Die Erhebung perzeptueller Prominenz auf Silben- und Wortebene: Der Einfluss von Bewertungsskalen, Bewertungsebenen und Normalisierung

Dissertation
zur Erlangung des Grades eines
Doktors der Philosophie
der Philosophischen Fakultäten
der Universität des Saarlandes

vorgelegt von

Denis Arnold

aus Bonn

Saarbrücken, 2013

Der Dekan: Univ.-Prof. Dr. R. Marti
Berichterstatter/innen: Univ.-Prof. Dr. B. Möbius und Univ.-Prof. Dr. P. Wagner
Tag der letzten Prüfungsleistung: 04.02.2013

Danksagung

Mein Dank gilt zunächst meinen Betreuern. Bernd Möbius hat mir die Chance gegeben, als wissenschaftlicher Mitarbeiter in Bonn zu arbeiten, und mein Dissertationsprojekt jederzeit unterstützt. Ich habe in der Zeit in Bonn viel über den Wissenschaftsbetrieb an sich lernen können. Petra Wagner hat schon meine Magisterarbeit betreut und war neben Bernd Möbius meine wichtigste Ansprechpartnerin. Sie hat mich schon während des Studiums wissenschaftlich gefördert und ohne sie hätte sich meine wissenschaftliche Ausrichtung sicherlich anders entwickelt. Stefan Breuer hat sich dafür eingesetzt, dass ich als studentische Hilfskraft in seinem Projekt arbeiten konnte. Bei der Arbeit im BOSS Projekt habe ich mir viele Fähigkeiten angeeignet, die ich bei der Umsetzung dieses Projektes gut gebrauchen konnte. Wolfgang Hess habe ich nicht nur eine solide phonetische Ausbildung zu verdanken, sondern auch eine wissenschaftlich und menschlich hervorragende Umgebung, in der ich mich entwickeln konnte. Meine (zeitweiligen) Kollegen in Bonn, Stefan Baumann, Christine de Bond, Donata Moers, Barbara Samlowski, Christopher Sappok und Charlotte Wollermann haben mir immer als gute Diskussionspartner gedient, wenn ich meine Gedanken ordnen musste und mich allesamt in meinem Vorhaben unterstützt. Berthold Crysmann, Jörg Mayer und Ulrich Schade haben mir immer wieder Denkanstöße gegeben. Mein Dank gebührt allen, die mich bei der Suche nach Probanden unterstütz haben. Hier gilt mein Dank vor allem Christopher Sappok, Lars Wallenborn, Nina Arnold, Timo Klein und Doris Mücke, die kräftig die Werbetrommeln gerührt haben und nicht nur Probanden, sondern auch Räume zu deren Erhebung organisiert haben. Hier sei auch meinen 216 Probanden des ersten Experiments und den 36 Probanden des zweiten Experiments gedankt. Ohne sie hätte ich keine Daten. Weiterhin möchte ich meiner Familie für die immer währende Unterstützung danken. Ganz besonderer Dank gilt meiner Frau Nina Arnold, die mich immer ermuntert hat, wenn ich mit meiner Arbeit unzufrieden war.

Inhaltsverzeichnis

1. Einleitung

Zur Wahrnehmung von Prominenz linguistischer Einheiten wird interdisziplinär geforscht. Seit über 30 Jahren schon wurde in den Bereichen der Phonetik, Psycholinguistik, Psychologie und Sprachtechnologie eine Vielzahl an Publikationen hervorgebracht, die verschiedene Aspekte von Prominenz betrachten. Im Theorieteil dieser Arbeit soll ein Überblick über die verschiedenen Arbeiten erstellt werden. Viele Studien versuchen dabei den Zusammenhang zwischen Akustik und der Perzeption von Prominenz aufzudecken. Einige Erkenntnisse hieraus werden in Kapitel 2.2.1 vorgestellt. In Kapitel 2.2.2 werden eine Reihe von Arbeiten vorgestellt, die den Anteil nicht akustischer Phänomene an der Wahrnehmung von perzeptueller Prominenz untersuchen. Hier finden vor allem Arbeiten Beachtung, die den Einfluss von linguistischem Wissen, wie beispielsweise Wortart, Erwartungen der Hörer[1] , aber auch multimodale Einflüsse, beispielsweise durch Gesten, untersuchen. Im Feld der Sprachtechnologie gibt es zum einen Ansätze, Prominenz automatisch zu erkennen und zum anderen Ansätze, bei denen versucht wird, das Konzept der Prominenz zur Verbesserung der Prosodie in der Sprachsynthese nutzbar zu machen. Kapitel 2.2.3 soll einen Überblick über diese Arbeiten geben.

Betrachtet man die Menge der Arbeiten, die sich bereits mit dem Gegenstand der Prominenz auseinandergesetzt hat, stellt sich unweigerlich die Frage, wofür es nun eine methodische Arbeit zur Erhebung von Prominenz braucht. Bei all den Studien, die im nächsten Kapitel beschrieben werden, wird sich einer Vielzahl verschiedener Skalen zur Erhebung von Prominenz von meist naiven Hörern bedient. In der Literatur finden sich hingegen nur wenige methodische Artikel, die sich mit der Erhebung mittels Skalen auseinandersetzen und die sich dazu auch noch in ihrer Kernaussage widersprechen. Diese Arbeit möchte also der Frage nachgehen, wie man idealerweise perzeptuelle Prominenz erhebt. Kapitel 2.3

[1] Die Arbeit benutzt zur besseren Lesbarkeit das generische Maskulinum. Selbstverständlich sind in den betreffenden Fällen jeweils beide Geschlechter gemeint.

soll ein Experiment motivieren, bei der verschieden Skalen hinsichtlich ihrer Eignung zu Erhebung von Prominenz untersucht werden. Die Durchführung und die Ergebnisse dieses Experiments werden in Kapitel 3 beschrieben.

Wenn man sich die Literatur zur Prominenz ansieht, fällt einem auf, dass sich die Arbeiten im Wesentlichen mit zwei verschiedene linguistische Einheiten beschäftigen. Der eine Teil der Artikel befasst sich mit Prominenz hauptsächlich auf Silbenebene, während sich der andere Teil mit der Prominenz auf Wortebene beschäftigt. Zu der Frage der Wahl der Einheit gibt es von jeder Seite gute Argumente, aber eine Studie, die sich mit den Unterschieden beschäftigt, die sich aus einer Erhebung von Wortprominenz versus der Erhebung von Silbenprominenz ergibt, fehlt weitgehend in der Literatur. In Kapitel 2.4 soll diese Darstellung etwas detaillierter ausgearbeitet werden und ein Experiment motivieren, welches in Kapitel 4 dargestellt wird.

Eine weitere Fragestellung, die im Diskurs bisher wenig Beachtung gefunden hat, ist die Frage nach der Normalisierung von Prominenzurteilen. Wie in Kapitel 2.5 dargestellt, berichten einige Autoren das Problem, dass ihre Probanden die Skalen nicht wie von den Versuchsleitern gewünscht ausnutzen. Lediglich wenige Studien bemühen Normalisierungsverfahren, um diese Effekte auszugleichen. Dies motiviert, die für die eigenen Experimente gesammelten Daten mittels zwei verschiedener Verfahren zu normalisieren und die Auswirkung von Normalisierung auf verschiedene Zusammenhänge zu untersuchen. Die Ergebnisse werden in Kapitel 5 dargestellt.

Neben dem vorrangigen Ziel der Arbeit, die Erhebung von Prominenz mittels naiven Hörern methodisch zu beleuchten, fand sich in den erhobenen Daten noch einige Evidenz dafür, dass der Kontext bei der Wahrnehmung von Prominenz eine erhebliche Rolle spielt. Es werden aber durch die Daten auch neue Fragen aufgeworfen. Eine kurze Zusammenfassung aller Ergebnisse und eine Aufstellung von zukünftigen Fragen finden sich im Kapitel 6, mit dem diese Arbeit abschließt.

2. Theorie

Dieses Kapitel soll einen Überblick über die Forschungslage im Bereich der perzeptuellen Prominenz geben. Zunächst soll definiert werden, was in dieser Arbeit mit dem Begriff Prominenz gemeint ist. Einer Zusammenfassung der Forschung über die Zusammenhänge zwischen Prominenz und Akustik und dem Einfluss von nicht akustischen Parametern auf die Wahrnehmung von Prominenz folgen einige Beispiele über die Anwendung des Konzepts von Prominenz in der Sprachtechnologie. Es folgt die Motivation für zwei Experimente und die Frage nach der Normalisierung von Prominenzurteilen.

2.1. Eine Definition von Prominenz

Wenn man sich mit einer Fragestellung beschäftigt, so sollte man den Gegenstand seiner Fragestellung definieren. Mit einer klaren Definition verdeutlicht man, worüber geredet wird und worüber nicht, und man verhindert, dass sich im Diskurs mehrere mit dem selben Wort auf unterschiedliche Dinge beziehen. Diese Arbeit möchte sich mit der Erhebung von Prominenz in gesprochener Sprache beschäftigen. Der Begriff der Prominenz reicht an verschiedene Gebiete der Prosodie heran. In Arbeiten zur Betonung, zum Fokus und weiteren prosodischen Phänomenen wird von prominenten Einheiten gesprochen. Es gibt auch eine große Anzahl an Arbeiten, die das Wort Prominenz im Titel tragen und den Zusammenhang der Prominenz zu akustischen, linguistischen oder anderen Kenngrößen, wie beispielsweise Erwartungen oder visuelle Cues in der multimodalen Verarbeitung untersuchen. Einen Überblick hierzu soll Kapitel 2.2 bieten.

Was bezeichnet aber Prominenz genau und wie grenzt sich Prominenz beispielsweise von Betonung ab? Hierzu sollen verschiedene Definitionen aus der Literatur betrachtet werden.

Prominenz ist ein perzeptuelles Konstrukt, dass das Empfinden beschreibt, dass eine Einheit aus ihrer Umgebung hervorsticht. In vielen Arbeiten wird Pro-

minenz über diese Empfindung definiert. Terken (1991) beginnt mit der folgenden Definition von Prominenz:

> „Prominence is the property by which linguistic units are perceived as standing out from their environment."

Streefkerk (2002) schreibt auf Seite 2 ihrer Dissertation folgendes über Prominenz:

> „When we listen to speech some parts seem more prominent than others. In other words, we perceive specific parts of the speech signal as uttered with more 'emphasis' than other parts. This emphasis is called 'prominence'."

Hierbei ist zu sagen, dass die Wahrnehmung von Prominenz sich aus mehreren Quellen speisen kann, die in komplexer Weise interagieren. Die verschieden Reize werden dann zu einer Gesamtwahrnehmung integriert. Goldman et al. (2010) verwenden hierfür den Begriff *binding* und referieren für den Gebrauch des Terminus auf Bache (2005) und Fauconnier und Turner (2002). Watson (2010) benutzt zur Beschreibung der Integration mehrerer Quellen den Terminus *Multiple Source view of prominence*. Prominenz ist ein gradueller Parameter. Streefkerk verweist zum Beleg auf die Arbeiten von Terken (1996) und Rietveld und Gussenhoven (1983). Aber auch noch in jüngster Zeit betonen Autoren, dass Prominenz graduell ist, wie beispielsweise Watson (2010) in seinen Ergebnissen schreibt, nachdem er in seinen Hypothesen noch vermutet, dass Prominenz kategorial ist:

> „We have found that 1) prominence is not categorical and can vary continuously with discourse structure [...] "

Trotzdem finden sich in der Literatur zahlreiche Verwendungen des Terminus *Prominenz* als binäre Eigenschaft - also eine Einheit ist entweder prominent oder nicht. Siehe hierzu auch Kapitel 2.3.

Wagner und Portele (1999) und Cole et al. (2010) weisen darauf hin, dass die Verwendung des Begriffs *Prominenz* auf zwei Ebenen verweisen kann. Wagner und Portele (1999) sprechen hierbei von *Emphase* und *Akzent*. Cole et al. (2010) unterscheiden zwischen *struktureller* und *akustischer* Prominenz. Hierbei fasst Cole alle Phänomene, die aufgrund der phonologischen Struktur vorgegeben sind, zum Beispiel die Silbe, die den Wortakzent trägt, unter *struktureller* Prominenz zusammen und alle Phänomene, die eine Hervorhebung durch Veränderung

verschiedener akustischer Parameter machen, um beispielsweise einen Fokus zu erzeugen oder verschiedene Stufen von Gegebenheit im Diskurs zu markieren mit *akustischer* Prominenz.

Alle gerade genannten Eigenschaften von Prominenz finden sich in der Definition von Wagner (2002), S.11, die auch die Definition von Prominenz für diese Arbeit sein soll :

> „Prominenz bezeichnet die graduell wahrgenommene Stärke einer prosodischen Einheit, die mindestens eine Silbe umfassen muß, relativ zu ihrer Umgebung. Kommt der Prominenz innerhalb einer prosodischen Einheit eine bedeutungsrelevante Funktion innerhalb der Kommunikationskette zu, erhält sie den linguistischen Status einer Betonung."

2.2. Forschungslage zur Prominenz

Im Folgenden wird zunächst die Forschungslage im Bereich der Prominenz vorgestellt. Als erstes soll die umfangreiche Forschung umrissen werden, die sich mit dem Zusammenhang zwischen der wahrgenommen Prominenz und verschiedenen akustischen Maßen beschäftigt. Als zweites folgt eine Zusammenstellung von Studien, die Evidenz dafür liefern, dass die Wahrnehmung von Prominenz nicht nur aus der Verarbeitung akustischer Hinweise allein generiert wird. Schlussendlich soll ein kleiner Überblick über Anwendungen des Prominenzkonzeptes in der Sprachtechnologie die Forschung an der Prominenz auch jenseits der Grundlagenforschung motivieren.

2.2.1. Prominenz und akustische Korrelate

Ein großer Teil der Forschung zur perzeptuellen Prominenz beschäftigt sich mit der Frage, welche akustischen Parameter in welcher Weise für den Eindruck von Prominenz verantwortlich sind. Die Parameter, die dabei die meiste Aufmerksamkeit erfahren haben, sind der Verlauf der Grundfrequenz, die Dauer der betrachteten Einheit oder Teile der Einheit, wie zum Beispiel die Dauer des Silbenkerns, die Intensität und spektrale Maße, wie beispielsweise das Verhältnis der Energie in verschiedenen Frequenzbändern. Einen Überblick über die verschiedenen Erkenntnisse zum Einfluss der verschiedenen akustischen Merkmale auf die wahrge-

nommene Prominenz zu geben fällt schon deshalb schwer, weil die verschiedenen Studien nicht nur auf unterschiedlichen Sprachen und Maßen (z.B. Dauer der Silbe vs. Nucleus vs. logarithmierte Werte), sondern auch auf unterschiedlichen Herangehensweisen basieren. Dies drückt sich in den Fragestellungen, aber auch in der Präsentation der Ergebnisse aus, und so geben nur verhältnismäßig wenige Studien Korrelationen zwischen Prominenzurteilen und akustischen Werten an. Wie schon Watson (2010) feststellte, ergeben sich dabei viele Unterschiede in der Gewichtung einzelner Faktoren aus den unterschiedlichen Perspektiven und Methoden, mit denen die verschiedenen Studien sich der Fragestellung nähern.

2.2.1.1. Der Einfluss der Dauer auf die Wahrnehmung von Prominenz

Es gibt verschiedene phonologische Ebenen, auf denen man Dauern messen kann. Wenn man Silbenprominenz untersucht, sind zwei mögliche Dauern, die man messen könnte, die ganze Silbe oder der Silbenkern. Dabei ist die Segmentierung von Wörtern in Silben keinesfalls trivial. Bei den verschiedenen Prozessen der Wortbildung können Laute, die in der Grundform zu einer Silbe gehören, einer anderen Silbe zugeordnet werden. Wenn man dabei das Maximum-Onset-Prinzip bis in letzte Konsequenz durchführt, entstehen dabei schon mal Silben, die nach gängiger Auffassung zu Phonotaktischen Beschränkungen nicht zulässig sind. Ein Beispiel ist die Zerlegung von Deutschland in Deutsch.land vs. Deu.tschland. Ein weiteres Problem ist, dass es keine kanonische Segmentierung der Phone gibt und dass die Grenzen zwischen zwei Segmenten oder Silben im Signal von verschiedenen Personen oder Algorithmen mehrere Nulldurchgänge weit auseinander gesetzt werden können.

Beim Silbenkern kann man verschiedene Definitionen heranziehen und den Silbenkern beispielsweise phonologisch motivieren, oder über Bereiche mit einer bestimmten Intensität. Manche Autoren verzichten auch auf die Diskussion des Silbenkerns und messen stattdessen Vokaldauern. Auch hier bleibt das grundsätzliche Problem der Segmentierung bestehen. Wenn man Wortprominenz untersucht, kann man neben den genannten Dauern noch die Länge des ganzen Wortes messen. Fant und Kruckenberg (1999) stellen heraus, dass Dauer das robusteste Korrelat für Prominenz ist. Die Untersuchungen von Fant und Kruckenberg beziehen sich dabei auf das Schwedische. Auch Turk und Sawusch (1996) stellten bei einer Studie zum Einfluss von Dauer und Intensität auf Prominenz

fest, dass der Einfluss der Intensität marginal ist. Hierbei wurden Silbenkonstrukte Muttersprachlern von amerikanischem Englisch vorgespielt.

In den verschiedenen Studien werden fast alle Modelle von phonologischer Dauer verwendet. Die Silbendauer in Sekunden oder Millisekunden werden beispielsweise in Portele (1998), Goldman et al. (2007), Krahmer und Swerts (2007), Streefkerk (2002), Wang und Narayanan (2007) und Kohler (2008) analysiert. Normalisierte Silbendauern werden von Mixdorff und Widera (2001) und Streefkerk (2002) verwendet. Tamburini (2003) zeigt für das *TIMIT Korpus*, dass die Dauer des Silbenkerns und die Gesamtdauer der Silbe in gleicher Weise zur Bestimmung von Prominenz geeignet sind. Da die Dauer des Silbenkerns laut Tamburini einfacher automatisch bestimmbar ist als die Silbendauer, verwendet Tamburini im folgenden die logarithmierte Dauer des Silbenkerns. Goldman et al. (2007) merken dagegen jedoch an, dass die Dauer des Kerns zu sehr von Onset und Coda der Silbe beeinflusst wird und sprechen sich für die Verwendung der Silbendauer aus. In Liljencrants (1999) wird zum einen die Silbendauer, zum anderen die Dauer des Vokals mit den Prominenzurteilen von Probanden korreliert. Hierbei findet sich einen höhere Korrelation zur Vokallänge *(r=.69)*, als zur Silbendauer *(r=.54)*. Auch in den verschiedenen Bedingungen von Arnold et al. (2010) sind die Korrelationen zwischen Silbenkern und Prominenz durchweg höher als die Korrelationen zwischen Silbendauer und Prominenz.

2.2.1.2. Der Einfluss der Grundfrequenz auf die Wahrnehmung von Prominenz

Bei der Untersuchung des Einfusses der Grundfrequenz auf die wahrgenommene Prominenz gibt es viele mögliche Werte, die man hierzu heranziehen kann. Man kann beispielsweise den Mittelwert, das Maximum oder die Spanne in der betrachteten Einheit angeben. Als Skalierung der Grundfrequenz kommt die Angabe in *Hz* oder logarithmische Darstellungen, beispielsweise in Halbtonschritten - in der englischsprachigen Literatur meistens mit *st* für *semitone* bezeichnet - in Frage. Es gibt auch verschiedene Verfahren zur Parametrisierung der Grundfrequenz, wie zum Beispiel das Model von Fujisaki (1983) oder TILT (Taylor, 2000), die in der Forschung zur Prominenz eingesetzt werden. Die Bestimmung der Grundfrequenz als Korrelat der Stimmlippenbewegung im Signal ist generell schwierig und sämtliche Verfahren sind mehr oder weniger fehleranfällig. Einen umfassen-

den Überblick über die Problematik der Grundfrequenzbestimmung liefert Hess (1983).

Streefkerk (2002) sagt, dass Veränderungen der Grundfrequenz für das Niederländische den wichtigsten Hinweis für Prominenz liefern. Sie untersucht hierbei Prominenz auf Wortebene im Niederländischen. Zum Einfluss der Grundfrequenz gab es zahlreiche Studien, die mit künstlichen Grundfrequenz Konturen gearbeitet haben. Unter anderem sind hier Terken (1991), Gussenhoven et al. (1997) und aus jüngerer Zeit Kohler (2008) zu nennen. Diese Arbeiten benutzen vornehmlich eine Trägerphrase und verschiede F0 Konturen, die von den Probanden verglichen oder angepasst werden sollen. Dabei kommen sowohl natürlichsprachliche Sätze (z.B. Gussenhoven et al. (1997)) als auch Silbenfolgen ohne Bedeutung (z.B. Terken (1991)) zur Anwendung. Die Ergebnisse dieser Studien zeigen schön, dass die Prominenz einer Einheit auch von ihrer Position in der Phrase und von der Prominenz ihrer Nachbarn abhängt. Gussenhoven und Rietveld (1998) benutzen manipulierte Sprachdaten, um zu zeigen, dass bei der Bewertung von Grundfrequenzmaxima hinsichtlich der Prominenz sprecherspezifische Charakteristika eine entscheidende Rolle spielen. Die künstlich erzeugten Daten sollten hierbei einmal von einem Mann und einmal von einer Frau stammen. Die „weiblichen" Stimuli wurden bei gleicher Auslenkung der Grundfrequenz weniger prominent bewertet, als „männliche" Stimuli. Die Autoren folgern daraus, dass die Hörer bei der Frau eine größere Spanne der Grundfrequenz annehmen und ihre Bewertungen entsprechend anpassen.

Die Grundfrequenz von natürlichen Stimuli wird in weiteren Studie behandelt. Mixdorff und Widera (2001) benutzen für ihre Studie eine Parametrisierung mittels des Fujisaki-Modells (Fujisaki, 1983). Sie finden dabei eine starke Korrelation zwischen der wahrgenommenen Silbenprominenz und dem Akzentkommando Aa. Der Einfluss ist hierbei stärker als der Einfluss der normalisierten Silbendauer. Auch Vainio und Järvikivi (2006) benutzen eine Parametrisierung nach Fujisaki. In ihrer Studie zur Evaluation verschiedener Skalen zur Beurteilung von Prominenz benutzen Jensen und Tøndering (2005) den Anstieg der Grundfrequenz in Halbtonschritten. Sie erhalten Korrelationen zwischen 0.59 und 0.62. Da sie keine weiteren akustischen Parameter betrachten, kann man davon ausgehen, dass sie den Parameter Grundfrequenz für wichtig erachten. Insbesondere bei der automatischen Prominenzerkennung ist die Parametrisierung der Grundfrequenz verbreitet. Sowohl in den Arbeiten von Tamburini (2003), Tamburini und Caini

(2005) und Tamburini und Wagner (2007), die allesamt auf dem selben Prinzip beruhen, als auch bei Wang und Narayanan (2007) wird eine Parametrisierung der Grundfrequenz in Anlehnung an TILT (Taylor, 2000) verwendet. Wang und Narayanan (2007) benutzen zusätzlich lokale Werte der Grundfrequenz im Silbenkern. Hierbei finden Maximum, Minium, Mittelwert, Median, Spannweite und Standardabweichung Verwendung. Die Autoren stellen heraus, dass, wenn man nur die parametrisierte Grundfrequenz oder nur lokalen Grundfrequenzparameter zusammen mit Silbendauer und Intensität benutzt, die Parametrisierung bessere Ergebnisse erzielt als die lokalen Werte. Die Kombination verbessert die Gesamtleistung des Algorithmus.

2.2.1.3. Der Einfluss der Intensität auf die Wahrnehmung von Prominenz

Auch für die Intensität gibt es verschiedene Maße die zur Quantifizierung verwendet werden. Verbreitet sind Angaben in *Dezibel (dB)* oder als *Root Mean Square (RMS)*. Auch hier stellt sich die Frage, ob über die ganze Einheit, also beispielsweise die Silbe, oder nur den Kern gemessen wird.

Kochanski et al. (2005) stellen heraus, dass Lautheit das beste Korrelat zu Prominenz ist. Die Datenbasis bilden hierbei verschiedene Korpora englischer Sprache. Bei der Lautheit handelt es sich um eine in Sone gemessene Größe, die verschiedene Eigenschaften des Gehörs - beispielsweise die gesteigerte Empfindlichkeit bei 1000 Hz - berücksichtigt. Die Autoren betrachten auch RMS und stellen fest, dass sich die Ergebnisse nicht wesentlich unterscheiden. Die Verwendung von Lautheit wird auch in Streefkerk (2002) diskutiert. In der Arbeit wird jedoch von der Berechnung aufgrund verschiedener Schwierigkeiten abgesehen.
Die meisten Studien benutzen Intensität in *dB*, so beispielsweise Eriksson et al. (2002), Streefkerk (2002) und Krahmer und Swerts (2007). Tamburini (2003) verwendet für seinen automatischen Prominenzerkenner den RMS normalisiert an Mittelwert und Varianz in der jeweiligen Phrase. Liljencrants (1999) gibt für seine Daten eine Korrelation von $r = 0.43$ zwischen den Prominenzurteilen und der Intensität in Dezibel an.

2.2.1.4. Der Einfluss spektraler Maße auf die Wahrnehmung von Prominenz

In den Studien von Campbell und Beckman (1997), Portele (1998) und Eriksson et al. (2002) wird der Einfluss verschiedener spektraler Maße bei der Wahrnehmung von Prominenz untersucht. Campbell und Beckman (1997) untersuchen die Energieunterschiede zwischen den ersten beiden Harmonischen sowie die Intensitäten in verschiedenen Bändern eines ERB-Spektrums. Während sich das erste Maß nicht bewährte, zeigte sich ein Anstieg in den höheren Bändern des Spektrums bei prominenten Einheiten.

Portele (1998) untersuchte die Energie in den folgenden vier verschiedenen Frequenzbändern: 0-1 kHz, 1-2 kHz, 2-4 kHz und 4-8 kHz. Diese berechnete er nur für Vokale. Die Unterschiede zwischen prominenten und nicht prominenten Vokalen waren dabei stark von der Art des Vokals und vom jeweiligen Sprecher abhängig.

2.2.2. Beeinflussung der Wahrnehmung von Prominenz durch nicht akustische Faktoren

Die wahrgenommene Prominenz einer linguistischen Einheit wird nicht ausschließlich durch ihre eigenen akustischen Eigenschaften festgelegt. Wie verschiedenen Studien gezeigt haben (z.B. Gussenhoven et al. (1997)), beeinflusst auch der Kontext die Wahrnehmung von Prominenz maßgeblich. Darüber hinaus zeigen zahlreiche Studien, dass auch weitere Faktoren Einfluss auf die Wahrnehmung der Prominenz haben. Eine Studie, die einen möglichen Einfluss von Top-Down Prozessen zusätzlich zu Bottom-Up Prozessen bei der Wahrnehmung von Prominenz erstmals andeutet, ist die Untersuchung von Fant und Kruckenberg (1989). In ihrer Studie untersuchten Fant und Kruckenberg die Prominenz auf Silbenebene im Schwedischen. Hierzu verwendeten sie eine 31-Punkt-Skala, auf der jede Silbe der durch einen trainierten Sprecher gesprochenen Äußerungen hinsichtlich ihrer wahrgenommenen Prominenz bewertet werden sollte. Um die Versuchsteilnehmer an den Gebrauch der Skala zu gewöhnen, wurden die Probanden angewiesen, sich die Äußerungen auf den Bewertungsbögen still vorzulesen und ihre „innere Stimme" hinsichtlich der Prominenz zu bewerten. Im Anschluss folgte das eigentliche Experiment, bei dem die Probanden die aufgenommen Sätze bewerteten. Als Fant und Kruckenberg später die Bewertungen der Probanden von

ihren eigenen Erwartungen mit den Bewertungen der tatsächlichen Äußerungen verglichen, stellten sie fest, dass sich die introspektiv gewonnen Bewertungen und die Bewertungen der Aufnahme kaum unterschieden.

Wenn zur Kommunikation neben dem akustischen Kanal auch der visuelle Kanal zur Verfügung steht, haben visuelle Reize Einfluss auf die Wahrnehmung von gesprochener Sprache. Das berühmteste Beispiel ist sicher der so genannte McGurk Effekt, der in McGurk und MacDonald (1976) beschrieben wird. Hierbei werden Tonaufnahmen, welche zu unpassenden Artikulationsbewegungen in einem Film synchronisiert sind, als zu einer anderen Kategorie zugehörig wahrgenommen. In der Studie von Granström et al. (1999) wird mittels einer virtuellen Agenten gezeigt, dass Mimikgesten einen Einfluss auf die Wahrnehmung von Prominenz haben. Die Autoren untersuchten den Einfluss von Augenbrauenbewegung auf die Wahrnehmung von Prominenz im Schwedischen. Es zeigte sich, dass die Augenbrauenbewegung als eigenständiger Hinweis für Prominenz interpretiert wird und unabhängig von der in der Studie untersuchten Grundfrequenzbewegung Prominenz markieren kann.

In einer Folgestudie von House et al. (2001) wurde untersucht, ob Kopfbewegungen aufgrund der größeren Fläche, die sich bewegt, ein stärkerer Hinweis für Prominenz ist als Augenbrauenbewegung. Ein weiterer Untersuchungsgegenstand war, wie kritisch das *Timing* der Gesten für die Prominenzwahrnehmung ist. Die Ergebnisse zeigen, dass beide Gesten unabhängig zur Markierung von Prominenz benutzt werden können. Die Kopfbewegung ist hierbei ein stärker Hinweis für Prominenz und in akustisch ambigen Settings übertrifft sie die Augenbrauenbewegung als Hinweis für Prominenz. Das Timing der Gesten in Koordination mit der prominenten Silbe ist laut den Autoren zwar wichtig, aber nicht sehr kritisch und muss in einem Rahmen von etwa 100 ms liegen.

Eriksson et al. (2001) führten eine Studie an schwedischen Versuchsteilnehmern durch, bei der vorrangig der Einfluss von *vocal effort* auf die Wahrnehmung von Prominenz erforscht werden sollte. Hierzu wurde von Versuchsteilnehmern ein Korpus hinsichtlich der Prominenz auf Silbenebene mittels einer kontinuierlichen Skala bewertet. Die Autoren führten einige Regressionsanalysen mit verschieden akustischen Variablen als Prädiktoren für die wahrgenommene Prominenz durch. Zusätzlich implementierten die Forscher ein linguistisch geleitetes Modell, bei dem die Prominenz auf Silbenebene aus drei Faktoren vorhergesagt wurde. Diese Faktoren waren, ob die Silbe einen Hauptakzent tragen kann, ob die Silbe einen

Nebenakzent tragen kann und ob das Wort, dem die Silbe angehört, kontrastiv eingesetzt wird. Die erklärte Varianz aus dem letzten Modell überragte mit 57% erklärter Varianz die Aufklärung der Modelle mit akustischen Parametern, bei denen die höchste aufgeklärte Varianz bei 48% lag. Aus ihren Ergebnissen folgerten Eriksson et al., dass dies nicht bedeute, dass Probanden sich bei ihrer Bewertungen von Top-Down Prozessen leiten lassen, sondern dass man schlicht nicht sagen könne, welche Parameter die Probanden bei ihrer Bewertungen von Silbenprominenz benutzen.

Das Experiment motivierte eine Folgestudie Eriksson et al. (2002). Das Material aus Eriksson et al. (2001) wurde englischen Muttersprachlern ohne Kompetenz im Schwedischen vorgespielt. Um die Erwartung von schwedischen Muttersprachlern zu untersuchen, wurde den schwedischen Muttersprachlern das Material nur in schriftlicher Form präsentiert. Es zeigt sich, dass sich wie bei Fant und Kruckenberg (1989) nur wenige Unterschiede zwischen den Bewertungen der Erwartung und der tatsächlichen Realisierungen ergeben. Nur bei wenigen Silben unterschied sich die Erwartung der Versuchsteilnehmer von der tatsächlichen Realisierung. Die Autoren stellen keine großen Unterschiede in den Bewertungen durch die englischen und schwedischen Hörer fest. Es zeigt sich aber, dass dort, wo die Erwartung der Muttersprachler und die tatsächliche Realisierung auseinander gehen, Unterschiede zischen den englischsprachigen und den schwedischsprachigen Versuchsteilnehmern auftreten. Weiterhin gibt es kleine Unterschiede bei der Gewichtung der akustischen Merkmale durch die beiden Hörergruppen.

In ihrer Dissertation beschreibt Streefkerk (2002) neben akustischen Korrelaten von Prominenz und dem Aufbau eines neuronalen Netzes zur automatischen Etikettierung von Prominenz auch die Vorhersage von Wortprominenz im Niederländischen. Hierzu verwendet sie die Prädiktoren *Wortart*, *Anzahl der Silben im Wort*, *Adjektiv-Nomen Kombination* und *Position des Wortes im Satz*. Mit fünf Regeln, die zu einem Wort jeweils eine Prominenzstufe addieren oder abziehen wird die Prominenz der Wörter auf einer Skala von 0-4 vorhergesagt. Die prädizierten Werte wurden mit Prominenzbewertungen von Probanden auf gesprochener Sprache verglichen. Dabei wurde eine Akkuratheit von 81.2 % korrekten Vorhersagen erzielt, wenn die Stufen 0 und 1 mit nicht prominent und die Stufen 2, 3 und 4 mit prominent kodiert werden, um einen Vergleich mit den binären Urteil der Probanden zu ermöglichen.

Wagner (2005) untersuchte den Einfluss der Erwartungshaltung von Mut-

tersprachlern und Nichtmuttersprachlern auf die Wahrnehmung von Silbenprominenz im Deutschen. Hierbei verwendete sie Sprache mit normaler Sprechgeschwindigkeit und sehr schnell gesprochene Sprache. Die Idee dahinter war, dass Nichtmuttersprachler weniger Erwartungen in der Fremdsprache haben und somit ihre Beurteilung von Silbenprominenz mehr von den akustischen Realisierung der Äußerung geleitet sind Muttersprachler. Einer weiteren Gruppe wurden ausschließlich die Stimuli in geschriebener Form präsentiert. Diese sollten, ähnlich wie bei Fant und Kruckenberg (1989) ihre erwartete Realisierung bewerten. Wagner fand hohe Korrelationen zwischen den Bewertungen der geschriebenen und akustisch präsentierten Stimuli. Die Korrelationen waren höher ausgeprägt, wenn die gesprochenen Stimuli schnell gesprochen waren. Die Gruppe der Nichtmuttersprachler verließ sich bei schnell gesprochener Sprache mehr auf akustische Parameter als die Muttersprachler. Wagner folgerte hieraus, dass Hörer sich auf Introspektion verlassen, wenn sie keine reliablen akustischen Hinweise erhalten und ihre Erwartungen als "Fallback Strategie" benutzen.

Vainio und Järvikivi (2006) berichten von einer Studie mit vier Experimenten zur Prominenzwahrnehmung im Finnischen. Untersucht wurde der Einfluss von Tonakzent, Intensität und Wortstellung auf die Wahrnehmung der Prominenz. Als Basis diente der Satz „Menemme laivalla Lemille." („Wir fahren mit dem Boot nach Lemi."). Probanden waren jeweils zwischen zehn und vierzehn Studenten, die allesamt finnische Muttersprachler waren. Die Prominenz wurde hierbei nicht direkt gemessen, sondern die Probanden mussten mittels eines Forced-Choice Tests angeben, ob das Wort „Lemille" oder das Wort „laivalla" betont ist oder keines von beiden. Die Autoren manipulierten in den vier verschiedenen Experimenten den Grundfrequenzverlauf und die Intensität des Materials sowie die Reihenfolge der letzten beiden Wörter. Die Forscher schließen aus ihren Ergebnissen, dass die Wortstellung einen Einfluss auf die Prominenzwahrnehmung im Finnischen hat.

Mit der Frage, ob geringere Prominenzwerte auf intermediären Verben im britischen Englisch auf den Einfluss der Position oder auf die Wortart zurückgehen, beschäftigt sich Jensen (2006). Hierfür wurden zwölf Sätze von drei Sprechern eingesprochen, so dass es insgesamt 36 Testsätze gibt. Dabei wurden systematisch Prominenz und Wortklasse über die verschiedenen Wortpositionen variiert. Als Versuchsteilnehmer dienten 23 schwedische Englischstudenten. Es zeigt sich, dass prominente Wörter in intermediärer Position geringere Prominenzwerte erhalten

unabhängig von der Zugehörigkeit zur Wortklasse. Verben und Nomen erhalten weniger hohe Prominenzwerte als Adjektive.

Die Studie von Krahmer und Swerts (2007) beschäftigt sich mit dem Einfluss von visuellen Hinweisen auf die Produktion und Perzeption von Prominenz. Die Autoren führten drei Experimente durch. Im ersten Experiment wurden 11 Sprecher aufgefordert, den Satz „Amanda gaat naar Malta" - *„Amanda geht nach Malta."* - der unter anderen auch schon in Gussenhoven et al. (1997) verwendet wurde, auf unterschiedliche Arten zu realisieren. Dabei sollten die Probanden auf „Amanda" oder „Malta" eine visuelle Geste (eine Handbewegung, oder ein Kopfnicken oder ein schnelles Hochziehen der Augenbrauen) und einen Tonhöhenakzent auf einem oder keinem der beiden Worte ausführen. Die Probanden realisieren dabei Wörter, auf denen sie eine visuelle Geste ausführen sollen, mit einer längeren Dauer und einem niedrigeren zweiten Formanten. Um herauszufinden, ob die akustischen Unterschiede für die auditorische Wahrnehmung relevant sind, wurden die Signale im zweiten Experiment von drei unterschiedlichen Hörern auf einer dreistufigen Skala bewertet. Diese unterschied zwischen keinem wahrnehmbaren Tonhöhenakzent (Wert *0*), einem leichten Tonhöhenakzent (Wert *1*) und einem starken Tonhöhenakzent (Wert *2*) . Die Autoren addierten die Werte der drei Probanden zu einer sechsstufigen Prominenzskala. Die Ergebnisse zeigen, dass ein Wort, das zusammen mit einer Körperbewegung realisiert wird, auditiv deutlich prominenter ist als ein Wort, das nicht von einer solchen Bewegung begleitet wird. Im dritten Experiment wurde 20 Versuchsteilnehmern das Material von drei Sprechern mit und ohne Bild vorgespielt. Es zeigt sich, dass die Prominenz eines Wortes höher ausfällt, wenn die Probanden die Geste dazu sehen können, als wenn sie nur das Signal hören. Eine visuelle Geste hebt dabei nicht nur die Prominenz des Wortes, mit der sie realisiert wird. Die Prominenz des anderen Wortes wird dadurch von den Versuchsteilnehmern als weniger prominent beurteilt, als wenn sie nur das Signal hören.

Vorhersagbarkeit und Wichtigkeit einer Information sollen einen Einfluss auf Realisierung von Prominenz haben. Die Studie von Watson et al. (2008) benutzt eine Variante des Spiels *Tic Tac TOE*, um der Frage nachzugehen, wie sich diese beiden Faktoren auf die Realisierung von Prominenz auswirken. Hierbei spielen die Probanden nicht mit einem Stift, sondern müssen ihre Spielzüge sprechen, die dann in das Spielfeld eingezeichnet werden. Bei dem Spiel gibt es je nach Spielstadium mehr oder weniger einfach vorhersehbare Züge und mehr oder we-

niger wichtige Züge. Wichtige Züge sind dabei solche, die das Spiel gewinnen oder ein Gewinnen des Gegners verhindern. Jeweils 20 Probanden spielten paarweise 20 Spiele. Untersucht wurden die Unterschiede zwischen vorhersagbaren und nicht vorhersagbaren sowie wichtigen und unwichtigen Spielzügen. Es zeigt sich, dass die Vorhersagbarkeit der Information vor allem Einfluss auf die Dauer und die Variation der Grundfrequenz hat. Nicht vorhersagbare Spielzüge haben ein längere Dauer und größere Spanne der Grundfrequenz als leicht vorhersagbare Spielzüge. Die Wichtigkeit der Information hat vor allem Einfluss auf die Intensität. Wichtige Spielzüge werden mit einer höheren Intensität realisiert als unwichtige Spielzüge.

