

Studie:

Auswirkung von Position und Sequenz bei Lautsymbolik



Ludwig-Maximilians-Universität München
Institut für Deutsch als Fremdsprache
Hauptseminar Lautsymbolik
Dozenten: Prof. Ph. D. van Peer
Dr. Albers
Dr. Auracher

Andreas Neumann
Münchner Str. 7
85635 Höhenkirchen
Tel. 0176 / 220 50 984
andi@neumann.biz

Abstract

Ziel der Untersuchung ist es heraus zu finden ob Lautsymbolik prinzipiell durch einzelne isolierte Laute erzeugt wird oder ob Position und Sequenz entscheidend sind.

Dazu werden mit Hilfe zweier Korpora Wörter generiert. Ein Auszug aus diesen generierten Wörtern wird dann mit Hilfe einer Online-Befragung auf ihre Assoziation hin überprüft. Entspricht die Assoziation der Ausrichtung des Ursprungskorpus kann man von einem lautsymbolischen Zusammenhang ausgehen.

Je nach gewähltem Generierungsverfahren werden nur Lautinventar, Lautinventar und Sequenz oder Lautinventar, Sequenz und Position berücksichtigt.

Als Ergebnis lässt sich festhalten dass die Sequenz der Laute in der Wortmitte kaum maßgebend ist. Die Zurordnung wird auf Basis einzelner ,isolierter Laute getätigt. Anfangs- und Endlautfolgen kommen aber besondere Bedeutung zu.

Inhalt

Auswirkung von Position und Sequenz bei Lautsymbolik

1. Aufbau der Korpora	5
1.1 Vorüberlegung.....	5
1.2 Die Quellen.....	5
1.3 Bereinigung der Korpora.....	6
1.4 Die Korpora.....	7
2. Generierung der Testwörter	8
2.1 Klasse 1 : Einzellaute, beliebig kombiniert.....	8
2.2 Klasse 2: Bigramme, beliebig kombiniert.....	10
2.3 Klasse 3 , <Anfangsbigramm><beliebiges Bigramm>+<Endbigramm>	11
3. Die Umfrage	13
3.1 Form der Umfrage.....	13
3.2 Aufbau des Fragebogens.....	13
3.3 Zahlen.....	14
4. Ergebnis	15
4.1 Vergleich der Wortklassen.....	16
4.2 Klasse I / Monogramme.....	17
4.3 Klasse 2/ Bigramme.....	17
4.4 Klasse 3/ Anfangs- und Endbigramme.....	17
4.5 Fazit.....	18

Anhang A - Die Korpora

Korpora	19
1. Korpus A - Wörter mit positiver Assoziation.....	19
2. Korpus B - Wörter mit negativer Assoziation.....	20

Anhang B - Die Programme

Auswertungsprogramme	22
1. Das Programm „auswertung.rb“.....	22
Generierungsprogramme	25
1. Das Programm „mono.pro“.....	25
2. Das Programm „bi.pro“.....	27
2. Das Programm „ame.pro“.....	29

Anhang C - Fragebögen

Der Fragebogen	31
Version I.....	32
Version II.....	34

Anhang D - Daten

Empirische Daten	36
Fragebogen I.....	37
Fragebogen II.....	39
Grafiken	41
Monogramme.....	42
Bigramme.....	43
Anfangs- und Endbigramme.....	44
Vergleich der Generierungsmethoden.....	45

Auswirkung von Position und Sequenz bei Lautsymbolik



I. Aufbau der Korpora

Für diese Studie wurden zwei Korpora erstellt. Korpus A mit 133 Einträgen bestehend aus Wörtern die positive Emotionen wecken und Korpus B bestehend aus 131 Wörtern die negative Emotionen wecken.

I.1 Vorüberlegung

Aufgrund der vielen, sich teilweise widersprechenden, Klassifizierungsschemata für Emotionen beschränke ich mich auf eine Aufteilung in positive und negative Emotionen.

I.2 Die Quellen

Als Ausgangspunkt diene „Verbal labels describing major affect states List of related terms in 5 languages“ (Hillsdale, NJ: Erlbaum 1988). Diese deutschen Wörter der Listen in eine Gruppe mit negativen und eine positiven Assoziationen eingeteilt, wobei nicht sicher zuordbare Wörter entfernt wurden.

Mit Hilfe des Wortschatzportals der Universität Leipzig (<http://wortschatz.uni-leipzig.de/>) wird die Liste um Synonyme erweitert.

I.3 Bereinigung der Korpora

Um eine Verfälschung der Daten zu vermeiden wurden folgende Ersetzungen und Bereinigungen vorgenommen:

Grammatische Morpheme wie „-ung“, „-heit“, „-lich“ usw. gelöscht:

Hemmung -> hemm
 kleinlich -> klein

Wörter die ihre Bedeutung vor allem Suffixen verdanken, wie „un-“ und „-los“ wurden wenn passend in die gegenteilige Klasse verschoben, ansonsten entfernt.

Unmoral (in negativer Klasse) -> moral (in positiver Klasse)

Einzelne Laute die als Folge von mehreren Zeichen geschrieben werden wurden durch ein an das IPA angelehntes Zeichen ersetzt:

sch -> ʃ
 c nach a,o,u -> X
 c nach i und e --> ç

Komposita wurden in ihre Bestandteile zerlegt.

Pesttot -> Pest und Tot
 Himmelsgabe -> Himmel und Gab

Nach diesen Schritten bestehen die Korpora größtenteils aus Wortstämmen, wobei ein Zeichen einem Laut entspricht.

I.4 Die Korpora

Korpus A hat 133 Einträge mit 844 Lauteinheiten, die positive Assoziationen wecken.

Korpus B hat 131 Einträge mit 867 Lauteinheiten, die negative Assoziationen wecken.

I.4.1 Auszug aus Korpus A

dusel	liçt	zärt
eben	lust	über
edel	lösen	üppig
ehr	maXt	farm

I.4.2 Auszug aus Korpus B

abfeu	horror	zöger
angreif	hölle	ärger
angst	jammer	über
anti	kalt	fauder



2. Generierung der Testwörter

Aus den Korpora werden im ersten Schritt mit Hilfe der Differenzlistentechnik klassentypische „Wortbausteine“ extrahiert. In einem zweiten Schritt werden diese Bausteine zu neuen Wörtern kombiniert.

Es werden drei Wortklassen erzeugt:

2.1 Klasse I : Einzellaute, beliebig kombiniert

Die für diese Klasse generierten Wörter beruhen auf der Annahme, dass die Reihenfolge und Position der Laute arbiträr ist.

2.1.1 Bausteine Klasse I

Die häufigsten Laute in Wörtern mit positiven Assoziationen sind:

15	i
14	g
13	h
10	l
7	o
7	t

Für die Wörter mit negative Assoziationen erhält man:

21	a
12	r

10	j
8	m
8	v
8	d
7	k

Für mehr Informationen zur Erkennung der Bausteine, siehe Anhang B: „auswertung.rb“.

2.1.2 Generierung neuer Wörter der Klasse I

Durch Permutation und Ausschluss jener Wörter die nicht aussprechbar sind erhält man aus den Ausgangsdaten für positive Wörter 88080 mögliche Wörter der Wortlänge drei bis sieben.

