

DYN HA / SS22 FUB / Liso

st. schwarz

20230101(11.45)

Contents

1	A. head	2
2	einleitung	2
2.1	task	2
3	zur autorin	2
4	work	3
4.1	corpus aufbereitung	3
4.2	basic statistics	3
4.2.1	ground truth	3
4.2.2	multiLX	3
4.2.3	similarities	4
4.3	sentiment analysis	6
4.3.1	in words	6
4.3.2	visualisation	8
4.3.3	dependencies	15
4.3.3.1	book/chapter dependency	15
4.3.3.2	multiLX dependency	17
4.4	nice network	21
4.5	corpus play	21
4.5.1	most probable text	21
B. REF:		22

1 A. head

Hausarbeit im Fach AVL, Modul “Vergleichende Literaturgeschichte”.

Seminar: “DIE SCHÖNSTEN LENGEVITCHES: DYNAMIKEN DER POST-DEUTSCHEN GEGENWARTSLYRIK”.

Dozentin: Chiara Liso.

SoSe 2022.

eingereicht: 2022-12-30

edited: 2023-01-01

anderes format: [view .html](#)

2 einleitung

Wir werden im Folgenden den Versuch unternehmen, aus einigen statistischen Berechnungen Aussagen zum lyrischen Werk Uljana Wolfs abzuleiten. Ob daraus Erkenntnisse hinsichtlich des Aspekts *postdeutsch* erwachsen, können wir noch nicht sagen. Die Arbeit wird explorativ vorgehen, dh. unser Erkenntnisinteresse ist durchaus ungerichtet. Wir wollen wesentlich einige Methoden zur Anwendung bringen, die geeignet erscheinen, weitere literaturwissenschaftliche Fragestellungen zu beantworten. Eine weiter gefasste Aufgabenstellung dieser Arbeit würde ca. lauten:

2.1 task

Bestimmung charakteristischer Merkmale im lyrischen Werk Uljana Wolfs mithilfe statistischer Methoden.

3 zur autorin

Uljana Wolf, der Öffentlichkeit seit 2005 durch ihre Gedichte bekannt, wurde 2006 für ihr Debüt *kochanie ich habe brot gekauft* (kookbooks 2005) mit dem Peter-Huchel-Preis ausgezeichnet und veröffentlichte seitdem neben Übersetzungen und essayistischen Schriften zwei weitere Gedichtbände, ebenfalls bei *kookbooks*. Dort ist sie in ein enges Netzwerk junger deutschsprachiger Autor:innen eingebunden, die sich auch (hier zum Aspekt *postdeutsch*) mehrheitlich durch ihre Affinität zu mehr- oder polylingualer Dichtung auszeichnen. Es gibt bei kook kaum Dichter:innen, deren Werk nicht irgendwie Mehrsprachigkeit künstlerisch ausdrückt, damit arbeitet. Nicht alle vereint eine wirklich (muttersprachliche Provenienz) in anderen europäischen u.a. Sprachen als dem Deutschen, die meisten jedoch deren unmittelbarer Einfluss auf ihre Dichtung. Uljana Wolf hält dazu in ihrem Aufsatz *ZU MEHRSPRACHIGKEIT IM GEDICHT* fest, dass Mehrsprachigkeit “ein ästhetisches, kein biografisches Moment” sei und “dass auch von einsprachigen oder einmuttersprachigen Autor:innen translinguale Literatur durch verschiedene Schreibstrategien erzeugt werden kann, um festgefügte Sprachordnungen zu verunsichern.” (Wolf 2021, 118). Wolf kommt nicht - sie meint, ihres *kochanie* wegen vielfach vermutet (Wolf 2021, 5) - aus dem polnisch-ukrainischen Grenzland, sondern aus Ostberlin, hat jedoch intensiv russisch und eben auch polnisch gelernt (und gelebt) und sich in alle Sprachen begeben, die für sie fruchtbar waren; zahlreiche Stipendienaufenthalte und nicht zuletzt ihr zwischen New York und Berlin pendelndes Arbeitswesen zeugen davon.

4 work

4.1 corpus aufbereitung

Das Korpus, welches nach Digitalisierung der Buchvorlagen aus einer Datenbank abgerufen wird, enthält 144 Einträge, nach Abzug der Kapitelüberschriften und Zitate 130 Datensätze (Texte), die zur Auswertung herangezogen werden können. Für die Analyse wurde das gesamte (publizierte) lyrische Werk Uljana Wolfs, bestehend aus, in der Reihenfolge des Erscheinens:

- kochanie, ich habe Brot gekauft (Wolf 2005)
- falsche freunde (Wolf 2009)
- meine schönste lengevitch (Wolf 2013)

herangezogen.

4.2 basic statistics

Wir werden versuchen, in der Arbeit einige Kennzahlen zu bestimmen, die charakteristisch für das Werk sein sollen. Basis sind hier Statistiken über Wortlängen und -frequenzen, Distribution multilingualer Elemente über das Korpus und Annäherungswerte zur Bestimmung des *sentiment*. Die Zahlen werden absolut und/oder relativ angegeben; absolut meint hier die konkrete Beziehung zu einer Position im Korpus, relativ meint jeweils die indexikalisierte, auf einer Gesamtheit v.H. angenommene Position oder Relation. Diese Relativierung ermöglicht eine gleichförmige Visualisierung der Daten in glatten Frequenzkurven, um die Verhältnisse schematisch abzubilden. (Zur Berechnung der Fourier-transformierten (FFT) Frequenzen cf. [Jockers \(2015\)](#))

4.2.1 ground truth

Die 130 Texte (Lyrik und lyrische oder experimentelle Prosa) haben einen Umfang von 10434 Wörtern (tokens), die sich in 3976 distinct types einteilen lassen, die type/token ratio, ein Indikator für *lexical diversity*, beträgt demnach 0.3810619. Die durchschnittliche Textlänge (median) beträgt 69 Wörter. Wir haben noch keine Vergleichswerte, die sinnvoll wären. . .

4.2.2 multiLX

Die multilingualen Elemente des Korpus (insgesamt 700 tokens) haben einen Anteil von 9% an den types. Mit der Textmatrix (cf. Table 1) lässt sich noch weiter rechnen.

Table 1: *simplest matrix of text beginnings*

gegen	vier	uhr	morgens	beobachte	ich	die	verschiebung
ach	wär	ich	nur	im	aufwachraum	geblieben	traumverloren
ach	wär	ich	nie	im	aufwachraum	gewesen	taub
schließ	mich	ein	liebe	ins	gebet	in	die
diese	kästen	enthalten	frauen	die	nicht	bearbeitet	werden
meine	väter	sind	einfache	männer	sie	haben	töchter
meine	väter	sind	keine	einfachen	männer	sie	haben
meine	münder	sind	keine	einfachen	väter	der	erste
meine	väter	sind	einfache	vermesser	der	erste	geht
meine	väter	sind	keine	einfachen	vermesser	der	erste