Unter der Annahme, dass die Erwartungshaltung von Hörern die Wahrnehmung von Silbenprominenz maßgeblich beeinflusst, führten Arnold und Wagner (2008) eine Primingstudie an deutschen Muttersprachlern im Deutschen durch. Hierbei wurde die Erwartungshaltung der Probanden hinsichtlich der Prominenz in bestimmten Sätzen geprimt. Die Probanden lernten dabei ein Prominenzmuster mit einer bestimmten syntaktischen und semantischen Struktur zu verknüpfen. Das Priming war grundsätzlich erfolgreich und führte zu signifikanten Unterschieden in der Beurteilung der Silbenprominenz durch die Probanden. In einer späteren Studie zeigten Arnold et al. (2010), dass die Unterschiede in den Beurteilungen der beiden Gruppen zu signifikanten Unterschieden in den Korelationen zwischen der bewerteten Silbenprominenz und den akustischen Maßen *Dauer*, *Grundfrequenz* und *Intensität* führten. In den Gruppen, in denen das Priming in Richtung des Testsatzes wirkte, fanden sich stärkere Korrelationen als in den Gruppen, in denen das Priming entgegen der Richtung des Testsatzes wirkte. Da für das Priming in den verschieden Bedingungen verschiedenes Material benutzt wurde, zeigten sich deutliche Unterschiede in der Stärke der einzelnen Korrelationen zwischen Prominenzbeurteilung und den akustischen Maßen.

Die Studie von Goldman et al. (2010) beschreibt Unterschiede zwischen der manuellen Annotation des französischen *C-PROM* Korpus, das in Avanzi et al. (2010) beschreiben wird. Leider wird der Algorithmus, den die Autoren benutzen, nicht näher beschrieben oder zitiert. Man kann leider nur mutmaßen, dass es sich um den in Avanzi et al. (2010) zitierten Algorithmus aus Goldman et al. (2007) handelt, was aufgrund der gleichen Erfassung von Prominenz mittels drei Kategorien sehr wahrscheinlich ist. Die Autoren gehen davon aus, dass bei der Wahrnehmung durch *binding*, also dem Zusammenführen verschiedener Information zu

einer Gesamtwahrnehmung, systematische Abweichungen zwischen der Akustik und der Wahrnehmung von Prominenz entstehen. Als Informationsstränge, die im *binding* zusammengeführt werden, sehen sie die linguistische Erwartung der Hörer auf der einen Seite und den akustischen Input durch das Signal auf der anderen Seite. Die Autoren der Studie beschreiben drei systematische Abweichungen, die auf, wie sie nennen, *auditory illusion* zurückzuführen sind. Als erstes beschreiben sie die *clitic negative illusion*. Hierbei nehmen Hörer akustisch prominente Klitika häufig als nicht betont wahr. Bei der *positive semantic quantity illusion* annotieren die Hörer „plein de“ - französisch für viel - in vier Fünfteln der Fälle als prominent. Die Autoren gestehen aber zu, dass die Daten mit fünf Ausprägungen sehr selten in ihrem Korpus vorkommen. Als letztes beschreiben die Autoren die *positive constructional hammock-pattern illusion*. Dies beschreibt die Tendenz der Hörer, das erste Element einer komplexen semantischen Konstruktion als prominenter wahrzunehmen. Die Autoren schließen mit der Schlussfolgerung, dass ein systematischer Bias bei der Beurteilung von Prominenz durch Hörer stattfindet, der durch das *binding* von der Wahrnehmung an das linguistische, lexikalische, syntaktische und semantische Wissen des Hörers hervorgerufen wird.

2.2.3. Prominenz in der Sprachtechnologie

Die Vorhersage von Wortprominenz anhand von Wortklasse, Wortposition und Wortklasse der Nachbarwörtern wird in Widera et al. (1997) beschrieben. Die Autoren vergleichen einen manuellen Ansatz mit vier *machine learning* Algorithmen. Hierbei kamen ein *information gain tree* und ein *semantic classification tree* und zwei *künstliche neuronale Netze* zum Einsatz. Die Daten zum Training der Algorithmen stammen aus der *Bonner Prosodische Datenbank*, die in Heuft (1996) beschrieben ist. Die Prädiktoren werden für die Synthese von fünf Testsätzen benutzt, die von elf Versuchsteilnehmern bewertet wurden. Bei dem Perzeptionsexperiment zeigt sich nur bei einem Satz eine signifikante Präferenz der Probanden, insgesamt ist keines der Verfahren dem anderen Verfahren signifikant überlegen. Die Autoren stellen fest, dass die Wortklasse und die Position im Satz die Prominenz am besten vorhersagen, während die Wortklasse der Nachbarn eine untergeordnete Rolle spielt.

Tamburini (2003) beschreibt einen Algorithmus zur automatischen Annotation von Silbenprominenz im Englischen. Hierbei werden nur signalinhärente In-

formationen verwendet. Als Datenquelle wurde das *TIMIT Acoustic − Phonetic Continuous Speech Corpus* (Garofolo, 1993) verwendet. Tamburini zeigt, dass sich die Dauer des Silbenkerns ähnlich zur Silbenprominenz verhält wie die Dauer der gesamten Silbe. Da die automatische Segmentierung in Silben wesentlich komplexer und damit anfälliger für Fehler ist als die automatische Segmentierung von Silbenkernen, verwendet Tamburini die Dauer des Silbenkerns für seinen Algorithmus. Tamburini kombiniert einen Erkenner für Betonung und einen Erkenner für Tonhöhenakzente zu einem Prominenzerkenner. Demnach sind prominente Silben entweder betont, oder sie tragen einen Tonhöhenakzent oder beides. Für den Betonungserkenner verwendet Tamburini die normalisierte Silbenkerndauer und die *RMS Energie* im Frequenzband von 500 bis 2000 Hz. Die Verwendung der Energie in dem Band sind aufgrund der Ergebnisse von Sluijter und van Heuven (1996) besonderes viel versprechend. Der Erkenner für Tonhöhenakzent benutzt das Produkt der Paramter *EvAmp* und *EvDur*, sowie den Parameter *EvRel* aus dem *TILT Model*. Für eine Beschreibung des Models siehe Taylor (2000). Insgesamt erreicht das System, das die beiden Erkenner kombiniert, eine korrekte Erkennungsrate von 81.44%. Hierfür wurde die Erkennung mit von einem Hörer annotierten Material verglichen.

Tamburini und Caini (2005) beschreiben zwei Algorithmen zur automatischen Erkennung von Prominenz im Englischen. Die Datenbasis ist hierbei das schon in Tamburini (2003) verwendete *TIMIT Acoustic − Phonetic Continuous Speech Corpus*. Der erste Algorithmus beruht auf einer multivariaten Gauß-Diskriminierung und klassifiziert die Silben in *prominent* und *nicht prominent*. Der Algorithmus benötigt ein Training auf von Hand annotierten Daten und gleicht dem in Tamburini (2003) vorgestellten Algorithmus stark, obwohl die Studie in Tamburini und Caini (2005) nicht zitiert wird. Als wichtigster Unterschied sticht die Anpassung des mittleren Frequenzbandes von 200 bis 500 auf 300 bis 2200 hervor. Der Algorithmus erreicht auf einem Teil des *TIMIT Acoustic − Phonetic Continuous Speech Corpus* eine Rate von 80.73 % korrekten Erkennungen. Als zweiter Algorithmus wird eine kontinuierliche „prominence function" implementiert, die einen kontinuierlichen Prominenzwert ausgibt. Im Wesentlichen werden hierfür die gleichen akustischen Maße wie für die anderen beiden Algorithmen herangezogen. Der Algorithmus klassifiziert 80.61 % der Silben korrekt. Tamburini und Caini schließen ihren Artikel damit, dass beide Algorithmen ähnlich gut abschneiden und auch im Vergleich zur Literatur - hier Jenkin und Scordilis (1996)

und Bagshaw (1993) - gute Ergebnisse liefern. Da der zweite Algorithmus den Kontext beachtet, wollen sie diesem Algorithmus in Zukunft den Vorzug geben.

In Tamburini und Wagner (2007) wird eine Anpassung des zweiten Algorithmus aus Tamburini und Caini (2005) an das Deutsche beschrieben. Als Datenbasis dient die *Bonner Prosodische Datenbank*. Der Algorithmus wurde dahingehend verändert, dass er nicht mehr das Maximum von zwei Komponenten ausgibt, sondern eine gewichtete Summe. Die Gewichte wurden hierbei durch parametrisches Suchen im Suchraum der Gewichtungskomponenten ermittelt. Mit den besten Parametern wird eine Spearman Korrelation von .71 zu den manuellen Annotationen erreicht. Die Autoren geben zwei Faktoren für Abweichungen zwischen den Prominenzurteilen von Hörern und Algorithmus an. Zum einen gibt es Silben, bei denen das Maximum des Tonhöhenakzentes erst nach der prominenten Silbe erreicht wird. Diese werden als schwächer prominent errechnet, als sie von Hörern wahrgenommen werden. Zum anderen tendieren Hörer dazu, rhythmisch alternierende Betonungen wahrzunehmen, auch dann, wenn diese keine akustischen Entsprechungen haben.

Wang und Narayanan (2007) beschreiben einen Algorithmus zur automatischen Etikettierung von Prominenz auf Wortebene in Spontansprache. Hierbei wird die Wortprominenz aus der Silbenprominenz generiert, indem der Wert der prominenteren Silbe als Wortprominenz genommen wird. Der Algorithmus bestimmt die Prominenz auf einem Kontinuum im Intervall [0,1]. Hierzu verwenden die Autoren die Parameter *Silbendauer*, *Intensität*, sechs lokale Grundfrequenzparameter (*Maximum*, *Minimum*, *Durchschnitt*, *Median*, *Spannweite* und *Standardabweichung*) und fünf Parameter, welche die *Kontur der Grundfrequenz* beschreiben. Die Autoren evaluieren den Algorithmus anhand von zwei Methoden. Zunächst korrelieren sie die Prominenzwerte mit Part of Speech Informationen. Hierfür wurde ein Subset des *ICSI Switchboard Data Corpus* verwendet, welches in Greenberg et al. (1996) beschrieben wird. Um diese Prozedur mit manuell gelabelten Prominenzwerten vergleichen zu können, wird der *SASO Dialog Corpus* verwendet, bei dem von drei englischen Muttersprachlern die prominentesten Wörter markiert wurden. Eine Beschreibung des Korpus findet sich in Traum et al. (2005). Der Algorithmus von Wang und Narayanan erreicht 76.8 % korrekte Klassifikation gemessen an dem *SASO Dialog Corpus*. Wang und Narayanan zeigen weiterhin, dass es statistisch signifikante Unterschiede in der Prominenz von Funktions- und Inhaltswörtern gibt. Inhaltswörter sind demnach prominen-

ter als Funktionswörter. Dieses Wissen könne nach Wang und Narayanan (2007) zur Verbesserung von automatischer Spracherkennung Verwendung finden. In der Diskussion findet sich der Hinweis, dass in Beachtung der Definition von Prominenz - die Autoren zitieren hierbei Terken (1994): „words or syllables that are perceived as standing out from their enviroment" - Prominenz gemessen werden sollte, indem man den Kontext mit einbezieht. In ihrer Studie betrachten sie die Parameter jedoch global und sehen diesen Aspekt als Arbeit für die Zukunft.

Goldman et al. (2007) beschreibt die automatische Prominenzannotation im Fran- zösischen. Als Basis dient ein 18 Minuten umfassendes Sprachkorpus, das aus simulierten Wegbeschreibungen und Radiointerviews besteht. Die sechs Aufnahmen sind von jeweils drei Sprecherinnen und Sprechern gesprochen worden. Für die manuelle Annotation des Korpus kam eine dreistufige Skala zum Einsatz. Das Programm zur Prominenzetikettierung basiert auf dem Programm, das in Mertens (2004) beschrieben wird. Der Algorithmus benutzt im Wesentlichen zwei Prädiktoren: Silbendauer und Grundfrequenzmaximum gemessen in Halbtönen. Die Autoren begründen ihre Bevorzugung der gesamten Silbendauer gegenüber des Silbenkerns damit, dass der Silbenkern zu sehr von der Stimmhaftigkeit der Konsonanten im Onset und der Coda abhänge. Der Erkenner liefert eine korrekte Erkennungsrate von 84.1 %. Die Autoren sind mit den Ergebnissen im Vergleich zu der von ihnen herangezogenen Studie von Tamburini und Caini (2005) sehr zufrieden. Die Klassifikation funktioniert dabei auf den beiden Textarten und auch zwischen den beiden Geschlechtern gleich gut.

Brenier et al. (2006) führten eine Studie zur automatischen Prädiktion von Prominenz durch. Es sei angemerkt, dass sie sich hierbei vornehmlich auf Tonhöhenakzent beziehen. Anhand der manuellen Annotationen von Tonhöhenakzenten, Informationsstatus, Kontrast (hier im Sinne von Fokus) und der Unterscheidung von konkreten und unkonkreten Aussagen im Switchboard Korpus (Ostendorf et al. (2001)), wollen die Autoren herausfinden, ob diese Features die Prädiktion von Prominenz gegenüber der Vorhersage anhand klassischer Werte wie z.B. *Part-of-Speech* verbessert. Als weiteren Prädiktor schlagen die Autoren die *Accent Ratio* vor, die sie mit mit Formel 2.2.1 angeben.

$$AccentRatio(w) = \begin{cases} \frac{k}{n} & \text{für } B(k, n, 0.5) =< 0.05 \\ 0.5 & \text{sonst} \end{cases} \tag{2.2.1}$$

Hierbei bezeichnen die Autoren mit k die Häufigkeit mit der das Wort w im Korpus akzentuiert vorkam, und mit n das gesamte Vorkommen des Wortes w im Korpus. B(k,n,0.5) bezeichnet die Wahrscheinlichkeit (unter Annahme einer Binomial Verteilung), dass k Auftreten bei n Versuchen auftreten, wobei die Wahrscheinlichkeit von Auftreten und Nichtauftreten gleich ist. Die Autoren halten fest, dass die weiteren Klassen die Prädiktion nicht maßgeblich erhöhen. Sie stellen aber auch heraus, dass verschiedene Wörter falsch klassifiziert werden, wie z.B. Pronomen und hochfrequente Verben. Hier sehen die Autoren das größte Potential für eine Verbesserung.

In Strom et al. (2007) wird eine Studie zur Prominenzgenerierung in der Unitselection Synthese anhand der *Multisyn Engine* (siehe Clark et al. (2005)) in *Festival* im Englischen beschrieben. Die Autoren benutzen die Kombination einer Tonhöhenakzent-Generierung mit einer für Emphase. Für die Vorhersage der Tonhöhenakzente wurde der Prädiktionsalgorithmus aus Nenkova et al. (2007) verwendet. Die Autoren testen die Synthese mit Aktivierung der Kostenfunktionen für Tonhöhenakzent und Emphase in einem Perzeptionsversuch an 52 Versuchsteilnehmern. Hierbei werden verschiedene Domänen untersucht. Die Probanden präferieren die Synthese, die eine Prominenzprädiktion benutzt. Dabei ergeben sich bei beiden Komponenten signifikante Verbesserungen. In der *child-directed* Domäne zeigt sich, dass sich die Emphase negativ auf die Präferenz der Probanden auswirkt. Die Autoren sagen, dass dies eventuell auf die Beurteilung durch Erwachsene zurückzuführen ist, dass Emphase hier offensichtlich eine andere Funktion als in den anderen beiden Domänen hat und dass die Ergebnisse weiterer Forschung bedürften.

In Windmann et al. (2010) wird die Anwendung des Algorithmus aus Tamburini und Wagner (2007) zur automatischen Prominenzetikettierung eines Korpus für die Integration ins *Bonn Open Synthesis System*, (*BOSS*) beschrieben. Eine Beschreibung von *BOSS* findet sich in Breuer (2009) sowie Breuer und Hess (2010). Der Algorithmus wird auf das bestehende Sprachkorpus angewendet. Die Gewichte werden dabei aus Tamburini und Wagner (2007) übernommen. Zur Evaluation werden zum einen die errechneten Prominenzwerte mit den im Korpus annotierten Betonungskategorien „*Primary*"„*Secondary*" und „*None*" korreliert. Die Ergebnisse weisen auf einen signifikanten Zusammenhang zwischen den berechneten Prominenzwerten und den linguistischen Kategorien. Als zweites wurde ein Experiment durchgeführt, bei dem deutsche Muttersprachler die

Prominenz auf einem kleinen Teil des Korpus mittels einer kontinuierlichen Skala auf einem modifizierten Interface, das schon bei Arnold und Wagner (2008) Verwendung fand, beurteilen. Die Prominenzwerte der Probanden korrelierten im Schnitt mit *r=.61*. Die Korrelation zwischen den mittleren Prominenzwerten der Versuchsteilnehmer und den durch den Algorithmus bestimmten Prominenzwerten lag bei *r=.62*. Es zeigten sich wie bei Tamburini und Wagner (2007) Effekte, die auf einen Rhythmus-Bias schließen lassen, bei dem Silben die auf rhythmisch schweren Zeiten liegen, höhere Prominenzwerte erhalten. Obwohl die Autoren den Ergebnissen ihrer Studie recht positiv gegenüberstehen, schließen sie, dass noch weitere Forschung notwendig ist, um die Unterschiede zwischen maschinellen und manuellen Prominenzannotationen zu ergründen.

2.3. Motivation Experiment zur Erhebung von Prominenz anhand unterschiedlicher Skalen

In den zahlreichen Studien zur Erforschung von Prominenz wurde eine große Bandbreite an verschiedenen Skalen verwendet. Angefangen bei binären Skalen über verschiedene Abstufungen von graduellen Skalen, bis hin zu kontinuierlichen Skalen, bei denen die Probanden die Prominenz auf freien Strecken annotieren, gab es eine Vielzahl an verwendeten Messinstrumenten. Im diesem Unterkapitel sollen die verschiedenen Ansätze, die man in der Literatur findet, dargestellt werden. Es wird gezeigt, dass die bisher vorhandenen methodischen Artikel zur Verwendung von Skalen zur Beurteilung von wahrgenommener Prominenz, widersprüchliche Erkenntnisse liefern. Aus der Darstellung soll ein Experiment motiviert werden, in dem verschiedene Skalen für die Bewertung von Prominenz auf Silbenebene evaluiert werden. Dieses Experiment wird dann im folgenden Kapitel 3 beschrieben.

2.3.1. Binäre Skalen

Zahlreiche Studien benutzen für die Prominenzetikettierung das binäre Merkmal prominent/ nicht prominent. Beispielhaft seien hier Streefkerk (2002), Mo et al. (2008) und Cole et al. (2010) genannt. Hierbei werden unter den Wörtern, die von den Probanden als prominent wahrgenommen werden, Markierungen vorgenommen. Streefkerk (2002) verwendet Punkte unter den betroffenen Wörtern, und bei Mo et al. (2008) werden die Wörter unterstrichen. McDowall (1975) benutzte diese Vorgehensweise bereits, um Betonung auf Wortebene zu untersuchen.

Die Skala ist für Probanden einfach zu gebrauchen, da die Probanden lediglich einen Schwellwert für sich definieren müssen. Man kann die Skala in eine graduelle Skala uminterpretieren, indem man die Anzahl der Bewertungen pro Einheit, die mit prominent bewertet wurde, kumuliert und so bei n Probanden eine (n+1)-wertige Skala erhält. Hierbei erhalten dann Einheiten die von niemanden als prominent bewertet wurden den Wert 0 und Einheiten die von $m \in [1, n]$ Probanden als prominent beurteilt wurden, den Wert m. Dies ergibt einen Wertebereich von 0 bis n, also n+1.

Man kann hier einen Kritikpunkt anfügen: Bei der Verwendung einer so konstruierten Skala nimmt man ein Maß, das eigentlich die Übereinstimmung der

Probanden angibt, als Prominenzwert. Wenn m von n Probanden die Einheit x für prominent halten, sind n-m Probanden anderer Meinung. Wenn man diesen Wert nun als Prominenzwert nimmt, verliert man Aussagen über die Übereinstimmung der Probanden. Man kann diese Kritik noch zuspitzen und die folgende Überlegung anstellen: Um ein möglichst differenziertes Ergebnis mit der so konstruierten Skala zu erzielen, müssen die Schwellwerte der einzelnen Probanden möglichst gleichmäßig über ein gedachtes Kontinuum von Prominenz streuen, um so eine gleichmäßige n+1 wertige Skala zu erhalten. Wenn man eine perfekte Interraterübereinstimmung hätte, so würde dies ausschließlich Werte von 0 und n erzeugen, wodurch man nichts gegenüber der Beurteilung durch einen einzelnen Annotator mit einer binären Skala gewonnen hätte. Eine hohe Interraterübereinstimmung, wie man sie sich sonst für alle Labelaufgaben wünscht, wäre für diese Interpretation also nicht wünschenswert.

2.3.2. Graduelle Skalen

In ihrer einflussreichen Studie zur Wahrnehmung von Silbenprominenz im Schwedischen benutzten Fant und Kruckenberg (1989) eine graduelle 31-Punkt-Skala von 0 bis 30. Den Versuchsteilnehmern wurde gesagt, dass eine unbetonte Silbe üblicherweise einen Wert von 10 hat und eine betonte Silbe etwa einen Wert von 20. Diese Skala wurde später von weiteren Studien, wie zum Beispiel Fant und Kruckenberg (1999) und Wagner (2005), benutzt. Heuft et al. (2000) wandelten die Skala leicht ab und verwendeten ein Skala von 0-31. Hierdurch wird die Skala symmetrisch, was die Versuchsperson dazu zwingt auch bei mittleren Werten eine Tendenz in Richtungen prominent oder nicht prominent zu wählen.

Grover et al. (1997) folgerten aus ihrer Studie, dass sie fortan eine 10-Punkt-Skala (von 0-9) benutzen wollten, um Prominenz auf Wortebene zu transkribieren. Auf diese Studie wird noch ausführlicher im Kapitel 2.3.4 eingegangen. Terken (1996) verwendet eine 11-Punkt-Skala. Er lässt seine Versuchsteilnehmer akzentuierte Silben beurteilen. Dabei soll ‘0’ „no prominence” und ‘10’ „strong prominence” entsprechen. Eine 10-Punkt-Skala wird in einem Experiment in Krahmer und Swerts (2007) verwendet. Die Autoren geben als Vorteil an, dass die Skala feinere Unterschiede erfassen könne und mit ihren 10 Stufen den niederländischen Schulnoten entspricht, mit denen die Probanden der Studie vertraut sind.

Jensen setzt in seinen Studien (Jensen, 2003) eine 4-Punkt-Skala ein. In Jensen

(2003) lässt er Wörter mittels dieser Skala beurteilen. Die einzelnen Skalenwerte sind wie folgt codiert: '0' „no stress"; '1' „weaker/reduced stress"; '2' „(normal) full stress"; '3' „(extra) strong stress". Jensen benutzt diese Skala nach der Evaluation durch sein Paper mit Tøndering (Jensen und Tøndering, 2005) (siehe auch hierfür Kapitel 2.3.4) in einer leichten Abwandlung für eine weitere Studie (Jensen, 2006). Er benutzt hier ebenfalls eine 4-Punkt-Skala. Anstatt Werte von 0 bis 3 benutzt er nun Werte von 1 bis 4. Er begründet diesen Schritt damit, anzuzeigen, das auch die Wörter mit einem minimalen Wert ein gewisses Maß an Prominenz aufweisen.

In Goldman et al. (2007) findet eine dreistufige Prominenzskala Verwendung. Hier wird bei der Annotation „P" für eine stark prominente Silbe , „p" für ein schwach prominente und „NP" für eine nicht prominente Silbe verwendet. Die Skala wird auch zur Annotation des *C-Prom* Korpus verwendet, welches in Avanzi et al. (2010) beschrieben wird.

Für die sechsstufige Skala, die aus den dreistufigen Antworten von drei Probanden in Krahmer und Swerts (2007) konstruiert wurden, gilt die gleiche Kritik wie bei der Konstruktion n-stufiger Skalen aus binären Urteilen von n Probanden. Prominenzurteil und Übereinstimmung der Urteile werden vermischt, so dass man bei einem Prominenzwert nichts mehr über die Übereinstimmung sagen kann.

2.3.3. Kontinuierliche Skalen

In einem Experiment aus Gussenhoven et al. (1997) wird eine 100 mm lange Skala für die Bewertung der relativen Prominenz verwendet. Die Skala war horizontal auf ein Blatt Papier gedruckt. Das linke Ende war mit „little emphasis", das rechte mit „much emphasis" beschriftet. Die Probanden markierten ihr Urteil mit einem Stift durch einen vertikalen Strich an der gewünschten Stelle der Strecke.

Eriksson verwendete für seine Studien (Eriksson et al., 2001, 2002) grafische Schieberegler. Hierbei sollten die Versuchsteilnehmer den Regler einfach in die gewünschte Position bringen, die grafisch den Verhältnissen gerecht wurde. Jeweils ein Regler sollte hierbei auf Maximum gestellt werden, einer auf Minimum verbleiben und die restlichen Regler entsprechend ihren Verhältnissen zu den anderen Reglern angeordnet werden. Eriksson bemerkte, dass nicht alle Versuchsteilnehmer den Anweisungen folgten und mindestens einen Regler auf Maximum stellten. Er berichtet, dass in diesen Fällen alle Werte linear in die gewünschte

Form gebracht wurden.

Windmann et al. (2010) verwendeten für ihre Studie ebenfalls eine grafische Oberfläche bei der Schieberegler für die Bewertung der Silbenprominenz zum Einsatz kamen. Auch hier wurde der Regelweg in 100 Schritte geteilt. Im Gegensatz zu den Studien von Eriksson waren die Probanden in ihrer Beurteilung jedoch frei, den prominentesten Silben verschiedener Sätze unterschiedliche Prominenzwerte zuzuordnen.

2.3.4. Evaluation von verschiedenen Skalen

Grover et al. (1997) führten eine Evaluation zweier verschiedener Ansätze durch. Zum einen benutzen sie eine 4-Punkt-Skala, zum anderen eine offene Skala. Die 4-Punkt-Skala hatte einen Wertebereich von 0-3, wobei drei gleiche Intervalle angenommen werden sollten. Die Rater sollten die 0 als *Default*-Wert annehmen, da die Autoren in einem Vortest zu dem Ergebnis gekommen waren, konsistentere Werte zu bekommen, wenn sie einen Wert als Default vorgeben. Die offene Skala hatte keinerlei Vorgaben bezüglich ihres Wertebereichs. Die Versuchsteilnehmer wurden angewiesen, sich gleich große Intervalle vorzustellen mit dem Wert 0 als kleinsten Wert und einen beliebigen Wert zu wählen, der die maximale Prominenz repräsentieren soll sowie ein Element, das halb so prominent ist, wie dieser Wert. Bei der Skala sollte der Wert 0 keine Prominenz und der maximale Wert die höchst mögliche Prominenz darstellen. Die Skala sollte über die Sätze hinweg konsistent sein. Das heißt, dass ein Wort, welches in seinem Satz maximal prominent ist, trotzdem einen niedrigen Prominenzwert erhalten kann als das prominenteste Wort in einem anderen Satz.

Insgesamt kamen Grover et al. zu dem Schluss, dass die offene Skala zu konsistenteren Ergebnissen bei der Beurteilung von Wortprominenz führt. Hierbei sollte die Skala von einer offenen Skala auf den Wertebereich von 0 bis 9 herunter gebrochen werden. Im wesentlichen wird also eine 10-Punkt-Skala propagiert. Als Kritik kann man einwenden, dass nur drei Versuchsteilnehmer an der Studie teilnahmen, wovon zwei aktiv an Prosodie forschen. Die Versuchsteilnehmer sind hierbei mit den Autoren der Studie identisch.

Jensen und Tøndering (2005) führten eine Evaluation von drei Skalen zur Messung von wahrgenommener Prominenz im Dänischen durch. Basis für ihre Untersuchungen ist das *DanPASS* Korpus, das von Nina Grønnum gepflegt wird. Jensen

und Tøndering untersuchten eine binäre-, eine 4-Punkt-Skala und eine 31-Punkt-Skala. Als linguistische Einheit für ihre Evaluation wählen sie wie Grover et al. (1997) das Wort. Für ihre Untersuchung ließen sie 57 dänische Muttersprachler zwei Monologe mit insgesamt 123 Wörtern hinsichtlich der Wortprominenz beurteilen. Hierbei bewertete jeweils eine Gruppe von Probanden das Material mit einer der drei Skalen. Zunächst verglichen Jensen und Tøndering die Reliabilität der Gruppen mittels *Cronbach's* α. Die Werte liegen zwischen 0.940 und 0.961 und zeigen keinen signifikanten Unterscheid. Bei einem Vergleich der Prominenzbeurteilungen transformieren Jensen und Tøndering alle Werte auf das Intervall [0,1]. Sie bemerken, dass man mit der 31-Punkt-Skala deutlich weniger extreme Werte erhält als mit der 4-Punkt-Skala, und dass man die extremsten Werte mit der binären Skala findet. Die Autoren stellen heraus, dass es für die Prominenzforschung wichtig sei, mit der verwendeten Skala statistisch signifikante Unterschiede zu finden. Sie benutzen hier zum einen den Einfluss von Part-of-Speech, Informationsstatus und die Korrelation zur Grundfrequenz. Im ersten Fall zeigen sich laut den Autoren nur geringe Unterschiede zwischen den Skalen, wobei die 4-Punkt-Skala ein wenig mehr signifikante Unterschiede zwischen den einzelnen Wortklassen aufweist. Bei der Informationsstruktur zeigen sich klar mehr signifikante Ergebnisse bei der 4-Punkt und 31-Punkt-Skala im Vergleich zur binären Skala. Bei der Korrelation zur Grundfrequenz erhalten die Autoren der Studie Werte zwischen *r=.593* (binäre Skala) und *r=.626* (4-Punkt-Skala). Auch hier liegt die 4-Punkt-Skala vorne.

Um ein Maß zu finden, das die Schwierigkeit der Aufgabe für die Probanden misst, erheben Jensen und Tøndering, wie lange die Probanden für die Bewertung eines Stimulus brauchen und wie oft sich die Probanden einen Stimulus während der Beurteilung anhören. Die Autoren finden, dass die benötigte Zeit für die Beurteilung eines Stimulus für die Skalen mit mehr Stufen höher ist. Der Unterschied ist zwischen allen drei Skalen ist signifikant. Der für die Zeit gefundene Effekt ist bei der Anzahl der Wiederholungen eines Stimulus während der Bearbeitung nicht so deutlich. Probanden, die die 31-Punkt-Skala benutzen, brauchen signifikant länger als Versuchsteilnehmer, die eine der anderen beiden Skalen benutzen. Zwischen der binären und der 4-Punkt-Skala gibt es jedoch keinen signifikanten Unterschied. Insgesamt schließen die Autoren aus den Beobachtungen, dass die Benutzung einer Skala mit mehr Stufen mit höheren „Kosten" einhergehe. Die Benutzung von Skalen mit mehr Stufen ist also ressourcenintensiv für die Pro-

banden. Jensen und Tøndering schließen ihren Artikel mit der Schlussfolgerung, das die 31-Punkt-Skala für naive Hörer schwierig zu benutzen sei. Die Autoren verweisen zwar auf ein Experiment, in dem fünf Experten wohl bessere Ergebnisse mit der 31-Punkt-Skala erreichten, dieses wird aber nicht berichtet. Die binäre Skala sei wohl für manche Zwecke ausreichend, der 4-Punkt-Skala aber bezüglich der Anzahl von signifikanten Zusammenhängen bei der Untersuchung von Part-of-Speech, Informationsstruktur und der Korrelation zur Grundfrequenz unterlegen. Die Autoren schließen damit, dass es keine Rechtfertigung für die Benutzung der 31-Punkt-Skala gebe, da sie die Aufgabe erschwere und keinerlei zusätzlich Auflösung brächte.

In ihrer Schlussfolgerung widersprechen sich die beiden Studien Grover et al. (1997) und Jensen und Tøndering (2005). Grover et al. sprechen sich für eine Skala mit mehr Stufen aus, während Jensen und Tøndering die 4-Punkt-Skala empfehlen. Während bei Grover et al. (1997) die drei Autoren die einzigen Versuchsteilnehmer waren, hatten Jensen und Tøndering (2005) 57 Probanden, die in die Auswertung einflossen. Es ist also interessant, diesem Widerspruch nachzugehen.

In Kapitel 3 wird daher ein Experiment durchgeführt, bei denen die 4-Punkt-Skala, die 11-Punkt-Skala und die 31-Punkt-Skala mit einer kontinuierlichen Skala, wie sie bei Eriksson et al. (2001), Eriksson et al. (2002) und Windmann et al. (2010) Verwendung findet, evaluiert und verglichen werden.

2.3.5. Skalenniveau

Bei der Verwendung einer Skala stellt sich die Frage nach ihrem Skalenniveau. Das Skalenniveau bestimmt, wie die Werte interpretiert werden dürfen, und welche statistischen Verfahren man anwenden darf.
Während die Frage bei einer binären Skala trivial ist, wird das Niveau der übrigen Skalen durchaus kontrovers diskutiert. Offensichtlich erfüllen die Skalen, wie sie hier vorgestellt wurden, alle Voraussetzung für eine Ordinalskala, da sich alle Werte in einer eindeutigen Reihenfolge befinden. Man könnte nun argumentieren, dass die Skalen zur Bewertung von Silbenprominenz Likertskalen (Likert (1932)) seien. In der Literatur gibt es zwei Lager, quer durch alle Fachbereiche, die darüber unterschiedlicher Meinung sind, ob ordinalskalierte Daten, oder Likertskalen im Speziellen, wie intervallskalierte Daten behandelt werden dürfen. Einen sehr kurz-

en Abriss zu diesem Thema findet man in Jamieson (2004). Jensen und Tøndering (2005) verweisen bei ihrem Paper auf Siegel und Castellan (1988) und entscheiden sich dafür, die Daten aus den graduellen Skalen als ordinalskalierte Daten zu behandeln.

2.3.6. Bewertung über verschiedene Sätze hinweg

Es gibt zwei verschiedene Strategien, was die Festlegung von maximal prominent angeht. Die meisten, wie zum Beispiel Grover et al. (1997), gehen davon aus, dass die prominenteste Einheit in verschiedenen Sätzen verschiedene Grade der Prominenz aufweisen kann und dies durchaus in der Bewertung deutlich werden soll.

Eriksson vertritt hier einen anderen Ansatz. In seiner Studie (Eriksson et al., 2001) sollten die Probanden jeweils mindestens einen Regler in minimaler Stellung belassen und einen Regler auf den maximalen Wert stellen. Bei allen Sätzen, die diese Vorgaben nicht erfüllten, wurden die Werte linear in die von Eriksson gewünschte Antwortform transformiert.

Es lässt sich darüber streiten, wie schwierig es für die Probanden ist, sich bei jedem Satz die Extremwerte für die Skala vorzustellen und ihr Urteil über den gerade zu Grunde liegenden Satz daran auszurichten. Demnach ist es sicher einfacher, wie bei Eriksson et al. (2001) die Einheiten, seien es jetzt Wörter oder Silben, nur in Relation der gerade zur Beurteilung anstehenden Äußerung zu beurteilen, da die Endpunkte der Skala nicht vorgestellt werden müssen. Über mehrere Äußerungen hinweg kann diese Vorgehensweise jedoch zu erheblichen Verzerrungen führen, da jeweils die prominenteste Einheit den gleichen Maximalwert erhält, obwohl es wahrscheinlich zwischen den unterschiedlichen Maxima von verschiedenen Äußerungen Unterschiede hinsichtlich der wahrgenommen Prominenz geben wird. Diese können sehr stark ausfallen, wenn man beispielsweise eine Äußerung mit einer starken Betonung durch einen kontrastiven Fokus, mit einer einfachen Aussage ohne irgendwelche Auffälligkeiten vergleicht.

2.4. Motivation Experiment zur Erhebung von Prominenz auf Silben vs. Wortebene

Die Literatur zur Prominenzforschung teilt sich im Wesentlichen in zwei Gruppen. Studien, bei der die Prominenz auf Wortebene betrachtet wird, und jene, welche die Prominenz auf Silbenebene untersuchen. Für beide Vorgehensweisen lassen sich viele Beispiele finden. Die Vorhersage von Prominenz wird von Widera et al. (1997) und Streefkerk (2002) auf Wortebene durchgeführt, während Wagner (2002) Prominenz auf Silbenebene vorhersagt. In Wang und Narayanan (2007) wird die automatische Prominenzetikettierung auf Wortebene beschrieben, in Tamburini (2003), Tamburini und Caini (2005) und Tamburini und Wagner (2007) auf Silbenebene. Die Evaluation verschiedener Skalen für Bewertung von Prominenz findet bei Grover et al. (1997) auf Silbenebene statt, Jensen und Tøndering (2005) hingegen führen ihre Evaluation verschiedener Skalen auf Wortebene durch. Auch für weitere Fragestellungen lassen sich jeweils Studien finden, die eine der beiden Ebenen untersuchen. Die Studien auf den verschiedenen linguistischen Ebenen beziehen sich dabei leider nie auf das gleiche Material. Es gibt nur sehr wenige Untersuchungen, die der Frage nachgehen, welche Unterschiede sich ergeben, wenn man die Prominenz auf Basis von verschiedenen linguistischen Einheiten untersucht.

In der Dissertation von Streefkerk (2002) findet sich eine kleine Studie zu genau dieser Fragestellung. Streefkerk benutzt eine binäre Skala und lässt eine Gruppe auf Wortebene und eine andere Gruppe das gleiche Material auf Silbenebene bewerten. Die Probanden wurden hierbei angewiesen, hervorgehobene Wörter (bzw. Silben) zu markieren. Im Anschluss wurden die Ergebnisse für alle acht Hörer eines Experiments aufsummiert.

Streefkerk stellt fest, dass sich bei der Bewertung von Prominenz auf Silbenebene eine detailliertere Verteilung ergibt. Dagegen ist die Übereinstimmung der Urteile der Probanden bei der Bewertung von Wortprominenz deutlich höher als bei der Beurteilung von Silbenprominenz. Streefkerk entschließt sich, für ihre Studie Prominenzurteile auf Wortebene zu benutzen.

Durch das Experiment von Streefkerk wurde also etabliert, dass die Prominenzbeurteilungen von Probanden auf Wortebene eine höhere Übereinstimmung erzielt als die Prominenzbeurteilung auf Silbenebene. Es gibt jedoch keine Aussagen darüber, wie sich beispielsweise die Verteilung der Urteile unterscheiden,

wenn die Prominenz auf Wort- und Silbenebene vergleichen wird. Zeigen sich beispielsweise extremere Werte, wenn die Versuchsteilnehmer nach der Wortprominenz gefragt werden, als wenn die Probanden nach der Silbenprominenz gefragt werden? Wie sieht es mit den Korrelaten zu den gängigen akustischen Parametern aus? Verschiebt sich vielleicht der Einfluss der Parameter Dauer, Intensität und Grundfrequenz je nachdem, ob Prominenz auf Wort- oder Silbenebene betrachtet wird?

2.5. Motivation Normalisierung

Die Frage, ob man die Beurteilungen der Hörer normalisieren sollte, taucht in der Literatur verhältnismäßig selten auf. Es gibt verschiedene Möglichkeiten, wie die Skalen nicht in der gewünschten Weise von den Versuchsteilnehmern benutzt werden. So nutzen beispielsweise verschiedene Probanden die Skala häufig nicht vollständig aus. Verschiedene Probanden benutzen so gegebenenfalls unterschiedliche Bereiche der Skala, die sich nicht nur in ihrer Ausschöpfung der Skala - beispielsweise vier Drittel der Skala gegen knapp die Hälfte - als auch hinsichtlich der Lage - eher im oberen Bereich der Skala gegen eher im unteren Bereich der Skala - deutlich voneinander unterscheiden können. In der Literatur finden sich zwei Artikel, die konkret eine Normalisierung von Hörerurteilen anwenden. In anderen Disziplinen, wie beispielsweise der Psychologie, werden Daten häufig z-transformiert, um sie hinsichtlich bestimmter statistischer Testverfahren vergleichbar zu machen. Dieses Verfahren wird in keinem der beiden Artikel verwendet.

Liljencrants (1999) führt in Anlehnung an Fant und Kruckenberg (1989) eine Studie durch, bei der 15 Probanden eine 220 Silben umfassende Äußerung hinsichtlich der Silbenprominenz bewerten. Die Bewertungen der Probanden werden mittels einer Regression zwischen ihren Bewertungen und den durchschnittlichen Werten aller Probanden normalisiert.