Auszug aus „good_mono.txt“:

goohl	liliglo	olotilg	tholi	higto
goohli	liligo	olotilh	tholig	higtog
goohlig	liligog	olotili	tholigg	higtogg
goohlih	liligoh	olotill	tholigh	higtogh

Durch Permutation und Ausschluss jener Wörter die nicht aussprechbar sind erhält man aus den Ausgangsdaten für negative Wörter 65520 mögliche Wörter der Wortlänge drei bis sieben.

Auszug aus „bad_mono.txt“:

aak	dschardasch	dschavmam	mscharkar	vakmad
aaka	dschardav	dschavmar	mscharkasch	vakmada
aakaa	dschark	dschavmasch	mscharkav	vakmadd
aakaad	dscharka	dschavmav	mscharm	vakmadk

Für genauere Informationen zum Vorgang der Erzeugung, siehe Anhang B: mono.pro

2.2 Klasse 2: Bigramme, beliebig kombiniert

Die Wörter der Klasse 2 beruhen auf der Annahme, dass die Abfolge der Laute Einfluss auf Lautsymbolik eines Wortes hat. Positionelle Informationen werden aber nicht berücksichtigt. Bigramme wurden außerdem gewählt da sie oft mit Silben korrelieren.

2.2.1 Bausteine Klasse 2

Die häufigsten Bigramme in Wörtern mit positiven Assoziationen sind:

9	fr
8	he
5	ed
5	nn
5	gl
5	iç
5	li
5	ie

Für Wörter mit negative Assoziationen erhält man:

15	er
8	ng
7	ve
7	an
6	ra
5	bi
5	we
5	rd
5	al
5	fe
5	au

Für mehr Informationen zur Erkennung der Bausteine, siehe Anhang B: „auswertung.rb“.

2.2.2 Generierung neuer Wörter der Klasse 2

Durch Permutation und Neukombination der Bigramme erhält man 4672 Wörter. Aufgrund des Erzeugungsverfahrens sind die Wörter vier, sechs oder acht Zeichen lang.

Auszug aus „bi_good.txt“:

edich	glichieli	fredfrich	glichiegl	liennnn
ediched	glichienn	fredfrie	glichiehe	lili
edicheded	glichli	fredfrli	glichieich	lilied
edichedfr	glichlied	fredfrnn	glichieie	lilieded

Durch Permutation und Neukombination der Bigramme erhält man 16093 Wörter. Auszug aus „bi_bad.txt“:

alanngal	birdvebi	raerwe	vebibowe	werabibi
alanngan	birdveer	raerweal	vebier	werabier
alanngau	birdvefe	raerwean	vebieral	werabife
alanngbi	birdveng	raerweau	vebieran	werabing

Für genauere Informationen zum Vorgang der Erzeugung, siehe Anhang B: bi.pro

2.3 Klasse 3 , <Anfangsbigramm><beliebiges Bi-gramm>+<Endbigramm>

Die Wörter dieser Klasse beruhen aus einer Zusammensetzung von Bigrammen. Dabei wird berücksichtigt an welcher Position im Korpus das Bigramm auftritt. Unterschieden wird zwischen Bigrammen die am Anfang stehen, Bigrammen in der Wortmitte und Bigrammen die am Wortende stehen.

2.3.1 Bausteine Klasse 3

Wörter mit positiven Assoziationen :

Anfang:		Mitte		Ende:	
8	^he	5	ed	4	en\$
7	^ge	5	nn	3	ig\$
7	^fr	5	iç	3	tz\$
5	^er	5	li	3	rt\$
4	^gl	5	ie	3	ll\$

Für die Wörter mit negative Assoziationen erhält man folgende Aufstellung:

Anfang:		Mitte:		Ende:	
6	^ve	6	ra	12	er\$
4	^fe	5	bi	4	al\$
3	^dü	5	we	4	ng\$
3	^dr	5	rd		
3	^an	5	au		

Für mehr Informationen zur Erkennung der Bausteine, siehe Anhang B: „auswertung.rb“.

2.3.2 Generierung neuer Wörter der Klasse 3

Wörter der Klasse 3 haben folgende Aufbau:

$\langle \text{Wort} \rangle ::= \langle \text{Anfangsbigramm} \rangle \langle \text{Mittelbigramm} \rangle^+ \langle \text{Endbigramm} \rangle$ (BNF)

Ein Wort besteht demnach aus einem Anfangsbigramm, einer beliebigen Anzahl Mittelbigrammen und einem Endbigramm.

Für positive Bigramme ergeben sich mit einer Mittelbigrammlänge von eins bis drei 3875 Möglichkeiten.

Auszug aus „good_ame.txt“

eredig	frediert	glichieli	geichededtz	heliig
eredlieden	fredietz	glichienn	geicheden	helilieden
eredliedig	fredig	glichli	geiche-	heliliedig
eredliedll	fredlieden	glichlied	dichen	heliliedll
			geiche-	

Für negative Bigramme ergeben sich mit einer Mittelbigrammlänge von eins bis drei Buchstaben 2325 Möglichkeiten.

Auszug aus „bad_ame.txt“

anaubial	drauweaual	dueaubibing	feauweraal	vebiwerang
anaubiaual	drauweauer	dueaubier	feauweraer	vebiwerdal
anaubiauer	drauweaung	dueaubing	feauwerang	vebiwerder
anaubiaung	drauwebial	dueaubiraal	feauwerdal	vebiwerdng



3. Die Umfrage

3.1 Form der Umfrage

Aufgrund der leichteren Durchführbarkeit habe ich mich für eine Onlineumfrage per Web-Formular entschieden. Ein Link zum Abfragesystem wurde an die Probanden verschickt.

Es wurden zwei verschiedene Fragebogen benutzt, bei denen die möglichen Antworten der Fragen 2 bis 4 variiert wurden. Ansonsten sind die Bögen identisch.

3.2 Aufbau des Fragebogens

Der Fragebogen gliedert sich in vier Teile:

Frage 1

Hier werden demographische Daten wie Geschlecht, Alter und Muttersprache abgefragt.

Frage 2

Hier werden zehn Wörter der Klasse 1 gegenübergestellt, fünf davon aus dem Korpus mit positiver Wörtern erzeugt und fünf aus dem Korpus negativer Wörter. Der Proband soll entscheiden ob er mit dem Wort positive oder neagtive Assoziationen verbindet.

Frage 3

Siehe Frage 2, die Wörter stammen aber aus Klasse 2.

Frage 4

Siehe Frage 2, die Wörter stammen aber aus Klasse 3.

Frage 5

Dient dazu eine subjektive Wertung der Qualität der Generierungsverfahren abzufragen.

Genaueres, siehe Anhang C

3.3 Zahlen

Teilgenommen an der Umfrage haben 53 Personen.



4. Ergebnis

Teilgenommen an der Umfrage haben 53 Personen. Version 1 des Fragebogens haben 39 Personen ausgefüllt. Version II wurde von 14 Personen beantwortet.

Zur Bestimmung der Signifikanz wurde der χ^2 -Test nach Pearson verwendet.

4.1 Vergleich der Wortklassen

Die besten Ergebnisse liefert die zufällige Kombination eigenständischer Laute (entspricht Klasse I). Diese Methode führt bei beiden Fragebögen zu einem signifikanten Ergebnis: Im ersten Fragebogen $\alpha < 0,001$, im zweiten $\alpha < 0,5$.