occurences of non german words

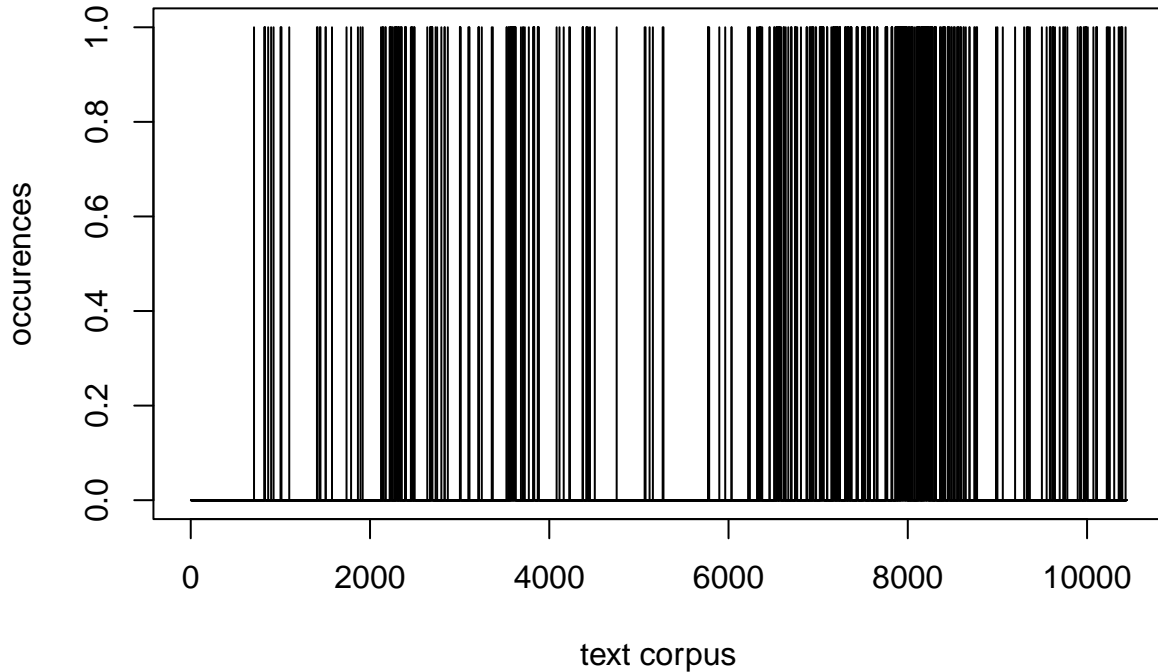


Figure 1: multilingual elements over corpus

4.2.3 similarities

Zum Beispiel lassen sich die Wortgleichungen visualisieren, die an bestimmten Positionen des Textes aufscheinen. Die Höhen in der folgenden Graphik markieren relative (Fourier-transformierte) Wortpositionen, an denen von Wolf die **wenigsten** analogen Wörter verwendet wurden. Es lässt sich erkennen, dass ein Text meist mit denselben Wörtern anfängt (baisse), die immer verschiedener werden, um sich bei der Hälfte des Textes über eine lange Strecke zu gleichen und ab 60% sprunghaft zu divergieren bis sie um 78% einen peak (hausse) erreichen an Verschiedenheit.

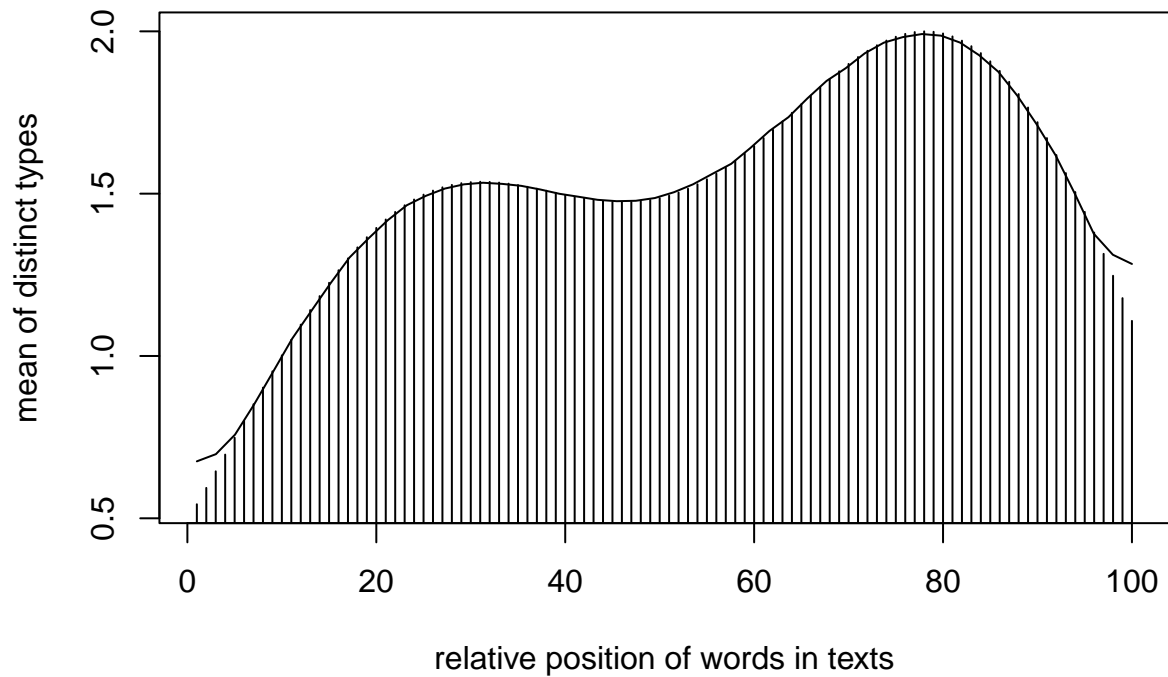


Figure 2: distinctness of word positions

Eine weiterhin schöne Graphik entsteht, wenn man die Matrix der Zeichenanzahl über die Gesamtheit der Wörter visualisiert. Hier zeigt sich, dass ein Text zwischen 21-31% die längsten Wörter (mean: 9.5 characters) enthält, diese zwischen 36 und 56% kürzer werden bis hin zu 7.76 Zeichen um bei 76% einen erneuten peak in der Zeichenanzahl (mean: 8.98 characters) zu erreichen.

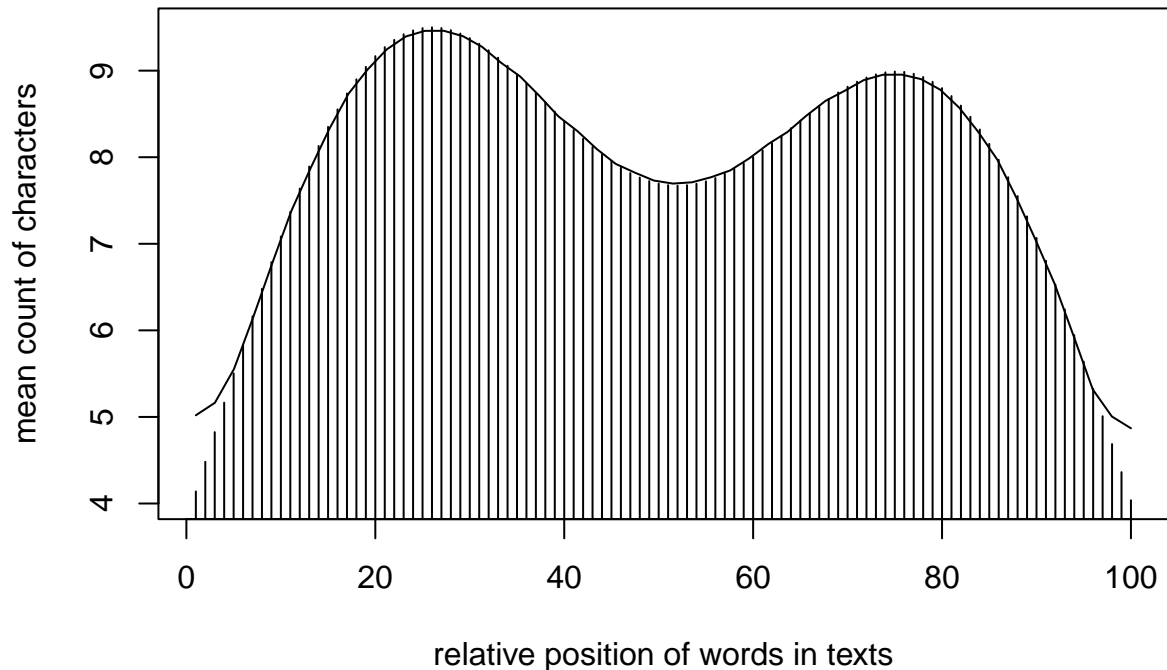


Figure 3: characters per position

4.3 sentiment analysis

Für jeden einzelnen Text kann ein Wert bestimmt werden, der Aussagen darüber zuliesze, wo im Spektrum positive/negative sentiment sich dieser verorten lässt. Die absoluten *sentiment values* werden weiterhin durch Fourier-Transformation auf die Gesamtheit des Korpus projiziert, um eine glatte, von der absoluten Position unabhängige Darstellung zu ermöglichen. (cf. Figure 10). Welche Aussagen sich aus diesen Erkenntnissen ableiten lassen sollen, ist mir noch nicht ganz klar. I will elaborate on that.