Eriksson et al. (2001) bemerken, dass die Probanden ihrer Studie zum Teil die Skala nicht wie in den Anweisungen vorgeschrieben verwenden. Die Versuchsteilnehmer wurden angewiesen, dass sie jeweils einen Regler der kontinuierlichen Skalen in Minimalstellung belassen sollten und einen Regler auf die Maximalstellung stellen sollten. Die restlichen Regler sollten sie nach ihrem Gefühl zwischen den beiden Extremeinstellungen verteilen. Die Autoren der Studie haben alle Da-

tensätze, die nicht diesen Vorgaben entsprachen, linear in die gewünschte Form transformiert.

In Kapitel 5 soll der Frage nachgegangen werden, ob man die Urteile von Probanden normalisieren sollte. Hierbei stellt sich die Frage, ob sich die Ausschöpfung der Skalen verbessert und ob sich die Korrelationen zu verschiedenen Maßen verändern. Eine weitere Frage ist, ob die Effekte des Primings auch nach einer Normalisierung bestehen bleiben.

3. Experiment zur Erhebung von Prominenz anhand verschiedener Skalen

Teile dieser Studie wurden auf der ICPhS 2011 in Hong Kong präsentiert und in Arnold et al. (2011b) publiziert.

3.1. Einleitung

In Kapitel 2.3 wurde dargestellt, dass obwohl das Konzept Prominenz vielfältig erforscht wurde (siehe hierzu Kapitel 2.2), dem Messinstrument zur Erhebung von Prominenz wenig Beachtung geschenkt wurde. Dieses Kapitel beschreibt ein Experiment, dass der Frage nach der optimalen Skala zur Erhebung von Prominenz nachgeht. Hierbei sollen den widersprüchlichen Ergebnissen von Grover et al. (1997) und von Jensen und Tøndering (2005) neue Daten gegenübergestellt werden und der Frage nachgegangen werden, in wieweit die Ergebnisse von Arnold und Wagner (2008) und Arnold et al. (2010) mit anderen Skalen reproduzierbar sind.

3.2. Forschungsfragen

In ihrem Artikel haben Jensen und Tøndering (2005) versucht die Schwierigkeit einer Skala zu erfassen, indem sie die benötigte Zeit zur Beurteilung durch die Versuchsperson und die Wiederholungen der Wiedergabe des zu bewertenden Stimulus festgehalten haben. In der Studie von Arnold und Wagner (2008) und Arnold et al. (2010) lagen die Wiederholungen unter allen gefunden Werten für die in Jensen und Tøndering (2005) getesteten Skalen. Die kontinuierliche Skala wurde voll ausgeschöpft. Leider sind die Werte für die wiederholte Wiedergabe

der Stimuli zwischen den beiden Studien nicht miteinander vergleichbar, da die Versuchsteilnehmer in der Studie von Arnold und Wagner (2008) und Arnold et al. (2010) angewiesen wurden, sich die Stimuli möglichst nur einmal anzuhören. Jensen und Tøndering (2005) geben nicht an, wie sie Ihre Versuchsteilnehmer instruiert haben. Sie geben lediglich an, dass die Versuchsteilnehmer die Stimuli erneut anhören konnten. In dieser Studie soll durch vergleichbare Anweisungen herausgefunden werden, ob die kontinuierliche Skala ähnlich schwierig zu bedienen ist wie die diskrete oder ob sie einen Vor- oder Nachteil gegenüber den diskreten Skalen darstellt.

Wenn sich herausstellen sollte, dass kontinuierliche Skalen leichter in der Anwendung für die Versuchsteilnehmer sind, so stellt sich die Frage, ob die Ergebnisse, die mit den verschiedenen Skalen gemessen wurden, auch zu den gleichen Werten führen. Wenn dies nicht der Fall ist, so stellt sich die Frage, welche Skala zu einheitlicheren und somit „besseren"Werten führt.

Wie bereits beschrieben, gaben Jensen und Tøndering (2005) nicht an, wie sie Ihre Versuchsteilnehmer hinsichtlich der Wiederholungen instruiert haben. Durch drei verschiedene Instruktionen soll herausgefunden werden, wie sich die Instruktionen auf die Anzahl der Wiederholungen und den Zeitbedarf auswirken und ob eine der Bedingungen zu einheitlicheren Bewertungen durch die Probanden führt.

Als letztes stellt sich die Frage, ob alle Skalen die Bewertung von Silbenprominenz gut auflösen und repräsentieren, um beispielsweise den in Arnold und Wagner (2008) und Arnold et al. (2010) gezeigten Primingeffekt abbilden zu können. Hierbei kommt es zu einem signifikanten Unterschied in der Prominenzbeurteilung des selben Satzes durch zwei Gruppen, die gelernt haben, das syntaktische und semantische Muster mit einem jeweils anderen Prominenzmuster zu assoziieren. Es wäre ja durchaus möglich, dass beispielsweise die 4-Punkt-Skala keine Unterschiede in der Bewertung durch das Priming abbilden kann. Andererseits könnte die kontinuierliche Skala Probleme bereiten, da sie weniger Punkte liefert, an denen sich die Probanden bei der Beurteilung orientieren können.

Die letzte Frage, die an das Experiment gestellt wird, ist, wie die Probanden bei ihren Bewertungen vorgehen. Suchen sie sich zunächst die prominentesten Silben heraus und bewerten die übrigen Silben dann in Relation dazu oder bewerten sie die Silben vielleicht einfach der Reihenfolge nach? Wenn die Versuchsteilnehmer verschiedene Strategien verwenden: Führt eine Strategie zu besseren Ergebnissen und gibt es einen Zusammenhang zu Zeitaufwand und Wiederholungen?

3.3. Versuchsaufbau

Für den Versuch sollen vier Skalen und drei Akkuratheitsbedingungen untersucht werden. Bei den Skalen werden die diskreten 4-Punkt, 11-Punkt, 31-Punkt und eine kontinuierliche Skala getestet. Bei der kontinuierlichen Skala werden intern 301 Punkte verwendet, die bei der Auswertung auch auf die Auflösungen der diskreten Skalen abgebildet werden. Es werden drei verschiedene Anweisungen benutzt, die im folgenden mit *schnell*, *ohne* und *akkurat* bezeichnet werden. In der Bedingung *schnell* soll die Versuchsperson möglichst vermeiden, sich die Stimuli wiederholt anzuhören. Für Bedingung *ohne* werden keinerlei Anweisungen zum erneuten Anhören der Stimuli gegeben. Die Anweisungen für die Bedingung *akkurat* sollen die Versuchsteilnehmer dazu ermuntern, möglichst genau bei der Bewertung der Stimuli vorzugehen und sich diese öfter erneut anzuhören. Die genauen Anweisungen finden sich in Anhang A. Beim Audiomaterial ergeben sich zwei Bedingungen, um das Priming durchführen zu können. Um Versuchsteilnehmer zu sparen, bewertet jede Versuchsperson jeweils das gleiche Material mit zwei unterschiedlichen Skalen. Die Skalen werden hierbei alle miteinander in allen Kombinationen und Reihenfolgen kombiniert. Es gibt somit 12 Kombinationen. Hierdurch ergeben sich 12 * 3 * 2 = 72 Bedingungen. Für jede Bedingung werden 3 Versuchsteilnehmer erhoben, was zu 216 Versuchsteilnehmern führt. Hierbei wird jede Skala in jeder Akkuratheits- und Primingbedingung von 18 Versuchsteilnehmern benutzt. Eine schematische Darstellung des Versuchsdesigns zeigt Abbildung 3.3.1. Als Versuchsmaterial dienen 15 Sätze. Hierbei dienen vier Sätze als Priming für den Testsatz. Diese unterscheiden sich für die beiden Primingbedingungen. Die übrigen 10 Sätze sind für alle Gruppen identisch. Somit werden diese Sätze von 2*18 = 36 Versuchsteilnehmern pro Skala und Akkuratheitsbedingung bewertet.

Der Versuch wurde am Computer mittels einer eigens dafür entwickelten Software durchführt. Die Probanden bekamen die Stimuli über Kopfhörer vorgespielt und bewerteten die Silben mittels grafischer Schieberegler am Bildschirm. Näheres zur Software findet sich in Kapitel 3.6.

Die Versuchsteilnehmer wurden randomisiert den einzelnen Gruppen zugewiesen. Hierbei wurde mit Hilfe des Pakets Random für R eine zufällige Reihenfolge für die 3*72 Gruppen generiert. Die Probanden wurden dann nach Reihenfolge ihrer Erhebung der entsprechenden Gruppe zugeteilt. Als Erstes wurden ihr

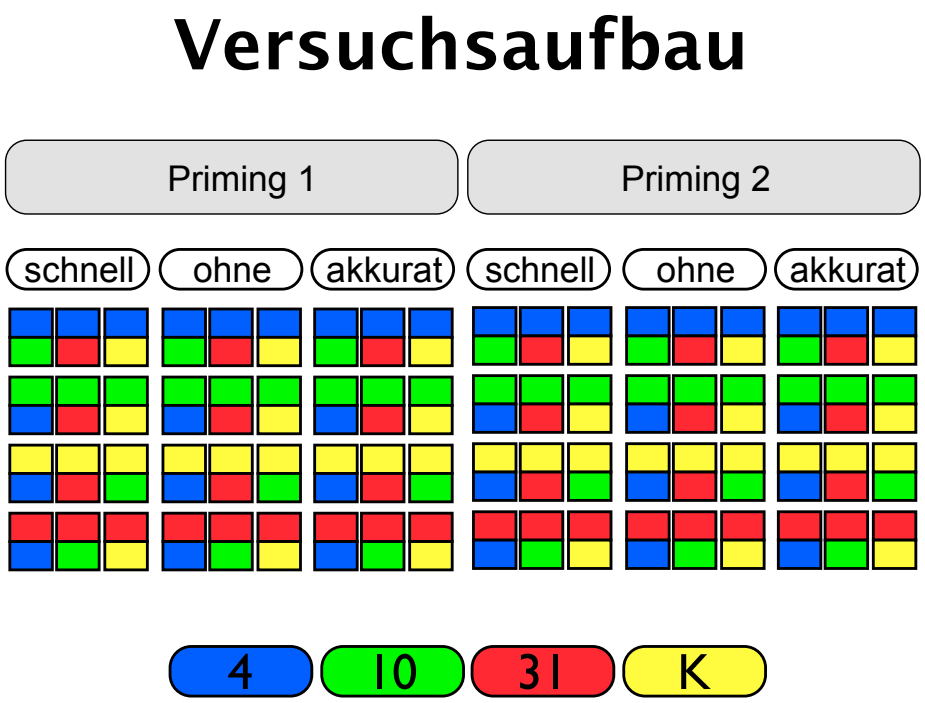


Abbildung 3.3.1.: Schematische Darstellung der 72 Versuchsbedingungen. Die Farbkodierungen stehen für die 4-Punkt-Skala (blau), die 11-Punkt-Skala (grün), die 31-Punkt-Skala (rot) und die kontinuierliche Skala (gelb). Diese lassen sich zwölf mal miteinander kombinieren. Es gibt zwei Primingbedingungen, in denen jeweils drei Akkuratheitsbedingungen mit den jeweils zwölf Skalaenkombinationen getestet werden. Dies führt zu insgesamt 72 Versuchsbedingungen.

Geschlecht, Alter und ihre Muttersprache abgefragt. Die Frage nach der Muttersprache sollte versehentlich erhobene Nichtmuttersprachler identifizieren und es ermöglichen, ihre Daten zu löschen und eine neue Versuchsperson in der gleichen Gruppe zu erheben. Hierzu wurde jeder Versuchsperson ein eindeutiger Code zugewiesen. Nach der Abfrage der Daten erhielten die Versuchsteilnehmer die auf die erste Skala zugeschnittene Anweisung auf dem Bildschirm. Nachdem die Versuchsperson bestätigt hatte, dass sie die Anweisungen verstanden hat, wurde der erste Block durchgeführt. Die Reihenfolge der Priming- und restlichen Sätze wurde für jede Versuchsperson und jede Bedingung neu randomisiert, um Reihenfolgeneffekte abzufangen. Als Letztes wurde immer der Testsatz für das Priming präsentiert. Nach dem ersten Durchgang wurden die Anweisungen für den zweiten Durchgang präsentiert. Für den zweiten Durchgang wurde exakt

das gleiche Material wie im ersten Durchgang mit einer anderen Skala bewertet. Auch der Anweisungsteil, der die Akkuratheit beeinflussen soll, war der gleiche wie im ersten Durchgang. Die Reihenfolge der Stimuli wurde neu randomisiert. Der letzte Satz war wieder der Testsatz für das Priming. Alle Daten wurden in eine MySQL-Datendank gespeichert. Deren Aufbau wird im Unterkapitel 3.6 Software beschrieben.

3.4. Material

In diesem Experiment soll unter anderem herausgefunden werden, ob alle getesteten Skalen den in Arnold und Wagner (2008) und Arnold et al. (2010) gezeigten Primingeffekt darstellen können. Hierzu wurde das beste Testset aus der Studie ausgewählt, um den Effekt zu replizieren. Im Folgenden sind der Testsatz und die zugehörigen Primingsätze für beide Bedingungen aufgeführt. Darstellung im **Fettdruck** soll eine kontrastive Betonung der Silbe darstellen.
Testsatz: Die **jun**ge Frau geht in das rote Haus.
Primingsatz 1: Der alte Mann stieg in den vollen Bus.
Primingsatz 2: Das kleine Kind ging in das kleine Haus.
Primingsatz 3: Die alte Frau steigt in den leeren Bus.
Primingsatz 4: Der junge Mann geht in das gelbe Haus.
Testsatz: Die **jun**ge Frau geht in das rote Haus.
Primingsatz 1: Der **al**te Mann stieg in den vollen Bus.
Primingsatz 2: Das **klei**ne Kind ging in das kleine Haus.
Primingsatz 3: Die **al**te Frau steigt in den leeren Bus.
Primingsatz 4: Der **jun**ge Mann geht in das gelbe Haus.

Die übrigen zehn Stimuli sind jeweils fünf Sätze beziehungsweise Phrasen mit einem Default Pattern und fünf Sätze beziehungsweise Phrasen, bei der eine Silbe - beispielsweise durch kontrastiven Fokus - hervorgehoben wurde. Die Stimuli wurden systematisch bezüglich ihrer Länge variiert. Hierbei wurden Sätze von drei bis zehn Silben benutzt. Eine vollständige Darstellung des Versuchsmaterials findet sich in Anhang B.

Alle Stimuli wurden im Studio der Abteilung für Sprache und Kommunikation in der alten Sternwarte in Bonn durchgeführt. Alle Stimuli wurden von der gleichen Sprecherin gesprochen und bei der Aufnahme auf die gewünschte Rea-

lisierung hin überprüft. Die Aufnahme erfolgte mit einem Neuman KM-100 mit einer AK 40 Kapsel, einem John Hardy M1 Vorverstärker, einem MOTU Travler Interface in Adobe Audition. Die Aufnahmen wurden im PCM-Format mono, mit einer Abtastrate von 44,1 kHz und einer Quantisierung von 16 bit aufgenommen. Anschließend wurden die Aufnahmen mit Hilfe der Analysesoftware Praat kontrolliert und die Teststimuli geschnitten. Die Stimuli wurden mit dem Programm textitnormalize in ihrer Lautstärke angeglichen.

Getestet wurden vier verschiedene Skalen. Es wurden drei diskrete Skalen und eine kontinuierliche Skala getestet. Hierbei gab es eine 4-Punkt-Skala, eine 11-Punkt-Skala, eine 31-Punkt-Skala und die kontinuierliche Skala. Diese wurde an Ihren Extremwerten mit den Labels „min“ und „max“ beschriftet. Abbildung 3.4.1 zeigt die Darstellung der verschiedenen Skalen im Versuch. Die technische Umsetzung der Skalen in der Software wird in Kapitel 3.6 beschrieben.

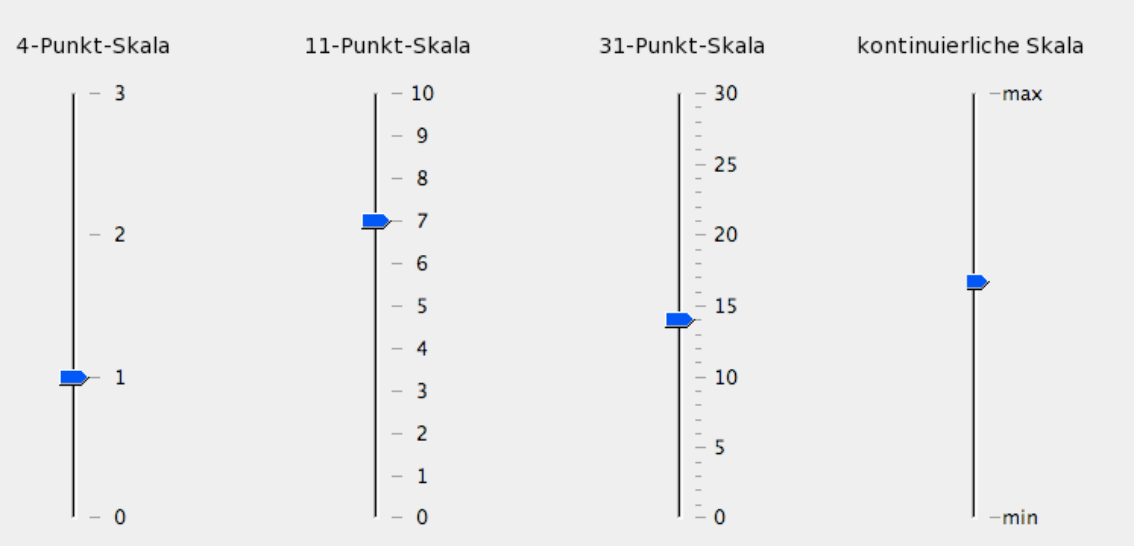


Abbildung 3.4.1.: Die vier verschiedenen Regler, die für die entsprechenden Skalen benutzt wurden, nebeneinander. Bei der 4-Punkt-Skala und der 11-Punkt-Skala sind alle Stufen mit dem zugehörigen numerischen Wert gekennzeichnet. Bei der 31-Punkt-Skala sind aus Platzmangel nur die Fünferschritte mit den numerischen Werten beschriftet, die übrigen Werte erhalten einfache Striche. Die kontinuierliche Skala trägt nur die Label „min“ und „max“ an den Enden der Skala.

3.5. Akustische Analyse der Stimuli

Um bei der späteren Auswertung Zusammenhänge zwischen den Prominenzbeurteilungen und akustischen Parametern bestimmen zu können, wurden die Stimuli einer akustischen Analyse unterzogen. Hierbei wurde die Länge der Silbendauern manuell gemessen. Weiterhin wurden die Intensität und die Grundfrequenz aller Silben bestimmt. Für das Labeln der Audiodateien kam Praat (Boersma und Weenink (2010)) zum Einsatz. Die Extraktion der Grundfrequenzmaxima und der durchschnittlichen Intensität erfolgte mit Hilfe von Praatskripten anhand der handgemessenen Labeldateien.

3.6. Software

Die Software zur Durchführung wurde in Java geschrieben und baut auf der in Arnold et al. (2010) verwendeten Software auf. Die Daten werden dabei über eine Java-Swing basierte grafische Benutzeroberfläche (Graphical User Interface,GUI) erhoben und in eine MySQL-Datenbank persistent gespeichert. Für die Bewertung der Silbenprominenz stehen in Anlehnung an Eriksson et al. (2001) und Eriksson et al. (2002) grafische Schieberegler auf dem Computerbildschirm bereit. Über den Schiebereglern findet sich eine orthographische Repräsentation der zu bewertenden Silbe. Bei Arnold und Wagner (2008) und Arnold et al. (2010) wurde der Schieberegler mit der Standard Java-Swing-Klasse JSlider() implementiert. Neben den Reglern wurde eine grafische Einteilung mittels .showTicks() erzeugt, die jedoch ohne Zahlen auskam und lediglich die Beurteilung der Reglerstellung erleichtern sollte. Nachteilig zeigte sich, dass in der Standardimplementierung von JSlider() eine Initialisierung eines Wertes voraussetzt wird. Dies wiederum bedeutet zum Einen, dass die Versuchsteilnehmer nicht unvoreingenommen bewerten können, und zum Anderen, dass man nicht ohne Weiteres überprüfen kann, ob es sich bei einem Wert, der dem Initialisierungswert entspricht, um eine Bewertung mit diesem Wert oder eine nicht vorgenommene Bewertung handelt. Um dieses Problem ein wenig abzumildern, wurden die Regler auf null initialisiert, einem Wert, der im Regelfall von den Versuchsteilnehmer nicht vergeben werden sollte. Um die Probleme aus Arnold und Wagner (2008) und Arnold et al. (2010) zu umgehen, wurde für die aktuelle Versuchsreihe eine Reihe von Verbesserungen vorgenommen. Es wurde eine Klasse VSlider() geschrieben, die von der

Basis-Klasse JSlider() erbt. Hierbei wurde eine Variable eingeführt, die die Sichtbarkeit des Reglerknopfes regelt und anhand derer man abfragen kann, ob eine Silbe bereits bewertet wurde oder nicht. Für die Darstellung des Reglerknopfes wurde die Klasse BasicSilderUI() überschrieben. Abbildung 3.6.1 zeigt die GUI zur Bewertung der Stimuli. Jensen und Tøndering (2005) gaben die durchschnitt-

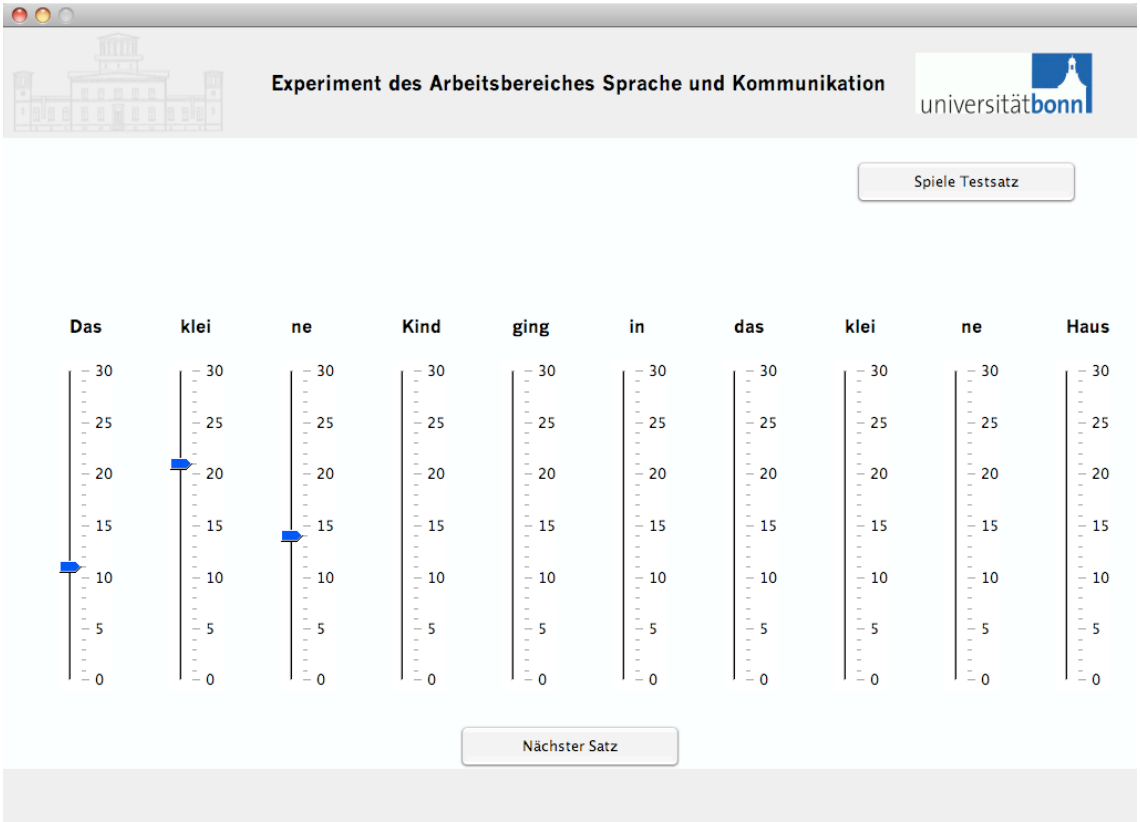


Abbildung 3.6.1.: Die grafische Oberfäche zur Bewertung der Stimuli. Hier abgebildet mit einer 31-Punkt-Skala. Der Testsatz lautet „Das kleine Kind ging in das kleine Haus“. Die ersten drei Silben sind bereits bewertet, die letzten sieben müssen noch bewertet werden. Oben rechts befindet sich ein Knopf, um den Sound erneut abzuspielen. Unten befindet sich der Knopf, um die Bewertung abzuschließen und mit dem nächsten Stimulus fortzufahren.

lich benötigte Zeit und die Anzahl der Wiederholungen als Maß der Schwierigkeit für die Versuchsteilnehmer an. Um Vergleichswerte zu erhalten, wurde ein Zähler für die Wiederholungen und eine Zeitmessung mit der Hilfe der Funktion System.nanoTime() implementiert, die die Zeit für die Bearbeitung eines Satzes misst.

Die Regler für die Bewertung werden alle mit der Klasse VSlider() erzeugt. Die Regler weisen in allen Fällen die gleichen Dimensionen auf, nur der Wertebereich

wird mit den jeweiligen Werten initialisiert. Die kontinuierliche Skala arbeitet hierbei mit 301 Punkten. Dies ist sinnvoll, da der Regler eine vertikale Größe von 301 Pixeln hat und somit jedem Pixel ein Wert entspricht. Die Label werden mit Hilfe von Hashtables erzeugt. Die 4-Punkt-Skala erhält vier Werte, die 11-Punkt-Skala erhält elf Werte. Die Werte werden zusätzlich mit Strichen versehen. Diese werden mit .setMajorTickSpacing(1) und .setPaintTicks(true) erzeugt. Bei der 31-Punkt-Skala werden nur die Fünferschritte beschriftet. Diese erhalten ein Label mit dem numerischen Wert und ein MajorTick. Die restlichen Schritte werden mit MinorTicks versehen. Die kontinuierliche Skala erhält nur zwei Markierungen. Am unteren Ende steht das Label „min“ und am oberen Ende das Label „max“. Dazwischen wird auf Markierungen verzichtet.

Für jede Versuchsperson wird ein eigener, achtstelliger alphanumerischer Code generiert. Dieser dient dazu, die Daten zuzuweisen. Für jede Versuchsperson wird die Reihenfolge der Stimuli eigens randomisiert. Hierbei wird für beide Durchgänge eine eigene Reihenfolge generiert. Alle Daten werden in eine MySQL-Datenbank gespeichert. Diese enthält vier Tabellen. In der Tabelle VpDaten werden Alter, Geschlecht und Muttersprache, Bedingung und der Code gespeichert. Die Muttersprache dient nur dazu, fälschlich erhobene Nichtmuttersprachler ausfindig zu machen. In der Tabelle ExpDaten werden die Bewertungen der einzelnen Silben, die Zahl der Wiederholungen des Stimulus, die benötigte Zeit, die Satzreferenz, die Anzahl der Silben des Stimulus, die Bedingung, die verwendete Skala, die Reihenfolge des Blockes, die Reihenfolge des Stimulus innerhalb des Bockes, sowie der Code der Versuchsperson gespeichert. Die Tabelle RF ist genauso aufgebaut wie die ExpDaten Tabelle. Hier werden alle Reglerbewegungen der Versuchsperson während der Bewertung gespeichert. In der Tabelle RF1 wird gespeichert an, mit welcher Reihenfolge die Regler das erste mal bewegt werden.

3.7. Durchführung

Alle Versuchsteilnehmer wurden an einem mit Kopfhörern ausgestatteten Computer erhoben. Die Erhebungen fanden im Computer-Pool des Arbeitsbereiches für Sprache und Kommunikation, in einem Computer-Pool der Leibnitz Universität Hannover, sowie in ruhigen Büroräumen statt. Die Versuchsteilnehmer rekrutierten sich hauptsächlich aus Studierenden und Doktoranden der Rheinischen Friedrich-Wilhelms Universität Bonn, der Heinrich Heine Universität Düsseldorf,

der Universität zu Köln, der Leibnitz Universität Hannover und deren Bekannten. Alle Versuchsteilnehmer nahmen freiwillig an dem Versuch teil und erhielten keine Bezahlung.

Aufgrund der verschiedenen technischen Gegebenheiten an den verschiedenen Orten stellte sich die Implementierung in Java als vorteilhaft heraus. Die Software lief sowohl unter SuSE, als auch Windows und Mac OS X. In jedem Fall wurde die jeweilige Hardware/Software-Kombination vor Ort eingehend getestet, um sicherzustellen, dass alle Sounds richtig abgespielt wurden und die Daten auf dem MySQL-Server ankamen. Die meisten Daten wurden auf einem zentralen Server gespeichert, nur die Daten der einzeln am MacBook erhobenen Probanden wurde in eine eigene Datendank geschrieben.

Nach jedem Versuch wurden die Datenbank auf Vollständigkeit der Daten geprüft. Bei Erhebung ganzer Kurse wurde zudem geprüft, ob versehentlich Nichtmuttersprachler mit erhoben wurden. Wenn Nichtmuttersprachler erhoben wurden, wurden die Daten umgehend aus der Datenbank entfernt und die dem Probanden zugeteilte Versuchsbedingung der nächsten Versuchsperson zugewiesen.

3.8. Ergebnisse

Die Auswertung aller Daten erfolgte mit Hilfe der Statistiksoftware R (R Development Core Team, 2010). Alle statistischen Tests und alle grafischen Darstellungen dieses Kapitels wurden mit R berechnet beziehungsweise erzeugt. Für das Auslesen der Daten aus der MySQL-Datenbank wurde ein R-Skript geschrieben, welches mit Hilfe der Pakete DBI (R Special Interest Group on Databases, 2009) und RMySQL (James und DebRoy, 2010) die Daten direkt aus der Datenbank in den Arbeitsbereich von R importierte.

3.8.1. Ausnutzung der Skalen

Jensen und Tøndering (2005) kommen zu dem Schluss, dass naive Hörer die Auflösung der 31-Punkt-Skala nicht vollständig ausnutzen können. Im Folgenden werden alle Bewertungen, die mit einer der vier Skalen abgegeben worden, betrachtet. Es wird also über alle Akkuratheits- und Priminggruppen hinweg kumuliert.

In Abbildung 3.8.1 ist die Anzahl der Bewertungen pro Skalenrang von der 4-

Punkt-Skala abgebildet. Bei der 4-Punkt-Skala wurde der Wert „1“ am häufigsten vergeben, gefolgt vom Wert „2“. Insgesamt werden mehr Urteile mit „0“ und „1“ beurteilt als mit „2“ und „3“. Die Verteilung ist deutlich linksschief, ein Shapiro-Wilk Test ist nicht signifikant (W = 0.88, p = 0.37). Somit kann eine Normalverteilung angenommen werden. Bei der 11-Punkt-Skala (3.8.2) ist der Shapiro-Wilk Test ebenfalls nicht signifikant (W = 0.91, p= 0.24). Die Verteilung ist im Gegensatz zur 4-Punkt-Skala rechtsschief. Insgesamt sind mehr Urteile auf der 11-Punkt-Skala in der oberen Hälfte der Skala, im Gegensatz zur 4-Punkt-Skala. Man kann also sagen, dass für die vorliegende Stichprobe Bewertungen auf der 11-Punkt-Skala mehr in Richtung prominent tendieren als auf der 4-Punkt-Skala.

Während bei der 4-Punkt-Skala und der 11-Punkt-Skala jeder Schritt an dem Regler mit einer Zahl beschriftet war, wurde bei der 31-Punkt-Skala aus Platzgründen nur jeder fünfte Schritt beschriftet. Wie man auf Abbildung 3.8.3 sehr gut sehen kann, wurden diese Anker von den Probanden bei ihrer Bewertung deutlich häufiger benutzt, als die Zwischenschritte, die nur mit Linien am Regler markiert waren. Wie bei der 11-Punkt-Skala wurden deutlich mehr Urteile über dem Mittelwert der Skala vergeben als darunter. Somit tendieren auch die Beurteilungen der 31-Punkt-Skala mehr in Richtung prominent als bei der 4-Punkt-Skala.

Bei der kontinuierlichen Skala wurde intern mit den Werten von 0 - 300 gerechnet. An den beiden Enden der Skala befanden sich die Label „min“ und „max“. Wie in Abbildung 3.8.4 zu sehen ist, wurde der Wertebereich der Skala gut ausgeschöpft. Es wurde jedes Pixel des Reglers verwendet. Die beiden Extremwerte bildeten, ähnlich wir die beschrifteten Schritte der 31-Punkt-Skala, Ankerwerte. Diese wurden von den Probanden offensichtlich bevorzugt ausgewählt. Was auffällt ist, dass auf der Seite des Extremums „prominent“ die Anzahl der Bewertungen zum Ende der Skala deutlich ansteigt, während dies auf der Seite des Extremums „Nicht prominent“ nicht der Fall ist. Hier fällt die Anzahl der Bewertungen in Richtung des Extremwertes deutlich ab.

Insgesamt zeigen sich deutliche Unterschiede in den Verteilungen der einzelnen Skalen. Sowohl bei der 4-Punkt-Skala als auch der 11-Punkt-Skala gibt es weitgehend normalverteilte Ratings, während die Probanden bei der 31-Punkt-Skala und der kontinuierlichen Skala deutlich häufiger zu den Werten mit einem Anker tendieren. Die 4-Punkt-Skala zeigt als einzige der getesten Skalen eine linksschiefe, während die übrigen Skalen rechtsschiefe Verteilungen aufweisen. Alle Skalen

werden von den Probanden gut ausgeschöpft.

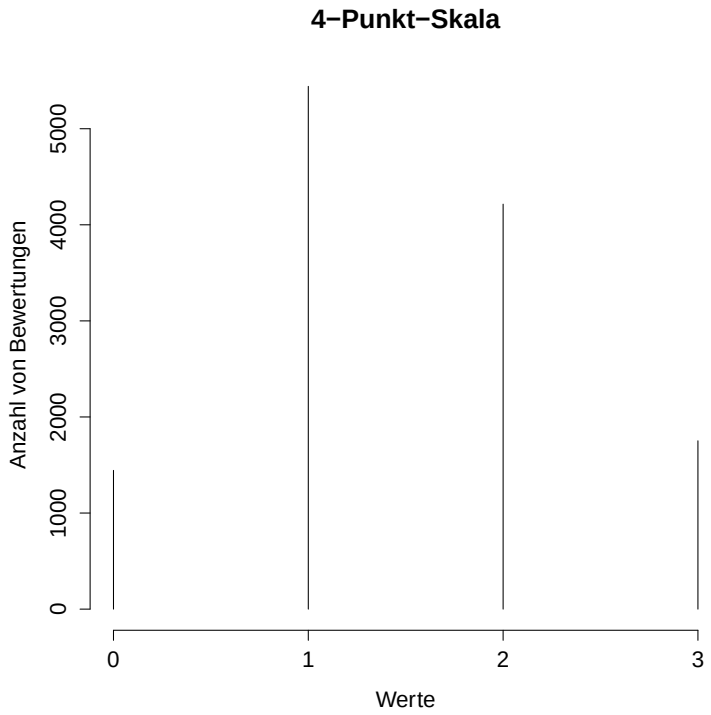


Abbildung 3.8.1.: *Ausnutzung der 4-Punkt-Skala.*

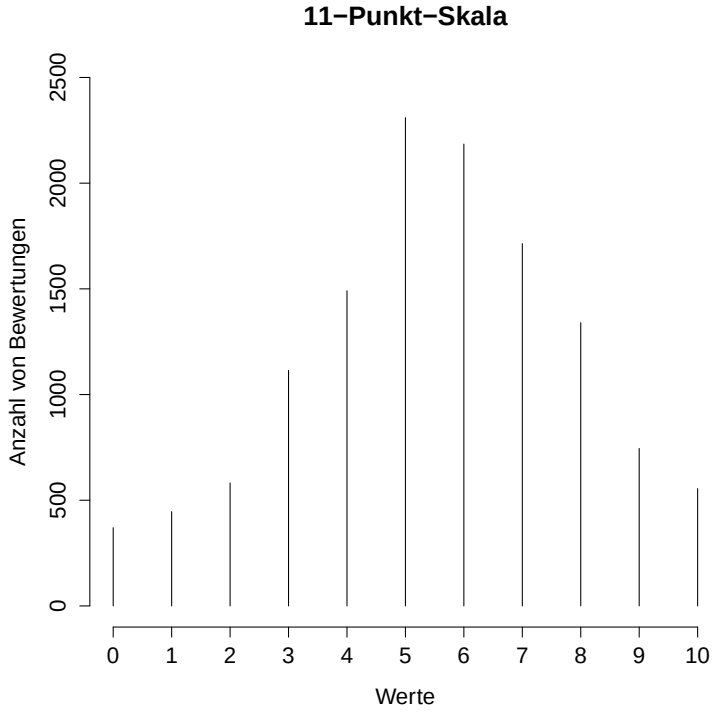


Abbildung 3.8.2.: *Ausnutzung der 11-Punkt-Skala.*

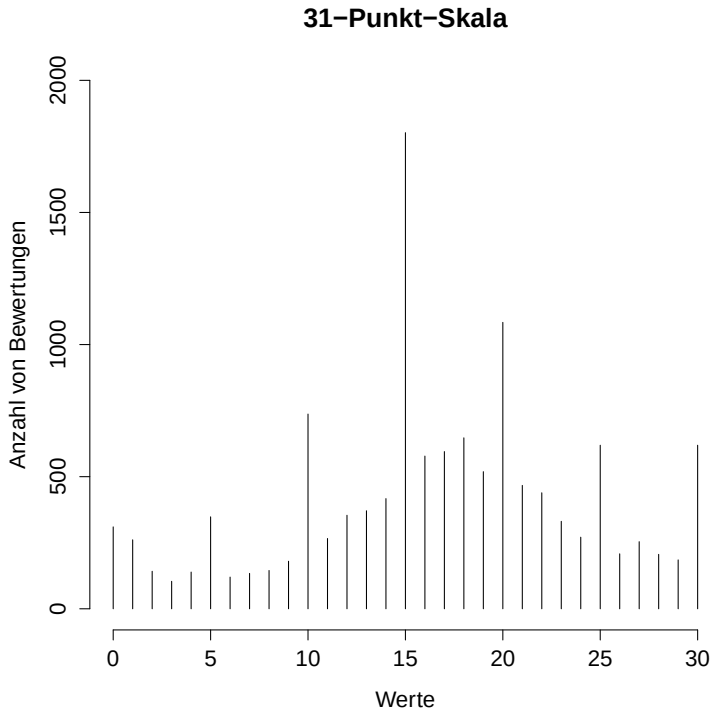


Abbildung 3.8.3.: *Ausnutzung der 31-Punkt-Skala.*

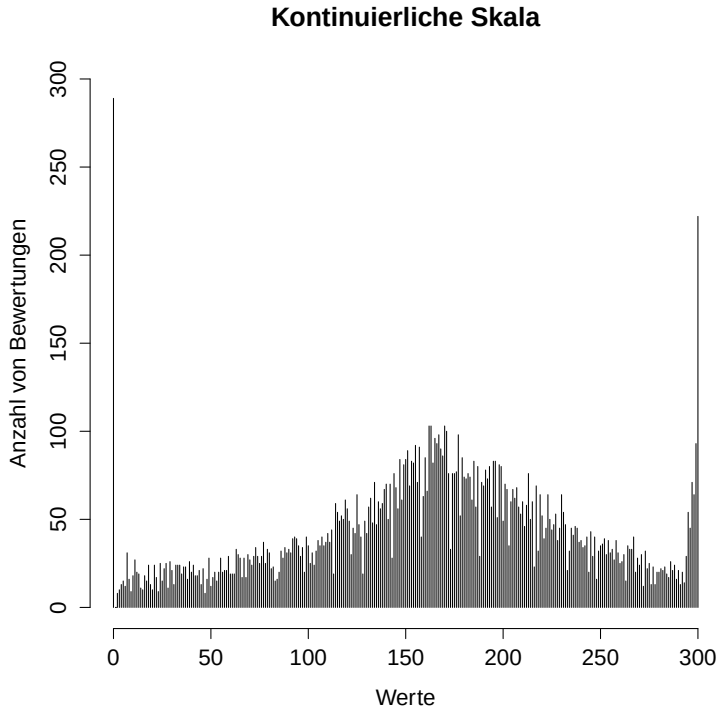


Abbildung 3.8.4.: *Ausnutzung der Kontinuierlichen Skala.*

3.8.2. Extremwerte und Verteilungen

Jensen und Tøndering (2005) beschrieben in ihrer Studie den Effekt, dass bei den Skalen mit mehreren Stufen weniger extreme Werte gefunden wurden als bei den Skalen mit weniger Stufen. Es ist also zu erwarten, dass in dieser Studie mit der 4-Punkt-Skala die extremsten Werte gefunden werden, gefolgt von der 11-Punkt und 31-Punkt-Skala. Bei der kontinuierlichen Skala hängt es vermutlich stark von der internen Unterteilung der Probanden ab, ob sie sich die Skala z.B in 4, 10 oder 100 Teilstücke denken.

