

KAMOKO

KAsseler MOrgenstern KORpus

Handbuch zur Nutzung des online-Korpus

Angela Schrott (Universität Kassel)

angela.schrott@uni-kassel.de

Jan Oliver Rüdiger (Universität Siegen)

e-mail@jan-oliver-ruediger.de

1. Was ist KAMOKO?

KAMOKO (KAsseler MOrgenstern KORpus) ist ein Korpus, das basierend auf Textbeispielen zentrale Strukturen sowie Formen und Funktionen des Französischen behandelt. Das strukturierte und kommentierte Korpus begleitet in verschiedenen Kontexten – Universität, Schule oder individuelles Sprachtraining – das Erlernen der französischen Sprache und richtet sich an alle, die Strukturen, Formen und Funktionen des Französischen vertieft kennenlernen wollen, sei es im Rahmen der Sprachpraxis oder der französischen Sprachwissenschaft.

Das Korpus geht auf eine umfangreiche Beispielsammlung des Linguisten und Sprachlehrers Claus Morgenstern (Ludwig-Maximilians-Universität München) zurück, die an der Universität Kassel digitalisiert und ausgebaut wurde. Es besteht aus 22 Einheiten, die jeweils zentrale Strukturen des Französischen phänomenbasiert auf der Grundlage authentischer Beispiele behandeln. Diese Textbeispiele stammen aus fiktionalen (Romane, Novellen) und nicht-fiktionalen Texten (journalistische Texte, Reiseführer).

Die in KAMOKO gesammelten Textbeispiele sind nach linguistischen Themen zu Lehreinheiten bzw. Subkorpora geordnet, in denen sprachliche Strukturen anhand von authentischen Textbeispielen erklärt und veranschaulicht werden. Jede Einheit widmet sich einem Thema des Sprachgebrauchs in Form einer strukturierten Abfolge von Beispielen, die Schritt für Schritt von einfachen Verwendungen zu immer komplexeren Funktionen führen. Während die Verwendungen am Beginn der Progression Fälle illustrieren, die geläufig in Grammatiken für Schule und Studium erläutert werden, illustrieren die komplexeren Verwendungen Funktionen, die man eher in sprachwissenschaftlichen Grammatiken findet. Während die ersten Beispielsätze einer Lehreinheit typische Verwendungen wiedergeben, stehen am Ende der Lehreinheit Verwendungen, bei denen es nicht mehr nur um sprachliche Korrektheit, sondern um feine Nuancen geht, die für das Textverstehen aber oft entscheidend sind.

Jedes Textbeispiel bietet neben dem Originaltext auch Variationen des Originals, in denen Formen ausgetauscht, Textelemente hinzugefügt oder weggenommen werden. Jedes Beispiel bietet damit mehrere

Varianten, von denen eine das Original ist. Jede Variante ist mit den Bewertungen französischer Muttersprachler verknüpft, die die Varianten als akzeptabel, bedingt akzeptabel oder nicht akzeptabel bewerten. Dieses Muster von Original und Variation in Verbindung mit Sprecherbewertungen bietet zwei Optionen der Nutzung.

Die **erste Option** ist die **Verwendung für die französische Sprachpraxis an Universitäten sowie die Nutzung der einfacheren Textbeispiele im Französischunterricht an Schulen**. Die didaktische Grundidee ist, dass die Studierenden die Varianten zunächst selbst bewerten und ihre Entscheidungen dann anhand der muttersprachlichen Bewertungen überprüfen und ggf. korrigieren. Auf diese Weise können Strukturen, Formen und Funktionen textbasiert in konkreten Verwendungskontexten erarbeitet werden. Die Progression der Textbeispiele veranschaulicht, wie grammatische Strukturen ineinandergreifen und wie die Kommutation einer Form durch eine andere einen Sinnunterschied erzeugt. Die Studierenden können nicht nur die sprachlich korrekten Verwendungen eigenständig erarbeiten, sondern auch die Spannung von grammatischer Regel und komplexer Sprachverwendung erfahren und als ein Charakteristikum von Sprache und Sprechen begreifen. Die Wahl zwischen den Varianten eröffnet die Möglichkeit, selbstständig über den eigenen Sprachgebrauch zu reflektieren. Für die Verwendung im schulischen Französischunterricht eignet sich KAMOKO vor allem als Grundlage für die Erstellung von Arbeitsmaterialien. KAMOKO kann von fortgeschrittenen Schülerinnen und Schülern genutzt werden, doch bietet sich vor allem eine durch die Lehrkräfte vermittelte Nutzung an, etwa wenn das Korpus als Datenbank für die Erstellung von Arbeitsblättern genutzt wird, die die Lehrwerke ergänzen.

Die zweite Option ist es, **KAMOKO als ein mit muttersprachlichen Bewertungen annotiertes linguistisches Korpus** zu sehen, das genutzt werden kann, um semantisch-pragmatische Profile von Formen und Strukturen korpusbasiert zu erarbeiten. Für diesen Zugriff sind vor allem die komplexeren Verwendungen relevant, die über die typischen Verwendungen hinausgehen und Nuancen des Sprachgebrauchs ausloten, die für linguistische Analysen interessant sind.

2. Technische Hinweise zur Nutzung

Das KAMOKO-Korpus kann kostenfrei genutzt werden.

Das KAMOKO-Korpus lässt sich am besten in Verbindung mit dem **KAMOKO-Editor** verwenden, der alle wichtigen Grundfunktionen für die Arbeit mit dem Korpus bietet. Für eine intensivere und detailliertere linguistische Nutzung kann später der **CorpusExplorer** hinzugezogen werden.

Beide Programme lassen sich nur mit Windows (ab Version 7) verwenden.

Sollten Sie mit Linux oder MAC arbeiten, können Sie die XML-basierten Dateien mit einem externen Programm (bspw. Atom oder VisualStudio, beide kostenlos) öffnen und so die Daten nutzen.

Folgende Tools bieten sich zur Arbeit mit XML-Dateien an:

<https://www.oxygenxml.com/> (Windows, MAC, Linux)

<https://panic.com/coda/> (MAC)

<https://www.sublimetext.com/> (Windows, MAC, Linux)

<https://atom.io/> (Windows, MAC, Linux; kostenlos)

<https://code.visualstudio.com/> (Windows, MAC, Linux; kostenlos)

Darüber hinaus werden die KAMOKO-Dateien in verschiedenen anderen Korpus-Formaten bereitgestellt und über die Projekthomepage zum Download angeboten. Folgende Formate werden unterstützt: TXM, CoNLL, CorpusWorkbench, TXT, SPEDDY, TreeTagger, WebLicht und XML.