4.3.1 in words

summary:

```
## $min
## $min$head
## [1] "kreisau, nebelvoliere"
##
## $min$book
## [1] "kochanie"
##
## $min$chapter
## [1] "krzyżowa, gefährten"
##
```

```

## $min$id
## [1] "142"
##
## $min$words
##   du  mich   ich  Brust  vogel  halb
##   4   3     3   2     2     2
##
##
## $max
## $max$head
## [1] "an die kreisauer hunde"      "nachtrag an die kreisauer hunde"
##
## $max$book
## [1] "kochanie" "kochanie"
##
## $max$chapter
## [1] "krzyżowa, gefährten" "krzyżowa, gefährten"
##
## $max$id
## [1] "143" "144"
##
## $max$words
##   euch   ihr gehört   echo   nur   nicht
##   7     4     3     2     2     2

```

Zum Beispiel: Die höchsten *sentiment* Werte, hier in [chapter:] *krzyżowa, gefährten, krzyżowa, gefährten*, lassen sich in [text:] *an die kreisauer hunde, nachtrag an die kreisauer hunde* finden. Die *most frequent words* dieses Abschnitts sind *euch, ihr, gehört, echo, nur, nicht*, die niedrigsten finden sich im Text *kreisau, nebelvoliere* mit *du, mich, ich, Brust, vogel, halb*.