Die Kombination von Anfangs- und Endbigrammen (entspricht Klasse 3) führt zu brauchbaren, aber zwiespältigen Ergebnissen. Beim ersten Fragebogen liegt der Wert bei $\alpha < 0,001$ beim zweiten ist er nicht signifikant.,

Die Ergebnisse der Neukombination von Bigrammen liefern zufällige Ergebnisse. Weder beim ersten noch zweiten Fragebogen wurden signifikante Werte erreicht.

Bei allen Generierungsmethoden gab es Ausreißer und Falscheinordnungen. Besonders Anfällig für die falsche Zuordnungen waren Wörter der Klasse 3. Die Probanden waren sich über die Ausrichtung des Wortes (positiv/negativ) meist einig, nur entsprach diese Zuordnung nicht dem erwarteten Werten sondern dem umgekehrten erwarteten Wert. Die geringste Streuung in der Zuordnung hatten die Wörter der Klasse 1.

4.2 Klasse I / Monogramme

Die Zuordnung der Probanden entsprach den Erwartungen. Im beiden Fragebögen wurden sieben von zehn Wörtern richtig zugeordnet.

Die negativ assoziierten Wörter wurden stets besser erkannt: vier von fünf im ersten Bogen, und sogar fünf von fünf im zweiten wurden Eindeutig zugeordnet. Bei den positive Assoziation war die Trefferquote geringer: drei von fünf im ersten und zwei von fünf im zweiten Bogen.

Hervorzuheben sind zwei Ausreißer: Dem Wort „iltit“ wurde trotz der Erwartung eines positiven Werts meist negative Eigenschaften zugeordnet (eventuell spiel die Ähnlichkeit zum Wort 'Itis' eine Rolle). Eine völlig einheitliche, richtige Zuordnung gab es beim Wort „hoohoi“ dem alle befragten Probanden den erwarteten positiven Wert zuordneten.

Für genauere Information siehe Anhang D Zahlen und Grafiken.

4.3 Klasse 2/ Bigramme

Von zehn Wörtern wurden pro Fragebogen jeweils vier richtig zugeordnet. Die Antworten scheinen eher dem Zufall als einem inhärenten Prinzip zu entspringen.

Interessant im ersten Fragebogen sind die Wörter „Edheich“, das als einziges eine hohe, der Erwartung entsprechende Zuordnung erfuhr und „Eralfe“ bei dem sich die Testpersonen in der Falschzuordnung extrem einig waren (eventuell positive Zuordnung aufgrund der Ähnlichkeit zum Wort "Elfe").

4.4 Klasse 3/ Anfangs- und Endbigramme

Das Ergebnis ist uneindeutig. Im ersten Fragebogen wurden sechs von acht Wörtern richtig zugeordnet, wobei Antworten eine starke Übereinstimmung aufwiesen (Eine Übereinstimmung zwischen 70 und 80 Prozent). Beim zweiten Fragebogen hält die starke Übereinstimmung der Antworten an. Richtig erkannt wurden aber nur fünf von zehn Wörtern.

Beachtenswert ist das Wort „anbing“ im ersten Fragebogen, das fast alle Probanden trotz erwarteten Negativwert mit positiven Assoziationen verbanden (hier könnte eine kinästhetische Beeinflussung der Grund sein). Im zweiten Bogen sind die Wörter „Geedlill“ und „Fredtz“ interessant denen trotz anderer Erwartung negative Eigenschaften und „Febier“ und „Anweer“ denen positive Eigenschaften zugesprochen wurden.

4.5 Fazit

Die hohe Übereinstimmung bei vielen Worten in der Testmenge die bei den Probanden herrschen deuten auf lautsymbolische Mechanismen in der Sprache hin. Ausreißer und Falschzuordnungen lassen sich eventuell auf darauf zurückführen, dass unsere Orthographie zwar Laute abbildet, aber nicht alle Feinheiten (v.a. Vokallängen) erfasst. Ein zweites Problem folgt auch aus orthographischen Konventionen und ließe sich durch Optimierung der Generierungsfunktionen minimieren: Stoßen zwei Vokale, oder bestimmte Konsonanten aufeinander (z.B. s und p , s und t oder p und p) erkennt der Leser nicht mehr zwei einzelne Symbole die für zwei Laute stehen. Sondern zwei Symbole die für einen Laut sehen (st,ei, scharf gesprochenes p usw.).

Auch im Bereich der Wortauswahl für die Korparaerstellung fehlt es an soliden theoretischen Grundlagen.

Liesen sich oben genannte Probleme lösen erscheint mir ein Kombination der Generierungsmethode 1 und Generierungsmethode 3 unter der Verwendung von feiner abgegrenzten Korpora als sehr vielversprechend Methode um lautsymbolische Qualitäten in beliebigen Wortfeldern zu ermitteln.

Anhang A - Die Korpora



Korpora

I. Korpus A - Wörter mit positiver Assoziation

amor	ergötz	gewiß
begeist	erheiter	gewähr
beglück	eros	glanz
behag	errung	glorie
behüt	fortuna	glätte
beifall	fortfritt	glück
belustig	forj	gnade
benedikt	frei	gunst
bergen	freud	gut
besel	freund	heil
bürg	fried	heim
courage	froh	heiter
dasein	fromm	held
deck	frucht	helf
durç	gabe	hell
dusel	gaudium	heroe
eben	geborg	herr
edel	gedeih	herz
ehr	geling	hilfe
eifrig	genehm	himmel
eigen	genuß	hingabe
entusias	gesichert	hoX
entzück	gesund	hymne
erfolg	gewinn	innig

ja	prunk	vergnügen
jauXzen	prägnanz	voll
jubel	respekt	waX
klar	rett	wahr
kraft	riçtig	warm
kuß	ruhig	wirk
kühn	ruhm	wohl
lass	segem	wonne
leb	sel	zart
leiçt	selbst	zufrieden
liçt	selig	zärt
lieb	sieg	über
liçt	sinn	üppig
lust	siçer	farm
lösen	sonne	fatz
maXt	spaß	firm
mann	stolz	futz
mumm	tapfer	
muse	tat	
mut	treu	
nett	triumph	
pläsier	trumpf	

2. Korpus B - Wörter mit negativer Assoziation

aber	bitter	feind
abfeu	böse	fessel
angreif	deprim	fluX
angst	droh	furçt
anti	druck	gefährd
apath	druß	gering
argwohn	dunkel	gewalt
arm	dünn	gram
ausgelaug	dürftig	grauen
aversion	düster	haß
bang	eineng	hemm
befangen	ekel	horror
beklemm	elend	hölle
bestraf	eng	jammer
bestürzt	entbehr	kalt
befränk	enttäuft	karg
befwer	feig	kette
bind	feige	knapp
		krank

kreuz
krise
kummer
kümmer
labil
last
leid
lump
läster
mangel
marter
martyr
matt
melanXol
misere
miß
muß
müde
mühsal
neid
nein
nervös
niçt
not
panik
pein

phobie
plage
pression
qual
raX
reißen
seuçe
sklav
sorg
strafe
strapaze
strick
suXt
sühne
tod
trauer
trübsal
un
ungeheuer
verdaXt
verdamm
verderb
verderben
verfem
vergelt
verhäng

weh
wenig
wider
wirr
wut
zag
zaudern
zwang
zweifel
zöger
ärger
über
faunder
fauer
feu
feusal
fleçt
fmerz
fmäh
furke
fwaX
fwer
fäbig
fäm

Anhang B - Die Programme



Auswertungsprogramme

I. Das Programm „auswertung.rb“

Dieses objektorientierte Programm benutzt eine Differenzlistentechnik um distinktive Merkmale aus zwei Listen zu extrahieren.