3. KAMOKO: Übersicht über die Teilkorpora

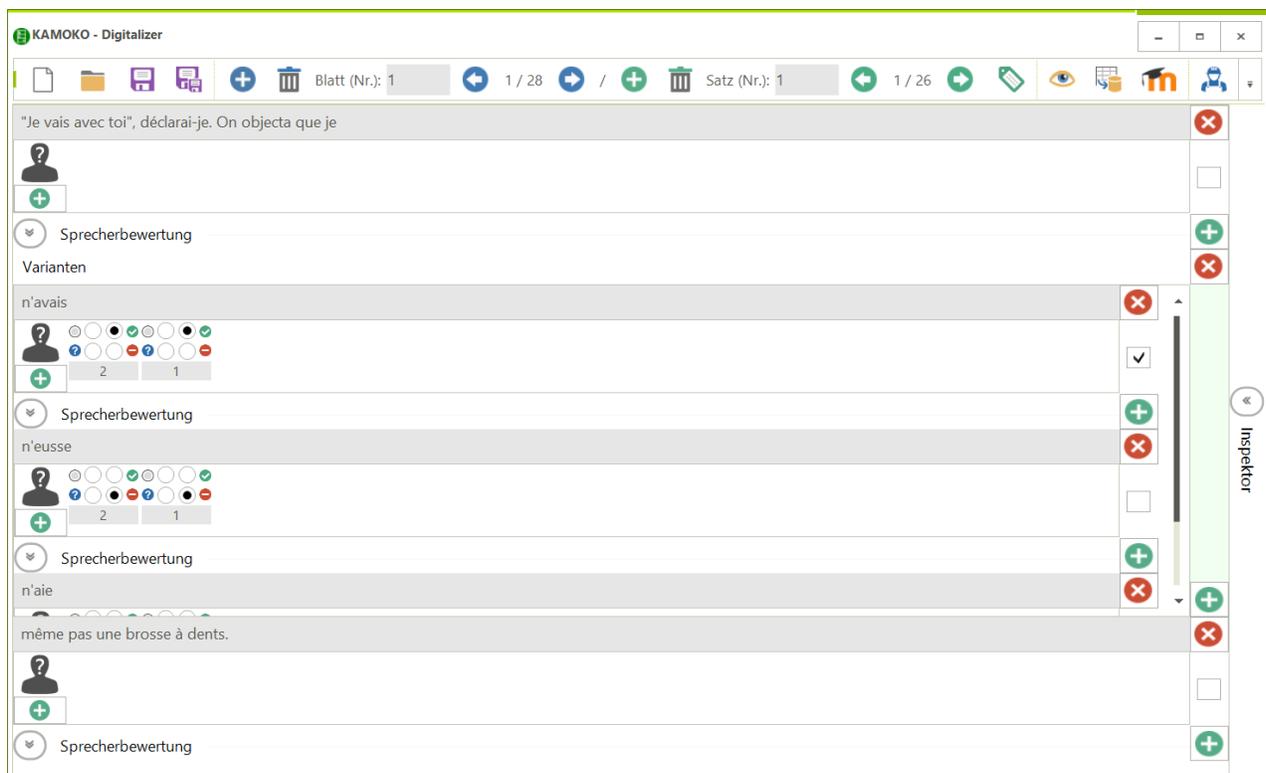
KAMOKO besteht insgesamt aus 22 Teilkorpora bzw. Kursen. Im Folgenden finden Sie eine tabellarische Auflistung mit Informationen zum Umfang (Blätter, Beispiele, Token) der einzelnen Einheiten.

Titel	Blätter	Beispiele	Token
Futur	36	826	415.455
Infinitiv	28	595	62.377
<i>ant</i> -Formen (<i>gérondif, participe présent</i>)	17	95	2.443
Modus I	28	752	49.571
Modus II	31	679	42.217
Tempus I	40	723	403.106
Tempus II	50	896	408.035
Tempus und Modus im Relativsatz	13	156	3.658
Verbale Konstruktionen I	36	820	47.266
Verbale Konstruktionen II	6	79	5.014
Interrogativsätze und Relativsätze	29	638	123.369
Infinitiv vs. Nebensatz	16	292	5.701
Negation	38	801	89.403
Adjektive	36	734	212.399
Adverbien	23	498	330.247
Präpositionen I	36	993	122.116
Präpositionen II	27	840	200.344
Präpositionen III	27	581	95.724
Präpositionen IV	8	175	7.305
Präpositionen V	19	195	4.715
Artikel	30	781	57.407
Personalpronomina	28	583	100.997
<i>Summe</i>	<i>591</i>	<i>12.732</i>	<i>2.791.571</i>

4. KAMOKO und die französische Sprachpraxis an Universitäten und Schulen

Mit seiner Progression von typischen Verwendungen einer Form oder Struktur bis hin zu feinen semantisch-pragmatischen Nuancen, die über das Konzept der sprachlichen Korrektheit hinausgehen, bietet sich KAMOKO vor allem als Korpus für die Sprachpraxis im Französischstudium an. Bei dieser Nutzung arbeiten Studierende mit dem Korpus und den Beispielvarianten, sie bewerten die Varianten und vergleiche ihre Auswahl mit den muttersprachlichen Kommentaren. Für diese Art der Nutzung haben wir ein Video-Tutorial zusammengestellt. Zum Nachlesen und Nachschlagen finden Sie in diesem Abschnitt eine Übersicht, die die wesentlichen Schritte dokumentiert

Voraussetzung für die Nutzung ist, dass Sie den [KAMOKO-Editor](#) von der Webseite herunterladen und auf einem Windows-PC installieren. Außerdem sollten Sie mindestens eine der [KAMOKO-XML-Dateien](#) herunterladen. Im Tutorial wird die Datei „Modus.kamoko.xml“ verwendet. Wenn Sie den KAMOKO-Editor öffnen, erscheint die Programmoberfläche. Klicken Sie auf das Ordner-Symbol in der Menüleiste um einen KAMOKO-Kurs zu laden. Es erscheint ein Dialog, in dem Sie die zuvor heruntergeladene KAMOKO-XML-Datei auswählen und öffnen können.



Jeder Kurs ist in Blätter unterteilt. Die einzelnen Blätter beinhalten Textbeispiele, die an der Programmoberfläche als „Satz“ bezeichnet werden. Der KAMOKO-Editor zeigt jeweils ein Beispiel aus einem Blatt an. Über die Steuerelemente des Editors können Sie durch die Blätter und Beispiele navigieren.

Zwei Steuerelemente sind für die Navigation grundlegend: Die blauen Elemente erlauben die Navigation durch Blätter und die grünen Elemente ermöglichen die Navigation durch die im Blatt enthaltenen Beispiele. Über die Pfeile können Sie vorwärts und rückwärts durch die Blätter eines Kurses und durch die Textbeispiele navigieren.

Die Textbeispiele sind so aufgebaut, dass der Originaltext – passend zum Thema des Kurses – an einer oder mehreren Stellen variiert wird. Diese Varianten wurden von französischen Muttersprachlern bewertet, deren Bewertungen im Korpus mit den Varianten verknüpft. Im für das Tutorial gewählten Kurs *Modus* werden unterschiedliche Verbformen (*subjunctif* vs. *indicatif*) als Varianten angeboten. Das Tutorial behandelt folgendes Textbeispiel:

“Je vais avec toi”, déclarai-je. On objecta que (je n’avais même pas – je n’eusse même pas – je n’aie même pas) une brosse à dents.

Das ausgewählte Beispiel wird im Hauptbereich des Editors angezeigt. Im oberen grauen Feld sehen Sie den Satzteil, der unverändert bleibt und in den Feldern darunter den Teil des Beispiels, der variiert wurde. In diesem Beispiel sind die Varianten drei Verbformen, die sich im Modus unterscheiden. Die Verbformen sind der *indicatif présent* (je n’avais même pas), der *subjunctif passé* (je n’eusse même pas) und der *subjunctif présent* (je n’aie même pas).

Auf der Programmoberfläche sind die drei Varianten am rechten Rand durch eine hellgrüne Leiste markiert, die in allen Textbeispielen den Teil markiert, der variiert wurde. Eine Variante ist das Original, zwei Varianten

wurden als Alternativen hinzugefügt, um zu testen, ob auch andere Formen möglich sind. Das Original erkennen Sie am Häkchen rechts. Die befragten Muttersprachler wussten natürlich nicht, welche Form das Original ist, sie sahen nur drei Varianten, die sie bewerten sollten.

