

# Einfache Korpusanalysen: Ein Schnelleinstieg

*Stefan Hartmann*

2019-06-10

## Inhalt

<b>1 Einstieg</b>	<b>1</b>
<b>2 Von der Fragestellung zur Konkordanz</b>	<b>1</b>
2.1 Eine Fragestellung formulieren . . . . .	2
2.2 Daten erheben . . . . .	2
<b>3 Von der Konkordanz zur Analyse</b>	<b>5</b>
3.1 Annotation . . . . .	6

## 1 Einstieg

Ziel dieses Tutorials ist es, Anfänger\*innen einen möglichst niedrigschwlligen Einstieg in einfache Korpusanalysen zu ermöglichen. Es ist insbesondere für Studierende gedacht, die z.B. für eine Seminararbeit eine Korpusrecherche durchführen möchten, aber bislang noch keine praktische Erfahrung mit korpuslinguistischen Methoden sammeln konnten. Das Tutorial bietet anhand eines konkreten Beispiels eine Schritt-für-Schritt-Anleitung, wie man von der Fragestellung zur Datengewinnung hin zur Analyse der Daten gelangen kann.

Um wirklich einen Schnelleinstieg bieten zu können, muss ich notwendigerweise vieles vereinfachen. Für Ihre konkrete Korpusstudie werden Sie daher wahrscheinlich nicht umhinkommen, sich an der einen oder anderen Stelle tiefer einzulesen. Dafür verweise ich im Text immer wieder auf weiterführende Ressourcen. Teilweise finden sich auch in diesem Tutorial vertiefende Passagen, die Sie aufklappen können:

### klick mich

Hallo, ich bin eine vertiefende Passage.

Sonst gibt es hier nichts zu sehen. Sie können mich gern wieder schließen. Danke.

Ein Hinweis vorab: Das Tutorial setzt keine Kenntnisse in der Korpuslinguistik oder im Umgang mit Tabellenkalkulationsprogrammen voraus, wohl aber grammatische Grundkenntnisse. Sollten Sie die Fachbegriffe nicht verstehen, empfehle ich sehr, sie nachzuschlagen und die entsprechenden Wissenslücken zu schließen.

## 2 Von der Fragestellung zur Konkordanz

Die meisten empirischen Studien lassen sich auf folgende Schritte herunterbrechen:

- Eine Fragestellung formulieren
- Daten erheben
- Daten auswerten.

## 2.1 Eine Fragestellung formulieren

Der erste Schritt ist wahrscheinlich der wichtigste. Nur wenn Sie eine gute Forschungsfrage haben, können Sie eine aussagekräftige empirische Analyse durchführen. Aus der Forschungsfrage ergibt sich die Methode: Für manche Fragestellungen bietet sich z.B. eine Fragebogenstudie an, für eine psycho- oder neurolinguistische Herangehensweise, für wieder andere eine Korpusrecherche.

Das heißt auch: Wenn Sie eine Korpusanalyse durchführen möchten, brauchen Sie eine Fragestellung, die körpuslinguistisch operationalisierbar ist. Beispielsweise lässt sich eine Frage wie “Welche Gehirnareale werden beim Hören von Bewegungsverben aktiviert?” natürlich nicht mit Hilfe von Korpusdaten beantworten.

Für unsere Beispielenalyse werfen wir einen Blick auf die prädiktative Verwendung der Partizipien *programmiert* und *vorprogrammiert*. Letzteres ist manchen Sprachpflegern ein Dorn im Auge: So bezeichnet es Batian Sick als

“umgangssprachliches Blähwort, über das schon Heerscharen von Sprachpflegern hergefallen sind – vergebens, denn es wird immer munter weiter vorprogrammiert. Dabei wissen nicht nur Programmierer: Man programmiert immer im Voraus, die Vorsilbe vor- ist daher pleonastisch, zu Deutsch: doppelt gemoppelt.”  
<https://bastiansick.de/kolumnen/abc/vorprogrammiertprogrammiert/>

Was Sprachpfleger wie Sick jedoch oft erkennen, ist, dass Sprache nicht immer “logisch” ist. Vielmehr suchen sich Wörter oft eigene Nischen. Beispielsweise ist mein Bürostuhl kein *Rollstuhl*, obwohl er Rollen hat - denn das Wort *Rollstuhl* hat eine eigene Bedeutung angenommen, die sich nicht kompositional aus seinen Einzelteilen ergibt. Im Falle von *vorprogrammiert* hingegen passt zwar die Paraphrase ‘im Voraus programmiert’. Aber trotzdem wäre denkbar, dass das Wort eine Spezialisierung erfahren hat: Wird *programmiert* möglicherweise eher dann verwendet, wenn ein Programmierungsvorgang im wörtlichen Sinn gemeint ist, und *vorprogrammiert* eher dann, wenn ein z.B. ein Skandal oder eine Katastrophe “vorprogrammiert” sind? Das ist die Fragestellung, der wir im Folgenden nachgehen möchten.

### Fragestellungen und Hypothesen

Die Unterscheidung von **Fragestellung** und **Hypothese** bereitet Anfänger\*innen oft Schwierigkeiten. Beide hängen eng zusammen. In unserem Beispiel könnte man die Frage in eine Hypothese umformulieren: “vorprogrammiert wird eher in metaphorischem und programmiert eher im wörtlichen Sinn verwendet.”

Hypothesen ergeben sich in der Regel aus konkreten Fragestellungen. Beispielsweise könnte in einer soziologischen oder politikwissenschaftlichen Studie die Fragestellung lauten: Welchen Einfluss hat das Alter auf das Wahlverhalten in Deutschland? Da man zu diesem Themengebiet aus der bisherigen Forschung und aus der Alltagserfahrung das eine oder andere schon weiß, kann man begründete Annahmen darüber treffen, wie die Antwort auf diese Frage aussieht. So könnte man davon ausgehen, dass z.B. ältere Menschen eher etablierte und vielleicht auch eher konservative Parteien wählen und dass außerdem bei Älteren eine höhere Wahlbeteiligung vorliegt. Diese Annahmen nennt man Hypothesen. Sie werden auf Grundlage der Daten, die man erhobt, überprüft.

Nicht immer ist es möglich oder notwendig, konkrete Hypothesen zu formulieren. Gerade bei Phänomenen, über die noch sehr wenig bekannt ist, bietet es sich manchmal an, **explorativ**, also “erkundend”, zu arbeiten. Auch dann gehe ich mit einer Fragestellung an meine Daten heran, ohne jedoch im Voraus eine Erwartung zu haben, wie die Antwort auf meine Frage aussehen wird.

## 2.2 Daten erheben

### 2.2.1 Suchsyntax

Für die Datenerhebung verwenden wir das DWDS-Kernkorpus des 20. Jahrhunderts, das über dwds.de zugänglich ist. Wir suchen auf der Wortebene mit Hilfe von regulären Ausdrücken nach den Formen *programmiert* und *vorprogrammiert*. Dafür benutzen wir den Suchstring @programmiert || @vorprogrammiert.