Das Programm gibt neben den automatisch erzeugten Ergebnissen noch eine Splitscreenliste aus mit der die Qualität/ Plausibilität der automatischen Ergebnisse überprüft werden kann.

I.1 Bedienung

Aufruf auf Konsolen/Terminalebene:

```
ruby auswertung.rb Liste1 Liste2 n-Grammlänge
```

Gibt die distinktiven Merkmale der Länge n-Grammlänge der Liste1 im Vergleich zu Liste2 aus.

```
ruby auswertung.rb Liste1 Liste2 n-Grammlänge Marker
```

Gibt die distinktiven Merkmale der Länge n-Grammlänge der Liste1 im Vergleich zu Liste2 aus. Markiert zusätzlich den Wortanfang mit ^ und das Wortende mit \$.

I.2 Quellcode

```
#!/usr/bin/ruby -w

#==Author: Andreas Neumann
#==Synopsis: Vergleicht zwei Listen auf Eigenschaften

$KCODE="utf8";

# Erzeugt sowohl Differenzlisten als auch duale Ausgaben von Listen um diese
# auf eigenschaf-
# ten zu vergleichen
class PosNegC

  def initialize(pos,neg,n,marker)
    @pos=pos
    @neg=neg
    @n=n
    @marker=marker
    @positiv=make_frq(@pos,@n,@marker)
    @negativ=make_frq(@neg,@n,@marker)
  end

  #Erstellt eine Frequenzliste aus <datei> mit Einträgen der Länge <n> die
  #aus den Wörtern des
  #Textes geschnitten werden
  #Gibt ein Hash <h> zurück
  def make_frq(datei,n,s)
    h=Hash.new(0)
    File.open(datei).each do |wort|
      wort='^'+wort+'$' if s
      a=wort.scan(/(?:\w|[$^])/u)
      i=0
      while i+n < a.length():
        h[a[i..i+n].to_s]+=1
        i+=1
      end
    end
    h
  end

  #Gibt ein Hash <h> nach Werten sortiert auf dem Terminal aus
  def sortierte_ausgabe(h)
    h.sort {|a,b| b[1] <=> a[1]}.each do |w|
      puts w[1].to_s+"\t"+w[0]
    end
  end

  #Gibt nach Buchstaben sortiert , nebeneinander aus
  def duale_ausgabe_char()
    uni=Hash.new()

    @positiv.each_key do |k|
      if uni[k]:
        uni[k] << @positiv[k]
      else
        uni[k]=[]
        uni[k] << @positiv[k]
      end
    end
    @negativ.each_key do |k|
      if uni[k]:
        uni[k] << @negativ[k]
      end
    end
  end
end
```

```

        else
            uni[k] = [" "] # Damit sortierung erhalten bleibt falls key
nicht vorhanden
            uni[k] << @negativ[k]
        end
    end

    uni.sort {|a,b| a[0] <=> b[0]}.each do |w|
        puts "#{w[0]}: #{w[1][0]} | #{w[0]}: #{w[1][1]}"
    end

end

#Gibt zwei Hashes <h1> <h2> nebeneinander auf dem Terminal aus,
#sortiert nach Anzahl des ersten Hashes <h1>
def duale_ausgabe_count()
    @positiv.sort {|a,b| b[1] <=> a[1]}.each do |w|
        puts "#{w[0]}: #{w[1]} | #{@negativ[w[0]] if @negativ[w[0]]}"
    end
end

def find_differences(h1,h2)
    d=Hash.new()
    h1_anz=0.0
    h2_anz=0.0

    h1.each_value {|v| h1_anz+=v}
    h2.each_value {|v| h2_anz+=v}

    h1.each_key do |k|
        d[k]=(h1[k] - (h2[k]/h2_anz*h1_anz).round())
    end
    d
end

def differenz_ausgabe()
    sortierte_ausgabe(find_differences(@positiv,@negativ))
end

# MAIN
#####

if ARGV.length() < 3 :
    puts "Zu wenig Argumente bei Aufruf:","auswertung.rb <positive_txt> <ne-
gativ_txt> <Länge der n-Gramme> optional <Wortanfänge und -enden markie-
ren>"
    exit
end

vergleicher=PosNegC.new(ARGV[0],ARGV[1],ARGV[2].to_i() -1,ARGV[3])

vergleicher.duale_ausgabe_count()
vergleicher.differenz_ausgabe()

```




Generierungsprogramme

I. Das Programm „mono.pro“

Diese Programm erzeugt mit Hilfe von Backtracking alle möglichen Permutationen einer Liste von Lauten. Nicht aussprechbare Lautkombinationen werden aussortiert.

I.1 Bedienung

Das Programm in einem Prolog-Interpreter laden. Das Prädikat `write_words/3` aufrufen.

```
write_words([Liste mit Lauten],Wortlänge,Zieldatei).
```

I.2 Quellcode

```
/*
Autor: Andreas Neumann
Synopsis:Permutiert einzelne Buchstaben, erstellt daraus Wörter und
schreibt sie in
eine Datei.
Erzeugt noch viele Dubletten, mit "sort | uniq" nachbearbeiten.
*/

/* Ermittelte Buchstaben */
positiv([i,g,h,l,o,t]).
negativ([a,r,sch,m,v,d,k]).

/*
Schreibt aus BUCHSTABEN=[Liste von Buchstaben], Wörter der Länge WORTLAENGE
und schreibt
dies in Datei IN_DATEI
*/
write_words(BUCHSTABEN,WORTLAENGE,IN_DATEI):-
```

```

current_output(Stream),
open(IN_DATEI,write,OUT),
set_output(OUT),
(
    (
        g_word(BUCHSTABEN,WORTLAENGE),
        !
    ) % hier die Wortlaenge bestimmen
    ;
    (
        set_output(Stream),
        close(OUT),
        true
    )
).

/* Ausgabe als Wort nicht als Liste */
pp([]):-nl.
pp([A|R]):-
    write(A),
    pp(R).

/* L gibt Wortlänge an */
g_word(BUCHSTABEN,Len):-
    generate_word(X,Len,BUCHSTABEN),
    sprechbar(X),
    pp(X),
    fail.

/* Generatorfunktion */
generate_word([],0,_).
generate_word([A|Rest],L,BUCHSTABEN):-
    L >= 0,
    L1 is L -1,
    member(A,BUCHSTABEN),
    generate_word(Rest,L1,BUCHSTABEN).

/* Testet ob Vokal */
is_vokal(A):-
    member(A,[a,e,i,o,u]).

/*Wortteil ist sprechbar wenn es entweder zwei Vokale und einen Konsonanten
enthält,
oder zwei Konsonanten und einen Vokal*/
sprechbar([]).
sprechbar([_]).
sprechbar([_,_]).

sprechbar([A,B,C|R]):-
    member(X,[A,B,C]),
    member(Y,[A,B,C]),
    is_vokal(X),
    not(is_vokal(Y)),
    append([B,C],R,Rest),
    sprechbar(Rest).

```

2. Das Programm „bi.pro“

Diese Programm erzeugt mit Hilfe von Backtracking alle möglichen Permutationen einer Liste von Bigrammen.