4.3.2 visualisation

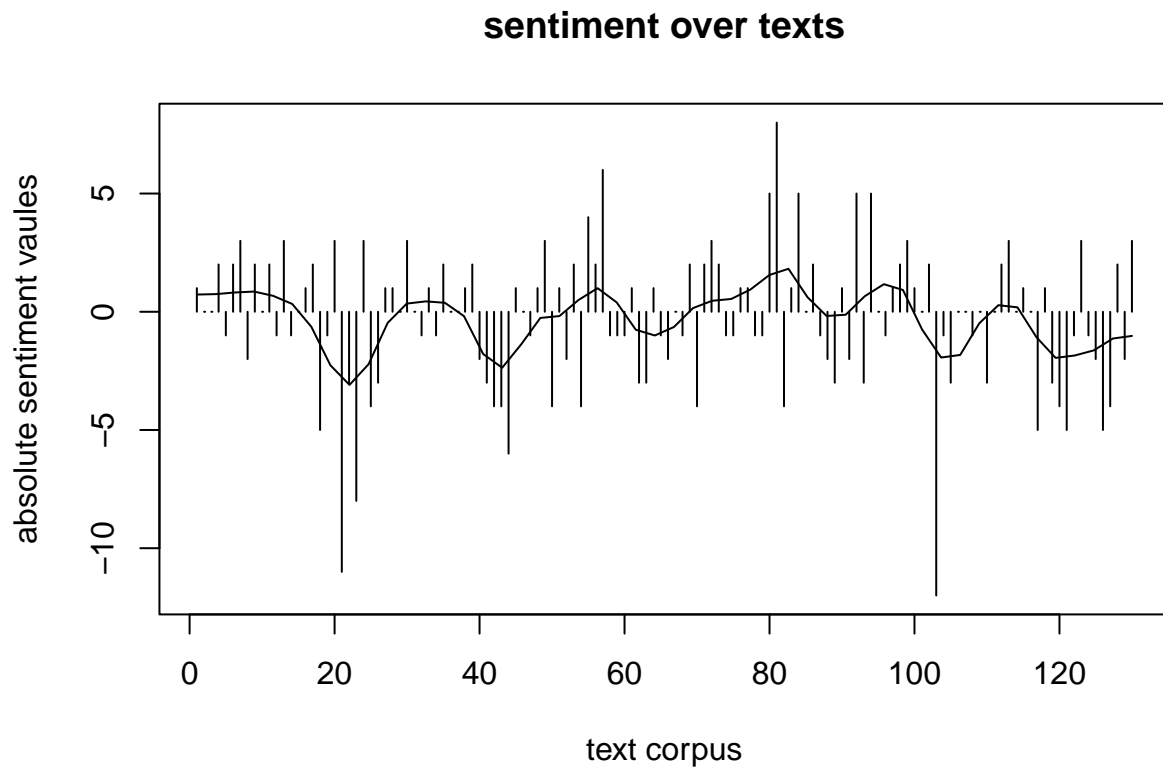


Figure 4: absolute *sentiment values* über das gesamte Textkorpus

sentiment over texts

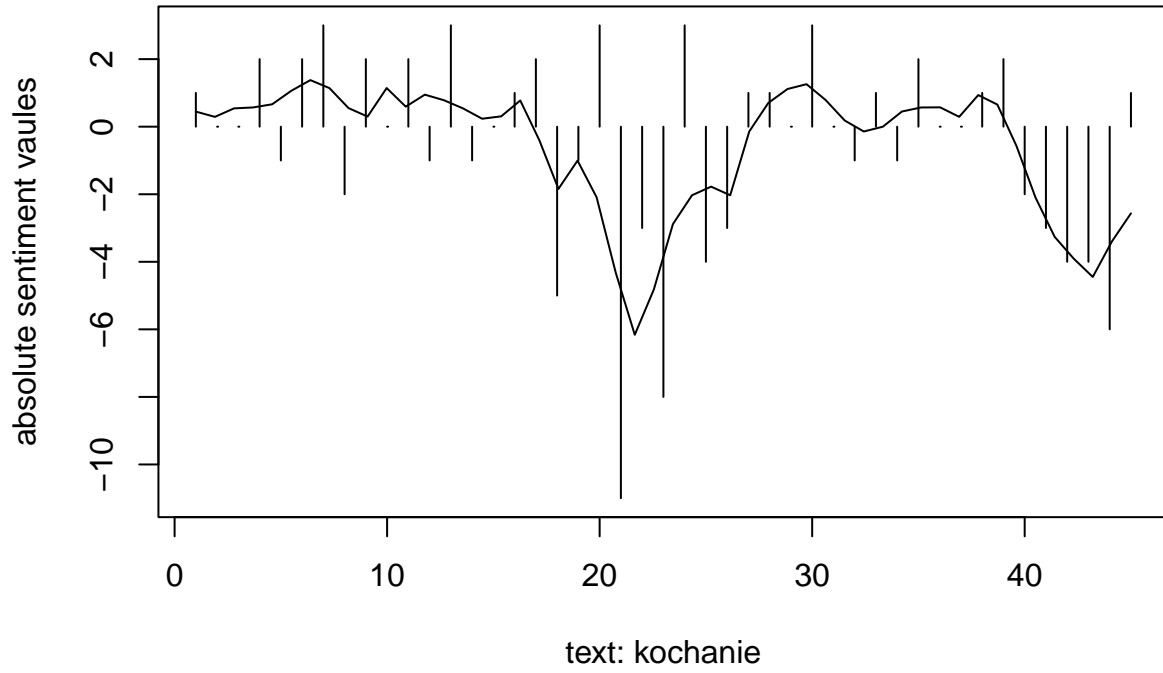


Figure 5: absolute *sentiment values* über *kochanie ich habe brot gekauft*

sentiment over texts

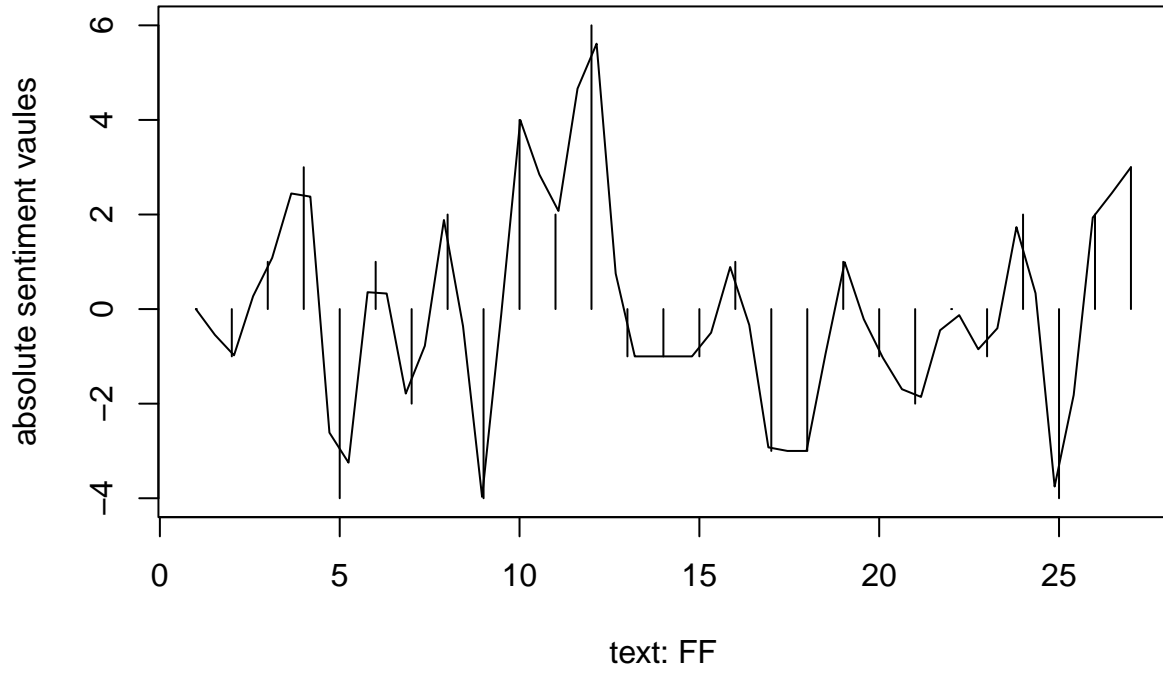


Figure 6: absolute *sentiment values* über *falsche freunde*

sentiment over texts

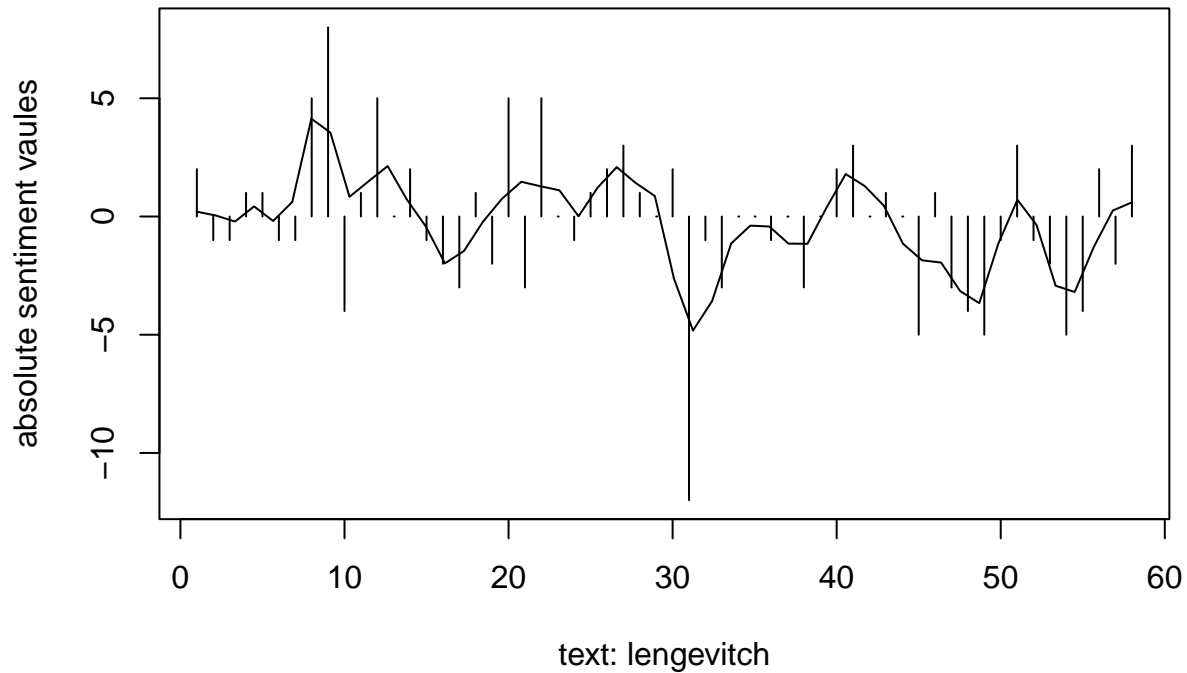
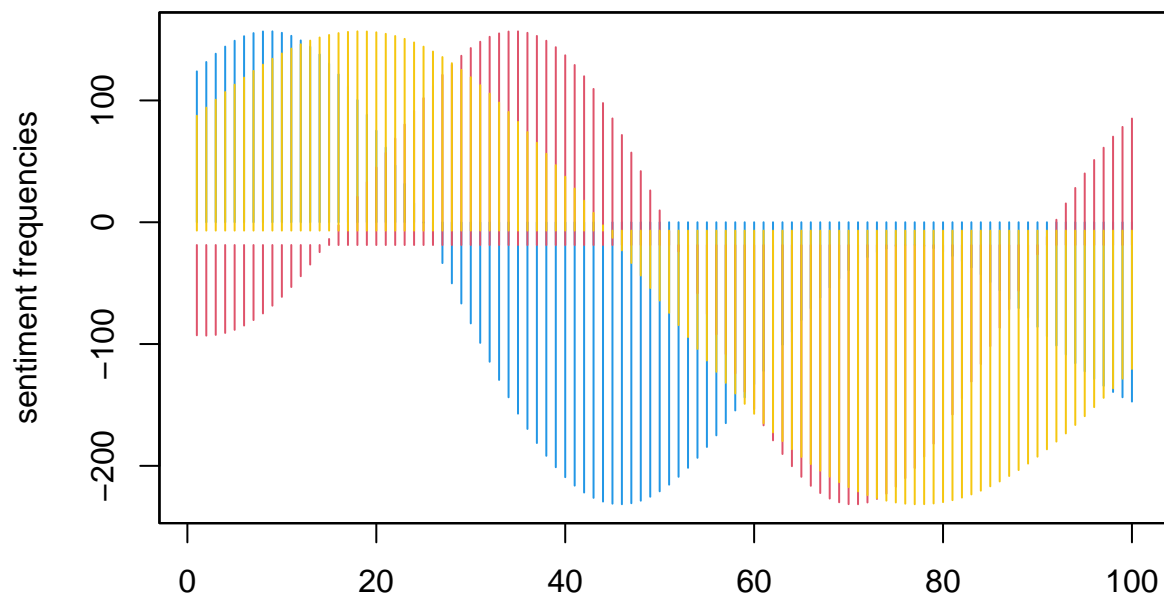