Ein Blick in Tabelle 3.8.1 zeigt, dass sich die Erwartungen, die sich aus der Studie von Jensen und Tøndering (2005) ableiten, nicht bestätigen. Wir finden keine abnehmende Spanne bei zunehmender Zahl der Stufen der Skalen. Auffällig hierbei ist, dass die durchschnittlichen Bewertungen die Skala nach oben fast voll ausschöpfen, während das jeweils untere Drittel der Skalen von den Probanden kaum genutzt wird. Es findet sich kein systematischer Einfluss der Ausprägungen von Akkuratheitsbedingung und Primingbedingung auf die Extremwerte oder auf den genutzten Teil der Skala.

Wir sehen für alle Gruppen, die die 4-Punkt-Skala benutzt haben, die gleichen Extremwerte und auch die gleiche Spanne. Im Mittel finden sich hier tatsächlich die größte Spanne und die extremsten Werte. Bei den Gruppen, welche die 11-Punkt-Skala benutzt haben, ist das Bild bereits uneinheitlicher. Drei Gruppen verwenden das Maximum der Skala, während drei Gruppen darunter bleiben. Die Gruppe mit der Akkuratheitsbedingung 0 und der Primingbedingung 1 hat die größte Spanne und mit einem umgerechneten Wert von 0.7 eine leicht größere Spanne als die Gruppen, welche die 4-Punkt-Skala verwendet haben.

Im Durchschnitt ist die Spanne der Gruppen, die mit der 31-Punkt-Skala gearbeitet haben größer, als die der Gruppen, welche mit der 11-Punkt-Skala gearbeitet haben. Es findet sich auch hier eine Gruppe, deren Spanne, die der Gruppen, welche die 4-Punkt-Skala verwendet haben übertrifft. Die Gruppen, die mit der kontinuierlichen Skala gearbeitet haben, haben im Schnitt eine größere Spanne als die Gruppen mit der 11-Punkt-Skala und 31-Punkt-Skala. Auch hier findet sich eine Gruppe, die die Werte der Gruppen mit den 4-Punkt Skalen übertreffen. Auch wenn sich keine Gruppe findet, bei der der maximale Median gleich dem Maximum der Skala ist, ist doch für alle Gruppen das Maximum nahe 1.

In den Abbildungen 3.8.5 - 3.8.9 sieht man beispielhaft die Bewertung von Satz

R8 mit den vier verschiedenen Skalen. Was beim Blick auf die Mittelwerte der einzelnen Ratings (Abbildung 3.8.5) verborgen bleibt, sind die Verteilungen, aus denen sich jedes Rating einer Silbe ergibt. In den folgenden Abbildungen sehen wir, dass die Verteilungen bei Skalen mit mehr Stufen deutlich variabler sind als bei der 4-Punkt-Skala. Diese Unterschiede zeigen, dass die Probanden bei unterschiedlichen Silben sehr stark in der Sicherheit ihres Urteils schwanken.

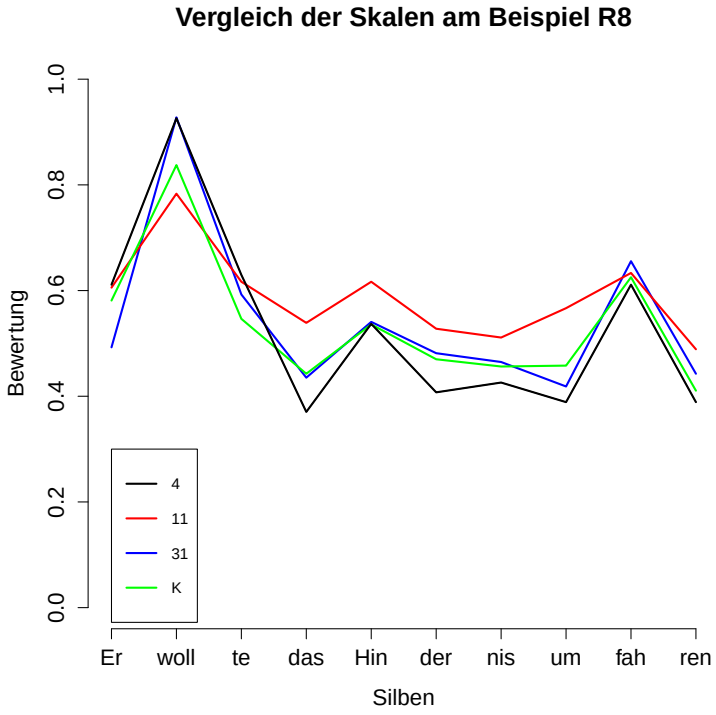


Abbildung 3.8.5.: *Bewertung mit den vier Skalen.*

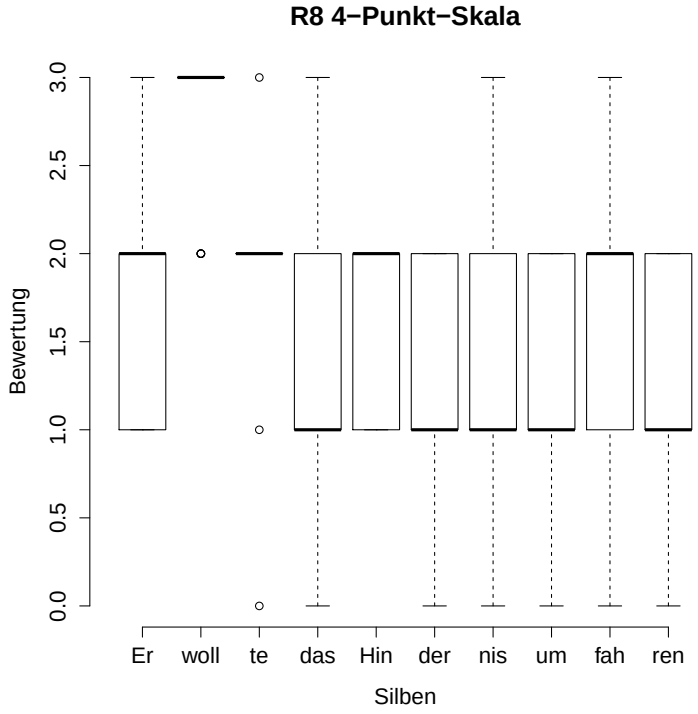


Abbildung 3.8.6.: *Bewertung mit der 4-Punkt-Skala.*

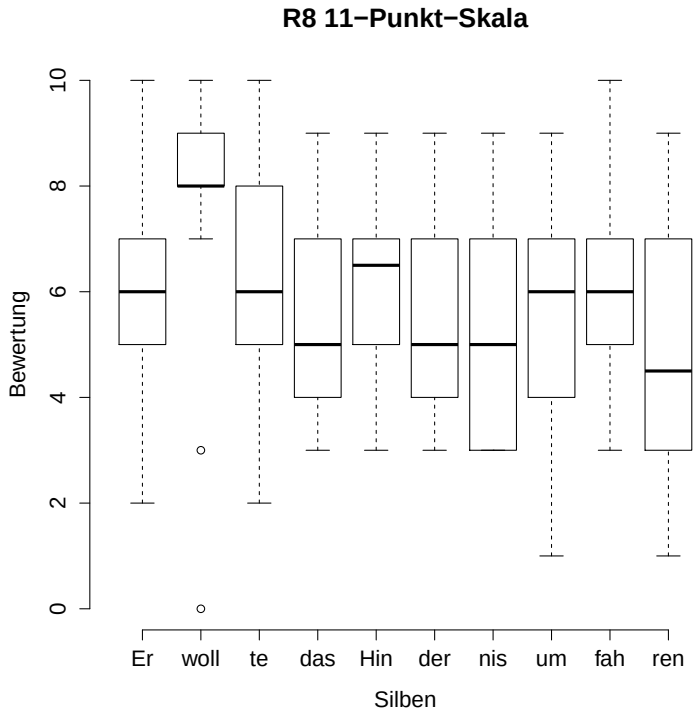


Abbildung 3.8.7.: *Bewertung mit der 11-Punkt-Skala.*

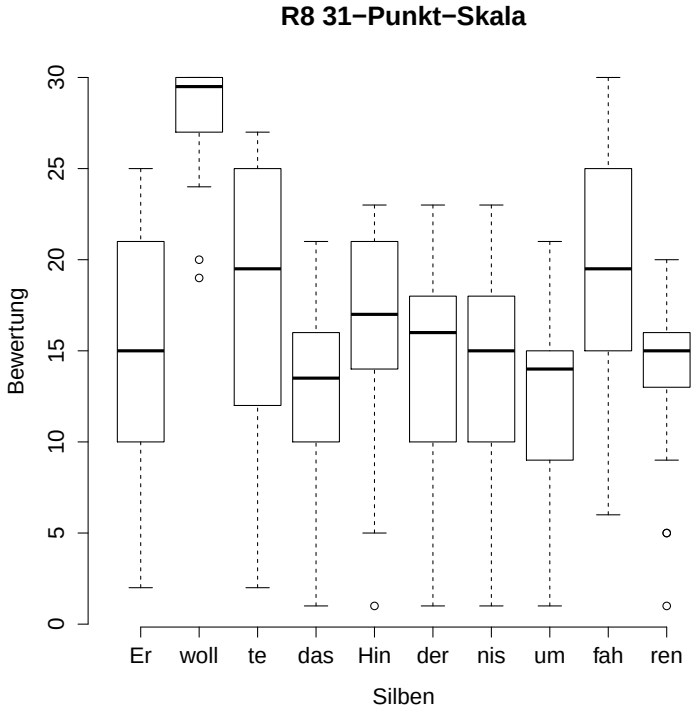


Abbildung 3.8.8.: *Bewertung mit der 31-Punkt-Skala.*

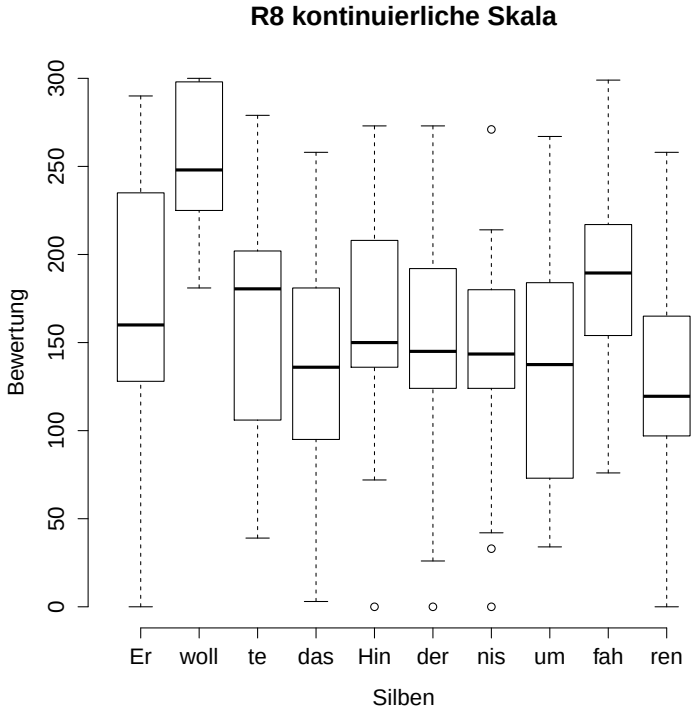


Abbildung 3.8.9.: *Bewertung mit der kontinuierlichen Skala.*

Tabelle 3.8.1.: *Maxima und Minima der durchschnittlichen Bewertungen des Materials mit den entsprechenden Skalen nach Transformation auf das Intervall [0,1], sowie die daraus resultierende Spanne. In der ersten Spalte sind die Skala mit 4-Punkt, 11-Punkt, 31-Punkt und kontinuierlicher Skala, die Akkuratheitsbedingung mit 0 für so schnell wie möglich, 1 für keine Vorgaben und 2 für möglichst akkurate Bearbeitungen und die Priminggruppen kodiert.*

Skala und Bedingungen	Maximum	Minimum	Differenz
Skala 4 Akk 0 Prim 0	1	0.33	0.66
Skala 4 Akk 1 Prim 0	1	0.33	0.66
Skala 4 Akk 2 Prim 0	1	0.33	0.66
Skala 4 Akk 0 Prim 1	1	0.33	0.66
Skala 4 Akk 1 Prim 1	1	0.33	0.66
Skala 4 Akk 2 Prim 1	1	0.33	0.66
Durchschnitt	1	0.33	0.66
Skala 11 Akk 0 Prim 0	0.9	0.3	0.6
Skala 11 Akk 1 Prim 0	0.9	0.45	0.45
Skala 11 Akk 2 Prim 0	1	0.45	0.55
Skala 11 Akk 0 Prim 1	1	0.3	0.70
Skala 11 Akk 1 Prim 1	1	0.4	0.60
Skala 11 Akk 2 Prim 1	0.95	0.4	0.55
Durchschnitt	0.95	0.38	0.57
Skala 31 Akk 0 Prim 0	1	0.36	0.63
Skala 31 Akk 1 Prim 0	1	0.32	0.68
Skala 31 Akk 2 Prim 0	1	0.47	0.53
Skala 31 Akk 0 Prim 1	0.97	0.45	0.51
Skala 31 Akk 1 Prim 1	0.98	0.36	0.61
Skala 31 Akk 2 Prim 1	0.95	0.43	0.51
Durchschnitt	0.98	0.4	0.58
Skala Kont Akk 0 Prim 0	0.94	0.44	0.50
Skala Kont Akk 1 Prim 0	0.98	0.39	0.59
Skala Kont Akk 2 Prim 0	0.99	0.39	0.61
Skala Kont Akk 0 Prim 1	0.95	0.37	0.57
Skala Kont Akk 1 Prim 1	0.96	0.20	0.76
Skala Kont Akk 2 Prim 1	0.96	0.30	0.66
Durchschnitt	0.97	0.35	0.62

3.8.3. Bearbeitungszeit und Anzahl der Wiederholungen als ein Maß für die Schwierigkeit der Beurteilungsaufgabe

Bei der Bewertung von Silbenprominenz durch naive Hörer sollte die Aufgabe für die Probanden möglichst leicht zu erledigen sein. Bei ihrer Studie argumentierten Jensen und Tøndering (2005), dass sich der Aufwand der Probanden in der Dauer, die der Proband für die Bewertung einer Phrase benötigt, und in der Häufigkeit, wie oft sich der Proband den zu bewertenden Stimulus erneut anhört, niederschlägt. Demnach wäre zu erwarten, dass Skalen, die für den Probanden einen höheren Aufwand in der Bewertung bedeuten, mehr Wiederholungen und mehr Bearbeitungszeit benötigen.

Tabelle 3.8.2 zeigt die durchschnittliche Bearbeitungszeit und die durchschnittliche Anzahl an Wiederholungen des Stimulus durch die Probanden. Hierbei sind die Skalen (4-Punkt, 11-Punkt, 31-Punkt und kontinuierliche Skala), die Manipulationsbedingung (Akk 0, Akk 1 und Akk 2) und die Priminggruppe (Prim 0 und Prim 1) angegeben. Durch die Instruktionen sollte die Anzahl der Wiederholungen der Stimuli durch den Probanden manipuliert werden. Hierbei sollten die Probanden in Bedingung 1 sich das Signal möglichst nur einmal anhören und nur, wenn es absolut erforderlich ist, ein zweites Mal. Bedingung 3 sollte sich das Signal möglichst mehrmals anhören, und bei Bedingung 2 wurde nur erwähnt, dass die Möglichkeit zum mehrmaligen Anhören der Aufnahmen besteht, ohne eine Anweisung darüber zu geben, wie oft sich die Probanden das Signal anhören sollten. Im Falle einer erfolgreichen Manipulation ist zu erwarten, dass die Anzahl der Wiederholungen bei Bedingung 3 signifikant über der Anzahl der Wiederholungen bei Bedingung 1 liegt, und dass die Anzahl der Wiederholungen bei Gruppe 2 zwischen den beiden anderen Werten liegt. Das Material der beiden Priminggruppen unterscheidet sich in vier von 15 Sätzen. Da die Sätze gleich viele Elemente aufweisen und bis auf die prosodische Realisation gleich sind, wird kein Einfluss auf die Bearbeitungsdauer und die Anzahl der Stimuluswiederholungen erwartet.

Wie wir in den Abbildungen 3.8.10 - 3.8.13 sehen können, kommt es in allen Bedingungen zu einer recht beträchtlichen Zahl an zum Teil sehr hohen Ausreißern. Auch bei den Wiederholungen gibt es zum Teil einzelne Bewertungen, bei denen die Anzahl der Wiederholungen zu den Ausreißern gezählt werden muss.

Wenn man Bedingung 1 und Bedingung 3 betrachtet, werden die Erwartungen hinsichtlich Bearbeitungsdauer und Stimuluswiederholungen deskriptiv bis

auf einen Fall erfüllt. Bei der kontinuierlichen Skala benötigt in Priminggruppe 1 die Gruppe mit Akkuratheitsbedingung 0 mehr Zeit als in Akkuratheitsbedingung 2, obwohl die Anzahl der Stimuluswiederholungen hypothesenkonform geringer ausfällt. Nicht alle Unterschiede fallen bei Überprüfung mittels t-Tests signifikant aus. Die Werte der Akkuratheitsbedingung 1 verhalten sich nicht hypothesengerecht. Die Bearbeitungsdauer liegen mal unter den Werten von Akkuratheitsbedingung 0 (z.B. 4-Punkt-Skala Priminggruppe 1), mal über den Werten von Akkuratheitsbedingung 2 (z.B. 31-Punkt-Skala Priminggruppe 0). Gleiches findet sich für die Stimuluswiederholungen. Hierbei verhalten sich die Anzahl der Stimuluswiederholungen und die Bearbeitungsdauer nicht immer gleich. Die Korrelation der beiden Ausprägungen korreliert mit Spearmans $\rho = 0.65$.

Wenn man die Skalen miteinander vergleicht, kommt man zu dem Ergebnis, dass sich weder für die Bearbeitungsdauer noch für die Anzahl der Stimuluswiederholungen eine eindeutige Reihenfolge ergibt. Sieht man die beiden Ausprägungen als Kosten, sind niedrigere Werte besser. Es finden sich Konstellationen von Priming- und Akkuratheitsgruppe, bei denen jeweils die 4-Punkt-Skala, die 11-Punkt-Skala und die kontinuierliche Skala die geringste durchschnittliche Bearbeitungsdauer haben. Bei den Stimuluswiederholungen finden sich Konstellationen, bei der jeweils eine der Skalen den niedrigsten Wert hat. Insgesamt lässt sich aus diesen Beobachtungen keine Präferenz für eine der vier Skalen ableiten.

Wie erwartet, zeigt die Zugehörigkeit zur Priminggruppe keinen systematischen Einfluss auf die Bearbeitungsdauer und die Anzahl der Stimuluswiederholungen. Unter gleichen Skalen und Akkuratheitsbedingungen sind die Werte mal in der einen, dann in der anderen Priminggruppe höher.

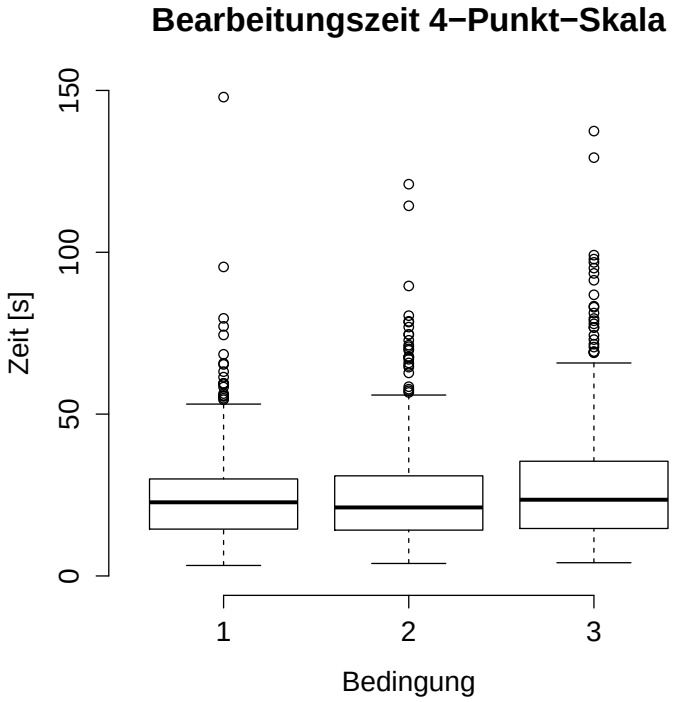


Abbildung 3.8.10.: *Bearbeitungszeit 4-Punkt-Skala.*

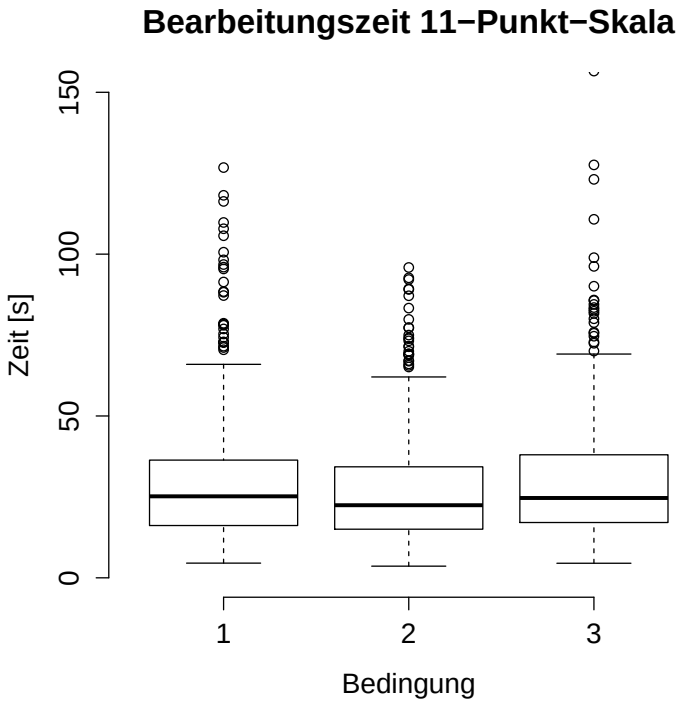


Abbildung 3.8.11.: *Bearbeitungszeit 11-Punkt-Skala.*

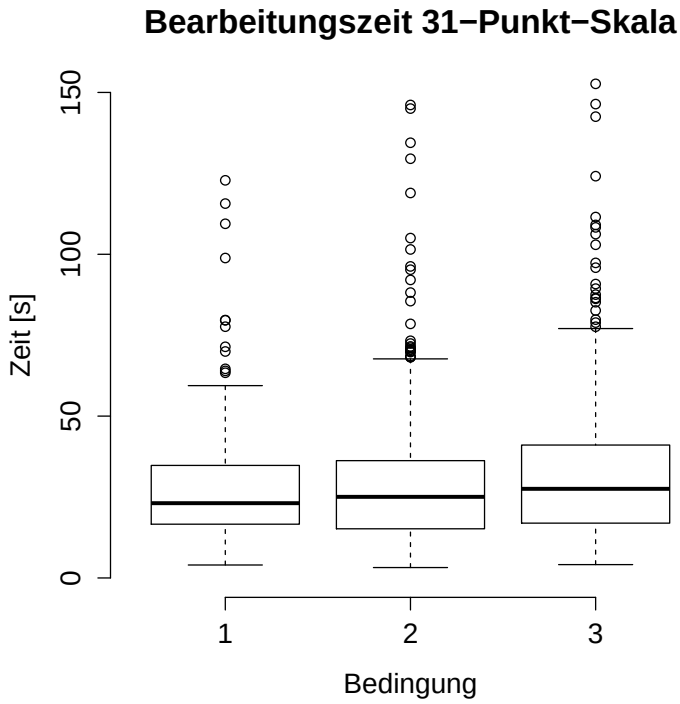


Abbildung 3.8.12.: *Bearbeitungszeit 31-Punkt-Skala.*

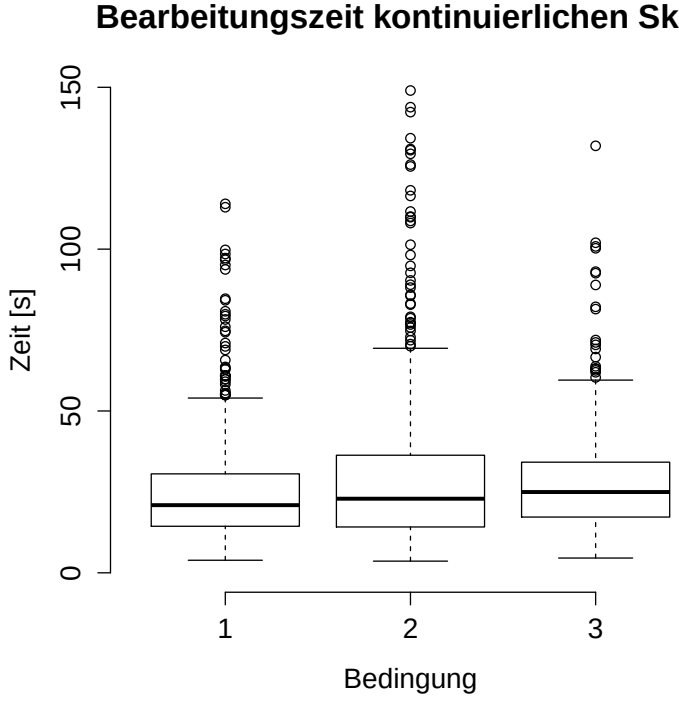


Abbildung 3.8.13.: *Bearbeitungszeit kontinuierliche Skala.*

Tabelle 3.8.2.: *Durchschnittliche Bewertungsdauer in Sekunden pro Satz und durchschnittliche Anzahl der Wiederholungen des Stimulus pro Satz durch den Probanden.*

Skala und Bedingungen	Bearbeitungsdauer (*SD*)	Stimuluswiedergaben (*SD*)
Skala 4 Akk 0 Prim 0	24.5 (15.1)	0.62 (1.4)
Skala 4 Akk 1 Prim 0	27.2 (16.4)	1.63 (2.2)
Skala 4 Akk 2 Prim 0	28.1 (18.6)	1.57 (1.9)
Skala 4 Akk 0 Prim 1	24.2 (12.9)	0.84 (1.4)
Skala 4 Akk 1 Prim 1	22.4 (14.6)	0.91 (1.5)
Skala 4 Akk 2 Prim 1	27.9 (18.6)	1.23 (1.6)
Skala 11 Akk 0 Prim 0	27.9 (20.3)	0.74 (1.2)
Skala 11 Akk 1 Prim 0	24.8 (15.4)	0.97 (1.3)
Skala 11 Akk 2 Prim 0	28.2 (21.7)	1.49 (2.1)
Skala 11 Akk 0 Prim 1	31.1 (18.4)	0.71 (1.3)
Skala 11 Akk 1 Prim 1	28.8 (18.6)	1.11 (1.6)
Skala 11 Akk 2 Prim 1	32.0 (20.1)	1.89 (2.7)
Skala 31 Akk 0 Prim 0	27.9 (15.4)	0.96 (1.4)
Skala 31 Akk 1 Prim 0	34.1 (29.7)	1.42 (2.2)
Skala 31 Akk 2 Prim 0	29.9 (20.2)	1.54 (2.0)
Skala 31 Akk 0 Prim 1	25.7 (14.9)	0.34 (0.7)
Skala 31 Akk 1 Prim 1	25.9 (17.5)	1.04 (1.7)
Skala 31 Akk 2 Prim 1	35.7 (25.8)	1.99 (3.0)
Skala Kont Akk 0 Prim 0	23.9 (14.9)	0.25 (0.6)
Skala Kont Akk 1 Prim 0	30.9 (24.7)	1.92 (2.8)
Skala Kont Akk 2 Prim 0	29.4 (18.2)	1.35 (1.9)
Skala Kont Akk 0 Prim 1	27.0 (19.2)	1.04 (1.8)
Skala Kont Akk 1 Prim 1	29.9 (26.2)	0.98 (1.5)
Skala Kont Akk 2 Prim 1	25.7 (12.7)	1.09 (1.4)

3.8.4. Interrater Reliabilität

Es ist wünschenswert, dass die Probanden beim Benutzen einer Skala zu möglichst ähnlichen, am besten natürlich den gleichen, Ergebnissen kommen. Um diese Eigenschaft der verschiedenen Skalen zu messen, betrachtet man die Interrater Reliabilität. Jensen und Tøndering (2005) benutzen in ihrer Studie Cronbachs α, um die Interrater Reliabilität der verschiedenen Skalen zu vergleichen. Da für das Priming fünf Sätze geurteilt wurden, die in den beiden Primingbedingungen unterschiedliche Bewertungen erzeugen sollten, wurde Cronbachs α auf den 69

Silben bestimmt, die von allen Probanden bewertet wurden. In Tabelle 3.8.3 ist die Interrater Reliabilität ausgedrückt in Cronbachs α für alle Skalen in den verschiedenen Priming- und Akkuratheitsbedingungen aufgeführt.

Die Werte liegen im guten bis akzeptablen Bereich. Besonders deutlich auffällig ist der Wert für die 11-Punkt-Skala in Priminggruppe 0 und Akkuratheitsgruppe 1. Dieser liegt deutlich unter allen übrigen Werten und kann als Ausreißer klassifiziert werden. Anscheinend ist hier eine besonders inhomogene Gruppe zusammen gekommen. Ein Zusammenhang zwischen der Interrater Reliabilität und der durchschnittlichen Bearbeitungszeit (r=0.050,p=0.81) oder der durchschnittlichen Stimuluswiederholung (r= 0.063, p=0.76) ist nicht festzustellen. Die Korrelationen sind äußerst gering und nicht signifikant.

Bezüglich einer Reihenfolge der Skalen ergibt ein ähnliches Bild, wie bei der Beurteilung hinsichtlich der Bearbeitungszeit. Obwohl die Werte für die 4-Punkt-Skala auf den ersten Blick am stärksten scheinen, ergeben sich jedoch auch Kombinationen von Priming- und Akkuratheitsbedingung, bei denen Cronbachs α bei der 11-Punkt-Skala und bei der kontinuierlichen Skala den höchsten Wert aufweist. Man kann aufgrund der gefundenen Werte also keiner Skala den Vorzug geben.

Tabelle 3.8.3.: *Die Interrater Reliabilität wurde auf den insgesamt 69 Silben bestimmt, die von allen Versuchsteilnehmer bewertet wurden. Als Maß dient Cronbachs α In der ersten Spalte sind die Skala mit 4-Punkt, 11-Punkt, 31-Punkt und kontinuierlicher Skala, die Akkuratheitsbedingung mit 0 für so schnell wie möglich, 1 für keine Vorgaben und 3 für möglichst akkurate Bearbeitungen und die Priminggruppen kodiert.*

Skala und Bedingungen	Interrater Reliabilität
Skala 4 Akk 0 Prim 0	0.936463
Skala 4 Akk 1 Prim 0	0.9226248
Skala 4 Akk 2 Prim 0	0.9255263
Skala 4 Akk 0 Prim 1	0.9285617
Skala 4 Akk 1 Prim 1	0.9020092
Skala 4 Akk 2 Prim 1	0.8969802
Skala 11 Akk 0 Prim 0	0.9257738
Skala 11 Akk 1 Prim 0	0.7744845
Skala 11 Akk 2 Prim 0	0.928246
Skala 11 Akk 0 Prim 1	0.9118242
Skala 11 Akk 1 Prim 1	0.890246
Skala 11 Akk 2 Prim 1	0.9313197
Skala 31 Akk 0 Prim 0	0.9186523
Skala 31 Akk 1 Prim 0	0.903531
Skala 31 Akk 2 Prim 0	0.9238671
Skala 31 Akk 0 Prim 1	0.8740874
Skala 31 Akk 1 Prim 1	0.9069991
Skala 31 Akk 2 Prim 1	0.858958
Skala Kont Akk 0 Prim 0	0.891102
Skala Kont Akk 1 Prim 0	0.8852171
Skala Kont Akk 2 Prim 0	0.9158729
Skala Kont Akk 0 Prim 1	0.8525094
Skala Kont Akk 1 Prim 1	0.9221993
Skala Kont Akk 2 Prim 1	0.9259735

3.8.5. Akustische Korrelate

Ein Hauptinteresse bei der Prominenzforschung sind die Zusammenhänge zwischen wahrgenommener Prominenz und den akustischen Parametern einer Äußerung. Jensen und Tøndering (2005) bestimmten in ihrer Studie die Korrelation zwischen der Grundfrequenz und der mittels der drei verschieden Skalen beurteilten Silbenprominenz. In der Literatur werden neben spektralen Eigenschaften des Signals vor allem Zusammenhänge zur Silbendauer, Intensität und der Grundfrequenz (f0) untersucht. Von daher ist es höchst interessant, ob mit allen vier Skalen gleich hohe Korrelationen zu den akustischen Parametern der Stimuli festgestellt werden.

In der Literatur werden anstatt einzelner Korrelationen auch lineare Modelle betrachtet z.B. Eriksson et al. (2002), bei denen verschiedene akustische Parameter als Prädiktoren für die wahrgenommene Silbenprominenz dienen. Aus dem Grund wurde zusätzlich zu den Korrelationen ein lineares Modell in R angepasst.

Für die akustische Analyse wurde Praat benutzt (Boersma und Weenink, 2010). Die einzelnen Stimuli wurden manuell auf Silbenebene annotiert. Die Silbendauern, Intensität und Maxima der Grundfrequenz wurden dann mit Hilfe von Praat-Skripten ausgelesen. In Tabelle 3.8.4 sieht man die Spearman Rangkorrelationen zwischen den Bewertungen der Silbenprominenz mittels der vier Skalen unter den verschiedenen Bedingungen und den Ausprägungen Silbendauer, Intensität und Grundfrequenz (f0). In der letzten Spalte der Tabelle sieht man die Modellpassung eines linearen Modells, bei dem die drei akustischen Parameter als Prädiktoren für die wahrgenommene Silbenprominenz benutzt wurden. Sämtliche Korrelationen fallen nicht sonderlich hoch aus. Insbesondere die Korrelationen zwischen Grundfrequenz und der wahrgenommen Silbenprominenz sind sehr gering und häufig auch nicht signifikant. Sie fallen damit sehr deutlich unter die von Jensen und Tøndering (2005) gefundenen Werte. Die besten Korrelationen finden sich zwischen Intensität und wahrgenommener Silbenprominenz. Insgesamt fällt eine hohe Streuung innerhalb der einzelnen Skalen, aber auch der einzelnen Bedingungen auf. Es gibt keine Skala und auch keine Bedingung, bei der kontinuierlich höhere Korrelationen zwischen der wahrgenommenen Silbenprominenz und den getesteten akustischen Parametern vorliegen.

Die Passung der linearen Modelle zeigt ein ähnliches Bild wie die akustischen Korrelate, wobei hier alle Werte in einem aussagekräftigen Rahmen bleiben. Innerhalb der Skalen und Bedingungen findet sich keine eindeutige Reihenfolge der Skalen. Die beste Passung wird bei der 11-Punkt-Skala erreicht, aber auch mit der kontinuierlichen Skala und der 4-Punkt-Skala gibt es jeweils ein Modell mit recht guter Passung. Insgesamt lassen sich aus den vorliegenden Daten keine Vorzüge für eine der vier Skalen erkennen.

Tabelle 3.8.4.: *In der ersten Spalte sind die Skala mit 4-Punkt, 11-Punkt, 31-Punkt und kontinuierlicher Skala, die Akkuratheitsbedingung mit 0 so schnell wie möglich, 1 keine Vorgaben und 2 möglichst akkurate Bearbeitungen und die Priminggruppen kodiert. In den darauf folgenden drei Spalten sind die Spearman Rankkorrelationen zwischen den Bewertungen und dem jeweiligen akustischen Parameter angegeben. In der letzten Spalte findet sich die Modellpassung r^2 eines linearen Modells, bei dem die drei akustischen Parameter als Prädiktoren für die wahrgenommene Silbenprominenz benutzt wurden.*

Skala und Bedingungen	Silbendauer	Intensität	f0	Modellpassung
Skala 4 Akk 0 Prim 0	0.339	0.353	0.172	.24
Skala 4 Akk 1 Prim 0	0.357	0.484	0.268	.35
Skala 4 Akk 2 Prim 0	0.442	0.308	0.179	.27
Skala 4 Akk 0 Prim 1	0.352	0.304	0.116	.22
Skala 4 Akk 1 Prim 1	0.291	0.298	0.107	.18
Skala 4 Akk 2 Prim 1	0.358	0.365	0.203	.30
Skala 11 Akk 0 Prim 0	0.303	0.484	0.284	.30
Skala 11 Akk 1 Prim 0	0.296	0.425	0.267	.25
Skala 11 Akk 2 Prim 0	0.337	0.376	0.191	.22
Skala 11 Akk 0 Prim 1	0.331	0.551	0.360	.40
Skala 11 Akk 1 Prim 1	0.306	0.400	0.212	.23
Skala 11 Akk 2 Prim 1	0.322	0.468	0.267	.35
Skala 31 Akk 0 Prim 0	0.322	0.424	0.253	.22
Skala 31 Akk 1 Prim 0	0.444	0.380	0.190	.32
Skala 31 Akk 2 Prim 0	0.346	0.429	0.256	.29
Skala 31 Akk 0 Prim 1	0.376	0.266	0.067	.20
Skala 31 Akk 1 Prim 1	0.295	0.520	0.374	.31
Skala 31 Akk 2 Prim 1	0.290	0.326	0.146	.19
Skala Kont Akk 0 Prim 0	0.314	0.522	0.300	.28
Skala Kont Akk 1 Prim 0	0.391	0.510	0.345	.36
Skala Kont Akk 2 Prim 0	0.399	0.429	0.231	.31
Skala Kont Akk 0 Prim 1	0.290	0.442	0.294	.27
Skala Kont Akk 1 Prim 1	0.355	0.451	0.274	.31
Skala Kont Akk 2 Prim 1	0.473	0.413	0.201	.37

3.8.6. Priming

Mittels eines Testsets aus Arnold und Wagner (2008) und Arnold et al. (2010) sollte überprüft werden, ob sich die gefundenen Primingeffekte mit allen vier Skalen replizieren lassen. Hierfür gab es die Akkuratheitsbedingungen, bei denen die Probanden angewiesen wurden, sich alle Sätze möglichst nur einmal anzuhören. Da die zehn übrigen Sätze, die jeder Proband bewerten sollte, einen deutlich anderen Kontext bereit hielten als die übrigen Testsets in Arnold und Wagner (2008) und Arnold et al. (2010), ist anzunehmen, dass der Primingeffekt gegenüber den genannten Studien schwächer ausfallen könnte.
Analog zu Arnold und Wagner (2008) und Arnold et al. (2010) wurde zur Auswertung die Formel 3.8.1 herangezogen. Hierbei stehe D_n für die Differenz zwischen der bewerteten Prominenz der Silbe n und der bewerteten Prominenz ihrer direkten Nachbarn. P_n steht für das Prominenzurteil der Silbe n.

$$D_n = \frac{2P_n - P_{n+1} - P_{n-1}}{2} \tag{3.8.1}$$

Die Annahme ist, dass sich an der manipulierten Silbe signifikante Unterschiede in D_n ergeben. Da nicht alle Skalen die Voraussetzungen für parametrische Tests erfüllen, wurde der nicht-parametrische Wilcoxon-Rangsummen Test verwendet. Wenn man entgegen der Vorraussetzungen einen Welch-Test benutzt, der auch in Arnold und Wagner (2008) und Arnold et al. (2010) verwendet wurde, ergeben sich die gleichen Ergebnisse.

Der Primingeffekt konnte nur mit der 31-Punkt-Skala repliziert werden, die auch in Arnold und Wagner (2008) und Arnold et al. (2010) verwendet wurde. Mit den übrigen Skalen war kein Effekt nachweisbar. Entgegen den Ergebnissen aus Arnold und Wagner (2008) und Arnold et al. (2010) trat bei dieser Studie der Effekt in die andere Richtung auf. D_n war bei der Gruppe größer, deren Primingsätze an der Stelle nicht prominent waren.