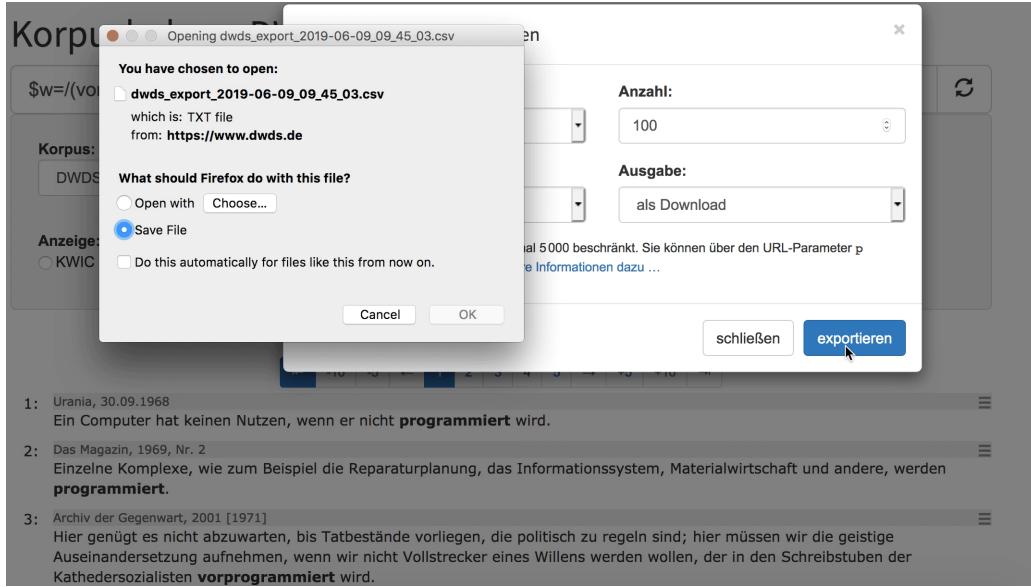


Fig. 1: Export aus dem DWDS

Das @-Zeichen bedeutet, dass wir genau diese Strings suchen und keine anderen Wortformen wie *programmierte*, *programmiertes* etc. Da uns nur die prädiktative Verwendung interessiert, brauchen wir die flektierten Wortformen nicht. Der horizontale Strich | ist der ODER-Operator; dass man ihn hier doppelt setzen muss, ist eine Besonderheit der DWDS-Suchsyntax.

**Alternative Suchabfrage mit regulären Ausdrücken** Alternativ können wir das gleiche Ergebnis auch durch Verwendung regulärer Ausdrücke erzielen: `$w=/(vor)?programmiert/g`. Ich ermutige alle, die sich mit Korpuslinguistik beschäftigen wollen, sehr, sich mit regulären Ausdrücken vertraut zu machen. Allerdings unterstützt die DWDS-Suchsyntax reguläre Ausdrücke derzeit nur in sehr beschränktem Maße. (Deutlich besser ist in dieser Hinsicht das alternative Abfrageportal Dstar, das jedoch für Anfänger\*innen nur bedingt geeignet ist.)

**Zur Suche im DWDS und anderswo** - Die Hilfe zur Suche im DWDS findet sich hier.

- Einen Einstieg in reguläre Ausdrücke bietet z.B. [regular-expressions.info](http://regular-expressions.info).
- In den Begleitmaterialien zu meiner “Deutschen Sprachgeschichte” finden sich ebenfalls einige Tutorials zur Suche in einschlägigen Korpora.
- Sehr empfehlenswert und erfreulich ausführlich ist außerdem die Korpuslinguistik-Seite von Noah Bubenhofen.

## 2.2.2 Export

Die Suche liefert uns 88 Treffer, die nun im Browser in ihrem jeweiligen Kontext dargestellt werden. Diese Daten wollen wir nun exportieren, und zwar im “Key Word in Context” (KWIC)-Format. Damit ist gemeint, dass der Suchtreffer zusammen mit seinem unmittelbaren Kontext dargestellt wird. Erfreulicherweise bietet das DWDS eine sehr gute Exportfunktion, die es erlaubt, Daten im CSV-Format zu speichern.

Eine solche Sammlung von Korpusbelegen, wie wir sie jetzt exportiert haben, nennt man in der Korpuslinguistik **Konkordanz**. Der Formatname “CSV” steht für “Comma-Separated Values”. Das heißt, in der Datei sind die einzelnen Werte durch Kommata voneinander abgetrennt. In einem Texteditor sieht das Ganze so aus wie in 2. Wie Sie sehen, enthält die Datei neben den Korpusbelegen selbst auch Metadaten zu den einzelnen Belegen, z.B. zu Autor\*in, Titel etc.