Figure 7: absolute *sentiment values* über *meine schönste lengevitch*

Wir sehen in den frequenzanalysierten Graphiken, dass sich die relative Verteilung der sentiment values über die Bücher in der Tendenz unterscheidet. Während in *kochanie* und *lengevitch* schon zu Beginn Höhen verzeichnet sind, die Werte dann zur Buchmitte hin sinken, zeigt sich bei *FF* ein erster Wechsel von negativ zu positiv schon im ersten Drittel. Die Stimmung aller drei Bücher ist in der Mitte gleich negativ (bewertet), in *kochanie* aber schon wieder aufsteigend, *falsche freunde* weist die grössten Schwankungen auf.

sentiment distribution over book



blue: kochanie, red: FF, yellow: lengevitch

Figure 8: relative sentiment values singled

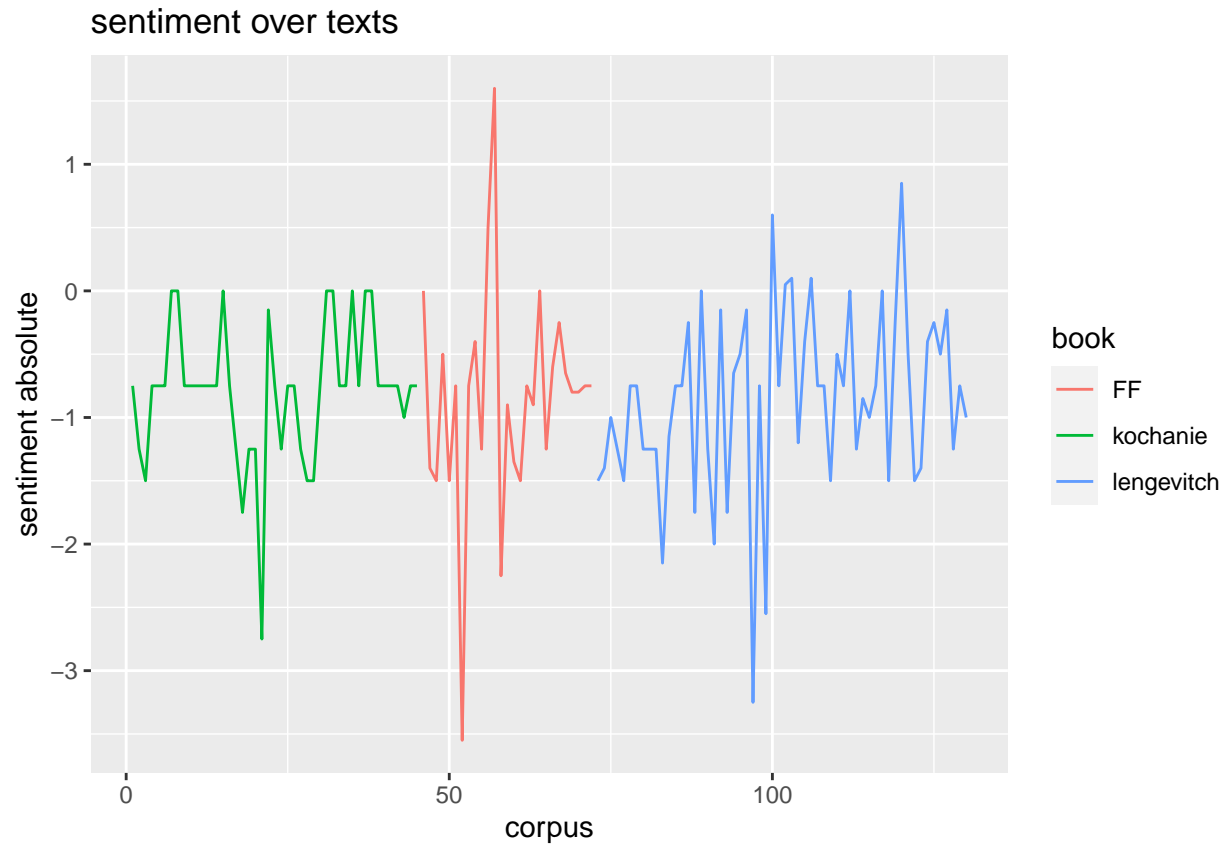


Figure 9: absolute *sentiment values* over corpus

In *kochanie* weisen die Werte die niedrigste, in *falsche freunde* die höchste Varianz auf.