Tabelle 3.8.5.: *Wilcoxon Rangsummen Test zur Überprüfung des Primingseffekts.*

	4-Punkt	11-Punkt	31-Punkt	Kontinuierlich
Wilcox	$W = 140.5$ $p = .49$	$W = 185.5$ $p = .46$	$W = 229$ $p < .05$	$W = 143.5$ $p = .56$

3.9. Diskussion

3.9.1. Ausnutzung der Skalen

Jensen und Tøndering (2005) behaupteten, dass nicht trainierte Probanden die Auflösung einer 31-Punkt-Skala nicht vollständig ausschöpfen können. Experten würden die Skala hingegen besser ausschöpfen. Im Gegensatz dazu haben Grover et al. (1997) festgestellt, dass Probanden mit Skalen mit mehr Stufen bessere Ergebnisse produzieren.

Bei allen Versuchsteilnehmern der vorliegenden Studie handelt es sich um nicht trainierte Muttersprachler. Die Abbildungen 3.8.1 - 3.8.4 zeigen sehr gut, dass alle Stufen der jeweiligen Skala durch die jeweiligen Probanden ausgeschöpft werden. Selbst bei der kontinuierlichen Skala, wo jedes von 301 Pixeln gezählt wird, wird jeder Wert von den Versuchsteilnehmern ausgenutzt. Das Minimum liegt hierbei bei acht Verwendungen des Wertes direkt neben dem Minimum der Skala. Man muss sich aber vor Augen halten, dass es sich hierbei um ein Pixel handelt, also die kleinste Einheit, die man auf einem Bildschirm ansteuern kann. Insgesamt muss man also sagen, dass naive Hörer sehr wohl in der Lage sind, viele Stufen einer Skala auszunutzen.

3.9.2. Extremwerte und Verteilungen

In der Studie von Jensen und Tøndering (2005) war ein Ergebnis, dass bei Skalen mit mehr Stufen weniger extreme Werte gefunden werden. Die Daten der vorliegenden Studie gehen hier in eine andere Richtung. Es lässt sich jeweils eine Bedingung finden, bei der mit der 11-Punkt, 31-Punkt und kontinuierlichen Skala ein größerer Umfang erreicht wird, als mit der 4-Punkt-Skala. In dieser Studie werden die Skalen bei gemittelten Werten vor allem nach unten nicht voll ausgeschöpft. Die Range der Skalen wird je nach Gruppe nur zwischen 45% und 76% genutzt. Aufgrund dieser Kompression, sind Skalen mit mehr Schritten ein Vorteil. Hier bleiben bei der durch die Probanden verursachten Verkürzung der Skala mehr Schritte übrig.

Bei den Verteilungen aller Urteile zeigt sich, dass die Probanden mit der 4-Punkt-Skala eine linksschiefe Verteilung erzeugen, während bei den drei anderen Skalen eine rechtsschiefe Verteilung erzeugt wird. Das bedeutet, dass Probanden mit ihren Urteilen verstärkt in Richtung „nicht prominent" tendieren, wenn sie

die 4-Punkt-Skala verwenden. Bei den drei übrigen Skalen tendieren die Versuchsteilnehmer eher in Richtung „prominent".

3.9.3. Bearbeitungszeit und Anzahl der Wiederholungen als ein Maß für die Schwierigkeit der Beurteilungsaufgabe

In ihrer Studie benutzen Jensen und Tøndering (2005) Bearbeitungszeit und Anzahl der Stimuluswiederholungen als ein Maß dafür, wie schwierig die Benutzung verschiedener Bewertungsskalen für ihre Probanden war. Die Überlegung dahinter ist, dass, wenn eine Aufgabe für die Versuchsteilnehmer schwieriger ist, die Probanden mehr Zeit benötigen, um die Aufgabe zu erledigen. Da sie dabei mehr gefordert sind, hören sie sich die Stimuli auch öfter an. Wie bei der Studie von Jensen und Tøndering (2005) werden auch hier benötigte Zeit und Anzahl der Stimuluswiederholungen als Maß für die Schwierigkeit bei der Benutzung der verschiedenen Skalen benutzt. Wie man in der Tabelle 3.8.2 sieht, zeigen sich bei den beiden Maßen nicht immer dieselben Tendenzen.

Bedingt durch die Aufgabenstellung erwarten wir, dass die Werte sowohl für die Bearbeitungsdauer als auch für die Wiederholungen über die drei Bedingungen hinweg steigen. Für die drei diskreten Skalen ist das der Fall, während bei der kontinuierlichen Skala die Gruppen aus Bedingung 2 deutlich länger brauchen und sich die Stimuli öfter anhören. Obwohl es nicht den direkten Erwartungen entspricht, dass die Gruppe, welche keine Anweisungen bezüglich der Anzahl der Wiederholungen bekommt, die Sätze besonders häufig wiederholt, verstößt es nicht gegen die Instruktionen und ein Blick auf die Passungen zwischen Prominenzbewertungen und akustischen Parametern bestätigt, dass sich durch Zufall in dieser Gruppe besonders motivierte Probanden gesammelt haben müssen, die ihre Aufgabe offenbar sehr ernst genommen haben. Positiv ist zu bemerken, das sich auch hier Bearbeitungsdauer und Anzahl der Stimuluswiederholungen ähnlich verhalten und für Bedingung 2 am höchsten und für Bedingung 1 am niedrigsten ausfallen.

Wenn man die vier Skalen bei jeweils gleichen Bedingungen vergleicht, findet sich keine eindeutige Reihenfolge der vier Skalen. Die Hypothese, dass Skalen mit mehr Schritten automatisch höhere Kosten bei den Probanden in Form von längerer Bearbeitungsdauer und einem häufigeres Anhören der Stimuli verursachen, muss anhand der vorliegenden Daten zurückgewiesen werden.

3.9.4. Interrater Reliabilität

Die Werte sind bis auf eine Gruppe allesamt als ordentlich einzustufen. Dies zeigt zum einen, dass das Konzept Prominenz eine recht reliable perzeptuelle Größe darstellt. Zum anderen, dass sich immer Versuchsteilnehmer in einer Gruppe zusammenfinden können, die sich in ihrem Urteil nicht besonders einig sind. Schaut man sich die akustischen Korrelate der Gruppe mit der niedrigsten Interrater Reliabilität an, stellt man fest, dass die Gruppe nicht die schlechtesten Korrelationen hat. Auch hat diese Gruppe nicht die geringsten Bearbeitungsdauern, so dass man nicht sagen kann, dass es sich bei der Gruppe um unkooperative Versuchsteilnehmer gehandelt hat.

Insgesamt sieht das Bild bei den Interrater Reliabilitäten ähnlich aus, wie bei den anderen Größen. Es gibt keine Skala, die konstant höhere Reliabilitäten bei den Urteilen der Probanden erzeugt. Man kann aus den vorliegenden Daten also keiner Skala den Vorzug geben.

3.9.5. Akustische Korrelate

Bei den vorliegenden Korrelationen zwischen bewerteter Prominenz und den akustischen Parametern Silbendauer, Intensität und Grundfrequenz ergibt sich kein einheitliches Bild. Häufig ist die Korrelation zwischen Intensität und Silbenprominenz am stärksten ausgeprägt. Dies entspricht den Ergebnissen von Kochanski et al. (2005). Es gibt aber auch drei Fälle, in denen die Korrelation zwischen Silbendauer und Silbenprominenz größer ist als die zwischen Intensität und Silbenprominenz. Die Passunger der linearen Modelle dieser drei Fälle sind dabei nicht die schlechtesten. Die Korrelationen zwischen Grundfrequenz und Silbenprominenz sind meistens sehr schwach ausgeprägt und zum Teil nicht signifikant. Bemerkenswert ist, dass in der Gruppe mit der besten Modellpassung die Korrelation zwischen Grundfrequenz und Silbenprominenz die höchste Ausprägung aller Gruppen hat und die Korrelation zwischen Silbendauer und Silbenprominenz leicht schwächer als die Korrelation zwischen Grundfrequenz und Silbenprominenz ist.

Die Frage, ob die Bewertung mit einer der vier Skalen systematisch bessere Korrelationen ergibt, muss verneint werden. Je nach Kombination aus Priming- und Akkuratheitsbedingung hat jeweils eine Gruppe mit einer anderen Skala die höchsten Korrelationen. Es findet sich auch kein systematischer Einfluss der

Priming- oder Akkuratheitsbedingung.

3.9.6. Priming

Der Primingeffekt aus Arnold und Wagner (2008) und Arnold et al. (2010) konnte mit der 31-Punkt-Skala erfolgreich repliziert werden. Zunächst ist die Replizierung des Effekts etwas sehr Positives. Bedingt durch die Einbettung des Materials in einen anderen Kontext war zu erwarten, dass das Priming unter Umständen nicht erfolgreich ist. Da der in Arnold und Wagner (2008) und Arnold et al. (2010) gefundene Effekt nicht übermäßig stark war und die Einbettung in das neue Versuchsmaterial, das einem anderen Zweck dient, den Effekt eher schwächt, ist die gelungene Replikation mit der Skala, die auch bei der originalen Studie verwendet wurde, ein guter Erfolg.

Mit dem vorliegenden Versuch sollte ja auch geklärt werden, ob alle Skalen den Primingeffekt zeigen können. Offensichtlich ist es mit den drei anderen Skalen nicht gelungen, den Effekt zu replizieren. Als kleiner Vorgriff auf Kapitel 5 sei gesagt, dass auch nach der Normalisierung mit zwei verschieden Verfahren der Unterschied, der mit dem Priming erzeugt werden sollte, signifikant bleibt. Zusätzlich wird mit beiden Verfahren der Unterschied bei der 11-Punkt-Skala signifikant.

Die beiden mittleren Skalen scheinen also grundsätzlich gut geeignet, um Unterschiede in der Beurteilung verschiedener Gruppen nachzuweisen. Die 4-Punkt-Skala hat offensichtlich nicht genug Stufen, um dies zu bewerkstelligen.

3.9.7. Fazit

Wenn man die Daten betrachtet, findet sich keine klare Rangfolge der vier getesten Skalen. Je nach Kombination der Parameter Priming- und Akkuratheitsbedingung hat jeweils eine andere Skala die stärksten Werte. Hierbei ergibt sich kein systematischer Einfluss der verschiedenen Bedingungen.

Hinsichtlich der Verteilung der Urteile einzelner Silben ergeben sich mit den mehrstufigen Skalen interessantere Einblicke als mit der 4-Punkt-Skala. Es zeigt sich, dass die Übereinstimmung der Probanden bei den einzelnen Silben stark schwankt. Hierbei finden sich einzelne Silben, bei denen sich die Probanden deutlich einiger über das Ausmaß an Prominenz sind, als bei anderen Silben, bei denen es zum Teil sehr große Abweichungen gibt. Es kann in der Zukunft sehr inter-

essant sein, zu untersuchen, was diese Unterschiede bedingt und ob beispielsweise Silben, bei denen die Urteile weiter auseinander liegen, „unkritischer“ hinsichtlich der Prominenz sind. Eine weiter Frage ist, ob die Probanden in solchen Fällen durch die verschiedenen Einflüsse unterschiedlich stark beeinflusst werden.

Die mehrstufigen Skalen haben bezüglich der Bearbeitungsdauer und Stimuluswiederholungen keinen Nachteil gegenüber der 4-Punkt-Skala, wie es in Jensen und Tøndering (2005) beschrieben wird. Es fallen also nicht automatisch höhere „Kosten“ bei der Verwendung von Skalen mit vielen Stufen an. Auch dass sich weniger extreme Werte bei der Beurteilung von Prominenz mit Skalen mit vielen Stufen zeigen, kann mit den vorliegenden Daten nicht belegt werden. Der Primingeffekt ist nur mit der 31-Punkt-Skala replizierbar gewesen. Wie wir in Kapitel 5 noch sehen werden, bestätigt sich dies auch bei einer Normalisierung der Prominenzurteile. Zusätzlich wird der Effekt durch das Priming noch bei der Gruppe signifikant, die mit der 11-Punkt-Skala bewertet hat.
Insgesamt kann man wohl zur Verwendung von diskreten Skalen mit vielen Stufen raten. Die Verwendung der 4-Punkt-Skala bietet keinerlei Vorteile gegenüber der 11-Punkt und der 31-Punkt-Skala und bietet eine schlechtere Auflösung. Mit der 4-Punkt-Skala und mit der kontinuierlichen Skala ist das Priming nicht repliziert worden. Die kontinuierliche Skala bietet hinsichtlich Bearbeitungsdauer keinen Vorteil gegenüber den anderen Skalen. Somit verbleiben alle Vorteile bei der 11-Punkt und der 31-Punkt-Skala.

4. Experiment zur Erhebung von Prominenz auf Silben- vs. Wortebene

Teile dieser Studie wurden auf der INTERSPEECH 2011 in Florenz präsentiert und in Arnold et al. (2011a) publiziert.

4.1. Einleitung

In der Literatur finden sich wenige Hinweise darauf, welche Unterschiede sich ergeben, wenn man die Prominenz in gesprochener Sprache auf Silbenebene oder auf Wortebene beurteilen lässt. Die meisten Studien lassen sich von theoretischen Überlegungen leiten, wenn es darum geht, zu begründen, auf welcher Ebene Prominenz bewertet werden soll. Das Ziel des im folgenden beschriebenen Experiments ist es, empirisch zu überprüfen, welche Unterschiede sich durch die Bewertung der Prominenz auf den beiden verschiedenen Ebenen ergeben.

Als erstes soll der Frage nachgegangen werden, ob die Bewertung auf Wortebene den Probanden leichter fällt, als die Bewertung auf Silbenebene. Hierzu soll wie in Kapitel 3 die Bewertungsdauer und die Anzahl der Stimuluswiederholungen herangezogen werden. Des weiteren soll zur Klärung der Frage die Interrater-Reliabilität in den beiden Gruppen herangezogen werden.

Im Fokus steht die Frage, ob es eine einfache Relation zwischen Silbenprominenz und Wortprominenz gibt. Man könnte sich beispielsweise vorstellen, dass die Wortprominenz ungefähr der maximalen Silbenprominenz der Silben des Wortes entspricht. In diesem Fall könnte man recht leicht Ergebnisse von Studien die auf Wortebene bewertet wurden mit Ergebnissen von Studien vergleichen, bei denen auf Silbenebene bewertet wurde.

Ein weiterer Fokus liegt auf der Frage, ob die gefundenen Korrelate zwischen

wahrgenommener Prominenz auf den beiden Ebenen und akustischen Merkmalen ähnlich ausfallen.

4.2. Versuchsaufbau

In diesem Versuch soll das selbe Material von zwei verschiedenen Gruppen hinsichtlich der Prominenz beurteilt werden. Die erste Gruppe soll hierbei die Prominenz der einzelnen Silben beurteilen, während die zweite Gruppe die Prominenz der Wörter beurteilen soll. Hierbei soll die Relation zwischen Wort- und Silbenprominenz untersucht werden. Auch ob sich unterschiedliche Ausprägungen in den Korrelationen zwischen akustischen Parametern und den Prominenzbeurteilungen ergeben ist Gegenstand der Studie.

Für die Studie wurde die selbst implementierte Software, die schon zur Evaluation der verschieden Skalen verwendet wurde, eingesetzt. Als Skala wurde die 31-Punkt-Skala verwendet. Für jede Versuchsperson wurden die 15 Stimuli in eine zufällige Reihenfolge gebracht. Die Versuchsteilnehmer hatten die Gelegenheit, sich die Stimuli beliebig oft anzuhören. Hierzu konnten sie selbstständig die Aufnahme mittels eines Knopfes am Computerschirm erneut starten. Der Versuchsperson stand für jede Einheit (also für jedes Wort für Gruppe 1 und jede Silbe für Gruppe 2) ein eigener Regler zur Verfügung. Über dem Regler befand sich eine orthographische Repräsentation der zu bewertenden Einheit. Jede Einheit eines Satzes musste von der Versuchsperson bewertet werden, um fortfahren zu können.

Wie bereits beim vorangegangenen Experiment zu den verschiedenen Skalen wurden auch hier neben den Bewertungen der einzelnen Einheiten die Reihenfolge in der die Versuchsperson die Einheiten bewertet hat, die Anzahl der Wiedergaben des Stimulus sowie die Dauer pro Bewertung eines Stimulus erhoben und in einer MySQL-Datenbank gespeichert.

4.3. Material

Als Stimuli wurden 15 Sätze konstruiert. Zunächst gab es zwei Sätze die komplett aus einsilbigen Wörtern bestanden. Diese sollten als Kontrollsätze dienen, um abschätzen zu können, ob die beiden Gruppen Prominenz in ähnlicher Weise bewerten. Hierbei sollten die Prominenzurteile der Gruppen hoch miteinander

korrelieren, da beide Gruppen die gleichen linguistischen Einheiten bewerten. Da alle Sätze für jede Versuchsperson neu randomisiert wird, ist davon auszugehen, dass sich Einflüsse von den eventuellen Unterschieden in den Beurteilungsstrategien der beiden Gruppen zeigen. Die beiden Sätze werden hier aufgelistet:

Das Kind schlief tief und fest.
Tom mag es wenn sein Tee heiß ist.

Es wurden weitere Sätze konstruiert, die aus einsilbigen Wörtern und einem mehrsilbigen Wort bestehen. Es wurden zwei Sätze mit einem zweisilbigen Wort konstruiert. Hierbei lag der Wortakzent der Wortes einmal auf der ersten und einmal auf der zweiten Silbe. Die übrigen Wörter waren in beiden Sätzen identisch, um weitere linguistische Einflüsse so gering wie möglich zu halten. Bei der Darstellung der beiden Sätze zeigt **Fettdruck** den Wortakzent an:

In **Lon**don ist es echt schön.
In Ber**lin** ist es echt schön.

Nach dem selben Prinzip wurden Sätze mit einem dreisilbigen Wort konstruiert. Auch hier wurde der Wortakzent systematisch variiert, während die übrigen Wörter in den drei Sätzen die gleichen waren. Nachfolgend ein Beispiel für drei Sätze mit einem dreisilbigen Wort. Darstellung im **Fettdruck** zeigt auch hier den Wortakzent an:

Der **Tech**niker lobt sein Team.
Der Mi**ni**ster lobt sein Team.
Der Präsi**dent** lobt sein Team.

Für die viersilbigen Wörter wurden vier weitere Sätze konstruiert, bei dem ein viersilbiges Wort mit dem Wortakzent jeweils auf einer anderen Position verwendet wurde. Aufgrund eines Fehlers war leider kein Wort dabei, bei dem die erste Silbe den Wortakzent trägt. Die übrigen Wörter waren bei diesen Sätzen nicht mehr identisch, jedoch alle einsilbig. Die Darstellung im **Fettdruck** zeigt den Wortakzent an:

A**me**rika ist sehr groß.
Er trägt viel Ver**ant**wortung.
Die Apo**the**ke ist schon zu.
Nomenkla**tur** ist auch nur ein Wort.

Neben den gerade beschriebenen Sätzen wurden zwei Sätze konstruiert, die zwei zweisilbige Wörter enthielten, welche ein einsilbiges Wort einschlossen. Die Idee hierbei war, dass nach der Definition von Prominenz in dieser Arbeit, sich eine signifikante Veränderung auf diesem Wort ergeben sollte, da die direkten Nachbaren unterschiedlich hoch ausgeprägte Prominenzbeurteilungen erhalten sollten. Darstellung im **Fettdruck** zeigt den Wortakzent an:

Er fährt im **Ju**li nach Lu**zern**.
Er fährt im Au**gust** nach **Zü**rich.

Das Material wurde von der selben Sprecherin aufgenommen, die schon die Stimuli für das Experiment für die Überprüfung der verschiedenen Skalen eingesprochen hatte. Die Aufnahmen fanden in einer Sprecherkabine an der Universität Bielefeld statt. Die Aufnahmen wurden im PCM-Format mono, mit einer Abtastrate von 44,1 kHz und einer Wortlänge von 16 bit aufgenommen.

4.4. Durchführung

An dem Experiment nahmen 36 deutsche Muttersprachler teil. Bei keinem der Versuchsteilnehmer lag eine bekannte Beeinträchtigung des Gehörs vor. Die Probanden waren größtenteils Studenten oder Doktoranden der Rheinischen Friedrich-Wilhelms-Universität Bonn, der Heinrich Heine Universität Düsseldorf und der Universität zu Köln. Alle Versuchsteilnehmer nahmen freiwillig an dem Versuch teil und erhielten keine Bezahlung. Alle Probanden wurden alleine in ruhigen Räumen erhoben. Jede Versuchsperson wurde zufällig einer der beiden Versuchsbedingungen zugewiesen. Die Versuchsteilnehmer wurden einzeln am Computer erhoben. Alle Anweisungen wurden auf dem Computerbildschirm ausgegeben. Die Stimuli wurden den Probanden mit Hilfe eines Kopfhörers vorgespielt. Die Probanden wurden dabei von der Software durch den Versuch geführt. Während die Versuchsteilnehmer ihr Tempo selber gestalten konnten, waren Pausen nicht

vorgesehen. Sämtliche Störungen wurden vom Versuchsleiter so gut es ging im Vorhinein vermieden.

4.5. Ergebnisse

Auch für die Auswertung dieses Versuches wurde wie schon in Kapitel 3.8 R (R Development Core Team, 2010) verwendet. Zur Kommunikation mit der MySQL-Datenbank wurden die Pakete DBI (R Special Interest Group on Databases, 2009) und RMySQL (James und DebRoy, 2010) verwendet.

4.5.1. Zeit und Wiederholungen

Bei der Erhebung wurden sowohl die Anzahl für das wiederholte Anhören des Stimulus als auch die Zeit für die Bewertung jedes einzelnen Satzes erhoben. Jensen und Tøndering (2005) verwendeten diese Ausprägungen in ihrer Studie zur Evaluation verschiedener Skalen zur Erhebung von Silbenprominenz als Maß für die Schwierigkeit und den Aufwand den ein Proband mit der Aufgabe hat.
Da die Zahl der Silben, die sich in einem Satz befinden, mindestens so hoch wie die Anzahl der Wörter, meist aber höher ist, musste die Gruppe, die die Silbenprominenz beurteilt hat, deutlich mehr Einheiten beurteilen. Während die Gruppe, die die Wortprominenz bewertete, insgesamt 80 Einheiten beurteilt hat, waren es bei der Gruppe, welche die Silbenprominenz beurteilt hat, 108 Einheiten. Dies sind 35%, also über ein Drittel mehr Einheiten. Aus diesem Grund wurden die Werte von Zeit und Wiederholungen für den Vergleich einmal als Rohdaten verwendet und einmal normalisiert. Für die Normalisierung wurden die Zeit und die Anzahl der Wiederholungen durch die Anzahl der Elemente im jeweiligen Satz geteilt. Die Ergebnisse sind in Tabelle 4.5.1 zusammengefasst.
Die Versuchsteilnehmer, die die Prominenz auf Wortebene beurteilen, brauchen weniger Zeit, um ihre Aufgabe zu erfüllen. Sie hören sich die Stimuli auch weniger oft an, im Vergleich zu den Probanden, die die Prominenz auch Silbenebene bewerten. Beide Unterschiede sind signifikant, wie eine Überprüfung mit einem t-Test zeigte. Nach der Normalisierung verringert sich der Unterschied in der benötigten Zeit beachtlich. Dieser ist nun nicht mehr signifikant. Der Unterschied in der Anzahl der Wiederholungen bleibt auch nach der Normalisierung signifikant. Zusammenfassend kann man sagen, dass sich die Versuchsteilnehmer,

die Silbenprominenz beurteilen, die Stimuli signifikant häufiger anhören, als die Probanden der anderen Gruppe. Der Unterschied in der Bearbeitungsdauer pro Einheit ist nicht signifikant.

Tabelle 4.5.1.: *Durchschnittliche Bearbeitungszeit und Anzahl von Wiederholungen des Stimulus bei Beurteilung auf Wort- und Silbenebene. Für die Normalisierung wurde die Bearbeitungszeit und die Anzahl der Wiederholungen jeweils durch die Anzahl der Elemente im Satz geteilt.*

	Wort	Silbe	t-test
Zeit [sec] (*SD*)	24.1 (13.7)	32.8 (21.3)	t(458.59)=-5.67; p < .001
Wiederholungen (*SD*)	0.48 (0.79)	1.16 (1.56)	t(398.76)=-6.38; p < .001
Zeit (norm.) (*SD*)	4.48 (2.32)	4.54 (2.84)	t(517.44)=-0.29; p = .77
Wdh. (norm.) (*SD*)	0.08 (0.15)	0.16 (0.21)	t(477.03)=-4.51; p < .001

4.5.2. Korrelation der Prominenzbewertungen in den beiden Kontrollsätzen

In den Stimuli gab es zwei Sätze, die ausschließlich aus einsilbigen Wörtern aufgebaut waren. Hierbei sind also die Anzahl und Identität der Einheiten für die beiden verschiedenen Gruppen identisch. Wenn beide Gruppen Prominenz in gleicher Weise bewerten, ist für diese beiden Kontrollsätze eine hohe Korrelation zwischen den Bewertungen durch die beiden Gruppen zu erwarten. Korrelieren die Bewertungen der beiden Gruppen bei diesen zwei Sätzen nicht hoch miteinander, ist anzunehmen, dass die beiden Gruppen grundsätzlich Unterschiede in ihrer Bewertungsstrategie aufweisen und dass die Unterschiede in den anderen Sätzen nicht primär aus den unterschiedlichen Bedingungen - der Bewertung von Prominenz auf Wort- versus Silbenebene - zurückzuführen sind, sondern auf die Unterschiede in den Gruppen. Tatsächlich korrelieren die Gruppen hoch miteinander. Für beide Sätze finden wir eine Pearson's Produkt-Moment Korrelation von $r=.94$ in beiden Sätzen.

In den Abbildungen 4.5.1 und 4.5.2 sieht man die Bewertung der Wort- und Silbenprominenz des ersten Kontrollsatzes. Wie man sieht, zeigen sich trotz der hohen Korrelation erhebliche Unterschiede in den Verteilungen der Beurteilungen in den beiden Gruppen. Es fällt auf, dass der höchste und niedrigste Wert, als auch der höchste und niedrigste Median bei der Beurteilung von Wortprominenz deutlich extremer ausfällt als bei der Beurteilung der Silbenprominenz, während

der Mittelwert für die beiden Gruppen etwa gleich ausfällt. Bei den Wörtern "Kind","tief" und "fest", die innerhalb des Satzes die höchste Ausprägung von Prominenz haben, sehen wir bei der Beurteilung von Wortprominenz eine deutlich geringere Streuung über die Probanden hinweg.

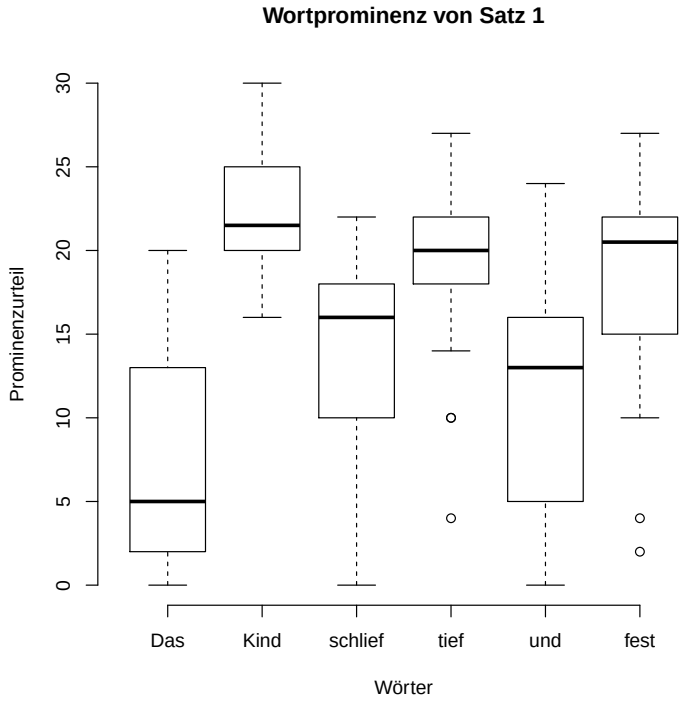


Abbildung 4.5.1.: *Beurteilung der Wortprominenz des ersten Kontrollsatz.*

Im Gegenzug beobachten wir bei der Beurteilung der Wortprominenz mehr Ausreißer, als bei der Beurteilung von Silbenprominenz.

Ein anderes Bild ergibt sich, wenn man sich den anderen Kontrollsatz ansieht (Abbildungen 4.5.3 und 4.5.4). Hier sind die maximalen Werte bei der Guppe, die die Silbenprominenz beurteilt hat, zu finden. Die minmalen Werte finden sich bei der Gruppe, die die Wortprominenz bewertet hat. Die Unterschiede in der Streuung sind nicht so stark wie beim ersten Kontrollsatz. Insgesamt kann gesagt werden, dass man anhand der beiden Kontrollsätze gut sehen kann, dass die beiden Gruppen vergleichbare Urteile fällen und dass somit die gefunden Unterschiede vor allem auf die Beurteilung der unterschiedlichen linguistischen Einheiten, nämlich Wort und Silbe, zurückzuführen sind.

4.5.3. Prominenzurteile auf Wort- und Silbenebene

Der Mittelwert der Ausprägung Prominenz aller 15 Sätze beträgt 14,2 bei der Gruppe, die Silbenprominenz beurteilt hat, und 14,3 bei der Gruppe, die Wortprominenz beurteilt hat. Die Werte unterscheiden sich nicht signifikant. Sie liegt einen Punkt unter dem anzunehmenden Mittelwert der Skala.

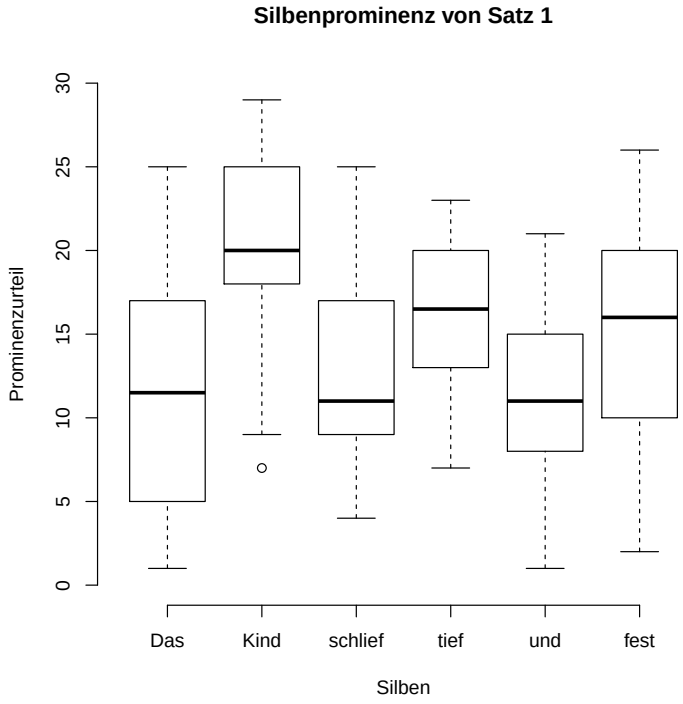


Abbildung 4.5.2.: *Beurteilung der Silbenprominenz des ersten Kontrollsatz.*

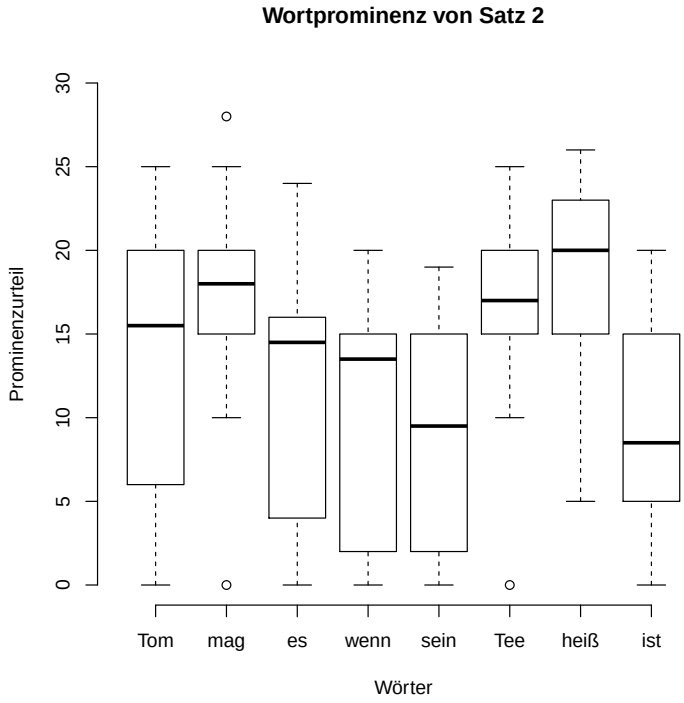


Abbildung 4.5.3.: *Beurteilung der Wortprominenz des zweiten Kontrollsatz.*

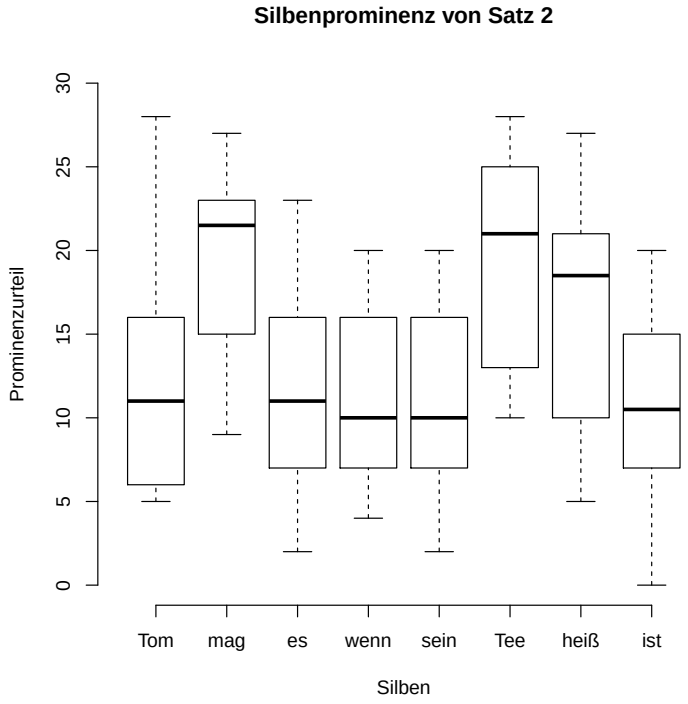


Abbildung 4.5.4.: *Beurteilung der Silbenprominenz des zweiten Kontrollsatz.*

Während bei den beiden Kontrollsätzen, die jeweils aus einsilbigen Wörtern konstruiert wurden, die Unterschiede nicht gravierend sind, stellt sich das Bild in den anderen Sätzen anders da. Die in der Einleitung geäußerte Vermutung, es könne eine einfache Relation zwischen Wortprominenz und Silbenprominenz geben, muss aus den vorliegenden Beobachtungen verneint werden. Bei mehrsilbigen Wörtern ist festzustellen, dass die Wortprominenz häufig höher ausfällt als die Prominez der Wortakzent tragenden Silbe. Dieses Phänomen kann man in den Abbildungen 4.5.5 und 4.5.6 sehen. Die Prominenz von "Berlin" liegt fast fünf Skalenpunkte über der Prominez der Silbe "lin". Bemerkenswert ist in diesem Beispiel insbesondere auch die Prominenz des ersten Wortes "In". Hier wurde die Prominenz des Wortes von der Gruppe, die auf Wortebene bewertet hat, fünf Skalenpunkte niedriger bewertet als von der Gruppe, die auf Silbenebene bewertet hat. Dies zeigt noch einmal sehr deutlich, wie wichtig der Kontext für die Prominenzbeurteilung einer Einheit ist.

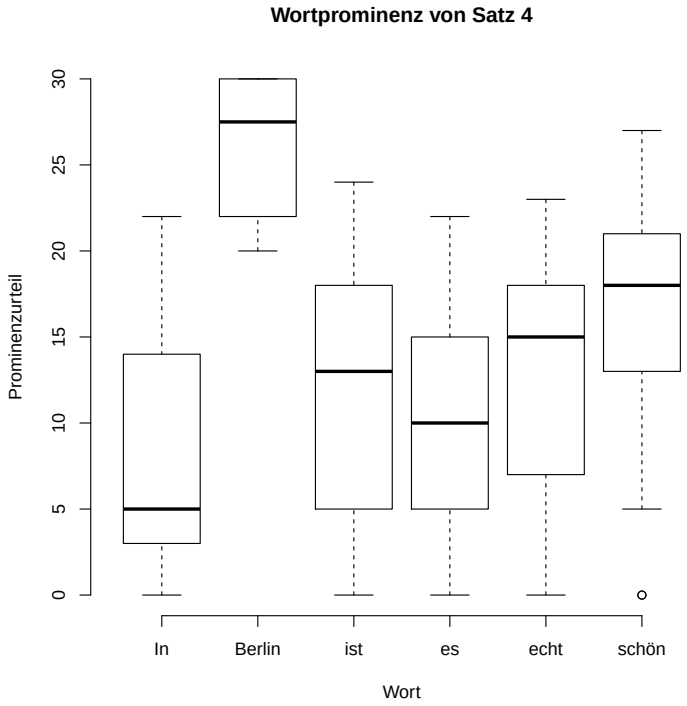


Abbildung 4.5.5.: *Beurteilung der Wortprominenz von Satz 4.*

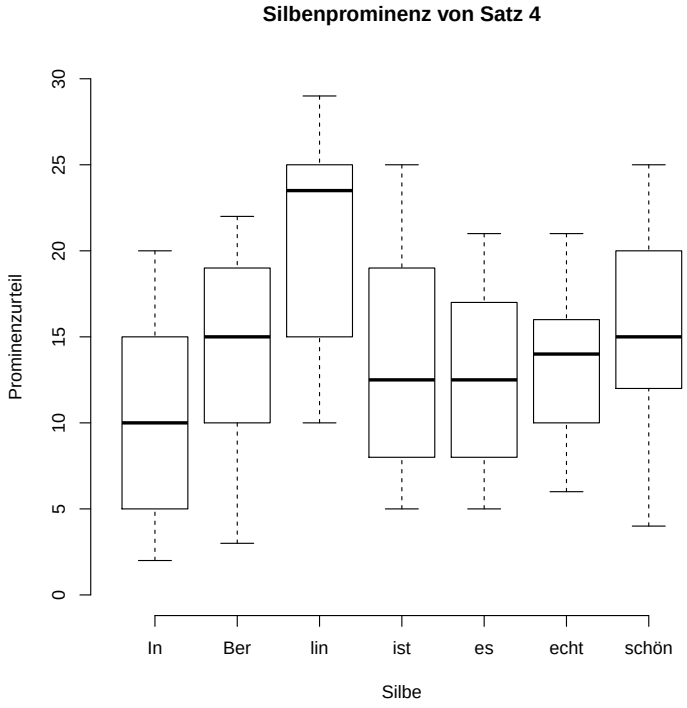


Abbildung 4.5.6.: *Beurteilung der Silbenprominenz von Satz 4.*

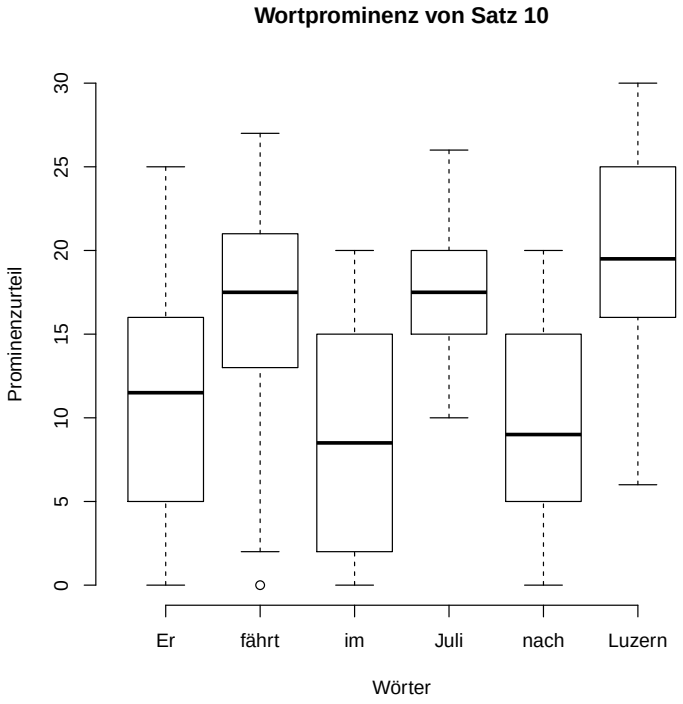


Abbildung 4.5.7.: *Beurteilung der Wortprominenz von Satz 10.*

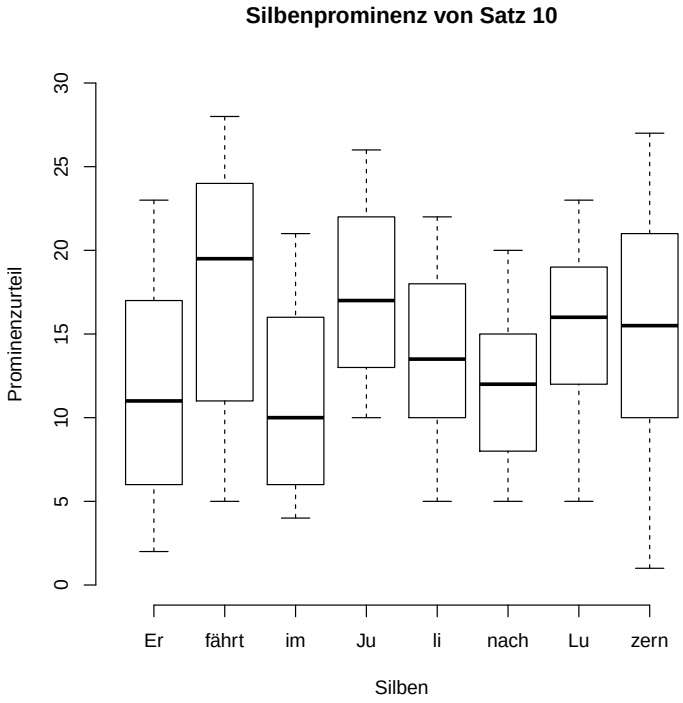


Abbildung 4.5.8.: *Beurteilung der Silbenprominenz von Satz 10.*

4.5.4. Akustische Korrelate

Eine zentrale Frage in der Prominenzforschung ist die Beziehung zwischen beurteilter perzeptueller Prominenz und akustischen Korrelaten, wie beispielsweise Dauer, Grundfrequenz, Intensität und spektrale Eigenschaften des Signals. Im folgenden werden beispielhaft die Korrelationen zwischen perzeptueller Prominenz auf Wort- und Silbenebene und den Parametern Dauer, Grundfrequenzmaxima und Intensität untersucht.

