinzufügen. Wir werden in einem späteren Video auf die vielfältigen Funktionen von Schnappschüssen detailliert eingehen, insbesondere weil Sie eine wichtige Rolle in der korpushermeneutischen Analyse spielen.

- Klicken Sie auf „Neue Analyse starten“ > „Volltextzugriff“ > „Text annotieren“. Wie Sie im Programm-Menü sehen, gelangen Sie auch hierüber zu dieser Analyse.
- Abschließend wollen wir uns noch ansehen, was es mit den Layern auf sich hat. Sie erinnern sich? Das gerade geladene EuroParl-Korpus hat 54 Mio. Token und drei Layer. Layer sind abhängig vom Annotations-/Import-Prozess. Wenn Sie z. B. den TreeTagger nutzen, dann erstellt dieser mindestens drei Layer: Wort, Lemma, POS.

Diese Daten splittet der CorpusExplorer automatisch auf und macht Sie unabhängig voneinander verfügbar. Hier rechts sehen Sie die verfügbaren Layer für dieses gerade angezeigte Dokument. Klicken Sie auf den Layer > POS.

- Es klappt eine Liste mit allen verfügbaren Layerwerten aus. In diesem Fall sind es Part-of-Speech-Informationen als Wortarten oder kurz POS. Setzen Sie einen Haken vor den Eintrag NN. Dadurch werden alle Begriffe farblich hervorgehoben, die mit diesem Wert annotiert wurden.
- Welche Layer und welche Layerwerte vorhanden sind, ist wie bereits erwähnt abhängig vom Annotations-Import-Prozess. Da das EuroParl-Korpus mit dem TreeTagger aufbereitet wurde, steht NN für ein „Normales Nomen“ – entsprechend dem Stuttgart-Tübingen-Tagset, das der TreeTagger von Helmut Schmid verwendet.
- Der zuletzt ausgewählte Layerwert kann in dieser Ansicht zur Annotation verwendet werden. Damit können Sie

z. B. bei Bedarf die automatische Annotation korrigieren. Außerdem können Sie eigene Layerwerte (Rechtsklick auf Layer) oder ganze Layer hinzufügen.

- Im nächsten Video erfahren Sie, wie Sie die Layer POS, Lemma und Wort auswerten können.

Video 7 Frequenztafel

- Nachdem Sie den CorpusExplorer gestartet und ein Korpus geladen haben, starten Sie die Analyse > Frequenzanalyse > Tabelle.
- Nach wenigen Sekunden erscheint die Frequenztafel. Von links nach rechts sehen Sie die Tabellenspalten POS – Lemma, Wort und Frequenz.
- Unterhalb der Tabellenspalten sehen Sie die Filterspalte, die sie nutzen können, um die Tabelle individuell zu filtern.
- Am Anfang und Ende jeder Tabelle ist jeweils grün hervorgehoben die Spaltenspalte. Dort wird

zusammengefasst, wie viele unterschiedliche POS-/Lemma-/Wort-Einträge existieren oder wie hoch die Summe der Frequenzen ist.

- Wenn Sie das hochfrequenteste Normale Nomen suchen, dann klicken Sie unterhalb der POS-Spalte auf „beinhaltet“ und geben Sie NN für den entsprechenden POS-Tag ein.
- Die Tabelle ist zunächst ungeordnet. Um die Tabelle zu sortieren, klicken Sie mit der rechten Maustaste auf die gewünschte Spalte, in diesem Fall auf „Frequenz“, und wählen Sie „sortieren in absteigender Reihenfolge“ aus. Dadurch wird die Tabelle anhand der Frequenz absteigend sortiert und Sie sehen, dass „Kommission“ das häufigste Nomen ist.
- Klicken Sie auf die Filterspalte „Lemma“ und geben Sie bei „beinhaltet“ „Freiheit“ ein. Sie sehen, dass Singular- und Pluralformen getrennt aufgelistet werden und dass verschiedene Freiheiten benannt werden. Grundfreiheit, Meinungsfreiheit, Religionsfreiheit, Pressefreiheit.