Users > stefanhartmann > Dropbox > Privat > Tutorials > korpus-schnelleinstieg > data > dwds\_export\_2019-06-09\_09\_44\_12.csv

```

1 "No.", "Date", "Genre", "Bibl", "ContextBefore", "Hit", "ContextAfter"
2 "1", "1968-09-30", "Gebrauchsleiteratur", "Urania", "30.09.1968", "Ein Computer hat keinen Nutzen, wenn er nicht", "programmiert", "wird."
3 "2", "1969-02-28", "Zeitung", "Das Magazin, 1969, Nr. 2", "Einzelne Komplexe, wie zum Beispiel die Reparaturplanung, das Informationssystem, Mater.
4 "3", "1971-01-27", "Zeitung", "Archiv der Gegenwart, 2001 [1971]", "Hier genügt es nicht abzuwarten, bis Tatbestände vorliegen, die politisch zu ri
5 "4", "1971-12-31", "Gebrauchsleiteratur", "Jung, Mathias: Der militärisch-industrielle Komplex. In: Haug, Hans-Jürgen u. Maessen, Hubert (Hgg.) Kr.
6 "5", "1971-12-31", "Gebrauchsleiteratur", "Jung, Mathias: Der militärisch-industrielle Komplex. In: Haug, Hans-Jürgen u. Maessen, Hubert (Hgg.) Kr.
7 "6", "1971-12-31", "Gebrauchsleiteratur", "Klix, Friedhart: Information und Verhalten, Berlin: Deutscher Verl. der Wissenschaften 1971, S. 731", "Diese M
8 "7", "1972-12-31", "Wissenschaft", "Offe, Claus: Strukturprobleme des kapitalistischen Staates, Frankfurt a. M.: Suhrkamp 1972, S. 90", "Das bedeutet
9 "8", "1973-10-17", "Zeitung", "Archiv der Gegenwart, 2001 [1973]", "70. Diesbezüglich verlangt die Konferenz die Einstellung der Kernversuche, die
10 "9", "1974-03-08", "Zeitung", "Die Zeit, 08.03.1974, Nr. 11", "Washington will Moskau sind, gewiß aus unterschiedlichen Interesse, auf einen israel.
11 "10", "1974-12-31", "Gebrauchsleiteratur", "Klee, Ernst: Behinderten-Report, Frankfurt a. M.: Fischer Taschenbuch-Verl. 1981 [1974], S. 38", "Die El
12 "11", "1974-12-31", "Gebrauchsleiteratur", "Klee, Ernst: Behinderten-Report, Frankfurt a. M.: Fischer Taschenbuch-Verl. 1981 [1974], S. 93", "Auch i
13 "12", "1974-12-31", "Gebrauchsleiteratur", "Klee, Ernst: Behinderten-Report, Frankfurt a. M.: Fischer Taschenbuch-Verl. 1981 [1974], S. 94", "Was ai
14 "13", "1975-05-09", "Zeitung", "Archiv der Gegenwart, 2001 [1975]", "Einleitend erläuterte der für die Landwirtschaft zuständige EG-Kommissar Petri
15 "14", "1975-09-05", "Zeitung", "Archiv der Gegenwart, 2001 [1975]", "Der Energieverbrauch und die Wasser-, Kohle-, Erdöl- und Gasreserven des Landes
16 "15", "1977-12-31", "Gebrauchsleiteratur", "Pilgrim, Volker Elis: Manifest für den freien Mann - Teil 1, Reinbek b. Hamburg: Rowohlt 1983 [1977], '
17 "16", "1978-05-11", "Zeitung", "Archiv der Gegenwart, 2001 [1978]", "Das auf OTS folgende Satellitensystem ECS (European Communications Satellites
18 "17", "1979-12-31", "Gebrauchsleiteratur", "Bädeker, Klaus: Werthers Freundin. In: Hoffmann, Raoul (Hg.) Auf Live und Tod, München: Dt. Taschenbuch-Ver
19 "18", "1979-12-31", "Gebrauchsleiteratur", "Bädeker, Klaus: Werthers Freundin. In: Hoffmann, Raoul (Hg.) Auf Live und Tod, München: Dt. Taschenbuch-Ver
20 "19", "1979-12-31", "Gebrauchsleiteratur", "Wilberg, Gerlinde M.: Zeit für uns, München: Frauenbuchverl. 1979, S. 127", "Wenn eine Geburt auf sieben
21 "20", "1979-12-31", "Gebrauchsleiteratur", "Wilberg, Gerlinde M.: Zeit für uns, München: Frauenbuchverl. 1979, S. 129", "Es gibt jedoch medizinisch
22 "21", "1980-05-05", "Zeitung", "Der Spiegel, 05.05.1980", "Dies unvermögen ist", "vorprogrammiert", "."
23 "22", "1981-10-19", "Zeitung", "Der Spiegel, 19.10.1981", "Sie können beispielweise als ""Demand-Schrittmacher"" programmieren", "werden, die ers
24 "23", "1981-12-03", "Zeitung", "Archiv der Gegenwart, 2001 [1981]", "Während sich BEGIN auf die in Camp David geplanten Autonomieverhandlungen ver
25 "24", "1983-12-31", "Gebrauchsleiteratur", "Ichenhäuser, Ernst Z.: Erziehung zum guten Benehmen, Berlin: Volk u. Wissen 1983, S. 68", "Ist es zu scl
26 "25", "1984-04-16", "Zeitung", "Der Spiegel, 16.04.1984", "Unter der Verwaltung kühler Monokratoren ist Phantasielosigkeit", "vorprogrammiert", "."
27 "26", "1984-05-21", "Gebrauchsleiteratur", "o. A. (Khl): Seesielflugkörper Harpoon. In: Aktuelles Lexikon 1974-2000, München: DIZ 2000 [1984]", "Die
28 "27", "1985-06-17", "Zeitung", "Der Spiegel, 17.06.1985", "Damit war für Südafrika die ""brasilianische Lösung"" einer ungehinderten Vermischung d
29 "28", "1985-12-31", "Gebrauchsleiteratur", "Sinn und Form, 1985, Nr. 2, 37", "Überanstrengte stilistische Sachlichkeit und Historizität sowie ai
30 "29", "1985-12-31", "Gebrauchsleiteratur", "Alt, Franz: Liebe ist möglich, München: Piper 1985, S. 129", "Wenn keine Seite bereit ist, den ersten k
31 "30", "1985-12-31", "Gebrauchsleiteratur", "Alt, Franz: Liebe ist möglich, München: Piper 1985, S. 175", "Die Katastrophen, vor denen wir heute stel
32 "31", "1985-12-31", "Gebrauchsleiteratur", "Zimmermann, Hartmut (Hg.): DDR-Handbuch - F. In: Enzyklopädie der DDR, Berlin: Directmedia Publ. 2000
33 "32", "1985-12-31", "Gebrauchsleiteratur", "Zimmermann, Hartmut (Hg.): DDR-Handbuch - M. In: Enzyklopädie der DDR, Berlin: Directmedia Publ. 2000
34 "33", "1985-12-31", "Gebrauchsleiteratur", "Zimmermann, Hartmut (Hg.): DDR-Handbuch - U. In: Enzyklopädie der DDR, Berlin: Directmedia Publ. 2000