Die Sprachdateien wurden manuell in Praat (Boersma und Weenink, 2010) auf Wort- und Silbenebene segmentiert. Anschließend wurden mit Hilfe von Skripten die Dauern aus den Annotationsdateien extrahiert. Auf Basis der extrahierten Dauern wurden mit Hilfe von Praat-Skripten die Maxima der Grundfrequenz in jeder Silbe und jedem Wort in Praat ermittelt. Ein weiteres Praat-Skript steuerte die Berechnung der durchschnittlichen Intensität jeder Einheit. Für die Auswertung wurden mit Hilfe von R Pearson's Produkt-Moment Korrelationen zwischen den Prominenzurteilen und den akustischen Parametern berechnet.

Die gefundenen Korrelationen zwischen der Wortprominenz und den akustischen Parametern und die Korrelationen zwischen der Silbenprominenz und den akustischen Parametern unterscheiden sich deutlich. Eine Aufstellung findet sich in Tabelle 4.5.2. Als erstes kann man festhalten, dass die Korrelationen zwischen der Wortprominenz und den akustischen Korrelaten höher sind, als die Korrelationen zwischen der Silbenprominenz und den akustischen Korrelaten. Bei der Silbenprominenz finden sich in etwa gleich hohe Korrelationen zu den drei gemessenen Parametern Silbendauer, Grundfrequenzmaximum innerhalb der Silbe und durchschnittliche Intensität von *r=.41 - .39*. Die Höhe der gefundenen Korrelationen liegt im Bereich der in der Literatur zu findenden Werte. Überraschend hingegen ist, dass die Werte fast gleich sind. Bei der Korrelation zwischen der wahrgenommenen Wortprominenz und den akustischen Parametern Wortlänge, Grundfrequenzmaximum innerhalb des Wortes und durchschnittliche Intensität ist ein deutlicher Unterschied festzustellen. Die Korrelation zwischen der Wortdauer und der Wortprominenz ist am höchsten ausgeprägt und liegt bei *r=.69*. Die Korrelation zwischen Wortprominenz und dem Grundfrequenzmaximum beträgt *r=.54* und liegt auf einer Höhe mit der Korrelation zwischen Wortprominenz und Intensität, welche bei *r=.53* liegt.

Als kleiner Vorgriff auf Kapitel 5 sei gesagt, dass sich an den Ergebnissen durch

die Normalisierung nichts Grundlegendes ändern wird. Auch wenn die Korrelationen leicht andere Werte annehmen, ändern sich die Verhältnisse nicht signifikant.

Tabelle 4.5.2.: *Pearson's Produkt-Moment Korrelationen zwischen Wort- bzw. Silbenprominenz und akustischen Parametern mit den zughörigen p-Werten.*

	Wortprominenz	Silbenprominenz
Dauer	*r=.69*	*r=.41*
	$p < .001$	$p < .001$
Maximum f0	*r=.54*	*r=.40*
	$p < .001$	$p < .001$
Intensität	*r=.53*	*r=.39*
	$p < .001$	$p < .001$

4.6. Diskussion

4.6.1. Zeit und Wiederholung

Wie schon in Kapitel 3 wurden die Zeit, welche die Probanden für die Bearbeitung eines Stimulus benötigten, und die Anzahl der Wiederholungen eines Stimulus erhoben. Wie bei Jensen und Tøndering (2005) wurden diese beiden Parameter als Maß für die Bearbeitungskosten verwendet. Dieses Konzept wurde auch auf die vorliegende Studie übertragen und somit Zeitbedarf und Anzahl der Wiederholungen eines Stimulus als Maß für die Schwierigkeit der Beurteilung von Prominenz auf unterschiedlichen linguistischen Einheiten erhoben. Es zeigte sich, dass in den Rohdaten sowohl die Bearbeitungsdauer als auch die Anzahl der Wiederholungen bei der Gruppe, welche die Prominenz auf Silbenebene beurteilt hat, signifikant über den Werten der Gruppe, die auf Wortebene geurteilt hat, liegt. Da die Anzahl der zu erbringenden Beurteilungen bei der Silbenbewertung 35% höher liegt als bei der Bewertung auf Wortebene, wurden die beiden Maße normalisiert. Nach der Normalisierung ist nur noch der Unterschied in den Wiederholungen signifikant. Offenbar wird die meiste Zeit bei der Bewertung der Prominenz durch die Probanden auf die Bedienung der Regler verwendet.

Die Unterschiede in Bearbeitungsdauer und Anzahl der Wiederholungen gehen in die gleiche Richtung. Der Unterschied in der Bearbeitungsdauer ist mit 60 ms allerdings äußerst gering. Es ist kaum anzunehmen, dass der nicht signifikante

Unterschied eine unterschiedliche kognitive Belastung bei der Beurteilung auf Wort- versus Silbenebene darstellt.

Im Gegensatz dazu ist der Unterschied in den Wiederholung auch in der normalisierten Form signifikant. Während bei der Gruppe, die auf Wortebene beurteilt, im Mittel fast jeder zweite Satz erneut angehört wird, hört sich die Gruppe, die die Prominenz auf Silbenebene beurteilt, im Schnitt mehr als jeden Stimulus erneut an. Die Probanden haben ein deutlich höheres Bedürfnis, die Stimuli mehrfach anzuhören, wenn sie Prominenz auf Silbenebene beurteilen sollen. Für dieses Experiment kann man sagen, dass sich ein höherer Aufwand für die Probanden gut in der Anzahl der Wiederholungen ablesen lässt. Dem entgegen scheint die normalisierte Bearbeitungsdauer in diesem Experiment kein gutes Instrument, um die höheren Kosten für die Bearbeitung auf Silben- versus Wort- ebene zu messen. Dies ist umso beeindruckender, da die Zeit, um den Stimulus anzuhören, voll in die Bearbeitungszeit einfließt. Da die Zeitdifferenz, die sich aus dem signifikant häufigeren Abspielen des Stimulus ergibt, sich nicht signifikant in der Zeitdifferenz niederschlägt, stützt dies weiter die oben getroffene Annahme, dass die meiste Zeit von den Probanden für die Bedienung der grafischen Oberfläche verbraucht wird.

Bei Jensen und Tøndering (2005) waren die Unterschiede der normalisierten Bearbeitungsdauer höher ausgeprägt als die Unterschiede in der Häufigkeit der Wiederholungen des Stimulus durch den Probanden. Bei diesem Experiment verhält es sich genau anderes herum. Jensen und Tøndering (2005) waren bei ihrem Experiment an der Schwierigkeit der Bedienbarkeit verschiedener Skalen interessiert. In diesem Experiment geht es um die Schwierigkeit, Prominenz auf zwei verschiedenen linguistischen Ebenen zu beurteilen. Die Interpretation, dass die Zeit hauptsächlich in das Ausführen der Bewertung geht, macht vor diesem Hintergrund Sinn, da bei Jensen und Tøndering (2005) unterschiedliche Skalen verwendet wurden und die Streuung in der Beurteilungszeit größer war als in diesem Experiment, bei dem beide Gruppen mit der gleichen Skala gearbeitet haben.

4.6.2. Prominenzurteile auf Wort- und Silbenebene

Es hat sich gezeigt, dass es keine einfache Relation zwischen Wort- und Silbenprominenz gibt. Es kann somit auch keine einfache Methode geben, die Promi-

nenz auf der einen Ebene aus der Prominenz der anderen Ebene abzuleiten. Die Wortprominenz ist mal deutlich höher als die Silbenprominenz der Wortakzent tragenden Silbe (zum Beispiel in Satz 4 “Berlin”), mal deutlich tiefer (zum Beispiel in Satz 4 “In”). Wie wir in verschiedenen Beispielen gesehen haben, ergeben sich hier stellenweise Verteilungen, die man aus theoriegeleiteten Überlegungen anderes herum vermuten würde. Hier sei als Beispiel Satz S10 genannt (siehe Abbildung 4.5.7 und 4.5.8), bei dem die letzte Silbe “zern”, welche den Wortakzent von “Luzern” trägt, eine niedrigere Prominenz zugewiesen bekommt, als die Silbe “Lu”, welche offensichtlich nicht den Wortakzent trägt.(Wir werden in Kapitel 5 sehen, dass sich die Daten nach einer Normalisierung anders verhalten. Hier hat in “Luzern” die Silbe “zern” die höhere Prominenz.)

Des Weiteren unterstreichen die vorliegenden Daten einmal mehr die Bedeutung des Kontextes bei der Wahrnehmung und Beurteilung von Prominenz.

4.6.3. Akustische Korrelate

Wir haben gesehen, dass die Korrelationen zwischen Silbendauer, Grundfrequenz Maximum, Intensität und der bewerteten Silbenprominenz etwa gleich hoch sind. Sie liegen zwischen *r=.39* und *r=.41*. Dass die drei gewählten Faktoren bei dieser Studie den gleichen Anteil an der Wahrnehmung von Prominenz haben, liegt höchstwahrscheinlich an der Zusammenstellung des Materials, bei dem eine gute Mischung getroffen wurde. Die gefunden Zusammenhänge sind nicht besonders stark, liegen aber durchaus im Rahmen der in der Literatur anzutreffen Werte.

Für die Korrelationen zwischen Wortdauer, Grundfrequenz Maximum, Intensität und der bewerteten Wortprominenz finden wir insgesamt höhere Werte. Diese liegen im Bereich von *r=.53* bis *r=.69*. Auch dieser Wertebereich liegt im Rahmen der berichteten Werte in der Literatur. Die gefunden Korrelationen für die Gruppe, welche die Wortprominenz beurteilt hat, sind deutlich höher als bei der Gruppe, die Silbenprominenz beurteilt hat.

Der Zusammenhang zwischen Wortprominenz und Wortdauer scheint hier besonderes stark zu sein. Die Korrelation ist deutlich höher als die Korrelation zwischen Wortprominenz und Intensität und als die Korrelation zwischen Wortprominenz und Grundfrequenz. Es ist jedoch höchst wahrscheinlich, dass es sich hierbei um ein Artefakt handelt. Fast alle prominenten Wörter im vorliegenden Material sind mehrsilbige Wörter. Diese sind naturgemäß deutlich länger als ein-

silbige Wörter. Da also mehr mehrsilbige und somit längere Wörter prominent sind als einsilbige und somit kurze Wörter, lässt sich die deutlich höhere Korrelation zwischen Wortlänge und Wortprominenz leicht erklären.

4.6.4. Fazit

Es zeigt sich, dass größte Vorsicht geboten ist, wenn Daten, bei denen Prominenz auf Wortebene bewertet wurde, mit Daten, bei denen auf Silbenebene bewertet wurde, verglichen werden sollen. Im vorliegenden Beispiel wird bei der Wortprominenz der Einfluss der Dauer deutlich stärker herausgehoben als bei der Silbenprominenz.

Was die Frage angeht, was für die Probanden leichter zu beurteilen ist, muss man ganz klar sagen, dass die Beurteilung von Wortprominenz den Versuchsteilnehmern deutlich leichter fällt als die Beurteilung von Silbenprominenz. Dies sieht man deutlich an der Anzahl der Stimuluswiederholungen. Die Passung zwischen bewerteter Prominenz und Akustik ist darüber hinaus bei der Beurteilung von Wortprominenz deutlich höher als bei der Beurteilung von Silbenprominenz.

5. Normalisierung von Prominenzurteilen

5.1. Einleitung

In diesem Kapitel sollen exemplarisch zwei Arten der Normalisation auf die Prominenzratings angewendet werden. Zum einen ist dies die lineare Normalisation, welche bei Eriksson et al. (2001) benutzt wird, bei der das Maximum und das Minimum eines jeden Ratings auf das Maximum und Minimum der Skala gebracht werden. Zum zweiten die gängige z-Transformation, bei der der Erwartungswert auf 0 und die Standardabweichung auf 1 gebracht werden. Hierbei gibt es zwei Fragestellungen an die Normalisation: Die erste Frage lautet, ob die Normalisation etwas an den Korrelationen zwischen den Prominenzurteilen und den akustischen Werten aus den Stimuli, denen sie zugrunde liegen, ändert. Ändert sich hierbei auch etwas an den Feststellungen aus Kapitel 3? Die zweite Frage lautet, ob die Normalisation den Ausgang des Primings verändert? Ist nach erfolgter Normalisation das Priming vielleicht mit allen Skalen nachweisbar? Oder erweist sich das Priming eventuell als Artefakt? Diesen Fragestellungen soll im folgenden nachgegangen werden.

5.2. Die Normalisationsverfahren

In ihrer Studie beklagen sich Eriksson et al. (2001), dass die Probanden nicht immer den Anweisungen zur Beurteilung der Stimuli nachgekommen seien. Die Autoren wiesen ihre Probanden an, jeweils eine Silbe eines zu beurteilenden Stimulus mit dem niedrigsten Wert und eine mit dem höchsten Wert der Skala zu belegen. In den Fällen, in denen die Probanden nicht den Vorgaben nachkamen, wurden die Daten durch lineare Transformation in die gewünschte Form gebracht.

Sei R_{max} das Maximum des Ratings, R_{min} das Minimum, $Skala$ die Zahl des Maximalwerts der Skala und R_n das Rating der Silbe n, so gibt die folgende Formel die gewünschte lineare Transformation in das Normalisierte Rating $RNorm_n$ der Silbe n:

$$RNorm_n = (R_n - R_{min}) * \frac{Skala}{R_{max} - R_{min}} \tag{5.2.1}$$

Wie man leicht sehen kann, ist die Formel 5.2.1 für $R_{max} = R_{min}$ nicht definiert. Man muss sich also dazu entschließen, eine Fallunterscheidung zu definieren. Damit muss man festsetzen, welchen Wert Urteile erhalten, bei denen alle Silben gleich bewertet wurden. Dies kann ein beliebiger, fester Wert sein, der tatsächliche Wert der Beurteilung oder man kann sich dazu entschließen, die Werte von der weiteren Beurteilung auszuschließen. Um keine Daten auszuschließen, soll für diese Arbeit die Normalisierung mit der Gleichung 5.2.2 erfolgen.

$$RNorm_n = \begin{cases} (R_n - R_{min}) * \frac{Skala}{R_{max} - R_{min}} & \text{für } R_{min} < R_{max} \\ R_n & \text{für } R_{min} = R_{max} \end{cases} \tag{5.2.2}$$

Im allgemeinen hat sich die z-Transformation bewährt, um Daten miteinander vergleichbar zu machen. Durch die z-Transformation werden die Daten in eine Form gebracht, bei der der Erwartungswert $\mu = 0$ und die Varianz $\sigma^2 = 1$ ist. Dadurch können die Bewertungen von zwei Probanden, deren Mittelwert und Streuung ihrer Ratings von einander abweichen, besser miteinander verglichen werden. Des Weiteren sind z-transformierte Daten für manche statistische Tests eine Grundvoraussetzung. Die Formel zu Berechnung der z-transformierten Werte ist in 5.2.3 angegeben.

$$Z_n = \frac{R_n - \mu}{\sigma} \tag{5.2.3}$$

Hierbei dient als Schätzer für μ der Mittelwert aller Ratings und für σ die Standardabweichung aller Ratings. In unserem Fall besteht die Grundgesamt jeweils aus allen Ratings eines Raters mit einer Skala.

5.3. Auswertung

5.3.1. Auswirkungen der Normalisierung auf die Prominenzurteile

Die Normalisierungen haben naturgemäß einen gravierenden Einfluss auf die Verteilungen der Prominenzurteile. Im Folgenden werden die Verteilungen der Prominenzurteile nach den Transformationen dargestellt. Die Verteilungen der Rohdaten finden sich in den Abbildungen 3.8.1 -3.8.4 in Kapitel 3.

Wie man in den Abbildungen 5.3.1 - 5.3.4 sieht, werden von der linearen Transformation deutliche Spitzen bei den beiden Extremen der jeweiligen Skalen erzeugt. Durch die Transformation werden zum Teil Werte erzeugt, die zwischen den Werten der zugrunde liegenden Skala liegen. Diese Fließkommazahlen sind bei normalem Gebrauch der Skala nicht zugelassen und lassen sich durch gängige Regeln zur Bildung eines Medians nicht bilden. Bemerkenswert ist, dass bei der 31-Punkt-Skala auch nach der Transformation lokale Maxima auf den Skalenschritten liegen, die während der Beurteilung mit einem Label versehen waren.

Die z-Transformation erzeugt linksschiefe Verteilungen. In den Abbildungen 5.3.5 - 5.3.8 sind die z-transformierten Werte, jeweils auf die erste Stelle nach dem Komma gerundet, zusammengefasst. Die Effekte, die sich in den Rohdaten der 31-Punkt-Skala und der kontinuierlichen Skala finden, verschwinden durch die z-Transformation.

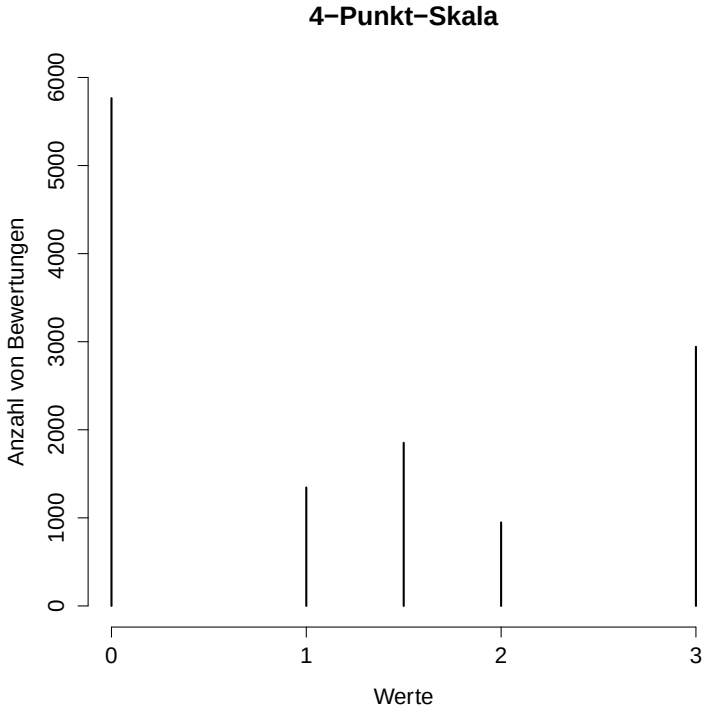


Abbildung 5.3.1.: *Ausnutzung der 4-Punkt-Skala nach der linearen Transformation.*

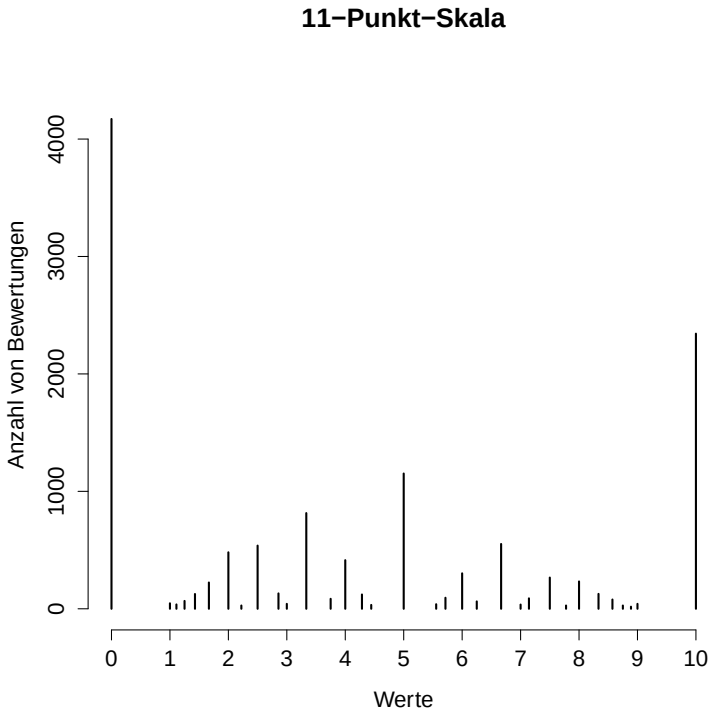


Abbildung 5.3.2.: *Ausnutzung der 11-Punkt-Skala nach der linearen Transformation.*

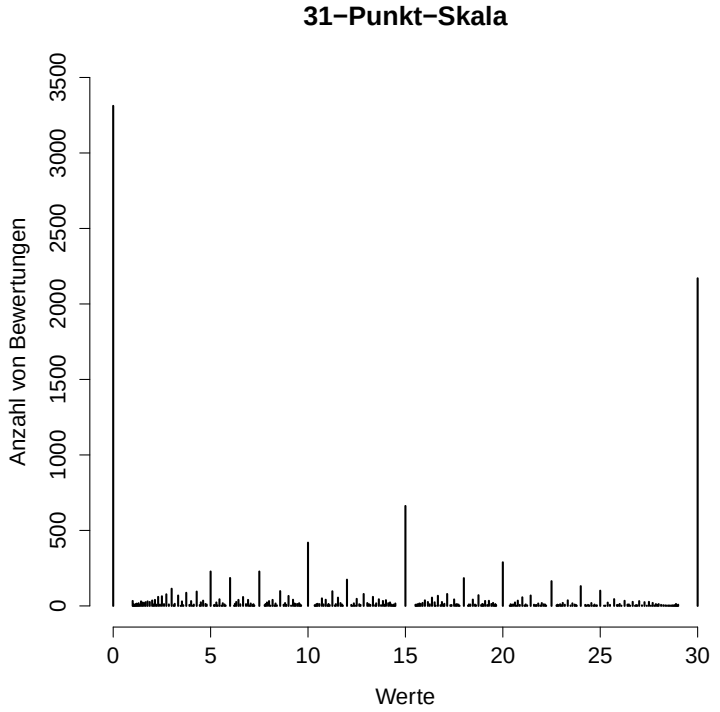


Abbildung 5.3.3.: *Ausnutzung der 31-Punkt-Skala nach der linearen Transformation.*

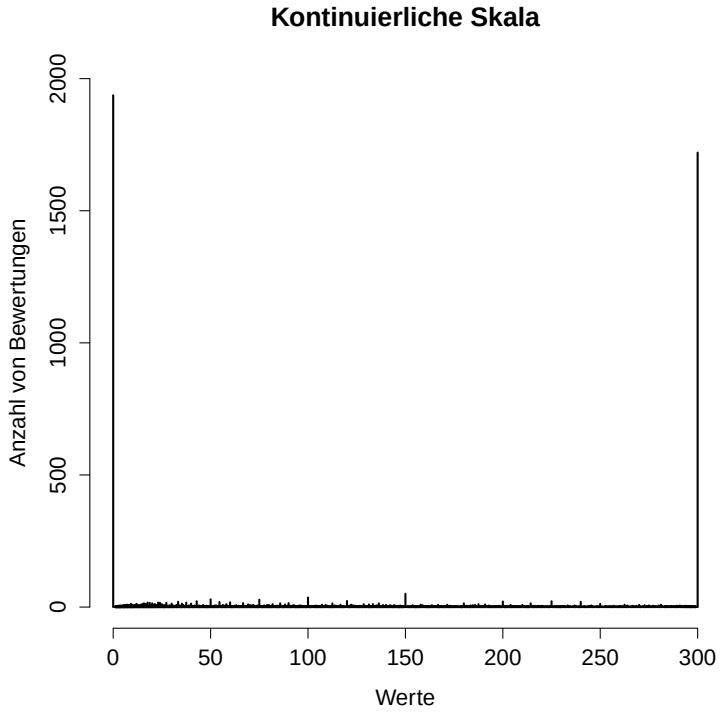


Abbildung 5.3.4.: *Ausnutzung der Kontinuierlichen Skala nach der linearen Transformation.*

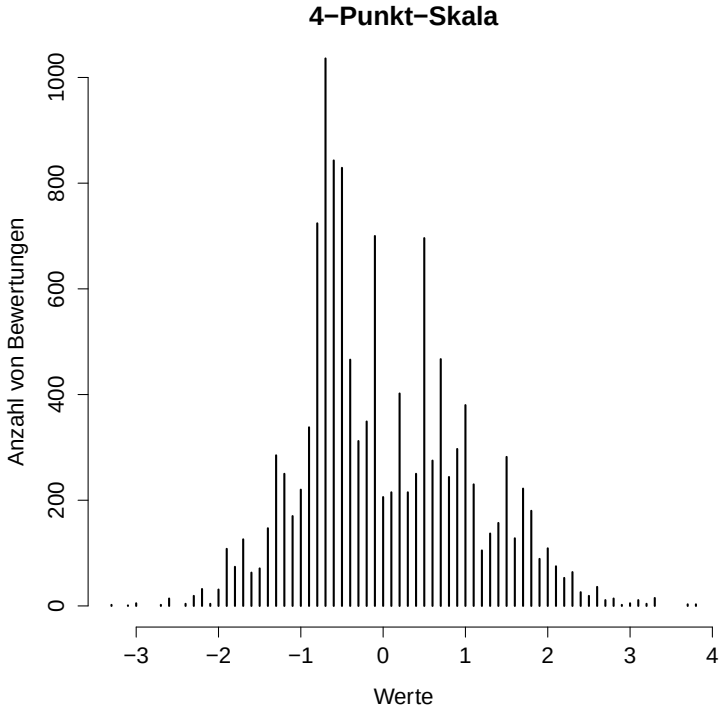


Abbildung 5.3.5.: *Ausnutzung der 4-Punkt-Skala nach der z-Transformation.*

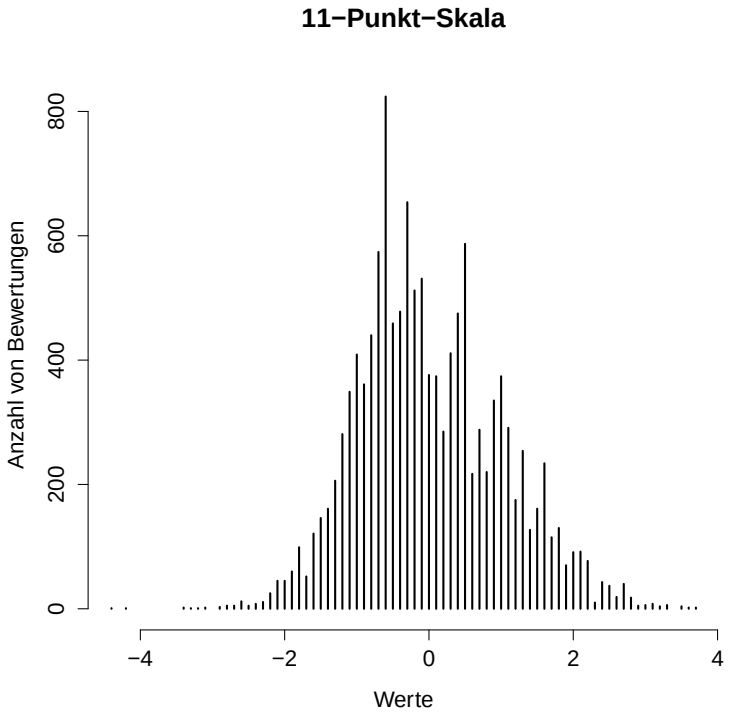


Abbildung 5.3.6.: *Ausnutzung der 11-Punkt-Skala nach der z-Transformation.*

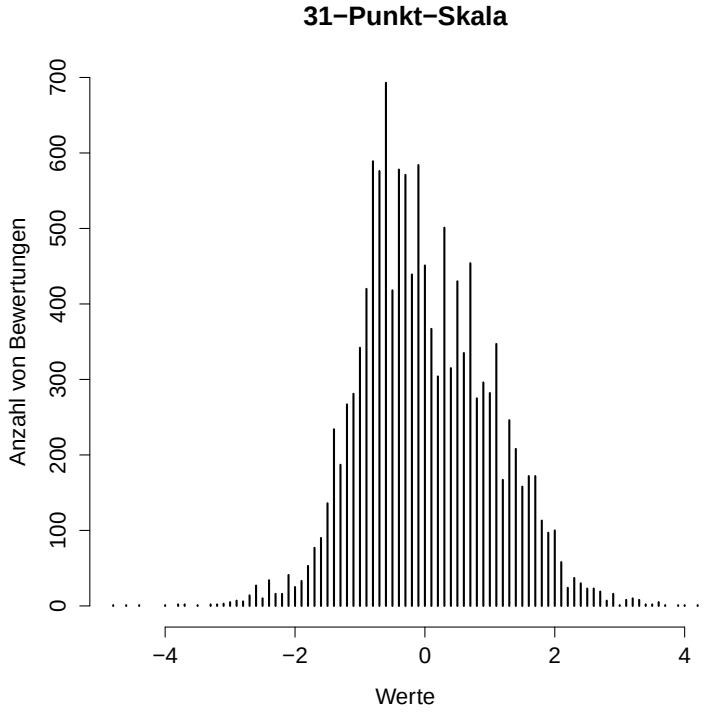


Abbildung 5.3.7.: *Ausnutzung der 31-Punkt-Skala nach der z-Transformation.*

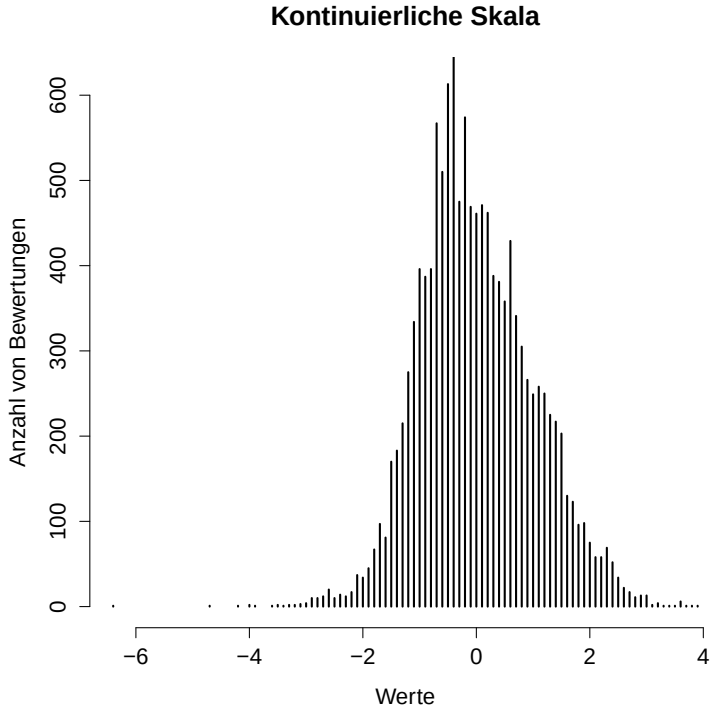


Abbildung 5.3.8.: *Ausnutzung der Kontinuierlichen Skala nach der z-Transformation.*

Betrachtet man die Prominenzbeurteilungen auf einzelnen Sätzen, ergeben sich deutliche Unterschiede zwischen den Rohdaten und den Daten, die mit den unterschiedlichen Verfahren normalisiert wurden. Ein sehr drastisches Beispiel ist die Beurteilung von Satz R4 mit der 31-Punkt-Skala bei der Akkuratheitsgruppe 0 und Priminggruppe 0.

Abbildung 5.3.9 zeigt die Rohdaten der Beurteilung des Satzes durch die entsprechende Gruppe. Wie man in der Abbildung 5.3.10 sehen kann, sind nach der Normalisierung mit der linearen Transformation die Beurteilungen, von wenigen Ausreißern abgesehen, der ersten drei Silben deutlich extremer. Die Streuung der vierten Silbe ist deutlich größer als bei den Rohdaten. Im Vergleich dazu sind die Modifikationen der z-Transformation deutlich moderater, wie man in Abbildung 5.3.11 sehen kann.

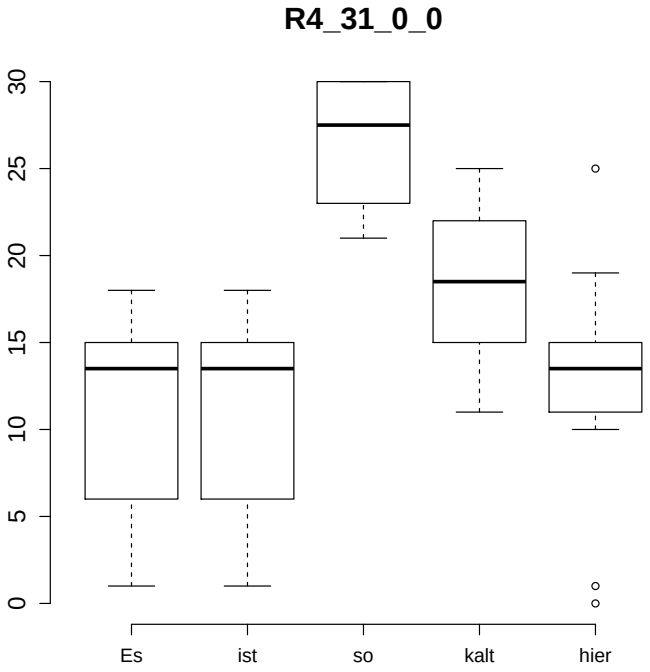


Abbildung 5.3.9.: *Prominenzurteile von Satz R4 mit der 31-Punkt-Skala, Akkuratheitsbedingung 0, Priminggruppe 0.*

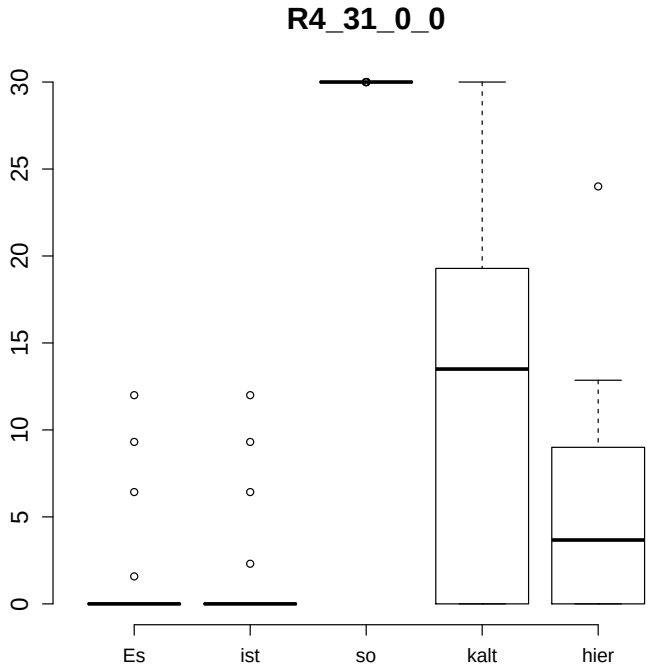


Abbildung 5.3.10.: *Linear transformierte Prominenzurteile von Satz R4 mit der 31-Punkt-Skala, Akkuratheitsbedingung 0, Priminggruppe 0.*

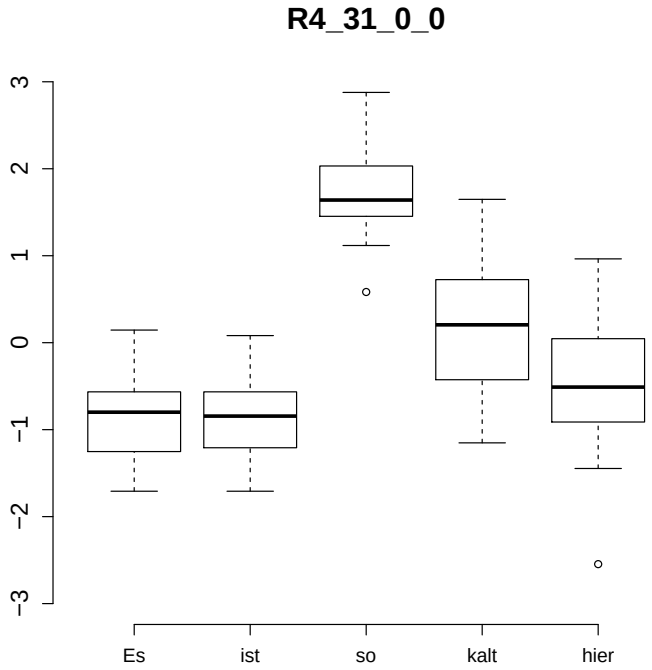


Abbildung 5.3.11.: *z-transformierte Prominenzurteile von Satz R4 mit der 31-Punkt-Skala, Akkuratheitsbedingung 0, Priminggruppe 0.*

In Kapitel 4 hat sich gezeigt, dass Silben deutlich weniger prominent von den Probanden beurteilt werden können, als vom linguistischem Wissen prädiziert. In Abbildung 4.5.8 haben wir gesehen, dass die Silbe "zern", welche den Wortakzent des Wortes Luzern trägt, deutlich weniger prominent beurteilt wurde als erwartet. Abbildung 5.3.12 und 5.3.13 zeigen, dass dieser Effekt durch beide Normalisierungsverfahren verschwindt.

In Abbildung 5.3.12 ist darüber hinaus zu sehen, wie ungünstig sich die lineare Transformation auf die Verteilungen der Prominenzurteile auswirkt. Die z-Transformation zeigt hier in Abbildung 5.3.13 ein besseres Bild.