2.1 Bedienung

Das Programm in einem Prolog-Interpreter laden. Das Prädikat `write_words/3` aufrufen.

```
write_words([Liste mit Bigrammen],Wortlänge,Zieldatei).
```

Die Wortlänge muss in diesem Fall ein vielfaches von Zwei sein.

2.2 Quellcode

```
/*
Autor: Andreas Neumann
Synopsis:Permutiert Bigramme, erstellt daraus Wörter und schreibt sie in
eine Datei.
*/

/* Bigramme */
positiv([fr,he,ed,nn,gl,ich,li,ie]).
negativ([er,ng,ve,an,ra,bi,we,rd,al,fe,au]).

/*
Erzeugt aus BIGR=[Liste von Bigrammen] Wörter der Länge WORTLAENGE und
schreibt
diese in Datei IN_DATEI
*/
write_words(BIGR,WORTLAENGE,IN_DATEI):-
    current_output(Stream),
    open(IN_DATEI,write,OUT),
    set_output(OUT),
    (
        (
            WL is WORTLAENGE /2, % Da ja Bigramme
            g_word(BIGR,WL),
            !
        )
        ;
        (
            set_output(Stream),
            close(OUT),
            true
        )
    ).

/* Ausgabe als Wort nicht als Liste */
pp([]):-nl.
pp([A|R]):-
```

```
    write(A),  
    pp(R).  
  
g_word(BUCHSTABEN, Len):-  
    generate_word(X, Len, BUCHSTABEN),  
    pp(X),  
    fail.  
  
/* Generatorfunktion */  
generate_word([], 0, _).  
generate_word([A|Rest], L, BUCHSTABEN):-  
    L >= 0,  
    L1 is L - 1,  
    member(A, BUCHSTABEN),  
    generate_word(Rest, L1, BUCHSTABEN).
```

2. Das Programm „ame.pro“

Dieses Programm erzeugt mit Hilfe von Backtracking alle möglichen Permutationen einer Liste von Bigrammen.

2.1 Bedienung

Das Programm in einem Prolog-Interpreter laden. Das Prädikat `write_words/3` aufrufen.

```
write_words([ [List mit Bigrammanfängen], [Liste mit Bi-
grammmitten], [Liste mit Bigrammenden]],Anzahl der Mittelbi-
gramme,Zieldatei).
```

2.2 Quellcode

```
/*
Autor: Andreas Neumann
Synopsis:Permutiert Bigramme, erstellt daraus Wörter und schreibt sie in
eine Datei.
Anfang, mitte und Ende werden unterschieden
*/

/* Bigramme
ART([ANFANG],[MITTE],[ENDE])
*/

positiv([[he,ge,fr,er,gl],[ed,nn,ich,li,ie],[en,ig,tz,rt,ll]]).
negativ([[ve,fe,due,dr,an],[ra,bi,we,rd,au],[er,al,ng]]).

/*
Erzeugt aus BIGN Wörter mit L Mittelbigrammen und schreibt sie in Datei
IN_DATEI
*/
write_words(BIGN,L,IN_DATEI):-
    current_output(Stream),
    open(IN_DATEI,write,OUT),
    set_output(OUT),
    (
        (
            g_word(BIGN,L),
            !
        )
        ;
        (
            set_output(Stream),
            close(OUT),
            true
        )
    ).

/* Ausgabe als Wort nicht als Liste */
pp([]).
```

```

pp([A|R]):-
    write(A),
    pp(R).

/* Erzeugt Wörter der Form <B_Anfang><B_Mitte>*LEN<B_Ende> */
g_word([B_Anfang,B_Mitte,B_Ende],Len):-
    member(A,B_Anfang),
    member(C,B_Ende),
    generate_word(X,Len,B_Mitte),
    write(A),
    pp(X),
    write(C),
    nl,
    fail.

/* Generatorfunktion */
generate_word([],0,_).
generate_word([A|Rest],L,BUCHSTABEN):-
    L >= 0,
    L1 is L - 1,
    member(A,BUCHSTABEN),
    generate_word(Rest,L1,BUCHSTABEN).

```

Anhang C - Fragebögen



Der Fragebogen

Hier finden sich die beiden Fragebögen die für die Onlinebefragung verwendet wurden. Die Position der Items innerhalb einer Frage wurde jedes mal zufällig neu verteilt.

Version I

Was bezeichnet diese Wort?[Exit this survey >>](#)**1. Umfrage****1. Allgemeines**

Geschlecht:

Alter:

Muttersprache:

2. Bezeichnet das folgende Wort eine positive Eigenschaft oder handelt es sich bei dem Wort um etwas verwerfliches ?

	gute Eigenschaft	schlechte Eigenschaft
toigo	<input type="radio"/>	<input type="radio"/>
vadmar	<input type="radio"/>	<input type="radio"/>
adkavsch	<input type="radio"/>	<input type="radio"/>
ramvark	<input type="radio"/>	<input type="radio"/>
iltit	<input type="radio"/>	<input type="radio"/>
ohlog	<input type="radio"/>	<input type="radio"/>
lioghi	<input type="radio"/>	<input type="radio"/>
schkad	<input type="radio"/>	<input type="radio"/>
mamdak	<input type="radio"/>	<input type="radio"/>
gotoh	<input type="radio"/>	<input type="radio"/>

3. Diese Liste führt verschiedene Tiernamen auf.**Welches dieser Tiere ist edel, hübsch und zutraulich?****Welches hässlich, aggressiv und gefährlich?**

	edel, hübsch	hässlich, rüdig
Edheich	<input type="radio"/>	<input type="radio"/>
Fealbing	<input type="radio"/>	<input type="radio"/>
Frichich	<input type="radio"/>	<input type="radio"/>
Bifera	<input type="radio"/>	<input type="radio"/>
Alaung	<input type="radio"/>	<input type="radio"/>
Linnhegl	<input type="radio"/>	<input type="radio"/>
Eralfe	<input type="radio"/>	<input type="radio"/>
Glichhe	<input type="radio"/>	<input type="radio"/>
Ichglich	<input type="radio"/>	<input type="radio"/>

Weveraal

4. Welches dieser Adjektive eignet sich einen einsamen, düsteren, kalten Arbeitstag zu beschreiben an dem alles schief geht?

Welches eignet sich besser für einen sonnigen Urlaubstag am Meer?

	schöner Urlaub	unangenehmer Arbeitstag
düweral	<input type="radio"/>	<input type="radio"/>
heliedig	<input type="radio"/>	<input type="radio"/>
frichededll	<input type="radio"/>	<input type="radio"/>
veraaauer	<input type="radio"/>	<input type="radio"/>
heglig	<input type="radio"/>	<input type="radio"/>
anbing	<input type="radio"/>	<input type="radio"/>
gennig	<input type="radio"/>	<input type="radio"/>
draufal	<input type="radio"/>	<input type="radio"/>

5. Bei welcher Frage war die Zuordnung am einfachsten?

- ☐ Frage 2
- ☐ Frage 3
- ☐ Frage 4

6. Anmerkungen

Done >>

Version II

Was bezeichnet diese Wort? v2

[Exit this survey >>](#)

1. Umfrage

1. Allgemeines

Geschlecht: Alter: Muttersprache:

2. Bezeichnet das folgende Wort eine positive Eigenschaft oder handelt es sich bei dem Wort um etwas verwerfliches ?

	gute Eigenschaft	schlechte Eigenschaft
verdal	<input type="radio"/>	<input type="radio"/>
linnich	<input type="radio"/>	<input type="radio"/>
ligl	<input type="radio"/>	<input type="radio"/>
hennheli	<input type="radio"/>	<input type="radio"/>
edhe	<input type="radio"/>	<input type="radio"/>
heichli	<input type="radio"/>	<input type="radio"/>
webirard	<input type="radio"/>	<input type="radio"/>
feanal	<input type="radio"/>	<input type="radio"/>
alvefe	<input type="radio"/>	<input type="radio"/>
eraungal	<input type="radio"/>	<input type="radio"/>

3. Diese Liste führt verschiedene Tiernamen auf.

Welches dieser Tiere ist edel, hübsch und und zahm?