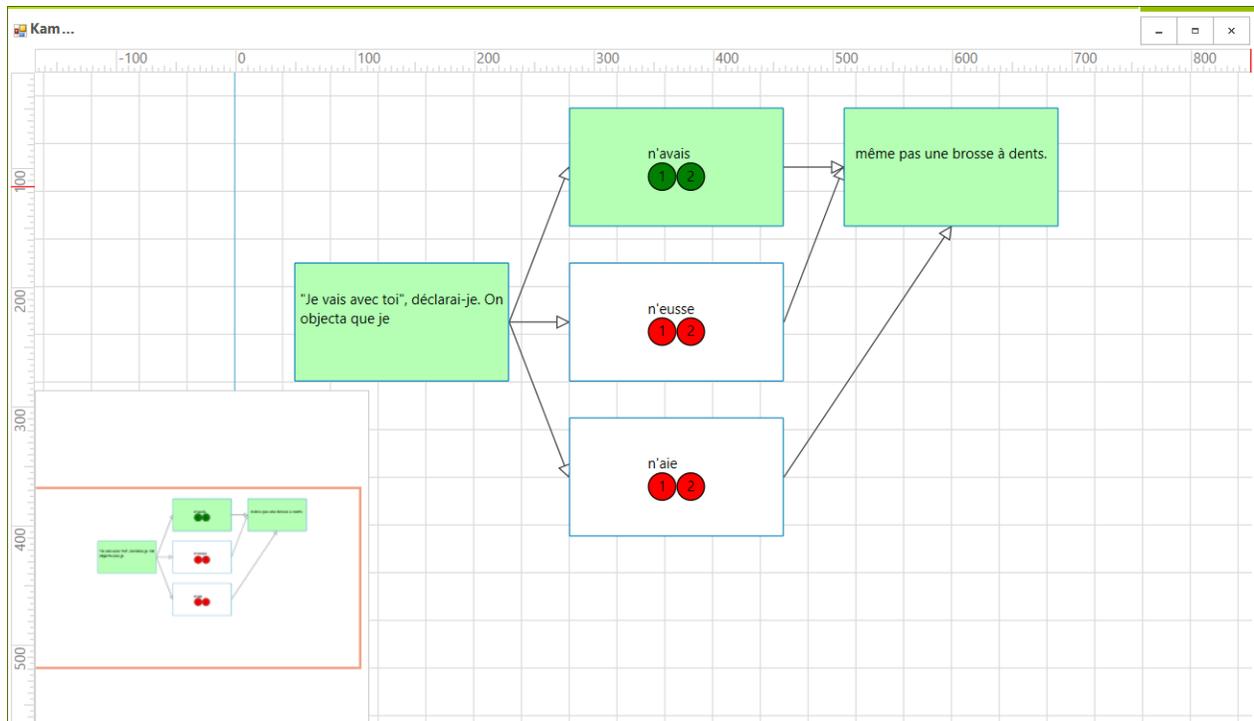
Alle die Varianten sind mit Kommentierungen hinterlegt. Die befragten Muttersprachler konnten zwischen drei Bewertungen wählen: Zustimmung, bedingte Zustimmung und Ablehnung. Zustimmung sagt aus, dass die Form als korrekt und sinnvoll bewertet wird, Ablehnung bedeutet meist, dass eine Form nicht korrekt ist, etwa wenn eine bestimmte Struktur den *subjunctif* fordert und der *indicatif* dort nicht verwendet werden kann. Die Ablehnung kann aber auch bedeuten, dass die Form in der Struktur vorkommen könnte, aber im vorliegenden Kontext den Sinn so verändert, dass das Beispiel sinnschwach oder unverständlich wird. Die bedingte Ablehnung kann mehrere Gründe haben: möglich ist, dass eine Form weniger gebräuchlich ist oder dass sie einen Sinnunterschied bewirkt, den die Muttersprachler nur bedingt passend finden. In der Notation im KAMOKO-Editor haben wir dafür folgende Symbole gewählt: Zustimmung wird durch das grüne Häkchen markiert, Ablehnung durch das rote Symbol und bedingte Zustimmung signalisiert das blaue Fragezeichen. In den Zeilen neben dem Kopf finden Sie die Bewertungen der Muttersprachler, die mit Ziffern unterschieden werden. Der schwarze Punkt markiert, für welche Bewertung sich die Muttersprachler entschieden hatten. Bei der ersten Variante (je n'avais même pas) setzen beide Muttersprachler ihren Punkt beim grünen Häkchen und stimmen damit der *indicatif*-Form zu, die übrigens auch das Original ist. Die zweite und die dritte Variante – also die beiden *subjunctif*-Formen – lehnen beide Muttersprachler ab und setzen ihren Punkt beim roten Symbol. Das Beispiel belegt damit eine sehr eindeutige Bewertung.

The screenshot shows a text input field containing the French phrase "n'avais". To the right of the text is a red 'X' icon. Below the text is a row of evaluation icons: a question mark icon, a green checkmark, a blue question mark, and a red 'X'. Below these icons are two columns of circles representing the number of speakers who chose each option: the first column has a '2' and the second has a '1'. At the bottom left is a plus sign icon, and at the bottom right is a green checkmark icon. Below the main interface is a section labeled "Sprecherbewertung" with a plus sign icon.

Um die muttersprachlichen Bewertungen auf einen Blick sehen zu können, ohne alle Notationen im Einzelnen durchgehen zu müssen, bietet der Editor ein Tool zur Visualisierung. Mit einem Klick auf das Auge-Symbol



starten Sie den Visualisierungsmodus zu starten und erhalten für das besprochene Beispiel folgende Grafik, in der die einzelnen Satzteile als Rechtecke und die Bewertungen als Kreise erscheinen. Wie gehabt, bedeutet grün Zustimmung, blau bedingte Zustimmung und rot Ablehnung. Die Visualisierung zeigt noch einmal deutlich, dass die befragten Muttersprachler einhellig dem Original zustimmen und die anderen Varianten ebenso einhellig ablehnen.



Für die Nutzung in Sprachpraxis und Französischunterricht bietet KAMOKO vielfältige Funktionen. Wir wollen Ihnen zwei Optionen näher vorstellen.

Die erste Funktion ist die **Möglichkeit einer Volltextsuche im jeweiligen Subkorpus**. Eine solche Volltextsuche kann sinnvoll sein, wenn nach der Arbeit mit dem zitierten Beispiel noch geprüft werden soll, ob es im Kurs noch weitere Texte gibt, in denen das Verb *objecter* vorkommt und wie in diesen Texten der Modusgebrauch ist. Klicken Sie zu diesem Zweck am rechten Rand auf „Inspektor“ und wählen Sie den Reiter „Suche“ aus. Den Reiter „Fehler“ dürfen Sie ignorieren, er betrifft diese Nutzung nicht.



Geben Sie einen Suchterm ein, der aus einem Wort oder einer Wortfolge bestehen kann. Die Suche eröffnet die Option, entweder die Fälle aufzulisten, bei denen das Wort oder die Wortfolge in der Position am Satzanfang steht oder die Fälle, bei denen Wort oder Wortfolge im innerhalb des Satzes und seiner Varianten vorkommen. Ein Klick auf die Lupe startet die Suche. Der Inspektor durchsucht nun das ausgewählte Teilkorpus und die Liste der Ergebnisse erscheint. Mit einem Doppelklick auf die Ergebnisse können Sie zu den einzelnen Beispielen springen.

Die zweite Funktion, die vor allem für die Erstellung von Lernmaterialien ergiebig und effektiv ist, ist die **Verknüpfung von KAMOKO mit der Lernplattform Moodle zur Generierung von Lückentexten**. In Moodle gibt es bekanntlich die Option, Lückentexte zu erstellen. Die KAMOKO-Texte eignen sich aufgrund ihrer Kommutationen und Varianten perfekt für die Erstellung von Lückentexten: Die einzelnen Beispiele können als Lückentext gespeichert werden und die Varianten erscheinen als Optionen zur Füllung der Lücke.