- Aktuell werden zwei Filter kombiniert. Mit einem Klick auf den kleinen Filter innerhalb der jeweiligen Spalte können Sie auch einen anderen Filteroperator auswählen. Wählen Sie für die Spalte „Lemma“ den Filteroperator „Ist gleich“. Dadurch werden nur die exakten Übereinstimmungen angezeigt.
- Um die Belege für einen bestimmten Eintrag anzeigen zu lassen, können Sie die Textlinse nutzen. Die Textlinse befindet sich vor jeder Tabellenzeile – der Kreis mit dem Dreieck. Klicken Sie auf die Textlinse vor „Freiheiten“.
- Eine weitere Tabelle klappt aus. Hierbei handelt es sich um eine KWIC-Ansicht, d.h. Sie sehen den jeweiligen Fundsatz – die Fundstelle in der Mitte – und was links sowie rechts von dieser im Satz steht.
- Durch das Ziehen der Trenner im Spaltenkopf können Sie die Ansicht individuell anpassen.
- Auch diese Tabelle kann durchsucht werden.

- Um den vollen Belegtext anzuzeigen, können Sie auf die Lupe klicken. Dann erscheint der komplette Text. Wenn Sie etwas herunterscrollen, finden Sie den Belegsatz farblich hervorgehoben.
- Rechts sehen Sie die Metadaten des jeweiligen Dokuments.

Video 8 Korpusverteilung

- Eine weitere Möglichkeit, sich einen Überblick zum vorliegenden Korpusmaterial zu verschaffen, besteht darin, dass die Metadaten des Korpus betrachtet werden.
- Falls noch nicht geschehen: Starten Sie den CorpusExplorer und laden Sie ein Korpus.
- Gehen Sie dann auf Analyse > Korpusverteilung > Tabelle.
- Zunächst einmal sollten Sie die Tabelle anhand der Tabellenspalte „Token“ sortieren.

- Klicken Sie mit der rechten Maustaste auf den Tabellenkopf „Token“ und wählen Sie die Option „absteigend sortieren“ aus.
- Sie werden feststellen, dass die Tabelle weiterhin sehr undurchsichtig ist. Das liegt an den unterschiedlichen Metadaten. Klicken Sie mit der rechten Maustaste auf die den Spaltenkopf „Kategorie“ und wählen Sie die Optionen „gruppieren“.
- Die so erstellten Gruppen schaffen gleich mehr Überblick und können mit einem einfachen Klick ausgeklappt werden. Klicken Sie auf die Gruppe „Speaker“. Vorgenommene Sortierungen und Filterungen gelten auch innerhalb der jeweiligen Gruppe. Daher sehen Sie zum Anfang der Speaker-Gruppe den Speaker mit dem meisten Token im gegenwärtigen Korpus.
- Ein guter Ausgangspunkt für eine tiefergehende Analyse wäre, dass unterschiedliche Speaker ausgewählt werden und diese dann separat betrachtet werden.