```

Fig. 2: Konkordanz im Texteditor

A	B	C	D	E	F	G	H	I	J	K	L
28	27	17/06/1985	Zeitung	Der Spiegel, Damit war fv9r Sv9dafrika die "brasilianische Programmie.							
29	28	31/12/1985	Gebrauchsleiteratur	Sinn und Forvüberanstrengte stilistische Sachlichkeit Programmie ist.							
30	29	31/12/1985	Gebrauchsleiteratur	Alt, Franz: Lii Wenn keine Seite bereit ist, den ersten vorprogrammier.							
31	30	31/12/1985	Gebrauchsleiteratur	Alt, Franz: Lii Die Katastrophen, vor denen wir heute programmie #NAME?							
32	31	31/12/1985	Gebrauchsleiteratur	Zimmermann: Zum anderen wird eine fortlaufende Div vorprogrammier.							
33	32	31/12/1985	Gebrauchsleiteratur	Zimmermann: FV9r verschiedene Gebiete des ML programmie (vgl. fv9r die gegenwv5rtig gv9tige Planperiode:							
34	33	31/12/1985	Gebrauchsleiteratur	Zimmermann: Als programmierter Unterricht (PU) wi programmie sind und der einem im Lehrprogramm gespeicherten Lehralgorithmus folgt, der die							
35	34	31/12/1986	Gebrauchsleiteratur	Ketman, Per So in manchen Spielen die Konfront vorprogrammier.							
36	35	31/12/1986	Gebrauchsleiteratur	Ketman, Per Als zuv9nfige Pfarrerin my9vite Miria programmie, noch ehe es v9berhaupt mv9dig ist: ein Dilemma, das keine einfache Lv9sung ke							
37	36	23/02/1987	Zeitung	Der Spiegel, Streit v9ber die parlamentarische Salon programmie:							
38	37	27/02/1987	Zeitung	Archiv der G Sollte HAMADEI FV9r die USA ausgeliefert vorprogramm zudem wv9re dies das Todesurteil fv9r den Entf9rten.							
39	38	05/10/1987	Zeitung	Der Spiegel, Doch die personellen Mv9griffe waren programmie.							
40	40	12/09/1988	Zeitung	Der Spiegel, Regelfrech programmie und in Szene gesetzt von staatlichen Instanzen war vor dem 9. November 1938 kein							
41	40	31/12/1988	Wissenschaft	Weizs9cker Das Altern dv9rfte darum genetisch programmie sein.							
42	41	07/04/1988	Zeitung	Archiv der G Somit wv9ren Krisen in den kommende vorprogramm.							
43	42	28/08/1988	Zeitung	Der Spiegel, Wie die Schwestern Voyager 1 war programmie.							
44	43	31/12/1988	Gebrauchsleiteratur	Brandt, Willy Auf Ablehnung - und sei es nur, dav9 au programmie, die Bv9ndnisfreiheit fv9r Deutschland keinesfalls in Erwv9gung ziehen wollten.							
45	44	31/12/1988	Wissenschaft	o. A.: Lexikor Literatur sowie im allg. Sprachgebrauch vorprogramm.							
46	45	26/02/1991	Zeitung	Archiv der G FV9r das Fiskaljahr 1991 sind Ausgaben programmie ist, das etwa dem Schuldendienst entspricht.							
47	46	26/09/1991	Gebrauchsleiteratur	o. A. [ley]: El Das System kann so programmie werden, dav9 es beispielweise Alarm gibt, wenn sich das Schiff einem Hindernis a							

Fig. 3: Konkordanz bei direktem Öffnen in Excel

Damit können wir zunächst noch wenig anfangen: Wir wollen die Konkordanz in ein Tabellenkalkulationsprogramm einlesen.

### 2.2.3 Import in ein Tabellenkalkulationsprogramm

Wenn Sie Microsoft Excel auf Ihrem Rechner installiert haben, sind die Default-Einstellungen höchstwahrscheinlich so gesetzt, dass CSV-Dateien in Excel geöffnet werden, wenn Sie darauf doppelklicken. Warum das keine gute Idee ist, zeigt der folgende Screenshot 3 (rote Hervorhebungen von mir nachträglich hinzugefügt).

Hier sind einige Sonderzeichen verlorengegangen, weil Excel die Kodierung der Datei nicht richtig erkannt hat. Es gibt mehrere Wege, diesem Problem zu begegnen. Ich empfehle hier zwei: Einen für Excel und einen für die freie Alternative Calc.

#### 2.2.3.1 Import in Excel

- Öffnen Sie die Datei in einem Texteditor. Für Windows empfehle ich Notepad++, für Mac die kostenlose (und für unsere Zwecke völlig ausreichende) Version von BBEdit, für Linux gibt es z.B. Notepadqq.

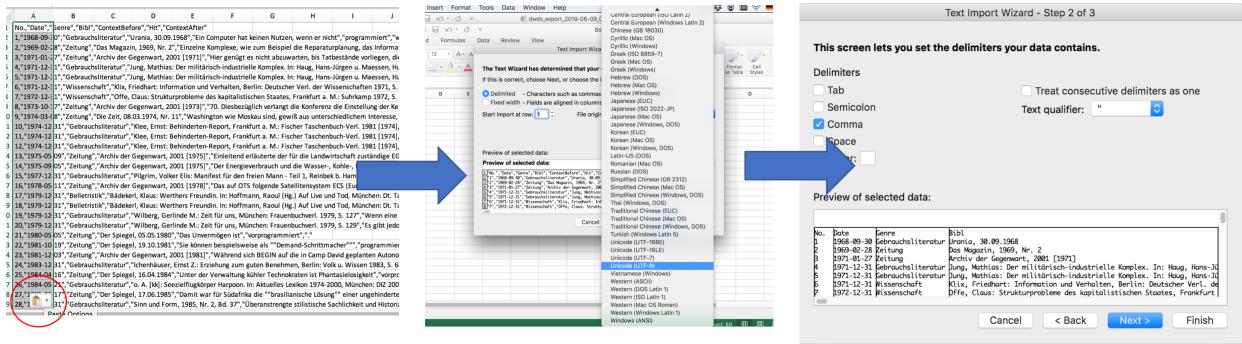


Fig. 4: Import in Excel

2. Markieren Sie mit Strg+A bzw. Cmd+A den gesamten Text.
3. Öffnen Sie ein leeres Tabellenblatt in Excel. Die nächsten Schritte, 4 bis 7, sind in 4 visualisiert.
4. In den meisten Fällen sollten Sie nun einfach mit Strg+V bzw. Cmd+V die Daten einfügen können. In manchen Fällen müssen Sie jedoch, wie im Screencast 4, die Option “Paste Special” verwenden (dt. “Inhalte einfügen”) und angeben, dass Sie den Unicode-Text einfügen möchten.
5. Mit Klick auf das kleine Klemmbrett-Symbol gelangen Sie zum Textimport-Assistenten. Hier müssen Sie Excel sagen, wie der eingefügte Text strukturiert ist. Auf der ersten Seite sagen Sie, dass es sich um einen Text handelt, bei dem die einzelnen Spalten durch ein Trennzeichen getrennt sind (“Delimited”) - diese Option ist in der Regel schon angewählt. Außerdem teilen Sie Excel hier mit, dass der eingefügte Text UTF-8-formatiert ist.
6. Auf der nächste Seite des Textimport-Assistenten geben Sie an, dass Kommata als Spaltentrenner benutzt werden. Bei den Textqualifizierern müssen Sie nichts ändern, da hier schon Anführungszeichen ausgewählt sind: Wie Sie in 2 sehen können, werden Anführungszeichen in der CSV-Datei genutzt, um zusammengehörigen Text zusammenzuhalten (denn wären sie nicht da, würde Excel jedes Komma im Text für einen Spaltentrenner handeln.)
7. Dieser letzte Schritt erübriggt sich meistens, kann aber nicht schaden: Zuletzt können Sie noch alle Spalten als “Text” formatieren. (Die Datumsspalte können Sie prinzipiell auch als “Datum” formatieren, falls Sie ausschließlich in Excel weiterarbeiten, aber tendenziell rate ich davon ab - gerade bei einer späteren Konversion in andere Dateiformate kann dabei alles mögliche schiefgehen...) Tipp: Um alle Spalten auf einmal als “Text” zu formatieren, einfach im Fenster ganz nach rechts scrollen und mit gedrückter Shift-Taste auf die letzte Spalte klicken, dann sind alle Spalten markiert.