Mit einem Klick auf das Moodle-Symbol können Sie ein Textbeispiel für den Lückentext auswählen. Kopieren Sie diesen Text in die Zwischenablage. Nutzen Sie dann im Moodle-Kurs die Optionen „Test“ und „Lückentext“ und fügen Sie unter "Fragentext" den zuvor kopierten Text aus der Zwischenablage ein. Die variablen Stellen werden als Lücken dargestellt. Die Varianten stehen als Auswahl zur Verfügung und sind mit den Bewertungen der Muttersprachler hinterlegt, sodass der Lückentext Lernern und Lernerinnen, die eine Variante gewählt haben, ein Feedback zu ihrer Entscheidung gibt. Wählt man die Variante aus, der die Muttersprachler zugestimmt haben und damit eine korrekte Lösung, dann erhält man einen Punkt. Wählt man die Variante aus, der die Muttersprachler bedingt zugestimmt haben, erhält man einen halben Punkt. Für die Auswahl abgelehnter Varianten gibt es keine Punkte.

Frage 1

Bisher nicht beantwortet

Erreichbare Punkte: 1,00

Frage markieren

Frage bearbeiten

"Je vais avec toi", déclarai-je. On objecta que je n'avais même pas une brosse à dents.

5. Korpusbasierte Linguistik und Korpuslinguistik.

KAMOKO bietet sich auch als Grundlage für linguistische Analysen an, da die Textbeispiele Verwendungen dokumentieren, die über die üblichen Beschreibungen in Grammatiken weit hinausgehen und in Fragestellungen linguistischer Untersuchungen hineinreichen.

Besonders wertvoll sind die Textbeispiele durch die Varianten und muttersprachlichen Wertungen. Greifen wir als Beispiel die Einheit zur Adjektivstellung im Französischen heraus, die Schülerinnen und Schüler, Studierende Linguisten und Linguistinnen beschäftigt. Das KAMOKO Subkorpus zur Adjektivstellung bietet hier 734 Textbeispiele, die nicht nur das authentische Original, sondern auch Kommutationen enthalten und die zu allen Varianten muttersprachlichen Wertungen bieten. Für textbasierte Analysen steht damit ein ergiebiges Korpus zur Verfügung, das im Zuge der Untersuchung noch erweitert und ausgebaut werden kann. Ideal wäre es, wenn Linguistinnen und Linguisten, die mit KAMOKO arbeiten, das Korpus ausbauen und erweitern. Die Möglichkeiten dazu sind technisch gegeben (cf. Abschnitt 5).

Wie alle digitalen Korpora eröffnet KAMOKO zudem korpuslinguistische Nutzungen, die über die in den Einheiten thematisierten Themen hinausgehen. Die korpuslinguistische Nutzung setzt voraus, dass Sie zuvor das Programm CorpusExplorer installiert haben. Dieses können Sie unter <http://CorpusExplorer.de> kostenfrei herunterladen. Klicken Sie auf das Export-Symbol



und wählen Sie die Sprache aus, im Fall von KAMOKO ist das natürlich Französisch. Danach beginnt der automatische Annotationsprozess, der die Beispiele mitsamt ihrer Variationen tokenisiert und mit Part-of-Speech und Lemma-Informationen verknüpft. Ist der Prozess abgeschlossen, können Sie das Korpus in verschiedenen Korpusformaten abspeichern. Mehr Informationen dazu finden Sie in Abschnitt 6.

6. Über das Korpus hinaus: Erweiterungen und Modifizierungen

Die Beschreibungen unter Abschnitt 3 und 4 leiten die Nutzung des bestehenden Korpus an, so wie Sie es online vorfinden. Vor allem Linguistinnen und Linguisten, die intensiv mit dem Korpus arbeiten, werden die Textbeispiele vermutlich weiter variieren und auch weitere Sprecher befragen. KAMOKO ist prinzipiell als offenes Korpus konstruiert, das von Nutzern erweitert werden kann. Solche Erweiterungen sind sogar erwünscht. Der Idealfall wäre aus unserer Sicht, dass die *scientific community* weitere Beispiele, Wertungen

und Kommentare einspeist und das Korpus auf diese Weise an Umfang und Differenziertheit gewinnt.

Der folgende Abschnitt stellt einige einfache Tools vor, mit denen das Korpus modifiziert und erweitert werden kann. Der erste Schritt ist wieder, dass Sie das Subkorpus, mit dem Sie arbeiten wollen, herunterladen. Änderungen und Erweiterungen können zunächst nur am heruntergeladenen Korpus durchgeführt werden, aus nachvollziehbaren Gründen nicht am online-Korpus.

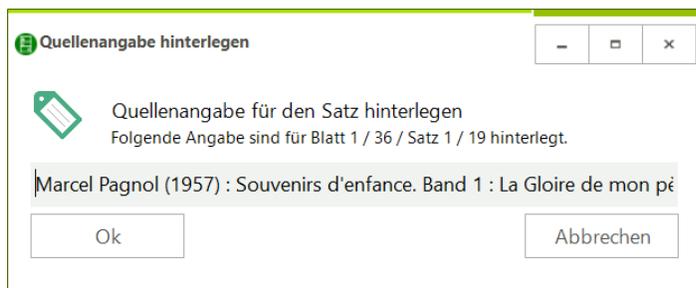
Der KAMOKO-Editor wurde speziell für die Bedürfnisse des KAMOKO-Projekts entwickelt. Der Editor bietet verschiedene Funktionen zur Verwaltung von Beispielen und zur Modifizierung von Beispielen. Zudem ermöglicht es der Editor, zu den Beispielen Metadaten zu hinterlegen. Die einzelnen Einheiten bzw. Subkorpora können bearbeitet und verändert werden. Es ist möglich, innerhalb einer Einheit Blätter hinzuzufügen oder zu löschen. Die Blätter, die aktuell lediglich durchnummeriert sind, können auch umbenannt werden, etwa entsprechend des jeweiligen thematischen Schwerpunkts. Auch die Textbeispiele können modifiziert werden; Beispiele können hinzugefügt oder gelöscht werden. Auf diese Weise können auch neue Varianten ergänzt werden, etwa indem man ein Textbeispiel noch einmal einfügt und in neuer Weise variiert. Mit dem Klick auf das Pluszeichen können Beispiele hinzugefügt werden, das Verschieben in den Papierkorb löscht ein Beispiel. Die blauen Symbole beziehen sich auf die Blätter, die grünen auf die Beispiele.



Ferner ist es möglich, einen Beispielsatz zu verändern. Sowohl der konstante Teil eines Textbeispiels als auch der variierte Teil können verändert werden und zwar über folgende Notation:



Textteile lassen sich über Plus- und X-Symbole hinzufügen bzw. entfernen. Bei einem Klick auf ein Plus erscheint eine Auswahl, ob ein konstantes oder ein variables Fragment erzeugt werden soll.



Zu jedem Beispiel können Metadaten annotiert werden. Das könnte beispielsweise eine weitere muttersprachliche Wertung sein, ein Kommentar aus einer Grammatik oder eine bibliografische Referenz auf linguistische Fachliteratur, die das jeweilige Phänomen behandelt.