sentiment over texts

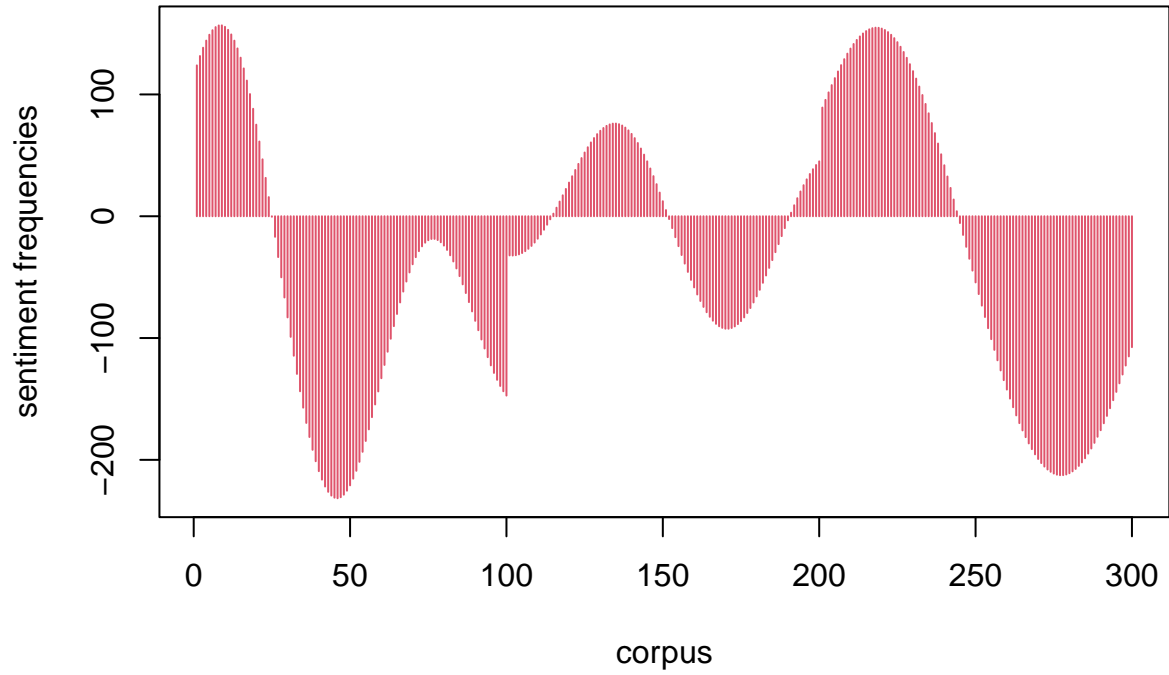


Figure 10: frequenzanalyierte (Fourier) *sentiment values* über Korpus

sentiment analysis over texts

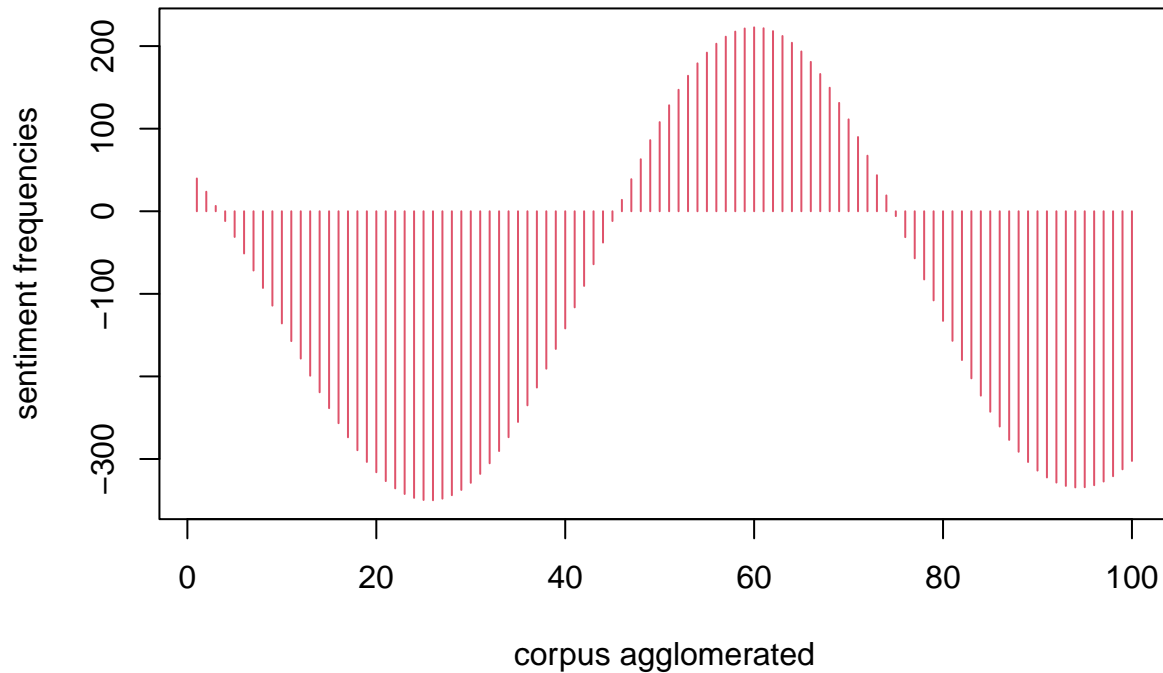


Figure 11: frequency analysis

4.3.3 dependencies

Mit der Regressionsanalyse des *R lme4 package* (Bates et al. 2015) lassen sich hier Abhängigkeiten (Korrelationen) der *sentiment values* von verschiedenen Faktoren (chapter, book, multilingual elements) aufzeigen.

4.3.3.1 book/chapter dependency summary:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: dta$sentiment ~ dta$book + (dta$book | dta$chapter)
## Data: dta
##
## REML criterion at convergence: 650.1
##
## Scaled residuals:
##   Min      1Q  Median      3Q      Max
## -3.9330 -0.5158  0.0577  0.6694  2.7037
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   dta$chapter (Intercept)          0.2986   0.5465
##                   dta$bookFF      8.2772   2.8770   0.35
##                   dta$bookkochanie 0.2987   0.5465  -1.00 -0.35
```