### 2.2.3.2 Import in Calc

Öffnet man die Datei im kostenlosen Tabellenkalkulationsprogramm Calc von LibreOffice (mit Rechtsklick > Öffnen mit), so öffnet sich zunächst automatisch der Textimportassistent. Hier muss man Calc mitteilen, welches Format die Datei hat. In unserem Fall ist der Text UTF-8-kodiert, wir haben Kommas als Spaltentrenner und Anführungszeichen als Textqualifizieren, wie in 5.

## 3 Von der Konkordanz zur Analyse

Nun haben wir die Konkordanz erfolgreich in ein Tabellenkalkulationsprogramm importiert. Hier können wir beliebig viele weitere Spalten hinzufügen. Das können wir nutzen, um die exportierten Belege mit **Annotationen** zu versehen.

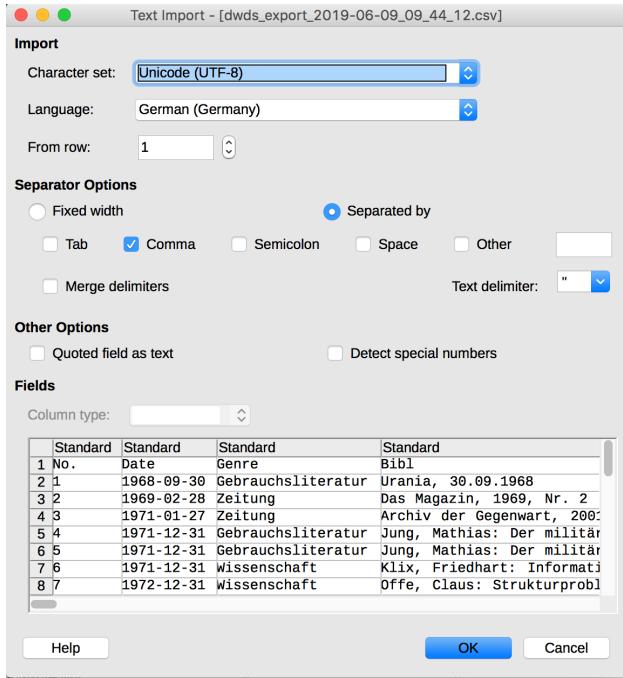


Fig. 5: Import in Calc

### 3.1 Annotation

Versieht man Daten mit zusätzlichen Informationen, so nennt man diesen Prozess Annotation. In der Korpuslinguistik stellt die Annotation einen ganz wesentlichen Schritt dar, der gewissermaßen die Brücke schlägt von der qualitativ-philologischen Analyse einzelner Belege zur quantitativen Auswertung.

Wir nutzen im Folgenden die Annotation, um unsere Daten in Kategorien zu unterteilen, die für unsere Fragestellung sinnvoll sind. Dafür müssen wir uns zunächst darüber im Klaren sein, was wir von unseren Daten überhaupt wissen wollen, d.h. wir müssen unsere eingangs genannte Fragestellung operationalisieren.

Zur Erinnerung: Unsere Fragestellung lautet, ob bei prädikativem Gebrauch *vorprogrammiert* gegenüber *programmiert* bevorzugt wird, wenn es sich um einen metaphorischen Kontext handelt.

Konkret bedeutet das, dass wir für jeden Datenpunkt folgende Fragen beantworten müssen:

1. Handelt es sich um eine prädiktative Verwendung? - Schon ein kurzer Blick auf die Daten zeigt, dass sich notwendigerweise einige **Fehltreffer** eingeschlichen haben: Häufig finden sich z.B. Passivkonstruktionen wie *Es gibt jedoch medizinische Gründe, aus denen eine Geburt eingeleitet oder sogar programmiert werden muß*. Uns interessieren aber nur Fälle, in denen das Partizip selbst das Prädikat bildet, also z.B. *Der Computer ist programmiert* und *Die Katastrophe war vorprogrammiert*.
2. Handelt es sich um eine metaphorische Verwendung? - Während beispielsweise Computer oder Roboter im wörtlichen Sinne programmiert werden, bezieht sich der Begriff bei Krisen und Katastrophen darauf, dass Voraussetzungen geschaffen wurden, die unausweichlich den thematisierten unschönen Ausgang zur Folge haben. Es liegt also ein metaphorischer Gebrauch vor, bei der Aspekte der Quelldomäne "Technik" auf eine abstraktere Zieldomäne übertragen werden.

In den nächsten Abschnitten wollen wir uns beiden Fragen etwas genauer zuwenden.

### 3.1.1 Annotation prädikativ vs. nicht-prädikativ

Wenn wir Daten annotieren, besteht eine wesentliche Herausforderung immer in der **Operationalisierung** konkreter Fragestellungen. In vielen Fällen ist es so, dass wir die Frage, die uns interessiert, auf den ersten Blick glauben für jeden Datenpunkt klar beantworten zu können. Bei genauerem Hinsehen ergeben sich dann aber doch einige Zweifelsfälle. So ist es auch hier: Um die Frage operationalisieren zu können, muss man zunächst einmal die Entscheidung treffen, ob man eine Struktur wie *Der Computer ist programmiert* als Zustandspassiv mit *sein* als Hilfsverb (analog zum Vorgangspassiv mit *werden* als Hilfsverb) oder als Konstruktion aus der Kopula *sein* und dem Partizip II *programmiert* interpretiert. Wir entscheiden uns hier für Letzteres. Jedoch zeigt dieses Beispiel: Wie wir Daten interpretieren, hängt oft genug von unserem theoretischen Zugang ab. Das ist nicht weiter schlimm, sondern liegt in der Natur der Sache - Wissenschaft kann nie ganz frei von Theorie und nie ganz frei von Interpretation sein. Wichtig ist, dass die Entscheidung, die wir treffen, sich gut begründen lässt und konsequent durchgehalten wird.

Wie setzen wir die Annotation nun in unserer Tabelle um? Auch hier zeige ich wieder Wege für Excel und Calc. Gerade die unten skizzierte Möglichkeit, Daten als "Tabelle" zu formatieren, finde ich persönlich an Excel sehr hilfreich, weshalb ich Excel i.d.R. bevorzuge. Allerdings halte ich es auch für wichtig, sich in der Wissenschaft nicht von proprietärer Software oder proprietären Datenformaten abhängig zu machen, und nicht jede Uni hat eine Office-Lizenz - deshalb zeige ich auch den Weg mit der freien Alternative auf.

#### 3.1.1.1 Umsetzung in Excel

Zunächst empfiehlt es sich, die Tabelle im Excel-Standardformat .xlsx zu speichern.