Video 9 Schnappschüsse

- Bisher wurden alle Analysen auf Basis aller in das Projekt geladenen Korpora durchgeführt. D.h. aber auch, dass bisher nur Aussagen über die Gesamttextmenge getroffen werden konnten. Im Folgenden werden Sie mehrere Möglichkeiten kennen lernen, Schnappschüsse zu erstellen und damit ihre Analyse zu präzisieren.
- Starten Sie den CorpusExplorer und laden Sie mindestens ein Korpus.
- Klicken Sie dann im Programm Menü auf das Schnappschuss-Symbol und wählen Sie Schnappschuss-Übersicht aus. Sie gelangen auf die Schnappschuss-Übersichtsseite, wo Sie die Schnappschüsse verwalten können.
- Der aktive Schnappschuss ist immer fett hervorgehoben und erscheint zudem jederzeit oben rechts im CorpusExplorer Fenster. Klicken Sie auf einen anderen Schnappschuss, um diesen auszuwählen. Für jedes geladene Korpus erstellt der CorpusExplorer automatisch einen Schnappschuss, der alle Texte des Korpus umfasst. Außerdem existiert ein dynamischer Schnappschuss. „Alle Texte & Korpora (dynamisch)“ – Dieser ist die Analysegrundlage, wenn Sie sich nicht um Schnappschüsse im CorpusExplorer kümmern... er soll den Einstieg erleichtern. Die Unterschiede zu einem normalen Schnappschuss: Der dynamische Schnappschuss verändert sich, wenn neues Textmaterial hinzugefügt wird und er wird nicht in der Projektmappe gespeichert.
- Klicken Sie im rechten Menü auf „Neu“. Es werden Ihnen drei Optionen angezeigt. Individuell, Autosplit und Zufällig. „Individuell“ erlaubt es, individuelle Filter zu definieren. Mit „AutoSplit“ können Sie Schnappschüsse automatisch auf Basis von Dokumentmetadaten erstellen. Die Option „Zufällig“ erlaubt es, ein zufälliges Sample zu generieren.
- Klicken Sie auf „individuell“. Vergeben Sie zuerst einen prägnanten Schnappschuss-Namen. „Freiheit(en)“
- Zur Auswahl stehen Ihnen drei Filterbausteine: Korpus, Meta-Bedingung und Volltext-Bedingung. „Korpus vollständig“ bedeutet, dass Sie nur bestimmte Korpora betrachten möchten. Die Volltext Bedingung erlaubt es Ihnen, den Inhalt der Dokumente zu durchsuchen. „Metabedingungen“ filtert hingegen Dokumente anhand ihrer Metadaten.
- Klicken Sie auf „Volltextbedingung“. In der linken Auswahlbox können Sie den abzufragenden Layer auswählen. In der rechten Auswahlbox stehen verschiedene Such-Operatoren zur Verfügung. In der Textbox darunter können Sie dann die Abfrage definieren.
- Geben Sie hier zum Beispiel „Freiheit“ ein. Bereits bei der Eingabe erscheint eine Liste - die Autovervollständigung. Diese Liste enthält nur verfügbare Begriffe. Um einen Begriff aus der Liste zu übernehmen, klicken Sie ihn entweder mit der Maus an, oder bestätigen Sie ihn mit einem Druck auf Enter (Freiheiten). Wenn Sie nicht durch die Liste scrollen wollen, können Sie

- auch ein Semikolon eingeben, um den Begriff zu übernehmen (Meinungsfreiheit). Nur grün hinterlegte Begriffe werden von der Suche verwendet. Um einen Begriff aus der Liste zu entfernen, können Sie ihn entweder markieren und löschen oder Sie klicken auf das X. (Meinungsfreiheit).
- Werfen wir abschließend einen Blick auf die gesamte Abfrage. Wir erstellen einen Schnappschuss mit folgendem Namen: „Freiheit(en)“. Dieser Schnappschuss fragt den Volltext jedes Dokuments einzelnen ab, ob in dessen Wort Layer mindestens ein Wert aus der Liste enthalten ist – in diesem Fall Freiheit und Freiheiten. Mit einem Klick auf „Schnappschuss erstellen“ wird der Schnappschuss generiert.
 - Neben der Möglichkeit, neue Schnappschüsse zu erstellen, besteht die Möglichkeit, bestehende Schnappschüsse noch weiter einzugrenzen.
 - Klicken Sie dazu auf „eingrenzen“. Auch hier sehen Sie die gleichen Funktionen, die Sie bei neuen Schnappschüssen
- haben, nur mit dem Unterschied, dass kein komplett neuer Schnappschuss erstellt wird, sondern der aktuell gewählte Schnappschuss weiter reduziert wird.
- Klicken Sie auf „Auto Split“. Im erscheinenden Dialog können Sie die Metaangabe auswählen.
 - Wählen Sie unter Dokument-Metadaten „Speaker“ aus. Außerdem sollten Sie angeben, um welchen Datentyp es sich handelt. In diesem Fall ist es eine Textinformation. Sie sehen aber auch, dass es für Zahlen und Datumsangaben weitere Optionen gibt. Klicken Sie auf „o. k.“ Es werden nun automatisch für jede Metanagabe eigene Schnappschüsse erstellt. Somit wird für jeden Speaker, der innerhalb des Korpus den Begriff „Freiheit“ verwendet, ein separater Schnappschuss angelegt.
 - Wählen Sie einen der Schnappschüsse aus und starten Sie die Analyse „Frequenzanalyse“ mit der „Tabellen“ Ansicht. Rechts oben sehen Sie, dass
- Sie auf dem gewählten Schnappschuss arbeiten. Über das Schnappschussmenü können Sie jederzeit in einen anderen Schnappschuss wechseln – Die Ansicht wird dann automatisch neu berechnet.
- Die meisten Analysen im CorpusExplorer bieten die Möglichkeit, einen Schnappschuss direkt aus der Analyseansicht heraus erstellen. Dazu klicken Sie zum Beispiel in der Tabelle auf die gewünschte Zeile, sodass sie hervorgehoben wird.
 - Klicken Sie dann im Tabellenmenü auf das Symbol für „Schnappschuss erstellen“ und tragen Sie dort den gewünschten Namen ein. Bestätigen Sie Ihre aus Eingabe mit „OK“ – Ein Blick in die Schnappschussübersicht zeigt das Ergebnis.
 - Schnappschüsse können außerdem untereinander kombiniert werden. Dafür stehen rechts die folgenden Mengenoperatoren zur Verfügung. Außerdem finden Sie hier die Funktion zum Importieren und Exportieren von Schnappschüssen. Exportieren erlaubt