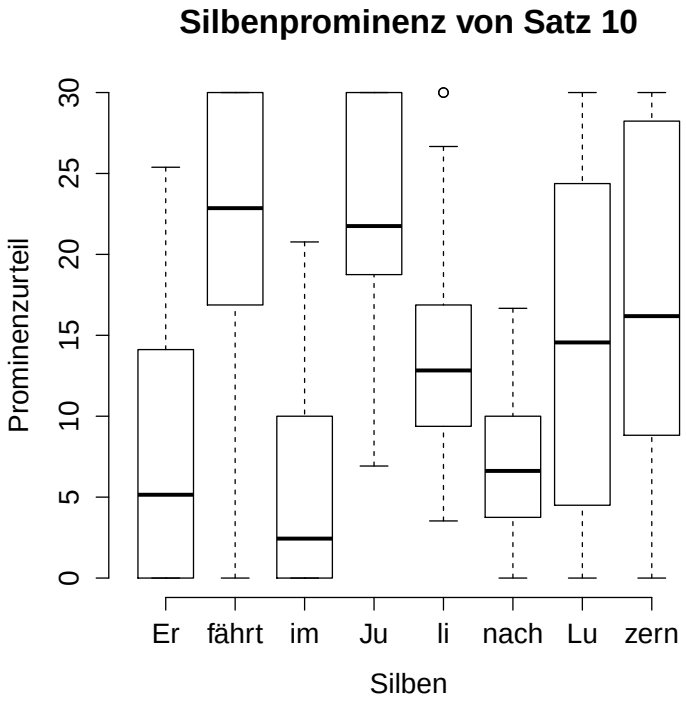


Abbildung 5.3.12.: *Linear transformierte Prominenzurteile von Satz S10.*

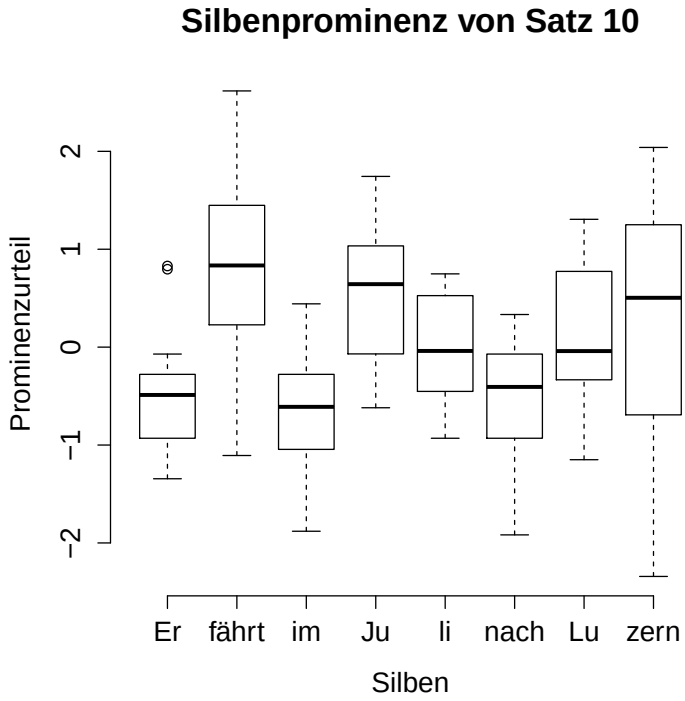


Abbildung 5.3.13.: *z-transformierte Prominenzurteile von Satz S10.*

5.3.2. Auswirkungen der Normalisierung auf Korrelate zwischen den Prominenzurteilen und den akustischen Merkmalen

Wenn man die Prominenzurteile der Probanden normalisiert sollte zu erwarten sein, dass sich die Korrelationen zwischen den gemittelten Prominenzurteilen und den akustischen Merkmalen Dauer, Intensität und Grundfrequenz verbessert. In Kapitel 3 wurde zudem ein lineares Regressionsmodell berechnet. Hier sollte sich die aufgeklärte Varianz verbessern. In Kapitel 3 konnte kein systematischer Einfluss der Akkuratheitsbedingung oder Priminggruppe auf die Höhe der Korrelationen nachgewiesen werden.

In Tabelle 5.3.1 sind die Korrelationen zwischen den gemittelten, nach Eriksson normalisierten Prominenzurteilen und den akustischen Merkmalen, sowie die Passung des linearen Modells aufgeführt. Im Vergleich mit der Tabelle 3.8.4, welche die Daten in gleicher Weise für die nicht normalisierten Prominenzurteile aufführt, sieht man, dass sich die Korrelationen mal verbessern, mal verschlechtern. Dabei überwiegen allerdings die Verbesserungen. Während bei der 4-Punkt-Skala alle bis auf drei Korrelationen besser werden und bei der 11-Punkt-Skala und der 31-Punkt-Skala alle bis auf 5 Korrelationen, verschlechtern sich die meisten Korrelationen bei der kontinuierlichen Skala. Die Modellpassungen verbessern sich durchgängig bis auf einen Fall (Skala 11, Akkuratheitsbedingung 0 und Primingbedingung 1). Auch mit den gewählten Normalisierung lassen sich keine systematischen Einflüsse der drei Faktoren Skala, Akkuratheits- und Primingbedingung ausmachen.

Die Tabelle 5.3.2 zeigt die Korrelationen und Modellpassung zwischen den akustischen Merkmalen und den z-transformierten Prominenzbeurteilungen. Wie auch bei der linearen Transformation zeigt sich kein konsistentes Bild. Bis auf die 4-Punkt-Skala überwiegen hier die Verschlechterungen der Korrelationen. Die Modellpassung liegt bis auf drei Fälle über den Passungen zwischen Akustik und den Prominenzurteilen ohne Normalisierung. Die Passungen sind aber bis auf zwei Fälle schlechter als bei den linear transformierten Daten.

Tabelle 5.3.1.: *Korrelation zu den akustischen Parametern und Passung der linearen Modelle zwischen Akustik und nach Eriksson normalisierten Prominenzurteilen. In der ersten Spalte sind die Skala mit 4-Punkt, 11-Punkt, 31-Punkt und kontinuierlicher Skala, die Akkuratheitsbedingung mit 0 so schnell wie möglich, 1 keine Vorgaben und 3 möglichst akkurate Bearbeitungen und die Priminggruppen kodiert. In den darauf folgenden drei Spalten sind die Spearman Rankkorrelationen zwischen den nach Erikson normalisierten Bewertungen und dem jeweiligen akustischen Parameter angegeben. In der letzten Spalte findet sich die Modellpassung r^2 eines linearen Modells, bei dem die drei akustischen Parameter als Prädiktoren für die normalisierte wahrgenommene Silbenprominenz benutzt wurden. Höhere Werte im Vergleich zu Tabelle 3.8.4 werden **fett**, niedriger Werte kusiv dargestellt.*

Skala und Bedingungen	Silbendauer	Intensität	f0	Modellpassung
Skala 4 Akk 0 Prim 0	**0.366**	**0.428**	*0.203*	**0.33**
Skala 4 Akk 1 Prim 0	*0.304*	**0.544**	**0.276**	**0.41**
Skala 4 Akk 2 Prim 0	*0.383*	**0.462**	**0.210**	**0.38**
Skala 4 Akk 0 Prim 1	**0.357**	**0.476**	**0.236**	**0.37**
Skala 4 Akk 1 Prim 1	**0.341**	**0.401**	**0.195**	**0.28**
Skala 4 Akk 2 Prim 1	**0.377**	**0.456**	*0.202*	**0.37**
Skala 11 Akk 0 Prim 0	**0.387**	*0.475*	*0.236*	**0.40**
Skala 11 Akk 1 Prim 0	**0.312**	**0.529**	**0.285**	**0.40**
Skala 11 Akk 2 Prim 0	**0.338**	**0.483**	**0.210**	**0.36**
Skala 11 Akk 0 Prim 1	**0.337**	*0.440*	*0.196*	*0.32*
Skala 11 Akk 1 Prim 1	**0.313**	**0.447**	*0.249*	**0.31**
Skala 11 Akk 2 Prim 1	**0.355**	**0.479**	*0.198*	**0.37**
Skala 31 Akk 0 Prim 0	**0.359**	*0.463*	**0.255**	**0.36**
Skala 31 Akk 1 Prim 0	*0.425*	**0.497**	**0.277**	**0.46**
Skala 31 Akk 2 Prim 0	**0.374**	**0.444**	*0.193*	**0.35**
Skala 31 Akk 0 Prim 1	**0.410**	**0.387**	**0.111**	**0.34**
Skala 31 Akk 1 Prim 1	*0.290*	*0.519*	*0.331*	**0.37**
Skala 31 Akk 2 Prim 1	**0.319**	**0.448**	**0.240**	**0.31**
Skala Kont Akk 0 Prim 0	*0.313*	*0.511*	*0.253*	**0.37**
Skala Kont Akk 1 Prim 0	*0.298*	**0.541**	*0.289*	**0.40**
Skala Kont Akk 2 Prim 0	*0.322*	**0.515**	**0.266**	**0.39**
Skala Kont Akk 0 Prim 1	**0.342**	**0.407**	*0.236*	**0.29**
Skala Kont Akk 1 Prim 1	*0.348*	*0.438*	*0.198*	**0.32**
Skala Kont Akk 2 Prim 1	*0.472*	*0.377*	*0.116*	**0.39**

Tabelle 5.3.2.: *Korrelation zu den akustischen Parametern und Passung der linearen Modelle zwischen Akustik und z-transformierten Prominenzurteilen. In der ersten Spalte sind die Skala mit 4-Punkt, 11-Punkt, 31-Punkt und Kontinuierlicher Skala, die Akkuratheitsbedingung mit 0 so schnell wie möglich, 1 keine Vorgaben und 3 möglichst akkurate Bearbeitungen und die Priminggruppen kodiert. In den darauf folgenden drei Spalten sind die Spearman Rankkorrelationen zwischen den z-transformierten Bewertungen und dem jeweiligen akustischen Parameter angegeben. In der letzten Spalte findet sich die Modellpassung r^2 eines linearen Modells, bei dem die drei akustischen Parameter als Prädiktoren für die normalisierte wahrgenommene Silbenprominenz benutzt wurden. Höhere Werte im Vergleich zu Tabelle 3.8.4 werden **fett**, niedriger Werte kusiv dargestellt.*

Skala und Bedingungen	Silbendauer	Intensität	f0	Modellpassung
Skala 4 Akk 0 Prim 0	*0.333*	**0.447**	**0.236**	**0.32**
Skala 4 Akk 1 Prim 0	*0.307*	**0.501**	**0.276**	**0.36**
Skala 4 Akk 2 Prim 0	*0.367*	**0.402**	**0.231**	**0.31**
Skala 4 Akk 0 Prim 1	*0.348*	**0.412**	**0.236**	**0.30**
Skala 4 Akk 1 Prim 1	**0.312**	**0.353**	**0.189**	**0.21**
Skala 4 Akk 2 Prim 1	**0.367**	**0.367**	*0.175*	*0.27*
Skala 11 Akk 0 Prim 0	**0.359**	*0.436*	*0.281*	**0.33**
Skala 11 Akk 1 Prim 0	*0.267*	**0.435**	**0.275**	**0.26**
Skala 11 Akk 2 Prim 0	*0.293*	**0.404**	**0.238**	**0.25**
Skala 11 Akk 0 Prim 1	**0.336**	*0.464*	*0.268*	*0.34*
Skala 11 Akk 1 Prim 1	*0.284*	**0.422**	**0.308**	**0.27**
Skala 11 Akk 2 Prim 1	**0.348**	**0.478**	*0.231*	**0.36**
Skala 31 Akk 0 Prim 0	*0.308*	*0.420*	**0.289**	**0.29**
Skala 31 Akk 1 Prim 0	*0.376*	**0.407**	**0.270**	**0.33**
Skala 31 Akk 2 Prim 0	**0.353**	*0.401*	*0.228*	0.29
Skala 31 Akk 0 Prim 1	*0.374*	**0.341**	**0.176**	**0.26**
Skala 31 Akk 1 Prim 1	*0.274*	*0.478*	*0.343*	**0.33**
Skala 31 Akk 2 Prim 1	*0.259*	**0.376**	**0.251**	**0.21**
Skala Kont Akk 0 Prim 0	*0.295*	*0.487*	*0.291*	**0.34**
Skala Kont Akk 1 Prim 0	*0.303*	*0.455*	*0.262*	*0.31*
Skala Kont Akk 2 Prim 0	*0.336*	**0.462**	**0.273**	**0.34**
Skala Kont Akk 0 Prim 1	**0.305**	*0.418*	*0.272*	0.27
Skala Kont Akk 1 Prim 1	*0.309*	*0.448*	*0.271*	0.31
Skala Kont Akk 2 Prim 1	*0.461*	*0.393*	*0.166*	**0.39**

5.3.3. Auswirkungen der Normalisierung auf die Ergebnisse des Primings

Das Priming gelang, wenn man wie in Kapitel 3 die Rohdaten betrachtet, nur mit der 31-Punkt-Skala. Nach der Normalisierung der Daten mit den zwei verschieden Methoden ist der Unterschied durch die Manipulation bei der 31-Punkt-Skala immer noch signifikant. Zusätzlich wird der Unterschied durch die Manipulation mit den beiden Normalisierungsmethoden auch in der 11-Punkt-Skala signifikant. Mit der 4-Punkt-Skala und der kontinuierlichen Skala tritt der Effekt weiterhin nicht auf. Die Ergebnisse der Wilcoxon Rangsummentests finden sich in 5.3.3.

Tabelle 5.3.3.: *Überprüfung des Primingeffekts mittels Wilcoxon Rangsummen Test für die verschiedenen Normalisierungen.*

	4-Punkt	11-Punkt	31-Punkt	Kontinuierlich
nicht normalisiert	W = 140.5	W = 185.5	W = 229	W = 143.5
	p = .49	p = .46	p < .05	p = .56
linear transformiert	W = 170.5	W = 244	W = 215	W = 190
	p = .39	p < .05	p < .05	p = .19
z-transformiert	W = 175	W = 342	W = 229	W = 171
	p = .34	p < .05	p < .05	p = .39

5.3.4. Auswirkungen der Normalisierung auf die akustischen Korrelate im Experiment zum Unterschied zwischen Wort- und Silbenprominenz

Auch die Prominenzdaten aus dem Experiment zum Unterschied zwischen Wort- und Silbenprominenz wurden auf die zwei Arten normalisiert. Hierbei ist natürlich von Interesse, ob sich etwas an den grundsätzlichen Aussagen zu den Korrelationen zwischen bewerteter Prominenz und Akustisk ergibt. Die Tabellen 5.3.4 und 5.3.5 zeigen die Korrelationen nach der linearen Transformation und der z-Transformation. Wie man gut sehen kann, ändern sich die Korrelationen nicht wesentlich. Die Korrelationen verbessern, bzw. verschlechtern sich auf der zweiten Stelle nach dem Komma. Die Rangfolge der Korrelationen wird dabei nur bei der Silbenprominenz verändert, bei die Korrelationen eng zusammen liegen.

Tabelle 5.3.4.: *Pearson's Produkt-Moment Korrelationen zwischen der linear transformierten Wort- bzw. Silbenprominenz und akustischen Parametern mit den zughörigen p-Werten.*

	Wortprominenz	Silbenprominenz
Dauer	*r=.71*	*r=.42*
	$p < .001$	$p < .001$
Maximum f0	*r= .53*	*r=.43*
	$p < .001$	$p < .001$
Intensität	*r=.48*	*r=.37*
	$p < .001$	$p < .001$

Tabelle 5.3.5.: *Pearson's Produkt-Moment Korrelationen zwischen der z-transformierten Wort- bzw. Silbenprominenz und akustischen Parametern mit den zughörigen p-Werten.*

	Wortprominenz	Silbenprominenz
Dauer	*r=.70*	*r=.39*
	$p < .001$	$p < .001$
Maximum f0	*r=.54*	*r=.43*
	$p < .001$	$p < .001$
Intensität	*r=.54*	*r=.44*
	$p < .001$	$p < .001$

5.4. Diskussion

5.4.1. Auswirkungen der Normalisierung auf die Prominenzurteile

Die lineare Transformation führt zu deutlichen Verzerrungen der Ratings. Dies ist sowohl in der Verteilung aller Ratings als auch in dem ausgewählten Beispiel eines Einzelratings deutlich zu sehen. Die Vorgehensweise erzeugt im Vergleich zu den Rohdaten und den z-transformierten Daten deutlich mehr Urteile an den beiden Extremen jeder Skala. Hierbei machen bei der 4-Punkt-Skala und der 11-Punkt-Skala die Summe der extremen Ratings über die Hälfte der gesamten Ratings aus, bei den beiden anderen Skalen deutlich mehr als ein Drittel. Es zeigt sich, dass die prominentesten Silben nun im Mittel den maximalen Wert der Skala erhalten, wenn man die einzelnen Ratings betrachtet. Eine Darstellung aller Ratings der Sätze R1-R10 unter den verschiedenen Bedingungen nach der

linearen Transformationen findet sich in Anhang D. Sehr häufig sind dabei die Urteile aller Versuchspersonen maximal. Dies zeigt auch die Auffassung, die hinter dieser Normalisierung steckt. Prominenz ist in dieser Auffassung nur im Kontext zu sehen und nicht absolut. Das heißt, dass die prominentesten Silben von zwei Sätzen jeweils den maximalen Skalenwert zugewiesen bekommen, auch wenn sie isoliert betrachtet als unterschiedlich prominent beurteilt würden.

Der im Experiment zur Erhebung von Prominenz auf Silben- vs. Wortebene befundene Effekt, dass eine Silbe im Vergleich zur Nachbarsilbe schwächer prominent beurteilt wird, obwohl sie den Wortakzent trägt, verschwindet durch beide Normalisierungen. Hier zeigt es sich, dass die Normalisierung wichtig sein kann, wenn man systematisch nach Abweichungen zwischen theoretisch vorhergesagter Prominenz und beurteilter Prominenz sucht.

Die z-Transformation führt erwartungsgemäß zu keiner Überbetonung der Extremen. Der visuelle Eindruck, dass die Verteilungen der Beurteilungen mit den vier Skalen recht nah an der Normalverteilung sind, liegt zu einem großen Teil an der Darstellung der Abbildungen, bei der alle Ratings im Raum einer Nachkommastelle zusammen gefasst werden. Der Eindruck würde durch Runden zur ganzen Stelle noch verstärkt. Die Ratings der einzelnen Sätze mit den verschiedenen Skalen sind sich nach der Normalisierung deutlich ähnlicher als bei den Rohdaten. Hier sticht nun die 4-Punkt-Skala nicht mehr so heraus. Um sich einen guten Überblick hiervon verschaffen zu können, findet sich in Anhang E die Darstellung aller Ratings der Sätze R1-R10 mit den verschiedenen Bedingungen nach der z-Transformation.

5.4.2. Auswirkungen der Normalisierung auf Korrelate zwischen den Prominenzurteilen und den akustischen Merkmalen

Die Auswirkungen der beiden Transformationsarten auf die Korrelationen zwischen den Prominenzurteilen und den akustischen sind recht unterschiedlich. Die lineare Transformation bewirkt, dass die meisten Korrelationen stärker werden. Hierbei profitiert vor allem die 4-Punkt-Skala, gefolgt von 11-Punkt-Skala und 31-Punkt-Skala. Auf die Korrelationen zwischen den Beurteilungen mit der kontinuierlichen Skala und den akustischen Merkmalen wirkt sich die lineare Transformation meist negativ aus. Bei der z-Transformation profitiert ebenfalls die 4-Punkt-

Skala. Bei der 11-Punkt-Skala und der 31-Punkt-Skala halten sich Verbesserung und Verschlechterung der Korrelationen in etwa die Wage. Bei der kontinuierlichen Skala führt die z-Tranformation vornehmlich zu schwächeren Korrelationen. Insbesondere der Zusammenhang zwischen Silbendauer und Prominenzurteil wird über alle Skalen hinweg meist schwächer.

Im Gegensatz dazu wird die Modellpassung mit beiden Transformationen in den meisten Fällen stärker. Bei der linearen Transformation wird lediglich eine, bei der z-Transformation drei von 24 Modellpassungen schwächer.

Zunächst scheint es nicht intuitiv, dass die lineare Transformation, die die Beurteilungen so stark verzerrt, die Korrelationen stärker verbessert als die z-Transforma- tion. Hier muss man sich das den Bewertungen zugrunde liegende Material vor Augen führen. In allen Sätzen ist mindestens eine Silbe besonders prominent hervorgehoben. Die lineare Transformation, verstärkt besonders die Silben, die sehr prominent sind. In der starken Repräsentation im Versuchsmaterial könnte also der Schlüssel zur vermeintlichen Stärke der linearen Transformation liegen. Würde man ein großes Korpus erstellen, bei dem die ganz stark prominenten Silben im Verhältnis weniger stark auftreten, könnte dieser scheinbare Vorteil verschwinden.

Letztendlich zeigen beide hier getesteten Verfahren zur Normalisierung einen positiven Effekt. Besonders die 4-Punkt-Skala profitiert von der Normalisierung. In Sappok und Arnold (2012) wurden verschiedene Normalisierungsmethoden auf Prominenzurteile untersucht. In der Studie zeigte sich ein positiver Einfluss durch die verschiedenen Normalisierungsverfahren auf die Korrelation zwischen akustischen Merkmalen und den Prominenzurteilen. Dieser Eindruck wird hier also auf einem anderen Material unterstützt. Insgesamt sprechen die hier diskutierten Daten für die Normalisierung von Prominenzurteilen.

5.4.3. Auswirkungen der Normalisierung auf die Ergebnisse des Primings

Ohne Normalisierung war der Unterschied zwischen den Ratings des relevanten Testsatzes nur bei der 31-Punkt-Skala signifikant. Da bei der 31-Punkt-Skala nur jeder fünfte Skalenschritt mit einem Label versehen war, wurden diese Werte bevorzugt von den Probanden gewählt, wie in Abbildung 3.8.3 dargestellt. Positiv war, dass dies die Skala war, mit der das originale Experiment durchgeführt wur-

de. Es lag nun jedoch der Verdacht nahe, dass sich der Effekt des Primings als Artefakt herausstellt. Nach der linearen Transformation blieb der Effekt bestehen, genauso wie nach der z-Transformation. Das Priming wurde also erfolgreich repliziert, obwohl aufgrund des anderen Kontextes eine Abschwächung des Effekts nicht unwahrscheinlich war. Mit beiden Transformationen wird auch der Unterschied bei der 11-Punkt-Skala signifikant. Die 4-Punkt-Skala und die kontinuierliche Skala zeigen weiterhin keinen signifikanten Unterschied.

Man muss schlussfolgern, dass eine Normalisierung der Prominenzurteile in Hinblick auf Untersuchung von Unterschieden in der Perzeption von Prominenz einen positiven Effekt hat.

5.4.4. Auswirkungen der Normalisierung auf die akustischen Korrelate im Experiment zum Unterschied zwischen Wort- und Silbenprominenz

Die Normalisierung hat keinen großen Einfluss auf die akustischen Korrelate in diesem Experiment. Im Wesentlichen würde man die gleichen Schlüsse, wie vor der Normalisierung ziehen. Im Gegensatz zum Experiment zur Erhebung von Prominenz anhand verschiedener Skalen, verbessern sich die Korrelationen nicht wesentlich. Wie in 5.4.2 festgestellt wurde, waren die Verbesserungen für die 31-Punkt-Skala relativ gering. Da bei diesem Versuch genau diese Skala verwendet wurde, liegt das Ergebnis im Einklang mit den Ergebnissen aus 5.4.2. Insgesamt entsteht also im Falle dieses Experiments kein Nachteil, aber auch kein wesentlicher Vorteil durch die Normalisierung.

5.4.5. Fazit

Die lineare Transformation, wie sie von Eriksson et al. (2002) benutzt wurde, verändert die Verteilung der Prominenzurteile massiv. Dies zeigt sich auch in den sehr deutlichen Veränderungen der Beurteilungen der einzelnen Sätze. Hierbei werden die Extremwerte der Skala deutlich betont. Die z-Transformation verändert die Beurteilungen nicht in solch radikaler Weise wie die lineare Transformation und hebt die Extremwerte der Skalen nicht hervor. Beide Transformationen verbessern die Passung zwischen Akustik und Prominenzurteilen deutlich. Von den Transformationen profitiert die 4-Punkt-Skala am meisten, gefolgt von

der 11-Punkt-Skala und der 31-Punkt-Skala. Die Korrelationen zwischen den einzelnen akustischen Merkmalen und den Ratings, die mit der kontinuierlichen Skala erfolgt sind, werden mit den Transformationen überwiegend schlechter, obwohl sich die Passung des linearen Modells verbessert. Nach den Transformationen ist der Unterschied, der durch das Priming erfolgen sollte, in der 11-Punkt-Skala und der 31-Punkt-Skala signifikant. Die Ergebnisse des Experiments zum Unterschied zwischen Wort- und Silbenprominenz werden durch die Normalisierung nicht wesentlich verändert. Die Korrelationen zwischen Silbenprominenz und Akustik bleiben ungefähr auf einem Niveau, und bei der Wortprominenz ist weiterhin die Korrelation zur Wortdauer am stärksten ausgeprägt. Unerwartete Unterschiede zwischen theoretisch erwarteten Prominenzurteilen und tatsächlichen Prominenzurteilen verschwinden durch die Normalisierungen.

Insgesamt kann man sagen, dass eine Normalisierung der Prominenzurteile anzuraten ist. Dieser Schluss steht in Einklang mit den Ergebnissen von Sappok und Arnold (2012), die auch deutliche Verbesserungen für verschiedene Normalisierungsverfahren gefunden haben.

6. Schluss

In diesem Kapitel sollen noch einmal die Ergebnisse der Arbeit zusammengefasst werden. Die Gliederung folgt dabei den Kapiteln der Arbeit. Den Abschluss der Arbeit bildet ein Ausblick, der einige Fragen für die künftige Forschung zur Wahrnehmung von Prominenz formuliert, die sich aus der Literaturlage, aber auch aus den Daten der zwei durchgeführten Experimente ergeben.

6.1. Zusammenfassung

6.1.1. Erhebung von Prominenz anhand verschiedener Skalen

In Kapitel 3 wurden, motiviert aus Grover et al. (1997) und Jensen und Tøndering (2005), vier Skalen zur Erhebung von Prominenz evaluiert. Hierbei wurden eine 4-Punkt-Skala, eine 11-Punkt-Skala, eine 31-Punkt-Skala und eine kontinuierliche Skala getestet. Der Verzicht auf eine binäre Skala wurde in Kapitel 2 ausführlich begründet.

Im Gegensatz zu Jensen und Tøndering (2005) konnten keine Nachteile hinsichtlich der Ausnutzung der Skalen und der Schwierigkeit in der Benutzung der Skalen mit vielen Stufen durch die Probanden gefunden werden. Hinsichtlich der Korrelation zwischen Prominenzurteilen und der Ausprägung von Silbendauer, Intensität und Grundfrequenz wurden bei keiner Skala durchgängig stärkere Korrelationen gefunden. Da auch die Anweisungen und Zugehörigkeit zu einer der beiden Priminggruppen keinen systematischen Einfluss auf die Ausprägung der akustischen Korrelate haben, kann geschlussfolgert werden, dass die Gruppeneffekte stärker wiegen als die Ausprägung der zur Beurteilung herangezogenen Skala.

Der Primingeffekt aus Arnold und Wagner (2008) und Arnold et al. (2010) konnte nur mit der 31-Punkt-Skala - nach der Normalisierung (s. Kapitel 5) auch mit der 11-Punkt-Skala - repliziert werden. Die 31-Punkt-Skala entspricht hierbei

der in den beiden genannten Studien verwendeten Skala. Die Skalen mit mehr Stufen ermöglichen interessantere Beobachtungen hinsichtlich der Verteilungen der Prominenzurteile. Hierbei stellt sich heraus, dass die Probanden bei bestimmten Silben deutlich größere Übereinstimmung zeigen als bei anderen Silben. Diese Ergebnis wird im Ausblick weiter diskutiert werden.

Insgesamt kann aus dem Experiment gefolgert werden, dass Skalen mit mehr Stufen, wie beispielsweise die 11-Punkt-Skala und 31-Punkt-Skala, Skalen mit wenig Stufen, hier der 4-Punkt-Skala, aber auch der kontinuierlichen Skala, vorzuziehen sind.

6.1.2. Erhebung von Prominenz auf Silben- vs. Wortebene

In Kaptitel 4 wurde untersucht, wie sich die Beurteilung von Prominenz auf Wort- und Silben- ebene unterscheidet. Die Beurteilung von Prominenz auf Wortebene fällt den Probanden dabei offensichtlich leichter als die Beurteilung auf Silbenebene. Dies zeigt sich zum einen dadurch, dass die Anzahl der Wiederholungen des Stimulus durch den Probanden bei der Beurteilung auf Silbenebene signifikant höher ausfällt, zum anderen dadurch, dass die Korrelationen zwischen Wortprominenz und den akustischen Merkmalen höher ausfallen als die Korrelationen zwischen Silbenprominenz und den akustischen Merkmalen. Die Ergebnisse sind in Einklang mit den in Streefkerk (2002) gefundenen Ergebnissen, gehen jedoch darüber hinaus, da Streefkerk (2002) nur untersucht hat, ob die Ratings bei Wort- oder Silbenebene besser übereinstimmen.

Es zeigt sich in den Daten, dass es keine einfache Relation zwischen Silbenprominenz und Wortprominenz gibt. Die Wortprominenz von mehrsilbigen Wörtern ist häufig höher als die Prominenz der Wortakzent tragenden Silbe. Bedingt durch den Kontext zeigen sich auch bei einsilbigen Wörtern zum Teil beträchtliche Unterschiede in der beurteilten Prominenz. Dies stärkt die Ansicht, nach der Prominenz abhängig vom Kontext ist.

Als Schlussfolgerung ist festzuhalten, dass beim Vergleich zweier Studien, bei denen Prominenz auf zwei verschiedenen Ebenen erhoben wurde, große Vorsicht geboten ist. Das man Prominenz grundsätzlich auf Wortebene beurteilen sollte, ist ein nicht zulässiger Schluss, da die Daten darauf hindeuten, dass sich die Prominenzverteilung auf Silbenebene nicht einfach aus der Wortprominenz ableiten lässt.

6.1.3. Normalisierung von Prominenzurteilen

In Kapitel 5 wurde der Frage nachgegangen, ob eine Normalisierung von Prominenzurteilen sinnvoll ist und ob die Ergebnisse aus den Kapiteln 3 und 4 durch eine Normalisierung verändert werden. Hierfür wurden zwei verschiedene Verfahren getestet, nämlich die z-Transformation und eine lineare Transformation, wie sie in Eriksson et al. (2002) verwendet wird.

Die Normalisierung der Prominenzurteile zeigt bei beiden getesteten Verfahren positive Effekte. Zum einen verbessert sich die Modellpassung der linearen Modelle, bei denen die Silbendauer, Intensität und Grundfrequenz als Prädiktor für Prominenz dienten. Nach der Normalisierung der Prominenzurteile zeigt sich der Primingeffekt sowohl mit der 31-Punkt und der 11-Punkt-Skala. In Kapitel 4 wurde beschrieben, dass bei einer Silbe, welche den Wortakzent trägt, die Prominenz niedriger ausfiel als bei der zweiten Silbe des Wortes. Nach der Normalisierung verschwand dieser Effekt, und die Beurteilung stimmte mit den linguistischen Voraussagen überein.

Unabhängig von der gewählten Methode bringt die Normalisierung Vorteile und ist damit bei der Untersuchung von perzeptueller Prominenz empfehlenswert. Die Ergebnisse weisen damit in die gleiche Richtung wie Sappok und Arnold (2012), bei dem mehrere Verfahren zur Normalisierung anhand eines Datensatzes evaluiert wurden.

6.2. Ausblick

Nachdem die Forschung viele verschiedene Quellen, die zum Eindruck von Prominenz beitragen, identifiziert hat, ist es nun an der Zeit, die Wechselwirkungen der verschiedenen Einflussfaktoren näher zu betrachten. In Kapitel 2 wurden verschiedene akustische und nicht akustische Parameter benannt, die einen Einfluss auf die Wahrnehmung von Prominenz haben. Hierbei ergeben sich Unterschiede in der Stärke der jeweiligen Faktoren in unterschiedlichen Untersuchungsanordnungen. Einzelne Parameter haben in bestimmten Sprachen unter bestimmten Versuchsanordnungen einen höheren Einfluss als in anderen Sprachen, Versuchsanordnungen und methodischen Zugängen zur Beurteilung der Prominenz.

Wie im Theorieteil der Arbeit dargestellt, gibt es zwei Ansätze, die Integration verschiedener Quellen zu einer Gesamtwahrnehmung von Prominenz zu beschrei-

ben. Goldman et al. (2007) verwenden hierfür das Konzept des *bindings*, Watson (2010) nennt es den *Multiple Source view of prominence*. Wenn die Forschung die Wahrnehmung von Prominenz möglichst umfassend verstehen will, wäre das Ziel ein umfassendes Modell, das die verschiedenen Parameter der einzelnen Informationsstränge erfasst und ihren individuellen Einfluss auf die Prominenz einer Einheit bestimmt und entsprechend der verschiedenen Einflussfaktoren gewichtet. Die in dieser Arbeit hervorgebrachten Erkenntnisse können dabei helfen, eine gute Datenbasis zu erstellen, die dann für die Erstellung eines umfassenden Modells dienen kann. Hier müsste in einem ersten Schritt ein umfassendes multimodales Korpus erstellt werden, dass mindestens hinsichtlich der im Theorieteil dargestellten Parameter annotiert werden müsste. Hierbei könnte zur Schonung von Ressourcen ein bestehendes multimodales Korpus hinsichtlich Prominenz annotiert werden. Auf dieser Datenbasis könnte dann ein multimodales Modell zur Prominenzwahrnehmung erstellt werden.

Dieses Modell könnte dann Vorhersagen dazu machen, wie einzelne Faktoren, zum Beispiel das Wegfallen des visuellen Kanals, die Wahrnehmung modifizieren. Diese Vorhersagen könnten dann experimentell überprüft werden, um das Modell zu validieren. Dabei sind der Modellierung natürlich Grenzen gesetzt, da sicherlich einzelne Gruppen einer Sprechergemeinschaft die freien Parameter unterschiedlich gebrauchen werden und auch einzelne Sprecher einen idiosynkratischen Gebrauch der Parameter pflegen. Ein solches Modell würde jedoch den bisherigen Kenntnisstand deutlich erweitern und ist somit der nächste große Schritt in der Erforschung der Prominenzwahrnehmung.

Literaturverzeichnis

Arnold, D., Möbius, B., und Wagner, P. (2011a). Comparing word and syllable prominence rated by naïve listeners. In *Proceedings of Interspeech 2011* (pp. 1877–1880). Florence, Italy.

Arnold, D. und Wagner, P. (2008). The influence of top-down expectations on the perception of syllable prominence. In *Proceedings of the ISCA Workshop on Experimental Linguistics* (pp. 25–28). Athens, Greece.

Arnold, D., Wagner, P., und Möbius, B. (2010). The effect of priming on the correlations between prominence ratings and acoustic features. In *Speech Prosody 2010, Satellite Workshop on Prosodic Prominence: Perceptual and Automatic Identification (Chicago, IL)* Chicago, USA.

Arnold, D., Wagner, P., und Möbius, B. (2011b). Evaluating different rating scales for obtaining judgments of syllable prominence from naïve listeners. In *Proceedings of the 17th ICPhS 2011* (pp. 252–255). Hong Kong.

Avanzi, M., Goldman, J.-P., und Simon, A. C. (2010). C-PROM. An Annotated Corpus for French Prominence Studies. In *Speech Prosody 2010, Satellite Workshop on Prosodic Prominence: Perceptual and Automatic Identification (Chicago, IL)* Chicago, USA.

Bache, C. (2005). Constraning conceptual integration theory: Levels of blending and disintegration. *Journal of Pragmatics*, 35, 1615–1635.

Bagshaw, P. (1993). An investigation of acoustic events related to sentential stress and pitch accents, in English. *Speech Communication*, 13, 333–342.

Boersma, P. und Weenink, D. (2010). Praat: doing phonetics by computer (version 5.1.31), computer program.

Brenier, J., Nenkova, A., Kothari, A., Whitton, L., Beaver, D., und Jurafsky, D. (2006). The (non)utility of linguistic features for predicting prominence in spontaneous speech. In *IEEE/ACL 2006 Workshop on Spoken Language Technology* (pp. 54–57). Aruba.

Breuer, S. (2009). *Multifunktionale und multilinguale Unit-Selection-Sprachsynthese*. Universität Bonn.

Breuer, S. und Hess, W. (2010). The Bonn Open Synthesis System 3. *International Journal of Speech Technology*, 13/2, 75–84.

Campbell, N. und Beckman, M. (1997). Stress, prominence, and spectral tilt. In *INT-1997* (pp. 67–70). Athens, Greece.

Clark, R. A., Richmond, K., und King, S. (2005). Multisyn voices from arctic data for the blizzard challenge. In *Proceedings of Interspeech 2005* (pp. 101–104). Lisbon, Portugal.

Cole, J., Mo, Y., und Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425–452.

Eriksson, A., Grabe, E., und Traunmüller, H. (2002). Perception of syllable prominence by listeners with and without competence in the tested language. In *Proceedings Speech Prosody 2002* (pp. 275–278). Aix-en-Provence.

Eriksson, A., Thunberg, G., und Traunmüller, H. (2001). Syllable prominence: A matter of vocal effort, phonetic distinctness and top-down processing. In *Proceedings of Eurospeech 2001* (pp. 399–402). Aalborg, Denmark.

Fant, G. und Kruckenberg, A. (1989). Preliminaries to the study of Swedish prose reading and reading style. *STR-QPSR*, 2/1998 KTH, Stockholm, 1–83.

Fant, G. und Kruckenberg, A. (1999). Syllable and word prominence in Swedish. In R. Andersson, Å. Abelin, J. Allwood, und P. Lindblad (Eds.), *Fonetik 99: Proceedings from the Twelfth Swedish Phonetics Conference*, number 81 in Gothenburg Papers in Theoretical Linguistics (pp. 57–60). Göteborg: Department of Linguistics, Göteborg University. ISSN 0349-1021.

Fauconnier, G. und Turner, M. (2002). *The way we think. Conceptual blending and the mind's hidden complexities.* New York: Basic Books.

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In P. MacNeilage (Ed.), *The The Production of Speech* (pp. 39–47). New York Heidelberg Berlin: Springer-Verlag.

Garofolo, J. (1993). *TIMIT Acoustic-Phonetic Continuous Speech Corpus.* Linguistic Data Consortium, Philadephia.

Goldman, J.-P., Auchlin, A., Roekhaut, S., Simon, A. C., und Avanzi, M. (2010). Prominence perception and accent detection in French. A corpus-based account. In *Speech Prosody 2010, Satellite Workshop on Prosodic Prominence: Perceptual and Automatic Identification (Chicago, IL)* Chicago, USA.

Goldman, J.-P., Avanzi, M., Lacheret-Dujour, A., und Simon, A. C. (2007). A Methodology for the Automatic Detection of Perceived Prominent Syllables in Spoken French. In *Proceedings of Interspeech 2007* (pp. 98–101). Antwerp, Belgium.

Granström, B., House, D., und Lundeberg, M. (1999). Procodic cues in multimodal speech perception. In *Proceedings of ICPhS '99* (pp. 655–658).: San Francisco, USA.

Greenberg, S., Hollenback, J., und Ellis, D. (1996). Insights into spoken language gleaned from phonetic transcription of the switchboard corpus. In *Proceedings of the International Conference on Spoken Language Processing* (pp. 24–27). Philadephia, USA.

Grover, C., Heuft, B., und Coile, B. V. (1997). The reliability of labeling word prominence and prosodic boundary strength. In *Proceedings of the ESCA Workshop on Intonation* (pp. 165–168). Athens, Greece.

Gussenhoven, C., Repp, B. H., Rietfeld, A., und Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, 102(5), 3009–3022.

Gussenhoven, C. und Rietveld, T. (1998). On the speaker-dependence of the perceived prominence of f0 peaks. *Journal of Phonetics*, 26, 371–380.

Hess, W. (1983). *Pitch determination of speech signals: Algorithms and devices*. Springer.

Heuft, B. (1996). *Eine prominenzbasierte Methode zur Prosodieanalyse und -synthese*. Frankfurt: Peter Lang.

Heuft, B., Portele, T., Wagner, P., Widera, C., und Wolters, M. (2000). Perceptual prominence. In *Forum Phoneticum, Speech and Signals - Aspects of Speech Synthesis and Automatic Speech Recognition* (pp. 97 – 116). Frankfurt: H.-W. Wodraz.

House, D., Beskow, J., und Granström, B. (2001). Timing and interaction of visual cues for prominence in audiovisual speech perception. In *Proceedings of Eurospeech 2001*, volume 1 (pp. 387–390). Aalborg, Denmark.

James, D. A. und DebRoy, S. (2010). *RMySQL: R interface to the MySQL database*. R package version 0.7-5.

Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical Education*, 38(12), 1217 – 1218.

Jenkin, K. und Scordilis, M. (1996). Development and comparison of three syllable stress classifiers. In *Proceedings of the International Conference on Spoken Language Processing* (pp. 733–736). Philadephia, USA.