Excel bietet die schöne Möglichkeit, Daten als Tabelle zu formatieren. Das ist über den Reiter Einfügen > Tabelle möglich, wie in 6 gezeigt. In der Regel erkennt Excel automatisch die Dimensionen der Tabelle, sodass Sie nur noch anklicken müssen, dass die Tabelle Überschriften hat, und dann auf "OK" klicken können, und schon sind alle Zellen schön formatiert, und vor allem kann man über die kleinen Pfeilsymbole oben die einzelnen Spalten nach bestimmten Werten filtern, was sich im weiteren Verlauf der Arbeit noch als nützlich erweisen kann. (Letzteres erreicht man auch über Daten > Filter, aber mit der Tabellen-Option wird das Ganze optisch noch ein bisschen hübscher, und vor allem muss man keinen neuen Filter setzen, wenn man eine neue Spalte hinzufügt.)

Um die Belege im Kontext besser lesen zu können, empfiehlt es sich, zunächst ein paar Feinjustierungen in der Formatierung vorzunehmen. So können wir Spalten, die wir derzeit nicht benötigen (z.B. alle Spalten mit Metadaten), zunächst ausblenden. (Nicht löschen! Im Zweifelsfall nie Spalten löschen, wer weiß, wozu man sie später noch benötigt...) Außerdem kann es hilfreich sein, den Text in der Spalte mit dem linken Kontext rechtsbündig zu formatieren und die Breite der einzelnen Spalten so anzupassen, dass man den Beleg und ausreichend viel Kontext lesen kann und doch alle derzeit wichtigen Spalten gleichzeitig auf dem Bildschirm zu sehen sind. Wenn Sie die HTML-Version dieses Dokuments lesen, sehen Sie im weiteren Verlauf von Screencast 6 (nach der Formatierung der Daten als Tabelle), wie eine solche Feinjustierung aussehen kann.

#### Zeilenumbruch innerhalb von Tabellenspalten

In einigen Fällen, in denen man sehr viel Text im linken und rechten Kontext hat und in denen man für die Annotation auch auf den weiteren Kontext angewiesen ist, kann es sinnvoll sein, die Tabelle so zu formatieren, dass innerhalb der Zelle ein Zeilenumbruch vorgenommen wird. Standardmäßig ist die Tabelle so formatiert, dass jede Zelle nur eine Zeile hat, und was über die Zelle hinausgeht, wird nicht angezeigt (ist aber trotzdem noch in den Daten vorhanden). Wenn man durch Klick auf den Buchstaben oberhalb der Spalte, die man formatieren möchte, zunächst die ganze Spalte markiert, kann man unter Rechtsklick > Zellen formatieren im Tab "Alignment" ("Ausrichtung") die Option "Wrap text" (Zeilenumbruch) aktivieren.

Als nächstes fügen wir eine neue Spalte rechts von der letzten existierenden Spalte hinzu, der wir die Überschrift "prädikativ" geben. (Wir könnten auch problemlos den Umlaut verwenden, aber ich neige dazu, aus Vorsicht alle Sonderzeichen, die Probleme bereiten könnten, wegzulassen.) Hier tragen wir nun für jeden Datenpunkt ein, ob es sich um eine prädiktative Verwendung handelt oder nicht. Ich verwende hierfür gern die Werte "y" und "n", weil sie schön kurz sind. j/n oder ja/nein gehen natürlich auch.

The screenshot shows the 'Create Table' dialog in Excel. The 'My table has headers' checkbox is selected, and the range '\$A\$1:\$G\$28' is specified. The 'OK' button is highlighted. To the right, the resulting table is shown with 28 rows of data. The last column, 'praedikativ', is circled in red.

No.	Date	Genre	Bibl	ContextBefore	Hit	ContextAfter	praedikativ
1	30/09/1968	Gebräuchlich	Urania, 30.0 Ein Computer programme wird.				
2	28/02/1969	Zeitung	Das Magazin Einzelne Kon programme.				
3	27/02/1971	Zeitung	Archiv der G Hier genutzt v programme wird.				
4	31/12/1971	Gebräuchlich	Der Spiegel, Der Verlag hat Programme für den Heimcomputer				
5	31/12/1971	Gebräuchlich Jung, Mathia Karl Steinbus	programmie, Stuttgart 1968				
7	31/12/1971	Wissenschaft Klix, Friedhai Diese Modelle programme und im Rechenzentrum der Universität Stuttgart sind die auf Muruura im Südpazifik durchgeführt werden.					
8	17/10/1973	Zeitung	Archiv der G Diesbezügliche programme sind die auf Muruura im Südpazifik durchgeführt werden.				
9	06/03/1974	Zeitung	Archiv der G 70. Die EEG-Daq programme.				
10	11/03/1974	Gebräuchlich	Klee, Ernst: Die EEG-Daq programme.				
11	31/12/1974	Gebräuchlich Klee, Ernst: Die EEG-Daq programme bleibt.					
12	31/12/1974	Gebräuchlich Klee, Ernst	I Was auf der vorprogramm ist, setzt sich auf Länderebene fort:				
13	09/02/1975	Zeitung	Archiv der G Erstellt er vorprogramm worden; Frankreich habe damals darauf bestanden, daß eine Qualitätskontrolle und eine Beschränkung				
14	23/02/1975	Zeitung	Archiv der G Anfangszeit der Vokabular und Bedürfnisse unterschieden sich von den in ihnen lebenden Menschen mit ihrer K				
15	31/12/1977	Gebräuchlich Pilgrim, Volk Wenn sie als programme werden, müssen die in ihnen lebenden Menschen mit ihrer Gesundheit oder ihrem Leben beziehen.					
16	11/05/1978	Zeitung	Archiv der G Das auf OTS programme ist.				
17	23/05/1978	Zeitung	Archiv der G Das auf OTS programme ist, die von ihm gewünschten Verhaltensweisen zu adaptieren.				
18	31/12/1979	Bellertistik	Bädeker, Kls Sie war nur f programme, und deren Vokabular und Bedürfnisse unterschieden sich beträchtlich von dem offenkundig ausgef.				
19	31/12/1979	Gebräuchlich Berg, Ger	I Was auf der vorprogramm wird, werden die individuellen Bedürfnisse des Babys nicht berücksichtigt.				
20	23/05/1980	Zeitung	Archiv der G Während der Geburt werden muste				
21	05/05/1980	Zeitung	Der Spiegel, Das Univers vorprogramm.				
22	19/10/1981	Zeitung	Der Spiegel, Sie können b programme werden, die erst bei Bedarf in Aktion treten:				
23	01/12/1981	Zeitung	Archiv der G Wenn sie sic programme sein könnte, ... - 2/X: BEGIN erklärt in einer Rede anlässlich der Eröffnung der Winteression der K				
24	23/05/1982	Zeitung	Archiv der G Der Spiegel, Es ist zu erwarten, daß es in Zukunft mehr				
25	16/04/1984	Zeitung	Der Spiegel, Unter der Ve vorprogramm.				
26	23/05/1984	Gebräuchlich o. A. [4]	Se Die Harpon programme und steuert sich in der letzten Phase ihres Fluges mit Hilfe eines eingebauten Radarsystems selbst				
27	17/06/1985	Zeitung	Der Spiegel, Damit war f programme,				

Fig. 6: Formatierung als Tabelle und Hinzufügen einer Annotationsspalte "praedikativ"

The screenshot shows the 'Format Cells' dialog for cell E1. The 'Text control' section is open, displaying two checkboxes: 'Wrap text' (checked) and 'Shrink to fit' (unchecked). The 'OK' button is visible at the bottom right of the dialog. To the left, a portion of the table is shown with several rows of text, some of which are wrapped within their respective cells.