```

##           dta$booklengevitch 0.4123   0.6421  -0.87 -0.33  0.87
## Residual                      8.9266   2.9877
## Number of obs: 131, groups:  dta$chapter, 12
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   9.795e-16  3.037e+00  0.000
## dta$bookFF    -1.004e-15  4.386e+00  0.000
## dta$bookkochanie -5.556e-01  3.070e+00 -0.181
## dta$booklengevitch -1.991e-01  3.065e+00 -0.065
##
## Correlation of Fixed Effects:
##              (Intr) dt$bFF dt$bkk
## dta$bookFF  -0.693
## dta$bokkchn -0.989  0.685
## dt$bklngvtc -0.991  0.686  0.980
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

```

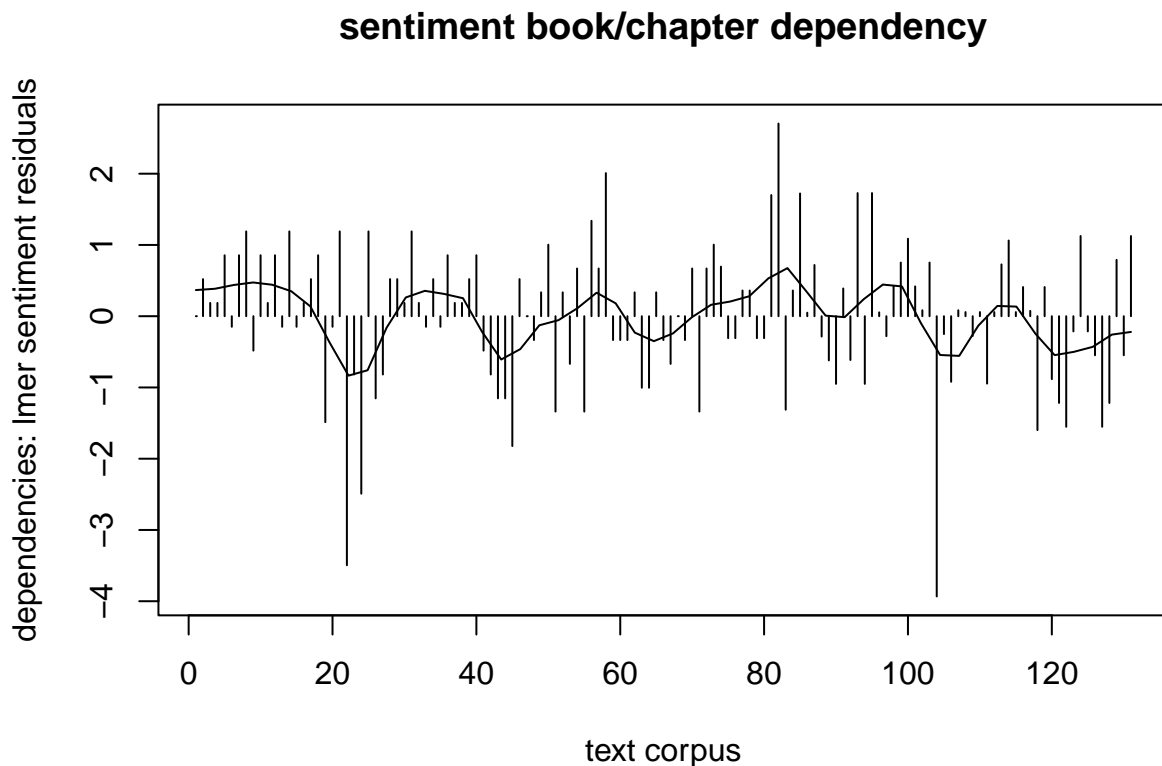


Figure 12: linear regression of sentiment dependencies, absolute

Es ist zu zeigen, dass die Abhängigkeit durchaus variiert. Die relativen Korrelationswerte schwanken zwischen -82 und 75 (bei $f(x)=x/100$), zwischen 40-80% lässt sich größere Abhängigkeit beobachten, dh. hier sind die sentiment values am stärksten vom Kapitel beeinflusst, beispielhaft beim Text[82:] , am wenigsten beim Text[104:] . (In der Regressionsanalyse wurden Veränderungen der Werte (Variablen) **sentiment** der Texte

auf einen Zusammenhang mit Buch und Kapitel der jeweiligen Messung geprüft, dh. bestimmt, ob Buch und/oder Kapitel einen Einfluss auf den Wert haben.)

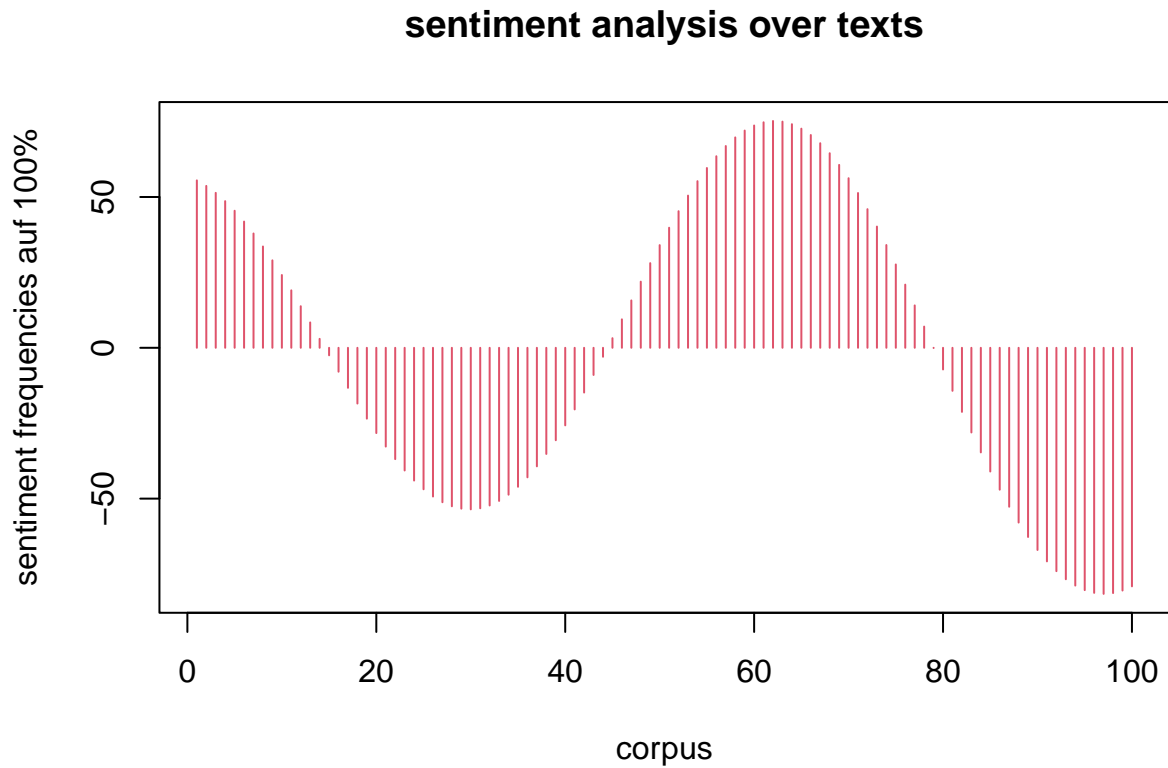


Figure 13: linear regression of sentiment dependencies, relative

4.3.3.2 multiLX dependency Weiterhin können wir versuchen, eine Abhängigkeit der sentiment values von der Verwendung multilingualer Elemente aufzuzeigen. Die verdichteten schwarzen Balken (cf. Figure 1) korrelieren hier mit der roten Linie der Stimmungswerte, was eine Abhängigkeit vermuten lässt.

sentiment/language correlation

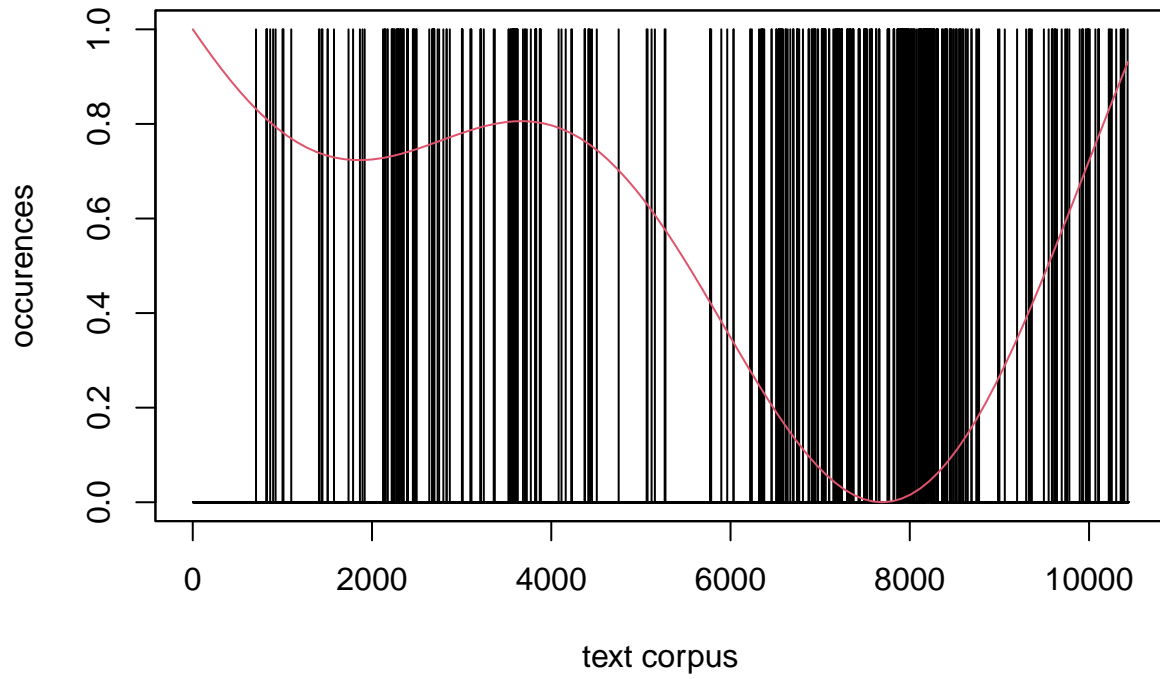


Figure 14: absolute positioned multilingual elements over sentiment

sentiment/language correlation

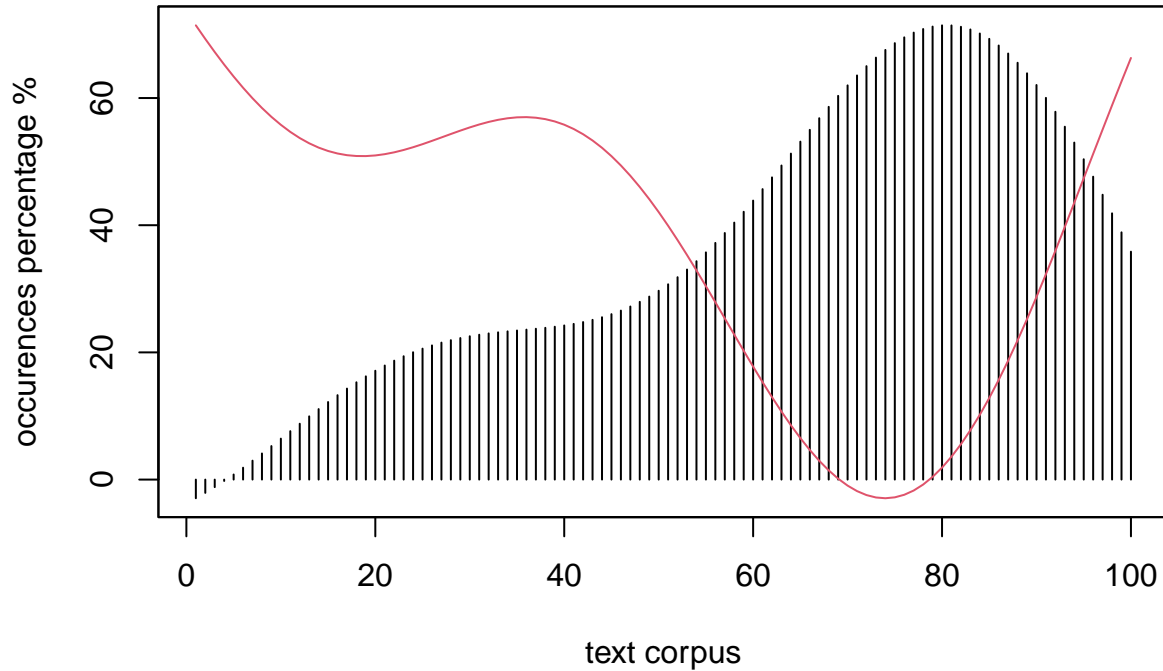


Figure 15: percentage of multilingual elements over sentiment

summary:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: dta_t$sentiment ~ dta_t$book + (dta_t$book | dta_t$chapter) +
##   (dta_t$lxp | dta_t$chapter) + (1 + dta_t$lxp) + (1 + dta_t$ttr)
## Data: dta_t
##
## REML criterion at convergence: 287.7
##
## Scaled residuals:
##   Min      1Q  Median      3Q      Max
## -3.9484 -0.5075  0.0293  0.5177  3.4563
##
## Random effects:
## Groups      Name                Variance Std.Dev. Corr
## dta_t.chapter (Intercept)          0.2844653 0.53335
##              dta_t$bookFF      0.2518810 0.50188 -0.03
##              dta_t$bookkochanie 0.2843860 0.53328 -1.00  0.03
##              dta_t$booklengevitch 0.2843925 0.53328 -1.00  0.03  1.00
## dta_t.chapter.1 (Intercept)      0.0394181 0.19854
##              dta_t$lxp           0.0001687 0.01299 -1.00
## Residual                        0.4794959 0.69246
## Number of obs: 131, groups: dta_t$chapter, 12
```

```

##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)      -1.391441   1.194420  -1.165
## dta_t$bookFF      -0.773382   1.166167  -0.663
## dta_t$bookkochanie -0.590773   0.918345  -0.643
## dta_t$booklengvitch -0.641043   0.917044  -0.699
## dta_t$lxp          0.005018   0.008272   0.607
## dta_t$ttr          1.391441   0.789468   1.763
##
## Correlation of Fixed Effects:
##           (Intr) dt_$FF dt_t$bkk dt_t$bkl dt_t$l
## dta_t$bokFF -0.624
## dt_t$bkkchn -0.820  0.761
## dt_t$bklngv -0.839  0.769  0.976
## dta_t$lxp   -0.083 -0.075  0.006  -0.057
## dta_t$ttr   -0.661  0.072  0.133   0.160   0.125
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