Welches hässlich, agressiv und gefährlich?

	edel, hübsh	hässlich, rädig
Fredtz	<input type="radio"/>	<input type="radio"/>
Drauraer	<input type="radio"/>	<input type="radio"/>
Febier	<input type="radio"/>	<input type="radio"/>
Vebial	<input type="radio"/>	<input type="radio"/>
Dueauer	<input type="radio"/>	<input type="radio"/>
Geedlill	<input type="radio"/>	<input type="radio"/>
Heliedlitz	<input type="radio"/>	<input type="radio"/>
Anweer	<input type="radio"/>	<input type="radio"/>
Eriell	<input type="radio"/>	<input type="radio"/>

Frieen

4. Welches dieser Adjektive eignet sich einen einsamen, düsteren, kalten Arbeitstag zu beschreiben an dem alles schief geht?

Welches eignet sich besser für einen sonnigen Urlaubstag am Meer?

	schöner Urlaub	unangenehmer Arbeitstag
ramamasch	<input type="radio"/>	<input type="radio"/>
ogihog	<input type="radio"/>	<input type="radio"/>
makdakk	<input type="radio"/>	<input type="radio"/>
aadkarr	<input type="radio"/>	<input type="radio"/>
giogigl	<input type="radio"/>	<input type="radio"/>
schmarka	<input type="radio"/>	<input type="radio"/>
hoohoi	<input type="radio"/>	<input type="radio"/>
gtio	<input type="radio"/>	<input type="radio"/>
tloglo	<input type="radio"/>	<input type="radio"/>
vakma	<input type="radio"/>	<input type="radio"/>

5. Bei welcher Frage war die Zuordnung am einfachsten?

☐ Frage 2

☐ Frage 3

☐ Frage 4

6. Anmerkungen

Done >>

Anhang D - Daten



Empirische Daten

Hier finden sich die in der Umfrage gewonnenen Ergebnisse:

Unter "Art" findet sich die Generierungsmethode des Worts, unter "erwartet" das erwartete Ergebnis (g für positiv, n für negativ) und unter "als_pos"/"als_neg" die Anzahl der Probanden die sich für positive/negative Zuordnung entschieden haben.

Darauf folgt die Ausgabe der statistischen Auswertung, sortiert nach Erzeugungsmethode.

Fragebogen I

Überblick Fragebogen I

	Wort	als_pos	als_neg	erwartet	Art
1	ohlog	17	22	g	mono
2	gotoh	24	15	g	mono
3	lioghi	33	6	g	mono
4	iltit	14	25	g	mono
5	toigo	29	10	g	mono
6	ramvark	8	31	b	mono
7	adkavsch	6	33	b	mono
8	mamdak	21	18	b	mono
9	schkad	11	28	b	mono
10	vadmar	16	23	b	mono
11	Edheich	31	8	g	bigram
12	Ichglich	13	25	g	bigram
13	Frichich	18	20	g	bigram
14	Glichhe	16	23	g	bigram
15	Linnhegl	24	15	g	bigram
16	Bifera	24	15	b	bigram
17	Eralfe	32	7	b	bigram
18	Fealbing	25	13	b	bigram
19	Weveraal	15	24	b	bigram
20	Alaung	19	20	b	bigram
21	heliedig	30	8	g	anf_u_ende
22	frichededll	29	10	g	anf_u_ende
23	heglig	18	21	g	anf_u_ende
24	gennig	30	8	g	anf_u_ende
25	d\xfcfcweral	5	34	b	anf_u_ende
26	veraauer	6	33	b	anf_u_ende
27	anbing	34	4	b	anf_u_ende
28	draufal	13	26	b	anf_u_ende

statistische Auswertung Fragebogen I

#-----

Monogramme:

positive Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.0	17.0	24.0	23.4	29.0	33.0

negative Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
18.0	23.0	28.0	26.6	31.0	33.0

gesamt:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.00	19.25	26.00	25.00	30.50	33.00

χ^2 -Test(Pearson):

X-squared = 15.2573, df = 1, p-value = 9.381e-05

#-----

Bigramme:

positive Assoziation:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	13.00	16.00	17.00	19.67	22.50	31.00
negative Assoziation:						
	7.0	13.0	15.0	15.8	20.0	24.0
gesamt:						
	7.00	14.00	16.00	17.91	22.00	31.00

χ^2 -Test(Pearson):

X-squared = 0.6873, df = 1, p-value = 0.4071

#-----
Anfang und Ende:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
positive Assoziation:						
	18.0	19.0	29.0	25.2	30.0	30.0
negative Assoziation:						
	4.00	20.50	29.50	24.25	33.25	34.00
gesamt:						
	4.00	19.00	29.00	24.78	30.00	34.00

χ^2 -Test(Pearson):

X-squared = 15.7005, df = 1, p-value = 7.42e-05

#-----

Fragebogen II

Überblick Fragebogen II

	Wort	als_pos	als_neg	erwartet	art
1	linnich	7	7	g	bigram
2	ligl	5	9	g	bigram
3	hennheli	9	5	g	bigram
4	heichli	7	7	g	bigram
5	edhe	9	5	g	bigram
6	eraungal	4	10	b	bigram
7	webirard	8	6	b	bigram
8	alvefe	9	5	b	bigram
9	feanal	6	8	b	bigram
10	verdäl	7	7	b	bigram
11	Heliedlitz	11	3	g	anf_u_ende
12	Geedlill	4	10	g	anf_u_ende
13	Frieen	9	5	g	anf_u_ende
14	Fredtz	2	12	g	anf_u_ende
15	Eriell	13	1	g	anf_u_ende
16	Dueauer	4	10	b	anf_u_ende
17	Febier	9	5	b	anf_u_ende
18	Vebial	6	7	b	anf_u_ende
19	Drauraer	6	8	b	anf_u_ende
20	Anweer	11	3	b	anf_u_ende
21	ogihog	8	6	g	mono
22	tloglo	5	9	g	mono
23	hoohei	14	0	g	mono
24	gtio	6	8	g	mono
25	giogigl	7	7	g	mono
26	vakma	5	9	b	mono
27	aadkarr	2	12	b	mono
28	makdakk	4	10	b	mono
29	ramamasch	6	8	b	mono
30	schmarka	5	8	b	mono

statistische Auswertung Fragebogen II

#-----
Monogramme

positive Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5	6	7	8	8	14

negative Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.0	8.0	9.0	9.4	10.0	12.0

gesamt:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.00	7.25	8.00	8.70	9.75	14.00

χ^2 -Test(Pearson):

X-squared = 3.9959, df = 1, p-value = 0.04561

#-----
 Bigramme

positive Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.0	7.0	7.0	7.4	9.0	9.0

negative Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.0	6.0	7.0	7.2	8.0	10.0

gesamt:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.00	6.25	7.00	7.30	8.75	10.00