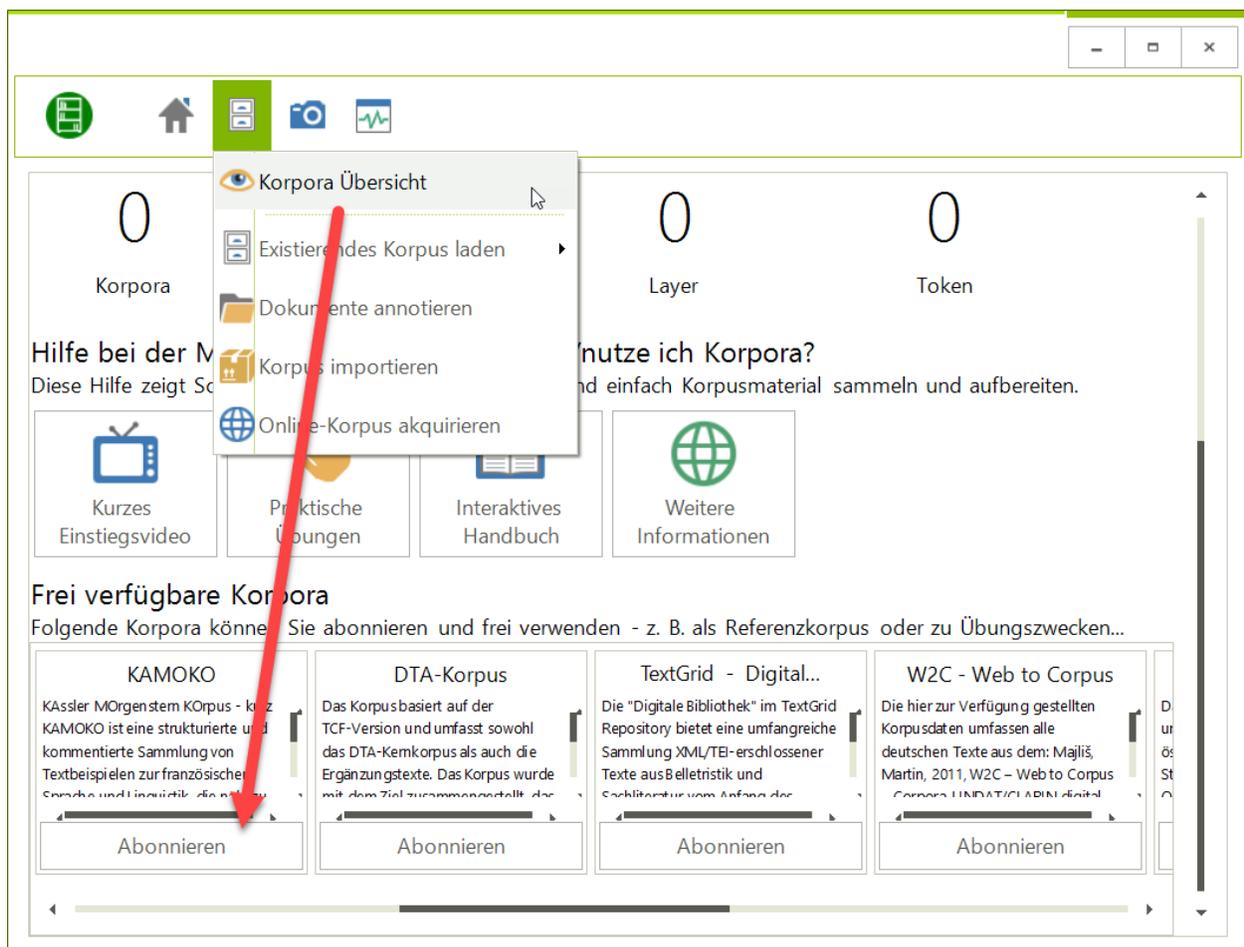
7. Der Corpus-Explorer – für Korpuslinguisten und solche, die es werden wollen

KAMOKO nutzt als Programm die digitale Plattform CorpusExplorer. Sie soll hier ebenfalls vorgestellt werden, da sie die Grundlage für KAMOKO ist und auch den Ausgangspunkt für die korpuslinguistischen Nutzungen des KAMOKO-Korpus bildet.

KAMOKO und der CorpusExplorer sind voneinander unabhängige Projekte, die an der Universität Kassel entstanden sind. Der [KAMOKO-Editor](#) basiert jedoch auf dem SDK (Software Development Kit) des CorpusExplorers. Dieser Abschnitt gibt einen Einblick in Analysemöglichkeiten, die sich mit den fertig annotierten Daten durchführen lassen. Zum besseren Verständnis ist diese Anleitung als *Hands-on Lab* konzipiert, so dass anhand einfacher Schritte die Funktionen nachvollzogen werden können. Dabei ist es durchaus erwünscht, den beschriebenen Weg an interessanten Stellen zu verlassen und eigene Parameter, Methoden und Analysen zu erkunden.

Laden Sie die kostenfreie Version des CorpusExplorers herunter (<http://www.CorpusExplorer.de>) und installieren Sie diesen. Nach der Installation startet der CorpusExplorer automatisch.

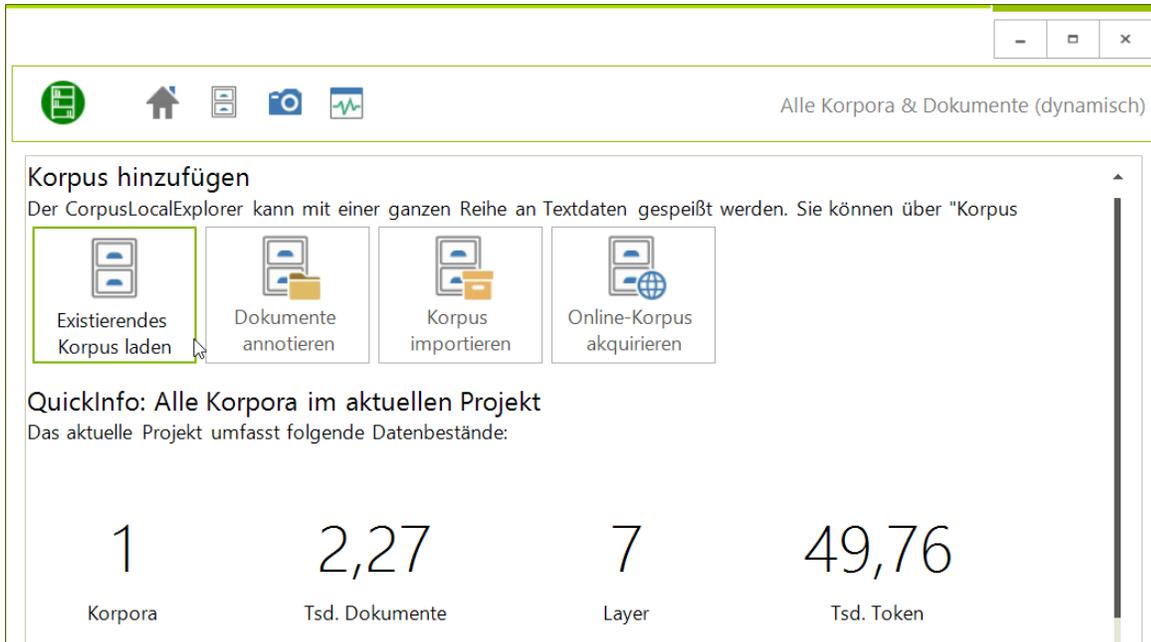
Klicken Sie im Hauptmenü auf das Korpus-Symbol und wählen Sie „Korpus-Übersicht“ aus. Scrollen Sie in der „Korpus-Übersicht“ ganz nach unten, dort finden Sie eine Liste mit frei verfügbaren Korpora. Suchen Sie in der Liste nach „KAMOKO“, klicken Sie auf „Abonnieren“ und bestätigen Sie die Installation.



The screenshot displays the Corpus Explorer web interface. At the top, there is a navigation bar with icons for home, corpus, and analysis. A dropdown menu is open, showing options: 'Korpus Übersicht', 'Existierendes Korpus laden', 'Dokumente annotieren', 'Korpus importieren', and 'Online-Korpus akquirieren'. A red arrow points from the 'Korpus Übersicht' option in the menu down to the 'KAMOKO' entry in the 'Frei verfügbare Korpora' section. Below the menu, there are statistics for 'Korpora', 'Layer', and 'Token', all showing '0'. There are also buttons for 'Kurzes Einstiegsvideo', 'Praktische Übungen', 'Interaktives Handbuch', and 'Weitere Informationen'. The 'Frei verfügbare Korpora' section lists several corpora with 'Abonnieren' buttons: KAMOKO, DTA-Korpus, TextGrid - Digital..., and W2C - Web to Corpus.

Starten Sie abschließend den CorpusExplorer neu. Abonnierte Korpora werden bei jedem Start des CorpusExplorers auf Updates überprüft. Bestätigen Sie die Installation des Updates.