```

Die Abhängigkeit der sentiment values vom Vorhandensein multilingualer Elemente lässt sich kurz umreißen: Wir stellen den grössten Zusammenhang mit [coefficient] $4.6e-05$ bei `dta_t$bookkochanie` fest, die Differenz zu `dta_t$bookFF` beträgt 7.74, zu `dta_t$booklengvitch` 4.75 Punkte, der Abstand der Abhängigkeit hier also 2.99 Punkte.

Die Abhängigkeit der sentiment values von der type/token ratio der Texte ebenfalls kurz umrissen: Wir stellen den grössten Zusammenhang mit [coefficient] 0.115873 bei `dta_t$booklengvitch` fest, die Differenz zu `dta_t$bookFF` beträgt 499.78, zu `dta_t$bookkochanie` 195.96 Punkte, der Abstand der Abhängigkeit hier also 303.82 Punkte.

4.4 nice network

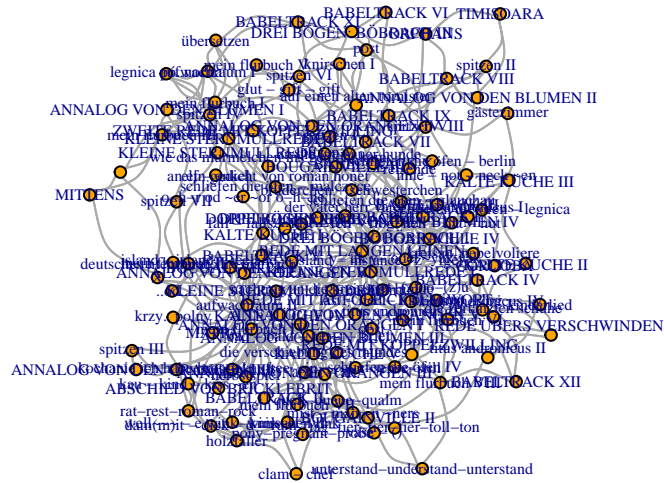


Figure 16: cooccurrences network of text elements

4.5 corpus play

Das Folgende zeigt einen Text, dem ein Algorithmus zugrundeliegt, der roughly die Wahrscheinlichkeit des Vorhandenseins eines Wortes an der jeweiligen Position im Text, über den gesamten Korpus betrachtet, bestimmt. Es tritt also jedes Wort dieses künstlichen Textes am wahrscheinlichsten an dieser Stelle auf, szsg. ein sehr simples *transformer* experiment, ohne jegliche Berücksichtigung semantisch-syntaktischer Kategorien und (noch) weit von einem lernfähigen Algorithmus entfernt

4.5.1 most probable text

13506.ST: Uljana Wolf / GPTtestee «< ein a a a einfachen a a i der und ich i in b in ein i a a er a es in in a u h in i a der in ein a der an in a i er er in es a a und in in in u er ich i in a in an in die i in er in u a in a den in in a a i in ich ein oder in i in in to in in a u wo das und in an ein der a es in der er der mit im ein die der es nicht die die b u t wir in die an ihre in ein ewig ich oder oder ich in nicht gehört ihr muttersprache lernen platz er herkunft is ich decke richtung u ich sich nicht dinge puckerten darin too small vom mesmer getroffen ungelenkten wagten war der anders anders anders er anders anders anders anders anders am es ein rage dann tee aber wenn es zeit ist für orangen ist es keine zeit no time at all für durst für wasser ich sich arrangieren lassen weil orangen sind ihr eignes rangiermaterial weil züge brücken ich schiffchen keeps me going unbeschwichtigt go nicht go lack use to her weil orangen kommunizieren durch die zeiten orangen oder rundum residenzpflicht orangen oder in eine grenzsicht its like were in prison es trees and animals were at

the für of the world wenn es die ist für tarnopol ist es keine zeit not time at all für weltenden schlossvorhänge
papiervorgänge weil orangen sind lebensmittelpunkt weil ein mangel mangel geh sei ihr von nutzen geh die
esse putzen dass der doktor wiederkommt dass nicht bis die türen aufgehen und man seine sippe auf die
straße setzt sie sagen a family is what one blumenschmuggler ihre dunklen wege durch europa ziehen sie
sagen we make a business with such bodies with blüten poesie und ich sage we will see what can all the
green stuff be recht auf seinen freigang hat aber dass orangen «< fin.

B. REF:

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Jockers, Matthew L. 2015. “Revealing Sentiment and Plot Arcs with the Syuzhet Package Matthew L. Jockers.” <https://www.matthewjockers.net/2015/02/02/syuzhet/>.
- Wolf, Uljana. 2005. *Kochanie Ich Habe Brot Gekauft: Gedichte / Uljana Wolf*. 1. Aufl. Reihe Lyrik 5. Idstein: Kookbooks.
- . 2009. *Falsche Freunde: Gedichte / Uljana Wolf*. 1. Aufl. Kookbooks : Reihe Lyrik 15. Idstein: kookbooks.
- . 2013. *Meine Schönste Lengevitch: Gedichte / Uljana Wolf*. 1. Auflage. Reihe Lyrik Band 32. Berlin: kookbooks.
- . 2021. *Etymologischer Gossip: Essays Und Reden / Uljana Wolf*. 1. Auflage. Kookbooks Reihe Essay 7. Berlin: kookbooks.