χ^2 -Test(Pearson):

X-squared = 0.0572, df = 1, p-value = 0.811

#-----
 Anfang und Ende

positive Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.0	4.0	9.0	7.8	11.0	13.0

negative Assoziation:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.0	5.0	7.0	6.6	8.0	10.0

gesamt:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.00	4.25	7.50	7.20	9.75	13.00

χ^2 -Test(Pearson):

X-squared = 0.0327, df = 1, p-value = 0.8566

#-----



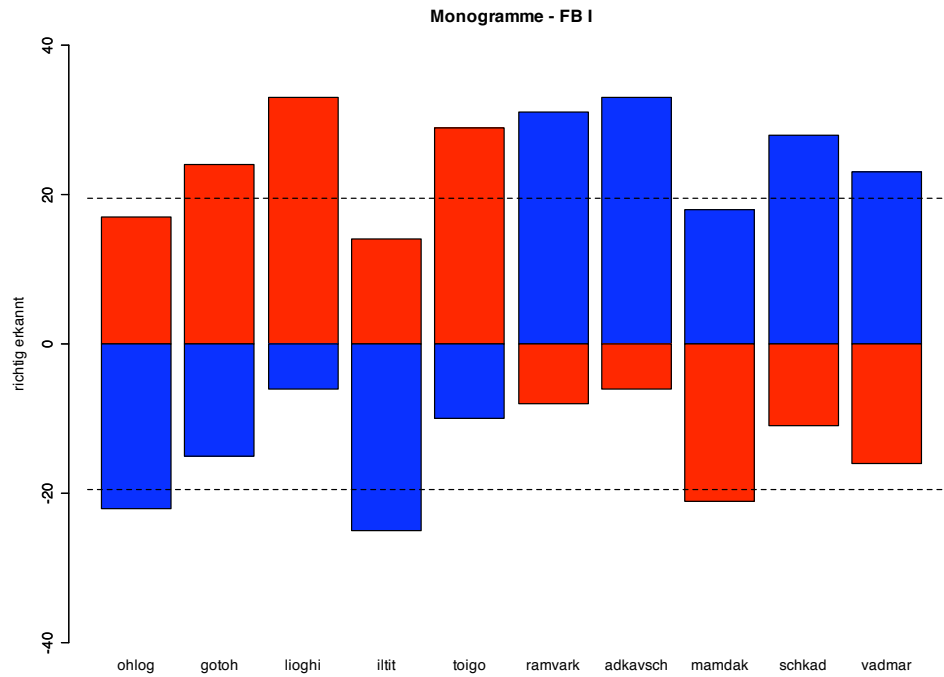
Grafiken

Die Grafiken sind folgendermaßen zu interpretieren:

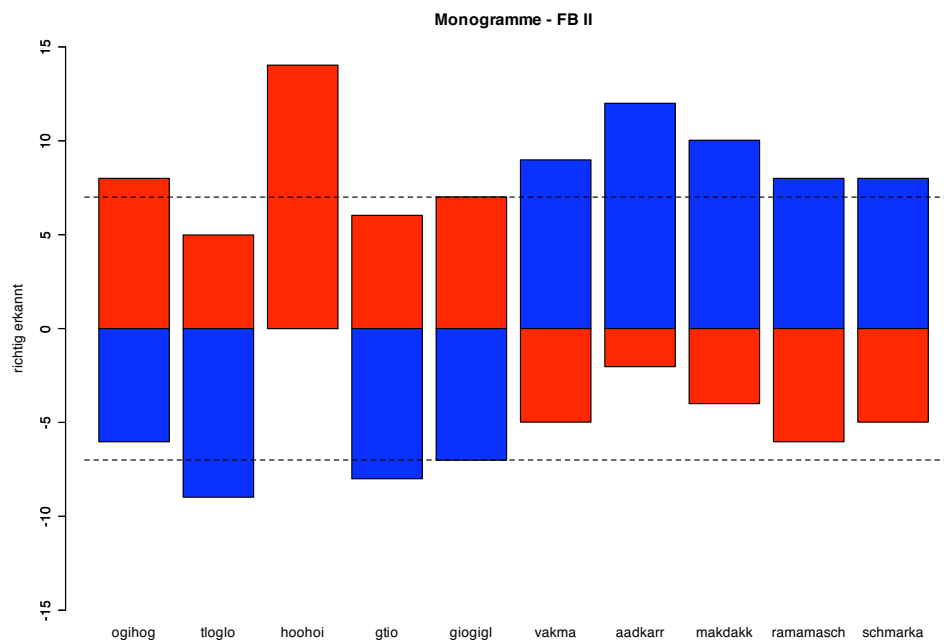
- Rot steht für Antworten mit positiver Assoziation, Blau für Antworten mit negativer Assoziation.
- Die Antworten der Probanden die der Erwartung entsprechen werden nach oben angetragen, die "falschen" Antworten nach unten.
- Die gestrichelten Linien begrenzen das Feld in dem sich der Balken bei einer zufälligen Verteilung bewegen würde.