es, einen Schnappschuss und damit alle zugrundeliegenden Korpusdaten in ein externes Format zu exportieren. Die Option „AntiClone“ ist besonders für Zeitungs- und Tweet-Korpora interessant. Mit dieser Option können Sie identische/ähnliche Texte aus dem Korpus eliminieren.

Video 10 NGramme

- Starten Sie den *CorpusExplorer* und laden Sie ein Korpus. Danach rufen Sie die Schnappschuss-Übersicht auf.
- Erstellen Sie einen neuen individuellen Schnappschuss mit dem Titel „Freiheit“. Dieser sollte eine Volltext Bedingung enthalten, die den Wort-Layer abfragt. Als Werte tragen sie „Freiheit“ und „Freiheiten“ ein. Erstellen Sie den Schnappschuss. Wechseln Sie zum gerade erstellten Schnappschuss.
- Wenn Sie an Sprachgebrauchsmustern interessiert sind, können Sie z. B. die N-Gramm-Analyse des *CorpusExplorer* nutzen, um Mehrworteinheiten zu erzeugen.

- Klicken Sie auf „Analysen“ > „Phrasen und Muster“ und dann auf „NGramm-Tabelle“.
- Die vorgegebenen Werte von „N = 5“ und „Musteranzahl = 0“ können Sie für einen schnellen Test belassen. Klicken Sie auf den grünen Startpfeil.
- Sortieren Sie zunächst die Tabelle nach absteigender Frequenz. Filtern Sie dann die Spalte N-Gramm nach „Freiheit“. Es ist zu erkennen, dass Freiheit mit Sicherheit, Rechten und Justiz häufig in einer Mehrworteinheit verbunden wird.
- Auch in dieser Tabelle haben Sie die Möglichkeit, Belege mittels Textlinse anzuzeigen.

Video 11 Kookkurenzen

- Starten Sie den *CorpusExplorer* und laden Sie ein Korpus. Danach rufen Sie die Schnappschuss-Übersicht auf.
- Erstellen Sie einen neuen individuellen Schnappschuss mit dem Titel „Freiheit“.

Dieser sollte eine Volltext Bedingung enthalten, die den Wort-Layer abfragt.

- Als Werte tragen sie „Freiheit“ und „Freiheiten“ ein. Erstellen Sie den Schnappschuss. Wechseln Sie zum gerade erstellten Schnappschuss. Klicken Sie auf Analyse > Kookkurrenz > Tabelle. Ordnen Sie die Tabelle nach absteigender Signifikanz und tragen Sie dann in die Filterspalte „Freiheit“ ein.
- Durch das Ausklappen der Text-Belege lässt sich z. B. folgende Phrase identifizieren: „Raum der Freiheit, der Sicherheit und des Rechts“ Sie ist maßgeblich dafür verantwortlich, dass Freiheit, Raum, Sicherheit sowie Recht starke signifikante Kookkurrenzpartner sind. Eine Möglichkeit besteht hierbei darin, dass alle Speaker, die diese Phrase verwenden, in einem Schnappschuss gebündelt werden. Dadurch können gruppierungstypische Sprachmuster identifiziert werden, ohne zuvor eine bestimmte Gruppe als solche zu definieren.