E	F	G	H	I
1	ContextBefore	ContextAfter	praedikativ	
2	In Computer haben Nutzen, weil	Cut	⌘X	wird.
3	ystem, Materialwirtschaft und andere	Copy	⌘C	n
4	in den Schreibstuben der Katheders	Paste	⌘V	wird.
5	ürgerlichen Kritiker, Karl Steinbus	Paste Special...	⌘V	y
6		Insert	.	Stuttgart 1968
7	ng von Puschet, Sydow und anderen	Delete	.	und im Rechenzentrum der Human
8	nets wieder durch objektivierende E	Clear Contents	.	und angeleitet werden können. n
9	onferenz die Einstellung der Kernen	Format Cells... ⌘1	.	sind und die auf Muruura im Südp
10	ress auf einen iranisch-arabischen	Column Width...	.	n
11	hostik wurde mit Hilfe eines	Hide	.	n
12	berplättung und Verkürzungswahl	Unhide	.	n
13	Was auf der Bun		.	ist, setzt sich auf Länderebene fort; v
14	n, seien die jetzt aufgetretenen Schw		.	worden; Franckreich habe damals d n
15	ren des Landes – und für die entsprechend pro grammert		.	werden, müssen die in ihnen leben
16	er, wenn sie nicht übersteigend pro grammert		.	n
17	ogen im Jahre 1981, auf 2000 im Jahre 1991 pro grammert		.	n
18	Jeder Freund war darum pro grammert		.	, die von ihm gewünschten Verhalt v
19	Sie war nur für den Verkehr mit Chiyamnen pro grammiert		.	, und deren Vokabular und Bedürfni
20	Wenn eine Geburt auf sieben Stunden pro grammert		.	wird, werden die individuellen Bed n
21	aus denen eine Geburt einfacher oder sogar pro grammiert		.	werden muß: n
22	Das Unvermögen ist vorprogrammiert		.	n
23	en beispielweise als "Demand-simulator" pro grammert		.	werden, die erst bei Bedarf in Aktion n

Fig. 7: Zeilenumbruch innerhalb von Zellen einschalten

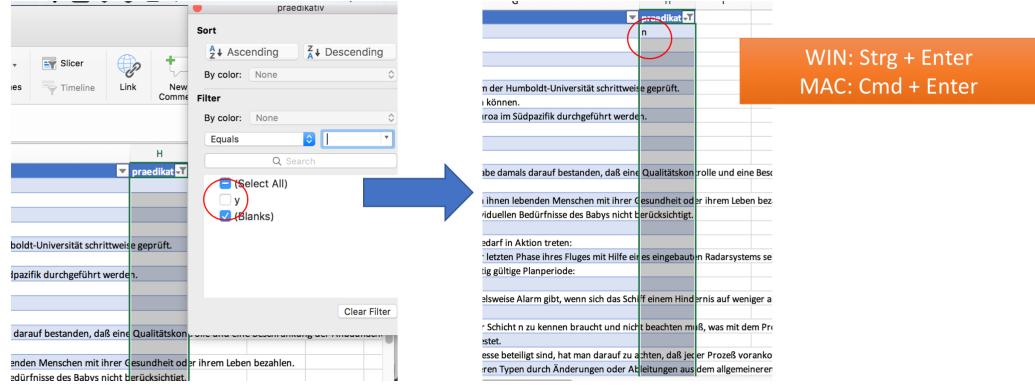


Fig. 8: Eine Tabellenspalte wird so gefiltert, dass nur noch die leeren Zellen zu sehen sind, und allen leeren Zellen wird mit Strg/Cmd+Enter derselbe Wert zugewiesen.

Um Zeit zu sparen, kann man auch nur einen der beiden Werte annotieren und dann die leeren Zellen einfach auffüllen, wie in 8 gezeigt: Hier sind die "y"-Werte schon annotiert, alle anderen Zeilen sind leer. Nun filtert man erst die "praedikativ"-Spalte so, dass nur noch die leeren Zellen zu sehen sind, indem man die Zellen mit dem Wert "y" abwählt. Dann markiert man die Spalte "praedikativ" von der ersten bis zur letzten Zeile (die Überschrift wird nicht mitmarkiert). Gibt man nun "n" ein (noch nicht Enter drücken!!), so erscheint der Wert zunächst in der ersten Zeile. Drückt man nun statt der Eingabetaste Strg+Enter (bzw. bei Mac Cmd+Enter), so wird der in der ersten Zeile eingegebene Wert auf alle folgenden Zellen übertragen.

Wenn wir nun den Filter herausnehmen, sehen wir, dass nun alle vorher leeren Zeilen ein "n" haben, während alle Zeilen mit "y" unverändert geblieben sind.

### 3.1.1.2 Umsetzung in LibreOffice Calc

In Calc empfiehlt es sich, zunächst einmal die Spaltenbreite anzupassen und nicht benötigte Spalten auszublenden (nicht löschen - im Zweifelsfall niemals Spalten oder Zeilen löschen, wer weiß, wofür man sie noch benötigt!). Ich selbst gehe in der Regel so vor, dass ich alle Spalten bis auf diejenigen mit den eigentlichen Belegen (linker Kontext, Treffer, rechter Kontext) ausblende und die Spalte mit dem linken Kontext so formatiere, dass der Text rechtsbündig angezeigt wird. So kann ich bequem den Beleg vom linken Kontext über den Treffer bis zum Keyword lesen. In der HTML-Version dieses Tutorials sehen Sie das in Screencast 9.

Wenn Sie die Formatierungsoptionen für zukünftige Sitzungen speichern möchten, müssen Sie die Datei in einem anderen Format, z.B. im Calc-Standardformat .ods, speichern. Prinzipiell können Sie aber auch einfach in der CSV-Datei weiterarbeiten. Wenn Sie die Datei zwischenspeichern, werden dann eventuell neu eingetragene Daten gespeichert, nicht aber die Formatierung, die Sie dann, wenn Sie die Datei schließen und wieder öffnen, noch einmal neu einstellen müssen.

Wir können nun eine neue Spalte rechts von der letzten existierenden Spalte hinzufügen, der wir die Überschrift "praedikativ" geben. (Wir könnten auch problemlos den Umlaut verwenden, aber ich neige dazu, aus Vorsicht alle Sonderzeichen, die Probleme bereiten könnten, wegzulassen.) Hier tragen wir nun für jeden Datenpunkt ein, ob es sich um eine prädiktative Verwendung handelt oder nicht. Ich verwende hierfür gern die Werte "y" und "n", weil sie schön kurz sind. j/n oder ja/nein gehen natürlich auch.