Rufen Sie erneut die „Korpus-Übersicht“ auf und klicken Sie auf den Button „Existierendes Korpus laden“ – eine Liste aller verfügbaren Korpora wird angezeigt. Wählen Sie aus der Liste einen beliebigen KAMOKO-Kurs aus. In diesem Beispiel wird der Kurs „Modus“ genutzt. Sie können aber auch gerne andere Kurse nutzen oder mehrere Kurse nacheinander laden. Wurde ein Kurs erfolgreich geladen, wird die Übersichtsstatistik automatisch aktualisiert.



Alle Korpora & Dokumente (dynamisch)

Korpus hinzufügen

Der CorpusLocalExplorer kann mit einer ganzen Reihe an Textdaten gespeißt werden. Sie können über "Korpus

- Existierendes Korpus laden
- Dokumente annotieren
- Korpus importieren
- Online-Korpus akquirieren

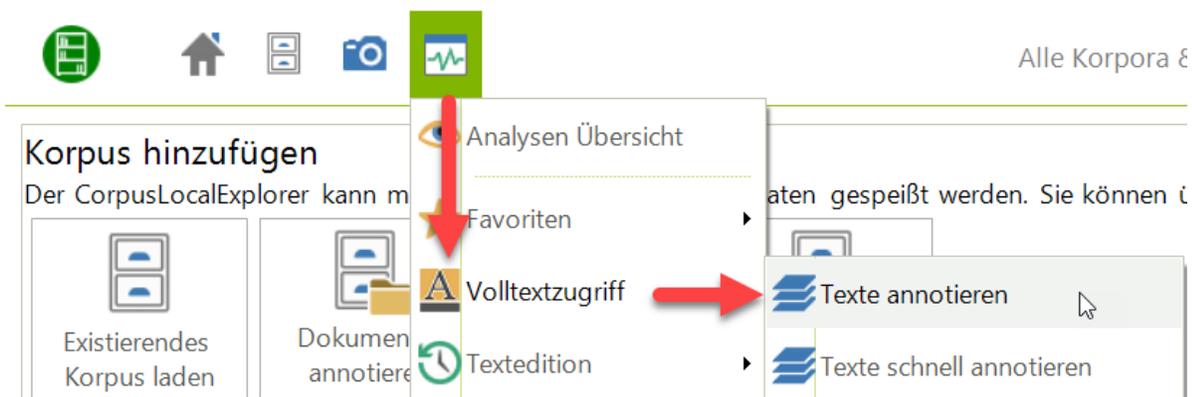
QuickInfo: Alle Korpora im aktuellen Projekt

Das aktuelle Projekt umfasst folgende Datenbestände:

1	2,27	7	49,76
Korpora	Tsd. Dokumente	Layer	Tsd. Token

Geladen ist der Kurs „Modus“, dieser umfasst 2,27 Tausend generierte Beispiele, mit 49,79 Tausend Token. Die Daten, wie z. B. Lemma, POS (Part-of-Speech) oder Sprecher/Muttersprachler-Bewertungen wurden in separaten Layern gespeichert.

Wählen Sie im Analyse-Menü „Volltextzugriff“ und dann „Texte annotieren“ aus. In diesem Analysemodul kann die Annotation (manuell: Muttersprachler / automatisch: POS/Lemma/Satz) betrachtet, verglichen und angepasst werden.



Alle Korpora & Dokumente (dynamisch)

Korpus hinzufügen

Der CorpusLocalExplorer kann mit einer ganzen Reihe an Textdaten gespeißt werden. Sie können über "Korpus

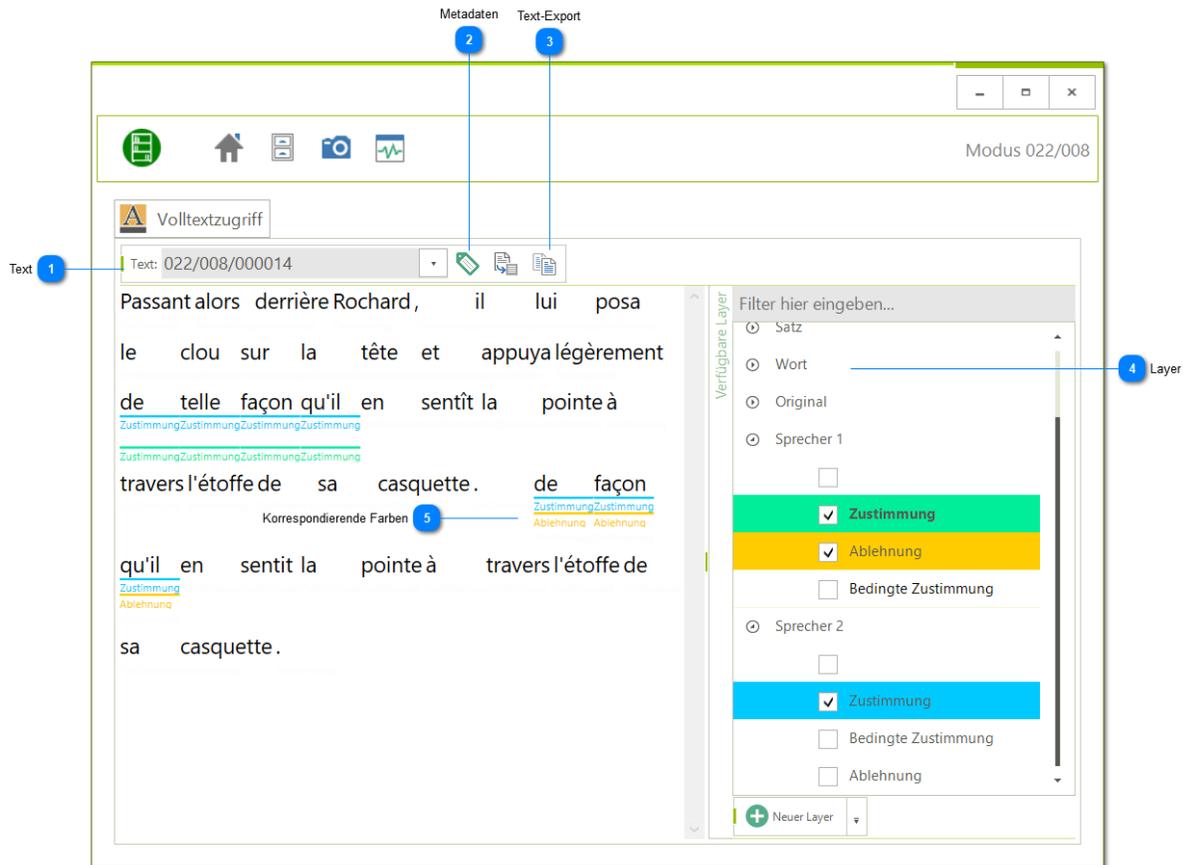
- Existierendes Korpus laden
- Dokumente annotieren
- Korpus importieren
- Online-Korpus akquirieren

Analyse

- Analysen Übersicht
- Favoriten
- Volltextzugriff
- Textedition

- Texte annotieren
- Texte schnell annotieren

Dieses Analysemodul erlaubt es, ein Beispiel mit allen Annotationen anzuzeigen.



Hier sind folgende Bereiche relevant:

1. Text

Text: 022/008/000014

Im KAMOKO-Korpus existieren eindeutige Text-IDs zur Identifikation der generierten Beispiele. Der erste Block (004) ist die Nummer des Blattes, der zweite Block (006) die Nummer des Satzes.

Während die ersten beiden Blöcke Bestandteil des KAMOKO-XML-Formats sind, wird der dritte Block (000353) erst bei der Erzeugung generiert. Er gibt an, welches Beispiel aus Blatt 4 / Satz 6 erzeugt wurde. Im Beispiel bedeutet dies, es gibt mindestens 352 weitere Beispiele mit ähnlichem Inhalt.