Monogramme

Monogramme Fragebogen I

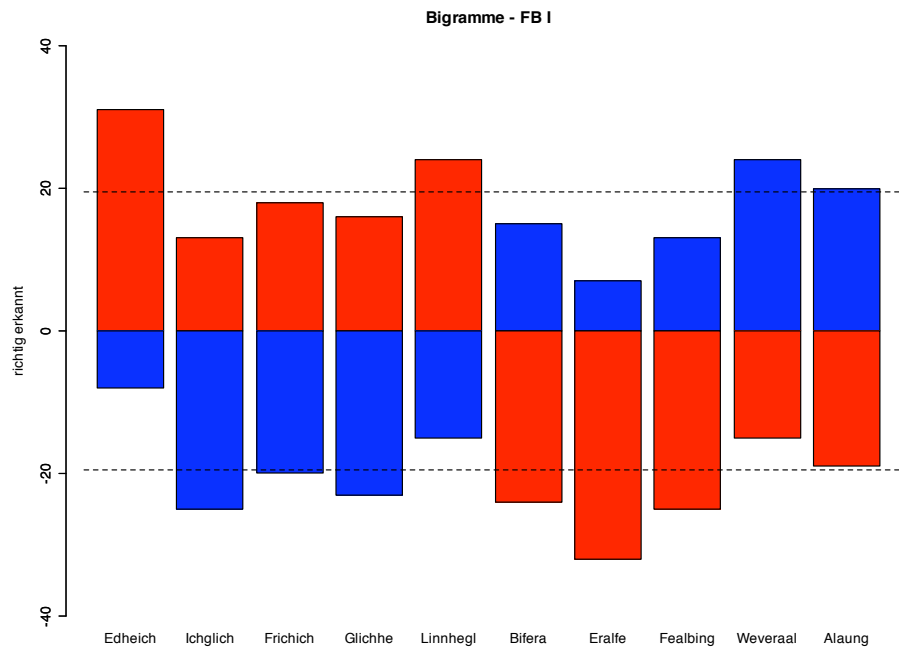


Monogramme Fragebogen II

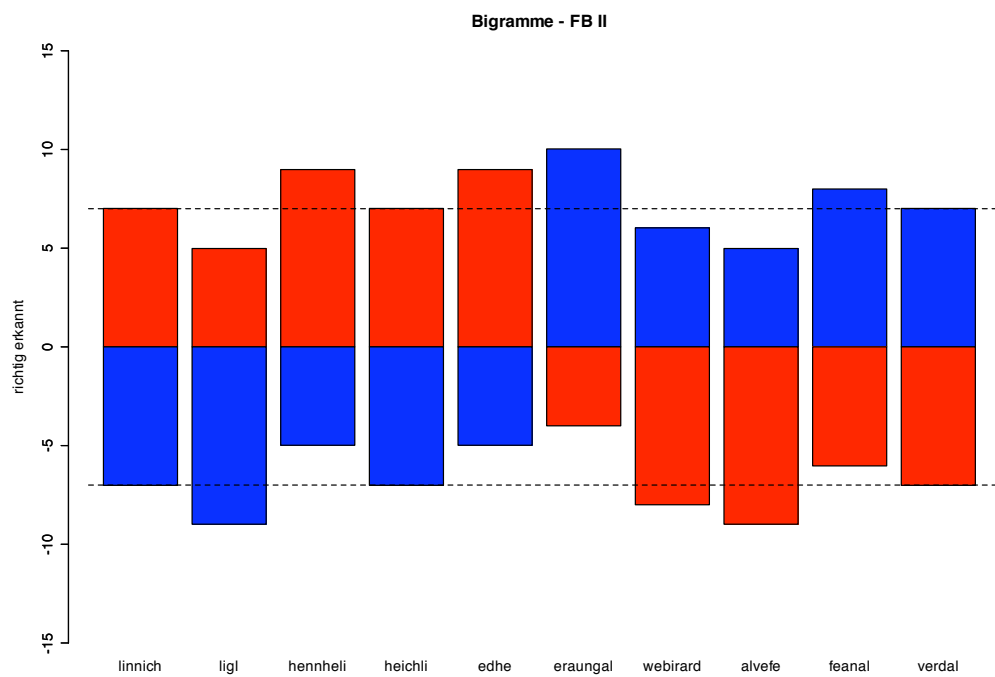


Bigramme

Bigramme Fragebogen I

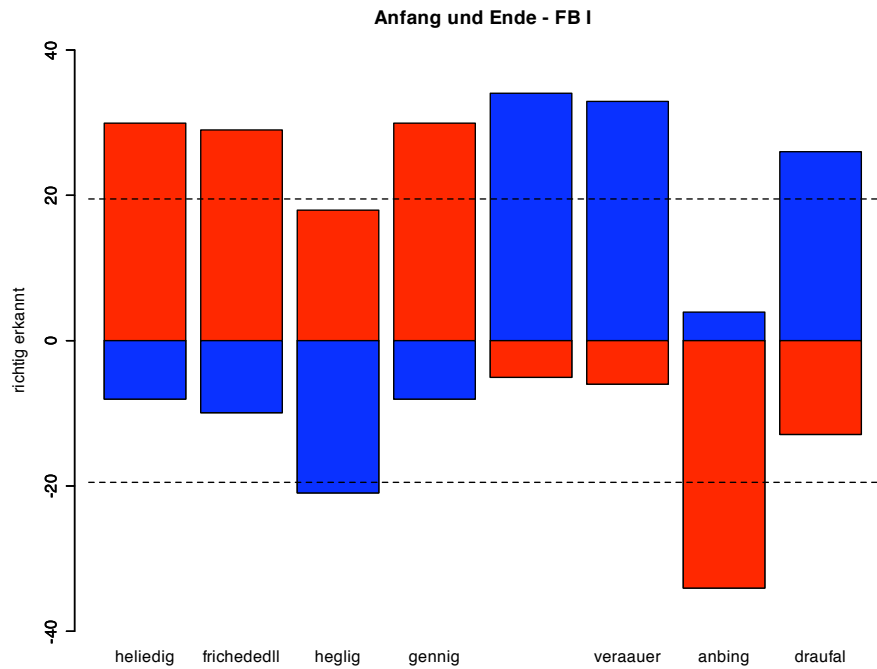


Bigramme Fragebogen II

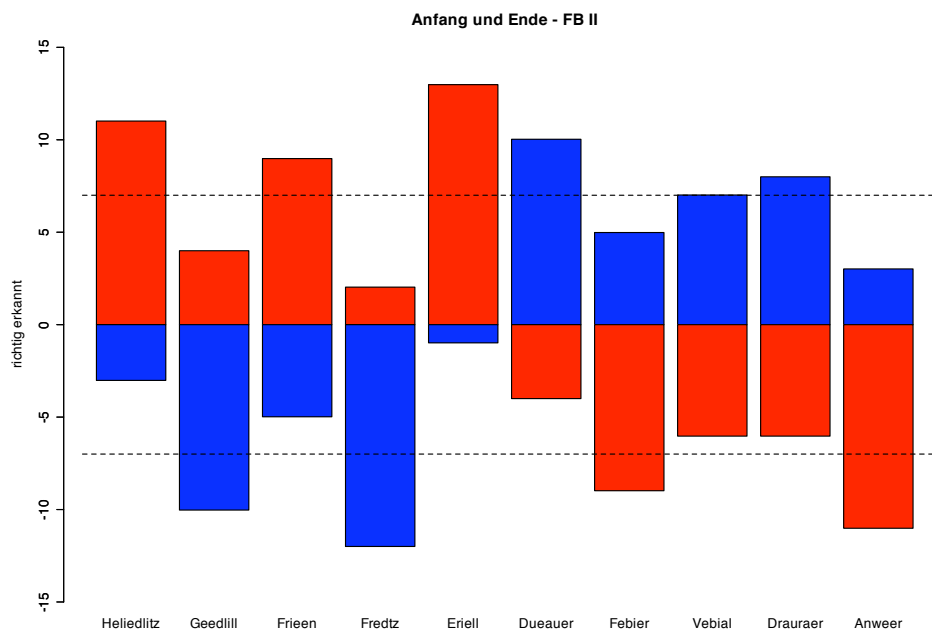


Anfangs- und Endbigramme

Anfangs- und Endbigramme Fragebogen I

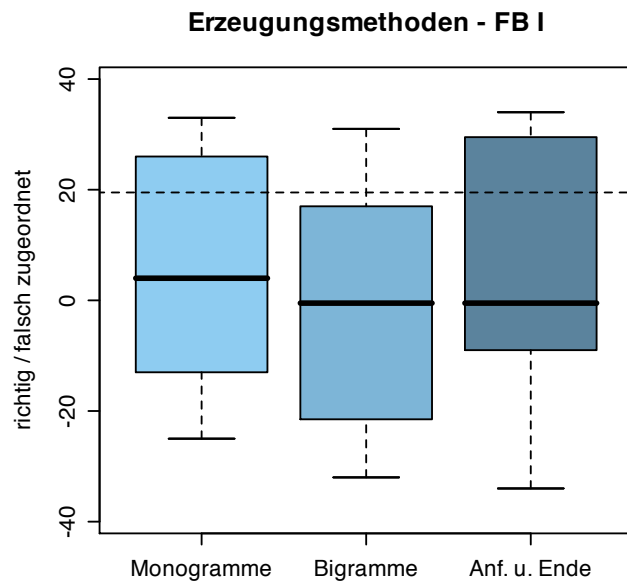


Anfangs- und Endbigramme Fragebogen II



Vergleich der Generierungsmethoden

Generierungsmethoden Fragebogen I



Generierungsmethoden Fragebogen II