Um Zeit zu sparen, kann man auch nur einen der beiden Werte annotieren und dann die leeren Zellen einfach auffüllen. Dafür müssen wir zunächst einen Filter setzen, wie in 9 gezeigt. Über diesen Filter können wir jetzt die leeren Zellen ausblenden. Hier sind die "y"-Werte schon annotiert, alle anderen Zeilen sind leer. Nun filtert man erst die "praedikativ"-Spalte so, dass nur noch die leeren Zellen zu sehen sind, indem man die Zellen mit dem Wert "y" abwählt. Dann markiert man die Spalte "praedikativ" von der ersten bis zur letzten Zeile (die Überschrift wird nicht mitmarkiert). Gibt man nun "n" ein (noch nicht Enter drücken!!), so wird der in der ersten Zeile eingegebene Wert auf alle folgenden Zellen übertragen.

The screenshot shows a LibreOffice Calc spreadsheet titled "dwds\_export\_2019-06-09\_09\_44\_12.csv". The table has the following structure:

A	E	F	G	H	I
1	Netz	ContextBefore	ContextAfter		
2	1 Ein Computer hat keinen Nutzen, wenn er nicht programmiert		wird.		
3	2aturplanung, das Informationssystem, Materialwirtschaft und andere, werden programmiert		.		
4	3es Willens werden wollen, der in den Schreibusen der Kathedersozialisten vorprogrammiert		wird.		
5	4nes seiner prominentesten bürgerlichen Kritiker, Karl Steinbuch, als "falsch programmiert		".		
6	5Karl Steinbuch Falsch programmiert		,		
7	6as anderer Form unter Anleitung von Peschel, Sydow durch Schmiehl (1967) programmiert		Stuttgart 1968		
8	7ernden Funktion nun nicht ihrerseits wieder durch objektivierende Erkenntnis programmiert		und im Rechenzentrum der Humboldt-Universität schrittweise geprüft.		
9	8Diesbezüglich verlangt die Konferenz die Einstellung des Kernes, die programmiert		und angeleitet werden können.		
10	9aus unterschiedlichem Interesse, auf einen israelischarabischen Ausgleich programmiert		sind und die auf Mururoa im Südpazifik durchgeführt werden.		
11	10Die EEG-Diagnostik wurde mit Hilfe eines Diplom-Ingenieurs programmiert		.		
12	11mehr zur Debatte, so daß die Zersplitterung und Wirkungslosigkeit weiterhin vorprogrammiert		.		
13	12Was auf der Bundesebene vorprogrammiert		bleibt.		
14	13marktordnung erlassen wurde, seien die jetzt aufgetretenen Schwierigkeiten vorprogrammiert		ist, setzt sich auf Länderebene fort:		
15	14-, Kohle-, Erdöl- und Gasreserven des Landes -- und für das Land -- werden programmiert		worin; Frankreich habe damals darauf bestanden, daß eine Qualitätskontrolle und eine werden.		
16	15Wenn sie als lebenslänglich programmiert		werden, müssen die in ihnen lebenden Menschen mit ihrer Gesundheit oder ihrem Leben ist.		
17	16ge Strecken von 5 000 Schaltungen im Jahre 1981 auf 20000 im Jahre 1991 programmiert		,		
18	17Jeder Freund war darauf programmiert		die von ihm gewünschten Verhaltensweisen zu adaptieren.		
19	18Sie war nur für den Verkehr mit Citymännern programmiert		, und deren Vokabular und Bedürfnisse unterschieden sich beträchtlich von dem offenkun		
20	19Wenn eine Geburt auf sieben Stunden programmiert		wird, werden die individuellen Bedürfnisse des Babys nicht berücksichtigt.		
21	20, jedoch medizinische Gründe, aus denen eine Geburt eingeleitet oder sogar programmiert		werden muß:		
22	21Das Unvermögen ist vorprogrammiert		.		
23	22Sie können beispielsweise als "Demand-Schreitmacher" programmiert		werden, die erst bei Bedarf in Aktion treten:		
24	23n in aller Stille die Einsicht um sich, daß darin der Weg in eine Sackgasse programmiert		sein könnte.] --> X1: BEGIN erklärt in einer Rede anlässlich der Eröffnung der Winterse		
25	24Ist es zu schaffen, oder ist die Nichteinlösung des Versprechens schon vorprogrammiert		?		
26	25Unter der Verwaltung kühler Technokraten ist Phantasielosigkeit vorprogrammiert		.		
27	26Die Harpoon wird vor ihrem Abschuß mit den Zielen daten programmiert		und steuert sich in der letzten Phase ihres Fluges mit Hilfe eines eingebauten Radarsyste		

Fig. 9: Formatierung der Tabelle in Calc und Setzen eines Filters

so erscheint der Wert zunächst in der ersten Zeile. Drückt man nun statt der Eingabetaste Alt+Enter, so wird der in der ersten Zeile eingegebene Wert auf alle folgenden Zellen übertragen.

Damit ist die Spalte nun vollständig ausgefüllt.

### 3.1.2 Annotation metaphorisch vs. nicht-metaphorisch

Für die weitere Annotation können wir die nicht-prädikativen Fälle außer Acht lassen.

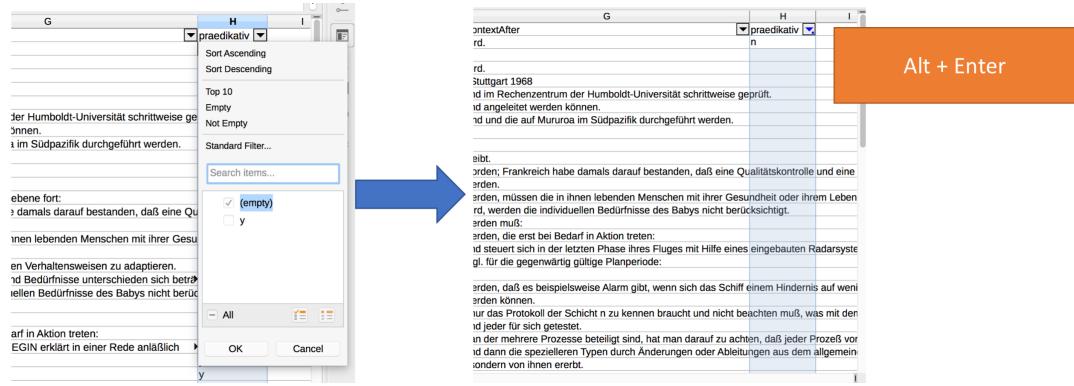


Fig. 10: Eine Tabellenspalte wird so gefiltert, dass nur noch die leeren Zellen zu sehen sind, und allen leeren Zellen wird mit Alt+Enter derselbe Wert zugewiesen.