2. Metadaten



Klicken Sie auf dieses Symbol, um die Metadaten zum aktuell gewählten Dokument vollständig anzuzeigen. Es erscheint der folgende Dialog:

Dokument-Metadaten

+ Hinzufügen

Blatt	004
Satz	006
Blatt/Satz	004/006
Quelle	
Original abgelehnt?	Ja
KAMOKO-DIDX	0ae62d3c-73a5-4bb5-a1
KAMOKO-SIDX	e03130be-45a0-4814-98
Bewertung 1	Ablehnung

Ok Abbrechen

Die Angaben Blatt/Satz und Quelle sind bereits im KAMOKO-Editor einsehbar. Die Angaben KAMOKO-DIDX und SIDX sind Teil des KAMOKO-XML-Formats und identifizieren eine(n) Blatt (DIDX) bzw. Satz (SIDX) eindeutig und über mehrere Korpusversionen hinweg. Die „Bewertung 1“ sowie weitere Bewertungen geben die aggregierte Bewertung des Sprechers (in diesem Fall 1) aus. Nur wenn alle Stellen als „Akzeptiert“ von einem Sprecher gekennzeichnet wurden, wird hier „Akzeptiert“ ausgegeben, bei min. einer „Bedingte Zustimmung“ bzw. min. einer „Ablehnung“ wird die Gesamtbewertung entsprechend herabgesetzt. Die Angabe „Original abgelehnt?“ zeigt an, dass im Beispiel eine Originalstelle von mindestens einem Sprecher abgelehnt wurde (wird im KAMOKO-Editor als Warnung hervorgehoben).

3. Text-Export



Über diese Symbole kann der gerade angezeigte Text als Korpus (Datei) oder als Plain-Text (Zwischenablage) exportiert werden.

4. Layer

- Wort
- Original
- Sprecher 1

Klicken Sie auf die Kreise vor den Layer-Namen, um alle Layer-Werte anzuzeigen. Vor jedem Layer-Wert ist ein Kästchen zu sehen. Setzen Sie in diesen Kästchen einen Haken, um den Wert anzuwählen. Die Farben werden dynamisch vergeben - parallel werden im Text (links) alle Token hervorgehoben, die mit diesem Wert annotiert wurden.

5. Korrespondierende Farben

de telle façon qu'il en sentît la pointe à
Zustimmung Zustimmung Zustimmung Zustimmung

travers l'étoffe de sa casquette. de façon
Zustimmung Zustimmung
Ablehnung Ablehnung

qu'il en sentit la pointe à travers l'étoffe de
Zustimmung
Ablehnung

Original
Sprecher 1

 Zustimmung
 Ablehnung
 Bedingte Zus

Die Farben im Text links entsprechen den automatisch zugeordneten Farben der Layer-Werte.

Nachdem Sie sich etwas mit den Analysemöglichkeiten vertraut gemacht haben, rufen Sie das Analyse-Menü erneut auf und wählen Sie „Frequenzanalyse > Tabelle“. Der CorpusExplorer führt automatisch eine Frequenzanalyse über die Layer POS, Lemma und Wort durch. Wie Sie gesehen haben, bietet KAMOKO weitaus mehr Möglichkeiten, als klassische Korpora, insbesondere durch die Sprecherbewertungen. Wir wollen diese Möglichkeiten nutzen. Klicken Sie daher in der Analyseansicht auf das blaue Layer-Symbol. Ein Dialog erscheint. Treffen Sie die gleichen Einstellungen wie im Screenshot und bestätigen Sie mit „Ok“.

Alle Korpora & Dokumente (dynamisch)

Frequenzanalyse

Ziehen Sie den Spaltenkopf hier her, um die Tabelle zu gruppieren.

	POS	Lemma	Wort	Frequenz
Beinhaltet:				Ist gleich:
	31			403286
ADV				507
ADV				64
ADV				151
ADV				115
ADV				228
ADV				102
ADV				3063
ADV				4
ADV				16
ADV		tout	tout	314

Layer auswählen

In diesem Dialog können Sie die Layerkonfiguration

Layer 1

Sprecher 1

Layer 2

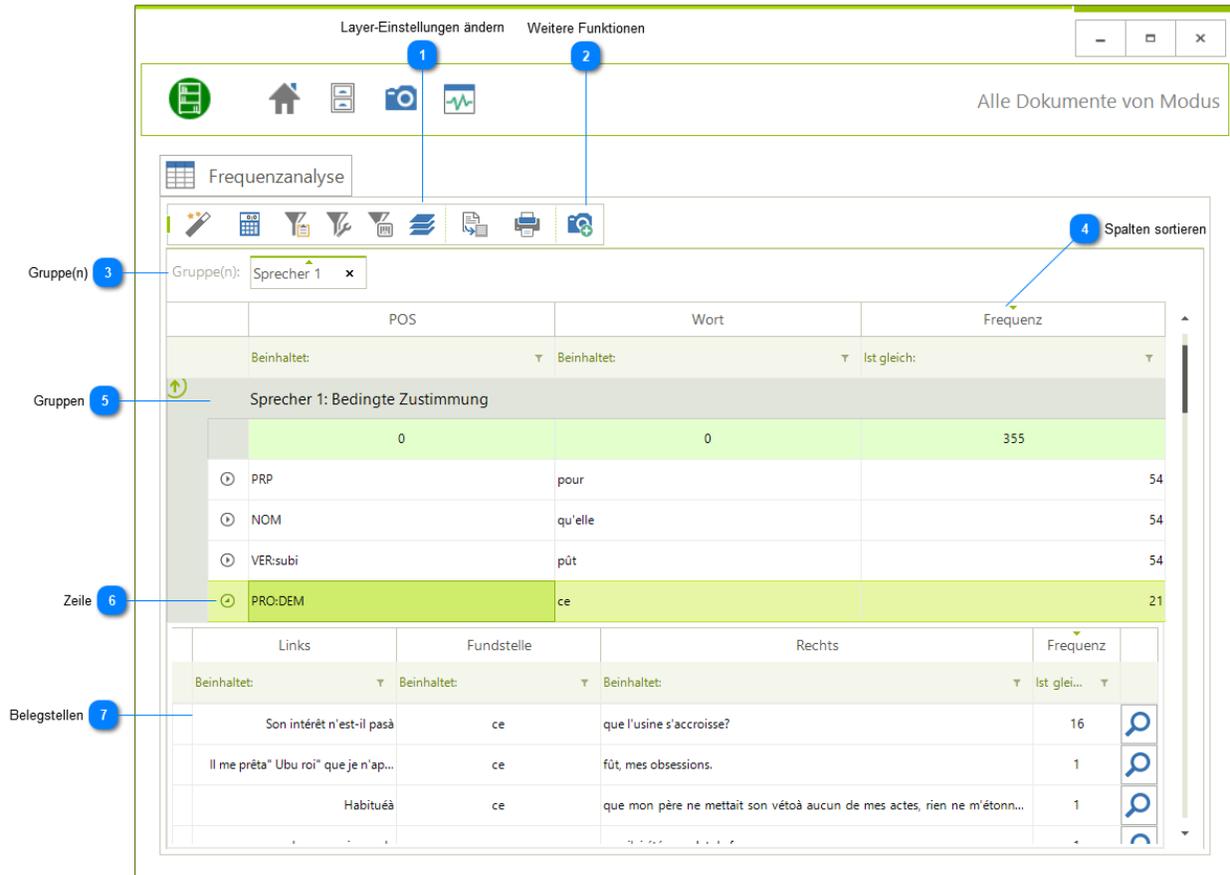
POS

Layer 3

Wort

Ok Abbrechen

Die Tabelle wird automatisch aktualisiert. Sie sehen eine neue Spalte "Sprecher 1" rechts. Durch Ziehen des Spaltenkopf, können Sie Spalten vergrößern und verschieben. Klicken Sie mit der rechten Maustaste auf den Spaltenkopf "Sprecher 1" und klicken Sie dann auf "Gruppieren". Die gruppierten Spalten können ausgeklappt werden. Im Beispiel wurde "Bedingte Zustimmung" ausgeklappt. Klicken Sie mit links auf einen Spaltenkopf, um die Daten auf- bzw. absteigend zu sortieren.



Folgende Bereiche sollten Sie sich näher ansehen:

1. Layer-Einstellungen ändern



Der CorpusExplorer erlaubt eine Frequenzanalyse über bis zu drei Layer.

2. Weitere Funktionen



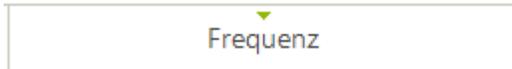
Erproben Sie diese Funktionen auf eigene Faust oder werfen Sie einen Blick in das CorpusExplorer-Handbuch.

3. Gruppe(n)



Tabellenspalten lassen sich gruppieren. Diese Option ist besonders hilfreich für Spalten, die wenig unterschiedliche Werte enthalten, wie z. B. Sprecher, Original oder POS. Um eine Spalte zu gruppieren, klicken Sie mit rechts auf den Spaltenkopf und wählen Sie „Gruppieren“ aus.

4. Spalten sortieren



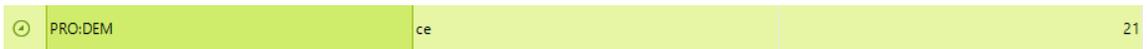
Klicken Sie mit links auf einen Spaltenkopf, um aufsteigende oder absteigende (gekennzeichnet durch kleinen grünen Pfeil) Sortierung zu aktivieren.

5. Gruppen



Gruppen können erweitert werden. Die gruppierte Spalte wird nicht angezeigt, da alle Zeilen innerhalb dieser Gruppen den identischen Wert beinhalten.

6. Zeile



Eine Zeile unterscheidet sich in min. einer Zeile von den anderen Spalten. Das Token „appartint“ (Wortform) wurde 32 Mal im Korpus als POS „VER:simp“ (verb simple past, siehe <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/french-tagset.html>) von Sprecher 1 mit bedingter Zustimmung akzeptiert. Auch die Frequenz-Zeilen lassen sich (wie die Gruppen) ausklappen. Dadurch erhalten Sie alle Belegstellen auf die diese Auszeichnung (Wort/POS/Sprecher1) zutrifft.

7. Belegstellen

Links	Fundstelle	Rechts	Frequenz
Beinhaltet: ▾	Beinhaltet: ▾	Beinhaltet: ▾	Ist glei... ▾
Son intérêt n'est-il pasà	ce	que l'usine s'accroisse?	16
Il me prêta" Ubu roi" que je n'ap...	ce	fût, mes obsessions.	1
Habituéà	ce	que mon père ne mettait son vétoà aucun de mes actes, rien ne m'étonn...	1

Die Belegstellen werden als KWIC (keyword in context) angezeigt. Die Spalte Frequenz gibt an, wie oft dieser Satz exakt im Gesamtkorpus vorkommt. Da die KAMOKO-Daten aus einem Ausgangsmaterial generiert wurden, sind Dopplungen normal. Klicken Sie auf die Lupe, um sich die kompletten Beispiele anzeigen zu lassen (Volltext).

Die Frequenzanalyse eignet sich bereits, um nach einfachen Belegstellen (zu Einzelbegriffen) zu suchen (Frequenzanalyse – Belegstellen anzeigen). Zur Übung können Sie versuchen, nach besonders häufigen (Frequenz absteigend sortieren) POS-Tags zu suchen z. B. nach "VER:simp" (verb simple past) und hierzu die Belegstellen auszuwerten. Komplexere Suchen können mittels der Analyse CutOff-Phrase realisiert werden. Wählen Sie im Analysemenü „Phrasen & Muster > CutOff-Phrasen“ aus. CutOff-Phrasen sind durch zwei Lexeme begrenzt, dabei spielt es keine Rolle welches Lexem zuerst in einem Beleg vorkommt. Für das folgenden Beispiel wurde die Negation *ne...plus* gewählt, die begrenzenden Lexeme sind in diesem Fall „ne“ und „plus“ (erproben Sie gerne auch andere Sprachmuster).

Alle Dokumente von Modus

Phrasen & Muster

Schnappschuss 3

Suchmaske 1

Begriff (A): ne Begriff (B): plus

Ziehen Sie den Spaltenkopf hier her, um die Tabelle zu gruppieren.

Pre	Match	Post	Spanne	Frequenz
Beinhaltet: 24	Beinhaltet: 25	Beinhaltet: 25	113	45
Je	ne rougis pas de ce que j'ai été: un valet de ferme, sans p...		14	1
" L'opération de cet hiver, l'ablation du rein dr...	ne puisse plus	s'illusionner sur la nature de la tu...	1	1
Je	ne rougis pas de ce que j'ai été: un valet de ferme, sans plus		14	1
Mary affectait de	ne plus	lui parler qu'en anglais.	0	2
J'ai voulu vous avertir que cette femme	ne doit plus	remette les pieds ici.	1	1
Elle	ne disait plus	qu'il ne fallait pas nous revoir.	1	1
Il	ne m'a plus	parlé de Dieu(.	1	2
Le plus étrange est qu'il	ne songeait plus	du tout au sujet, au rapport de ce...	1	1
Elle	ne disait plus	qu'il ne fallût pas nous revoir.	1	1
J'ai voulu vous avertir que cette femme	ne remet plus	les pieds ici.	1	1

KWIC 2

Die folgenden drei Bereiche ermöglichen einen schnellen Einstieg:

1. **Suchmaske**

Begriff (A): Begriff (B): 

Geben Sie hier zwei Suchbegriffe ein (Einzeltoken) und bestätigen Sie die Suche mit einem Klick auf den grünen Pfeil.

2. **KWIC**

Pre	Match	Post	Spanne	Frequenz
Beinhaltet: 24	Beinhaltet: 25	Beinhaltet: 25	113	45
Je	ne rougis pas de ce que j'ai été: un valet de ferme, sans p...		14	1
" L'opération de cet hiver, l'ablation du rein dr...	ne puisse plus	s'illusionner sur la nature de la tu...	1	1
Je	ne rougis pas de ce que j'ai été: un valet de ferme, sans plus		14	1

Die Tabelle zeigt alle Ergebnisse als KWIC. Dabei bilden die beiden Suchworte die Rahmung (unbestimmte Reihenfolge) der Fundstelle (Match). Die Spanne gibt an, wie viele Token zwischen den beiden Token in Match enthalten sind (nützlich für Sortieren oder Suchen - besonders langer/kurzer Matches). Die Frequenz gibt an, wie oft der Satz im Korpus vorkommt.

3. Schnappschuss



Dies ist das Ende des Hands-on Labs, aber vielleicht Ihr Anfang mit eigenen Analysen. Wählen Sie verschiedene Tabellen-Zeilen aus (per Linksklick - bei gedrückter STRG/CTRL-Taste). Klicken Sie danach auf das Symbol "Schnappschuss erstellen". Sie werden aufgefordert den Schnappschuss sinnvoll zu benennen. Bestätigen Sie den Namen mit einem Klick auf OK. Schnappschüsse funktionieren wie Subkorpora - sie haben soeben einen Schnappschuss angelegt, der alle Dokumente beinhaltet, in denen die gewählten Belegstellen vorkommen. Sie können diesen Schnappschuss nutzen, um darauf aufbauend eigene Analyse zu erstellen, oder nur die gewählten Dokumente zu exportieren - Weitere Details und Funktionsweisen finden Sie im Handbuch des CorpusExplorers unter: <http://www.CorporusExplorer.de>.