Jensen, C. (2003). Perception of prominence in Standard British English. In *Proceedings of the 15th ICPhS 2003* (pp. 1815–1818). Barcelona.

Jensen, C. (2006). Are verbs less prominent? In G. Amrazaitis und S. Schötz (Eds.), *Working Papers 52. 2006 Proceedings from Fonetik 2006* Lund.

Jensen, C. und Tøndering, J. (2005). Choosing a scale for measuring perceived prominence. In *Proceedings of Interspeech 2005* (pp. 2385–2388). Lisbon.

Kochanski, G., Grabe, E., und Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118(2), 1038–1054.

Kohler, K. (2008). Perception of prominence patterns. *Phonetica*, 65(4), 257–269.

Krahmer, E. und Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57, 396–414.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140).

Liljencrants, J. (1999). Judges of prominence. In R. Andersson, Å. Abelin, J. Allwood, und P. Lindblad (Eds.), *Fonetik 99: Proceedings from the Twelfth Swedish Phonetics Conference*, number 81 in Gothenburg Papers in Theoretical Linguistics (pp. 101–107). Göteborg: Department of Linguistics, Göteborg University. ISSN 0349-1021.

McDowall, J. J. (1975). The reliability of ratings by linguistically untrained subjects in response to stress in speech. *Journal of Psycholinguistic Research*, 3(3), 247–259.

McGurk, H. und MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.

Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception. In *Proceedings Speech Prosody 2004* Nara, Japan.

Mixdorff, H. und Widera, C. (2001). Perceived prominence in terms of a linguistically motivated quantitative intonation model. In *Proceedings of Eurospeech 2001*, volume 1 (pp. 403–406).: Aalborg, Denmark.

Mo, Y., Cole, J., und Lee, E.-K. (2008). Naïve listeners' prominence and boundary perception. In *Proceedings Speech Prosody 2008* Campinas, Brazil.

Nenkova, A., Brenier, J., Kothari, A., Calhoun, S., Whitton, L., Beaver, D., und Jurafsky, D. (2007). To memorize or to predict: Prominence labeling in conversational speech. In *NAACL-HLT 2007* (pp. 9–16). Rochester, NY, USA.

Ostendorf, M., Shafran, I., Shattuck-Hufnagel, S., Carmichael, L., und Byrne, W. (2001). A prosodically labeled database of spontaneous speech. In *ISCA Workshop on Prosody in Speech Recognition and Understanding* (pp. 119–121). Red Bank, NJ, USA.

Portele, T. (1998). Perceived Prominence and Acoustic Parameters in American English. In *Proceedings of ICSLP 1998*: Sydney, Australia.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

R Special Interest Group on Databases (2009). *DBI: R Database Interface*. R package version 0.2-5.

Rietveld, T. und Gussenhoven, C. (1983). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13, 299–308.

Sappok, C. und Arnold, D. (2012). On the normalization of syllable prominence ratings. In *Proceedings Speech Prosody 2012* Shanghai.

Siegel, S. und Castellan, N. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Sluijter, A. und van Heuven, V. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the International Conference on Spoken Language Processing*, volume 2 (pp. 630 –633).

Streefkerk, B. (2002). *Prominence - Acoustical and lexical/syntactic correlates*. LOT.

Strom, V., Nenkova, A., Clark, R., Vazquez-Alvarez, Y., Brenier, J., King, S., und Jurafsky, D. (2007). Modelling prominence and emphasis improves unit-selection synthesis. In *Proceedings of Interspeech 2007* Antwerp, Belgium.

Tamburini, F. (2003). Automatic prosodic prominence detection in speech using acoustic features: an unsupervised system. In *Proceedings of Eurospeech 2003* (pp. 129–132). Geneva, Switzerland.

Tamburini, F. und Caini, C. (2005). An Automatic System for Detecting Prosodic Prominence in American English Continuous Speech. *International Journal of Speech Technology*, 8, 33–44.

Tamburini, F. und Wagner, P. (2007). On Automatic Prominence Detection for German. In *Proceedings of Interspeech 2007* (pp. 1809–1812).: Antwerp, Belgium.

Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America*, 107(3), 1697–1714.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, 89(4), 1768–1776.

Terken, J. (1994). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, 95, 3662–3665.

Terken, J. (1996). Variation of accent prominence within the phrase: models and spontaneous speech data. In Y. Sagisaka, W. Campbell, und N. Higuchi (Eds.), *Computing Prosody for Spontaneous Speech* (pp. 95–116). Berlin: Springer-Verlag.

Traum, D., Swartout, W., Gratch, J., Marsella, S., Kenny, P., Hovy, E., Narayanan, S., Fast, E., Martinovsky, B., Baghat, R., Robinson, S., Marshall, A., Wang, D., Gandhe, S., und Leuski, A. (2005). Virtual humans for non-team interaction training. In *6th SIGdial Workshop on Discourse and Dialogue* Lisbon, Portugal.

Turk, A. E. und Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, 6(99), 3782–3790.

Vainio, M. und Järvikivi, J. (2006). Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics*, 34(3), 319–342.

Wagner, P. (2002). *Wahrnehmung und Vorhersage deutscher Betonungsmuster*. Universität Bonn.

Wagner, P. (2005). Great expectations - introspective vs. perceptual prominence ratings and their acoustic correlates. In *Proceedings of Interspeech 2005* (pp. 2381–2384). Lisbon.

Wagner, P. und Portele, T. (1999). Two dimensions of prominence. In *Proceedings of the ESCA Workshop on Dialogue and Prosody*: Eindhoven.

Wang, D. und Narayanan, S. (2007). An acoustic measure for word prominence in spontanious speech. *IEEE Transactions on Audio, Speech and Language Processing*, 15(2), 690–701.

Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. *The Psychology of Learning and Motivation*, 52, 163–183.

Watson, D. G., Arnold, J. E., und Tanenhaus, M. K. (2008). Tic tac toe: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, 106, 1548–1557.

Widera, C., Portele, T., und Wolters, M. (1997). Prediction of word prominence. In *Proceedings of Eurospeech 2007* (pp. 999–1003).: Rhodes.

Windmann, A., Wagner, P., Tamburini, F., Arnold, D., und Oertel, C. (2010). Automatic Prominence Annotation of a German Speech Synthesis Corpus: Towards Prominence-Based Prosody Generation for Unit Selection Synthesis. In *Proceedings of SSW7* (pp. 377–382). Nara, Japan.

Abbildungsverzeichnis

Tabellenverzeichnis

Anhang A. Instruktionen für beide Experimente

Im Folgenden sind Instruktionen für die verschiedenen Bedingungen im Experiment zur Erhebung von Prominenz anhand verschiedener Skalen angegeben. Die Anweisungen waren für den ersten und zweiten Teil gleich aufgebaut (erste Klammer). Je nach verwendeter Skala wurden die entsprechenden Werte für Minimum und Maximum (zweite und dritte Klammer) eingesetzt. Je nach Akkuratheitsgruppe wurde der erste Satz, nichts oder der dritte Satz eingesetzt (letzte Klammer). Die beiden Priminggruppen haben die jeweils gleichen Anweisungen erhalten.

Instruktion:

Im (ersten / zweiten) Teil unseres Experiments werden wir Ihnen ein paar Sätze vorspielen. Sie werden nun gebeten, die Prominenz der einzelen Silben des jeweiligen Satzes zu bewerten. Hierzu ist für jede Silbe ein Regler vorhanden. Stellen Sie den Regler auf **(0 / min)**, wenn Sie glauben, dass die Silbe **völlig unprominent** ist. Stellen Sie den Regler auf **(3 / 10 / 30 / max)** , wenn Sie glauben, dass die Silbe **maximal prominent** ist. Benutzen Sie die übrigen Werte für die Abstufungen dazwischen.
(Sie haben die Möglichkeit, sich die Sätze mehrmals anzuhören. Hören Sie sich die Sätze jedoch bitte nur dann erneut an, wenn es unbedingt sein muss. / - / Benutzen Sie die übrigen Werte für die Abstufungen dazwischen. Versuchen Sie hierbei so präzise wie möglich zu sein. Sie können sich hierzu die Sätze während der Beurteilung erneut vorspielen lassen.)
Sollte Ihnen irgendwas nicht klar sein, fragen Sie bitte den Versuchsleiter!

Für die Gruppe des Experiments zur Erhebung von Prominenz auf Silben- vs. Wortebene, die Prominenz auf Silbenebene bewertet hat, lautete die Instruktion:

Im ersten Teil unseres Experiments werden wir Ihnen ein paar Sätze vorspielen. Sie werden nun gebeten, die Prominenz der einzelen Silben des jeweiligen Satzes zu bewerten. Hierzu ist für jede Silbe ein Regler vorhanden. Stellen Sie den Regler auf **0**, wenn Sie glauben, dass die Silbe **völlig unprominent** ist. Stellen Sie den Regler auf **30**, wenn Sie glauben, dass die Silbe **maximal prominent** ist. Benutzen Sie die übrigen Werte für die Abstufungen dazwischen. Sie haben die Möglichkeit, sich die Sätze mehrmals anzuhören. **Sollte Ihnen irgendwas nicht klar sein, fragen Sie bitte den Versuchsleiter!**

Für die Gruppe des Experiments zur Erhebung von Prominenz auf Silben- vs. Wortebene, die Prominenz auf Wortebene bewertet hat, lautete die Instruktion:

Im ersten Teil unseres Experiments werden wir Ihnen ein paar Sätze vorspielen. Sie werden nun gebeten, die Prominenz der einzelen Wörter des jeweiligen Satzes zu bewerten. Hierzu ist für jede Wort ein Regler vorhanden. Stellen Sie den Regler auf **0**, wenn Sie glauben, dass die Wort **völlig unprominent** ist. Stellen Sie den Regler auf **30**, wenn Sie glauben, dass die Wort **maximal prominent** ist. Benutzen Sie die übrigen Werte für die Abstufungen dazwischen. Sie haben die Möglichkeit, sich die Sätze mehrmals anzuhören. **Sollte Ihnen irgendwas nicht klar sein, fragen Sie bitte den Versuchsleiter!**

Anhang B. Satzlisten für beide Experimente

Hier sind alle Sätze aus den beiden Experimenten aufgelistet. Die Markierung durch **Fettdruck** markiert eine starke Betonung. Die Aufnahmen auf der Begleit-CD-Rom folgen der hier verwendeten Kodierung.

Experiment 1:

P11 : Der alte Mann stieg in den vollen Bus.
P12 : Das kleine Kind ging in das kleine Haus.
P13 : Die alte Frau steigt in den leeren Bus.
P14 : Der junge Mann geht in das gelbe Haus.
T1 : Die **jun**ge Frau geht in das rote Haus.
P21 : Der **al**te Mann stieg in den vollen Bus.
P22 : Das **klei**ne Kind ging in das kleine Haus.
P23 : Die **al**te Frau steigt in den leeren Bus.
P24 : Der **jun**ge Mann geht in das gelbe Haus.
T2 : Die **jun**ge Frau geht in das rote Haus.
R1 : Tim geht Heim.
R2 : Lass ihn rein.
R3 : Geht doch nach Hause.
R4 : Es ist **so** kalt hier.
R5 : **Ich** fliege morgen nach Rom.
R6 : Jan fährt heute mit dem Rad.
R7 : Die Sterne funkeln am Firmament.
R8 : Er **woll**te das Hin der nis um fahren.
R9 : Karin fährt zum Bergsteigen nach Tirol.
R10 : **Das** Motorrad ist dreizehn Jahre alt.

Experiment 2:

Satz 1: Das Kind schlief tief und fest.

Satz 2: Tom mag es wenn sein Tee heiß ist.

Satz 3: In Lon**don** ist es echt schön.

Satz 4: In **Ber**lin ist es echt schön.

Satz 5: Der Techniker lobt sein Team.

Satz 6: Der Minister lobt sein Team.

Satz 7: Der Präsident lobt sein Team.

Satz 8: Bring die **A**nanas mit.

Satz 9: Bring die Ba**na**nen mit.

Satz 10: Er fährt im Juli nach Luzern.

Satz 11: Er fährt im August nach Zürich.

Satz 12: Amerika ist sehr groß.

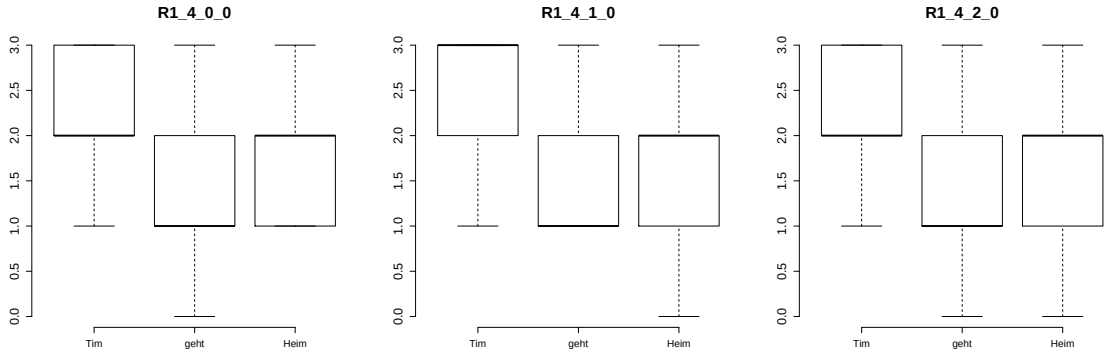
Satz 13: Er träg viel Verantwortung.

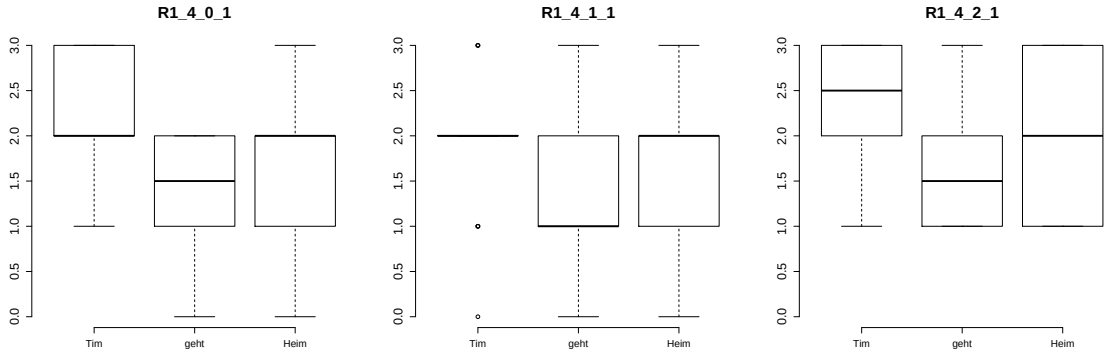
Satz 14: Die Apotheke ist schon zu.

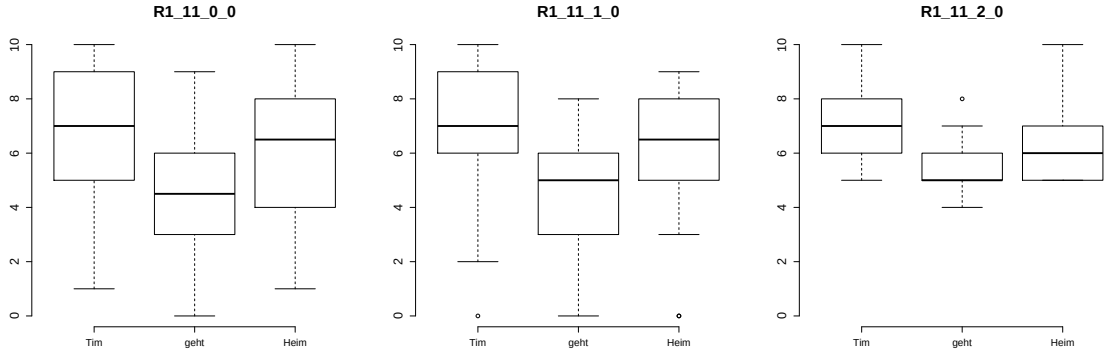
Satz 15: Nomenklatur ist auch nur ein Wort.

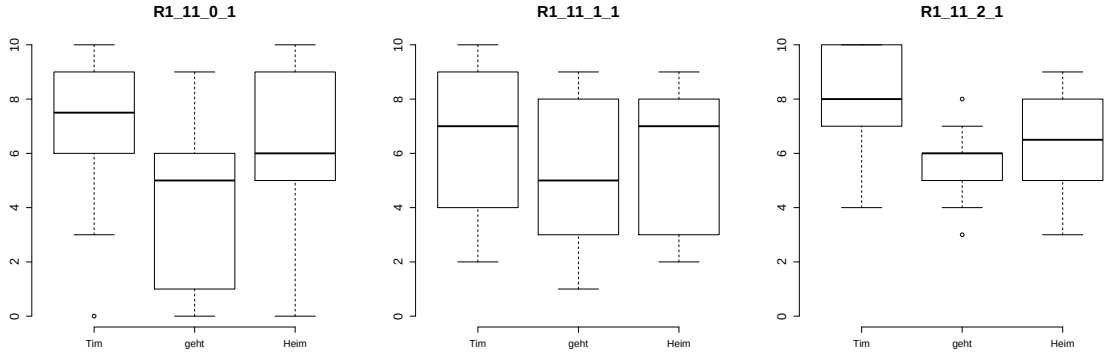
Anhang C. Boxplots Experiment 1

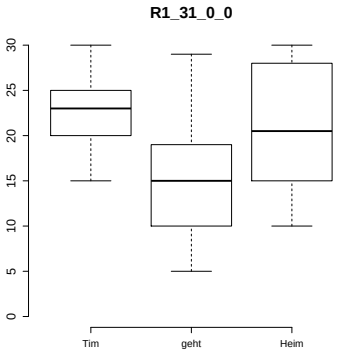
Auf den folgenden Seiten werden die Ratings aller Sätze durch die Probanden des ersten Experiments in Boxplots dargestellt. Hierbei setzt sich der Name der Abbildung wie folgt zusammen: Satzreferenz_Skala_Akkuratheitsbedingung_Priminggruppe. Boxplot R10_4_0_0 zeigt also die Bewertungen des Satzes R10 mit der 4-Punkt Skala unter Akkuratheitsbedingung 0 und Priminggruppe 0.

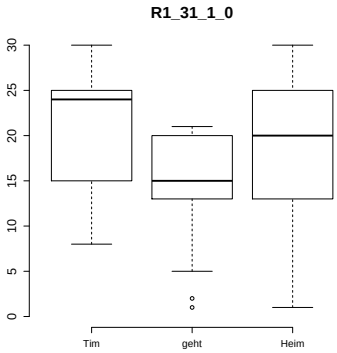

R1_4_0_0
R1_4_1_0
R1_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Tim
geht
Heim

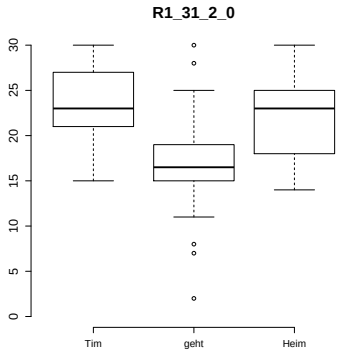

R1_4_0_1
R1_4_1_1
R1_4_2_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Tim
geht
Heim

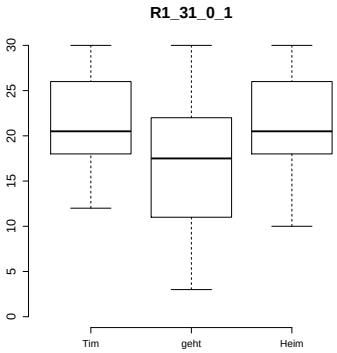

R1_11_0_0
R1_11_1_0
R1_11_2_0
0
2
4
6
8
10
Tim
geht
Heim

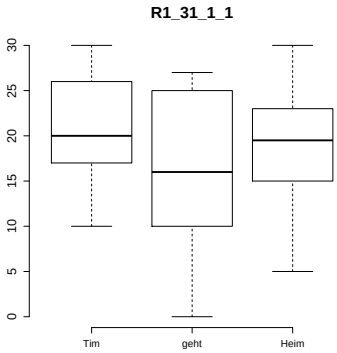

R1_11_0_1
R1_11_1_1
R1_11_2_1
0
2
4
6
8
10
Tim
geht
Heim

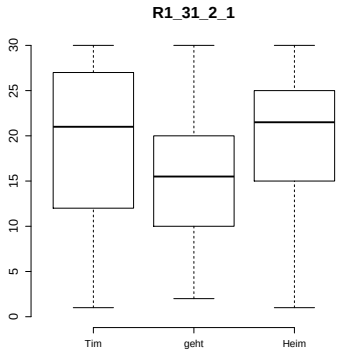

R1_31_0_0
0
5
10
15
20
25
30
Tim
geht
Heim

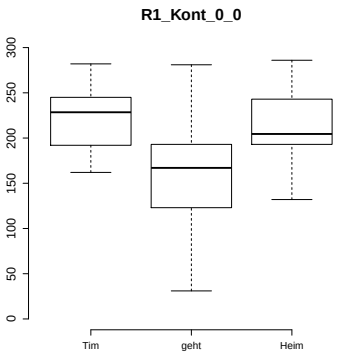

R1_31_1_0
0
5
10
15
20
25
30
Tim
geht
Heim

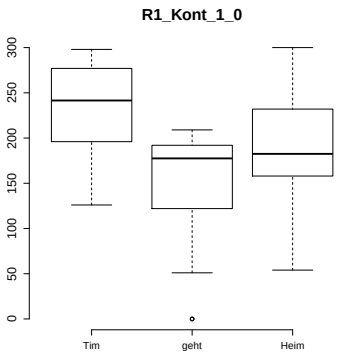

R1_31_2_0
0
5
10
15
20
25
30
Tim
geht
Heim

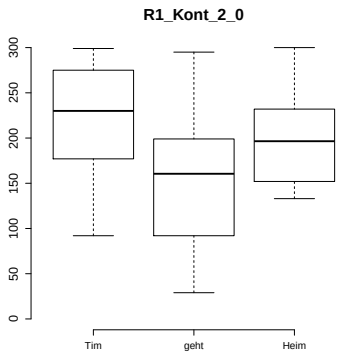

R1_31_0_1
0
5
10
15
20
25
30
Tim
geht
Heim

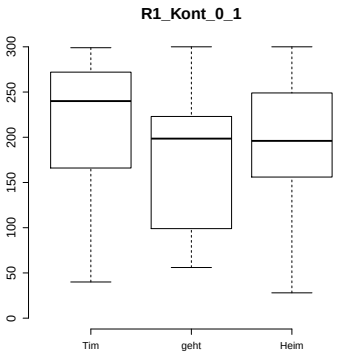

R1_31_1_1
0
5
10
15
20
25
30
Tim
geht
Heim

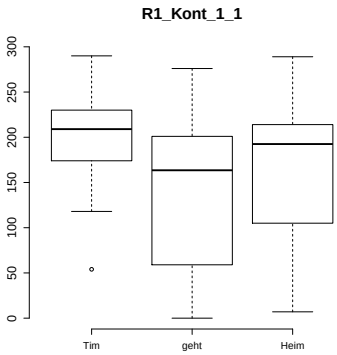

R1_31_2_1
0
5
10
15
20
25
30
Tim
geht
Heim

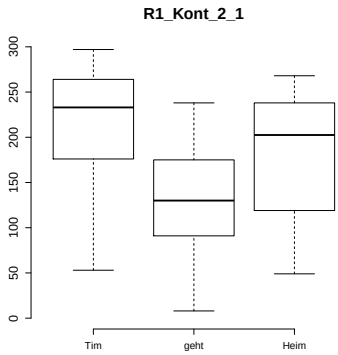

R1_Kont_0_0
0
50
100
150
200
250
300
Tim
geht
Heim

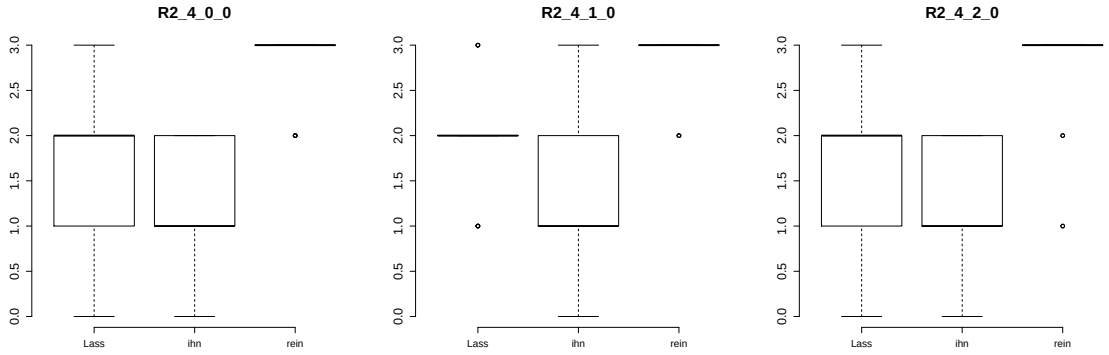

R1_Kont_1_0
0
50
100
150
200
250
300
Tim
geht
Heim

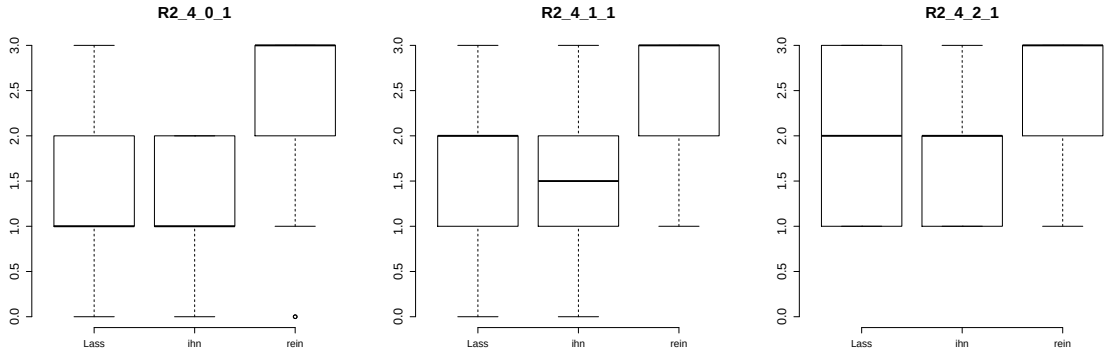

R1_Kont_2_0
0
50
100
150
200
250
300
Tim
geht
Heim

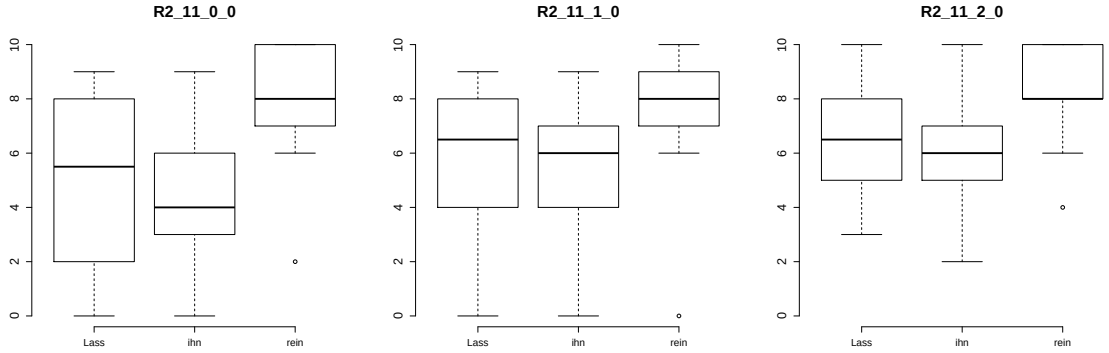

R1_Kont_0_1
0
50
100
150
200
250
300
Tim
geht
Heim

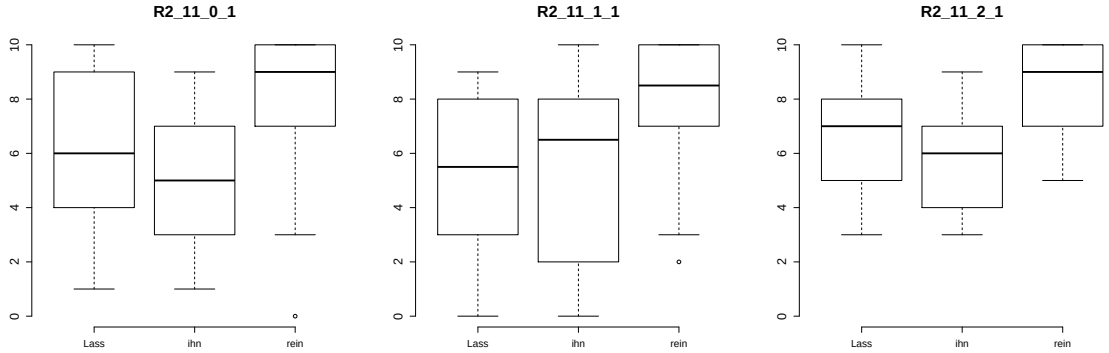

R1_Kont_1_1
0
50
100
150
200
250
300
Tim
geht
Heim

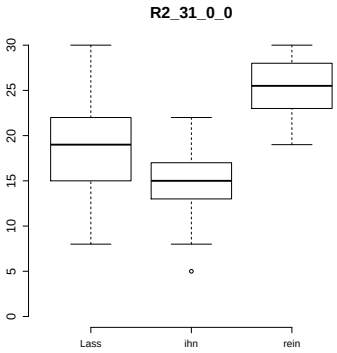

R1_Kont_2_1
0
50
100
150
200
250
300
Tim
geht
Heim

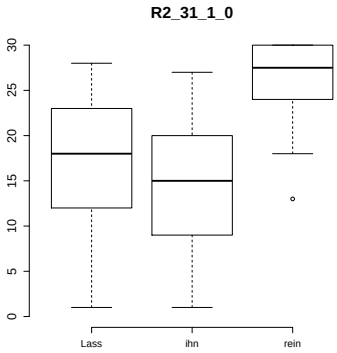

R2_4_0_0
R2_4_1_0
R2_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lass
ihn
rein

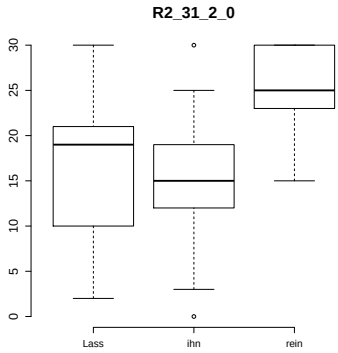

R2_4_0_1
R2_4_1_1
R2_4_2_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lass
ihn
rein

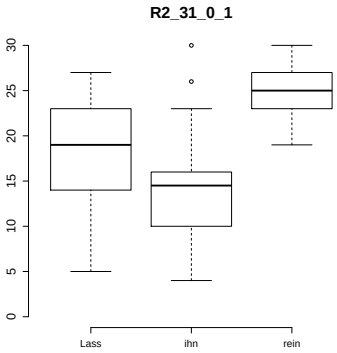

R2_11_0_0
R2_11_1_0
R2_11_2_0
0
2
4
6
8
10
Lass
ihn
rein

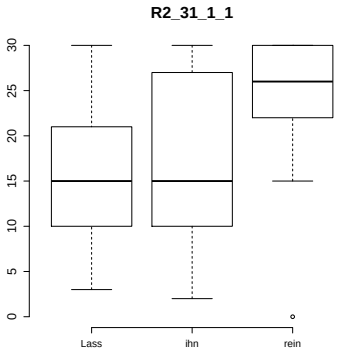

R2_11_0_1
R2_11_1_1
R2_11_2_1
0
2
4
6
8
10
Lass
ihn
rein

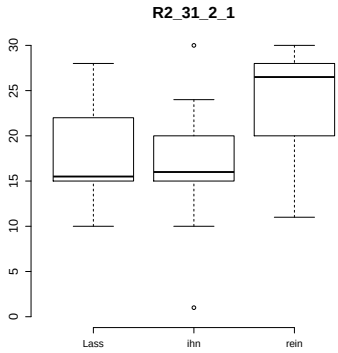

R2_31_0_0
0
5
10
15
20
25
30
Lass
ihn
rein

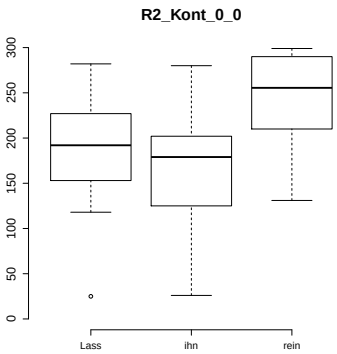

R2_31_1_0
0
5
10
15
20
25
30
Lass
ihn
rein

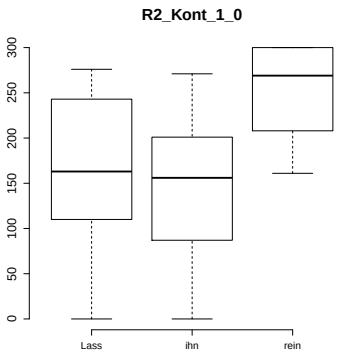

R2_31_2_0
0
5
10
15
20
25
30
Lass
ihn
rein

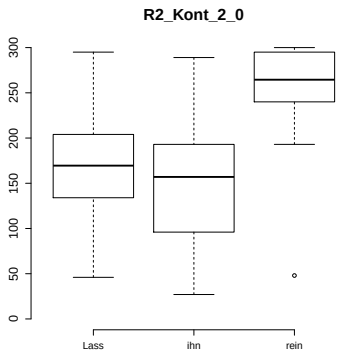

R2_31_0_1
0
5
10
15
20
25
30
Lass
ihn
rein

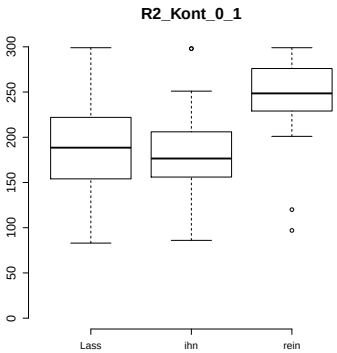

R2_31_1_1
0
5
10
15
20
25
30
Lass
ihn
rein

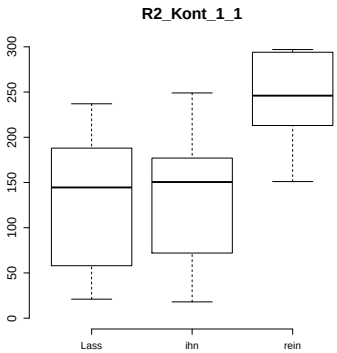

R2_31_2_1
0
5
10
15
20
25
30
Lass
ihn
rein

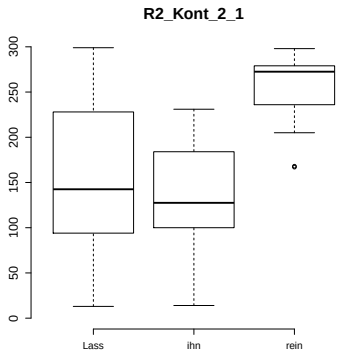

R2_Kont_0_0
0
50
100
150
200
250
300
Lass
ihn
rein

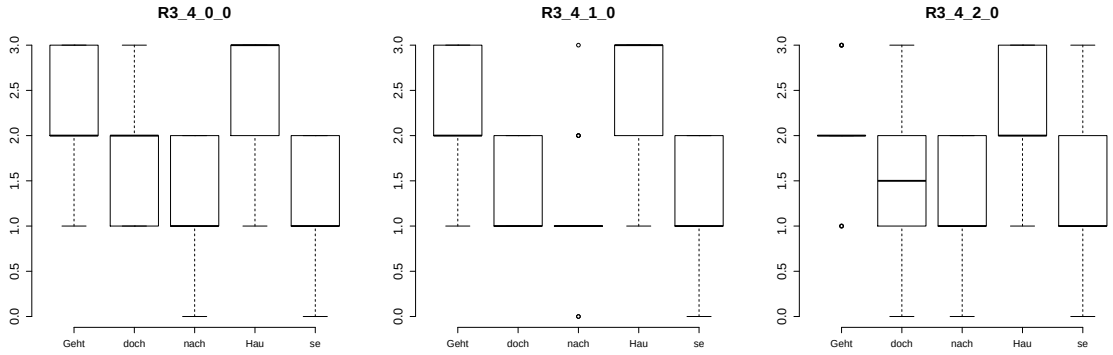

R2_Kont_1_0
0
50
100
150
200
250
300
Lass
ihn
rein

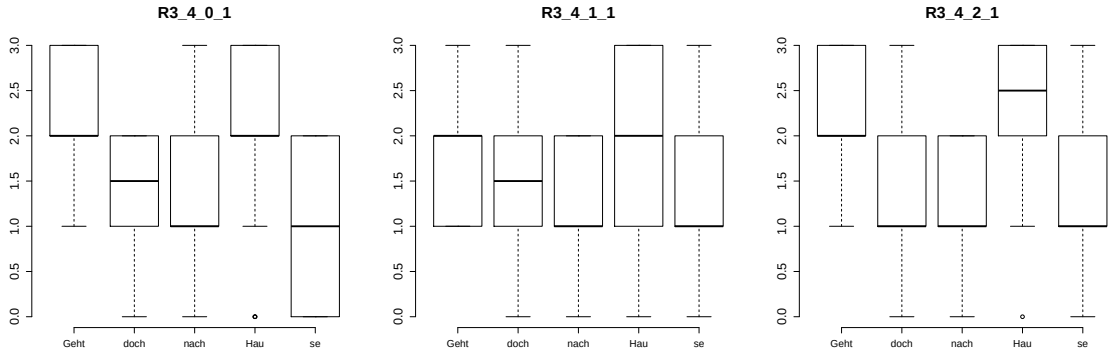

R2_Kont_2_0
0
50
100
150
200
250
300
Lass
ihn
rein

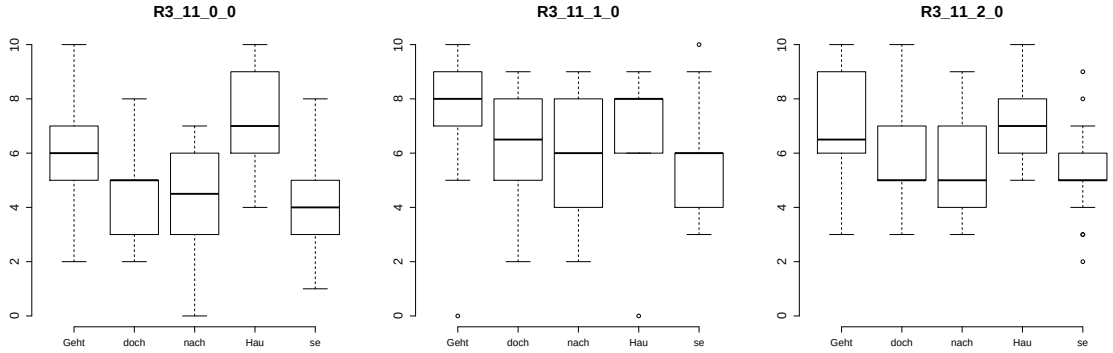

R2_Kont_0_1
0
50
100
150
200
250
300
Lass
ihn
rein

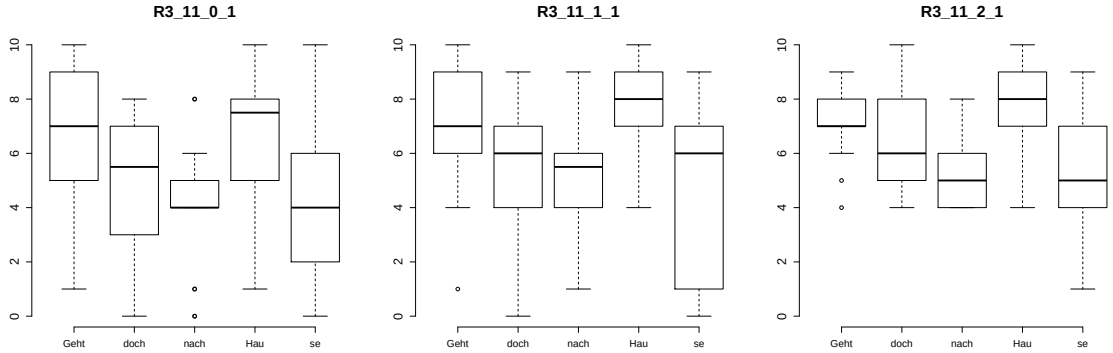

R2_Kont_1_1
0
50
100
150
200
250
300
Lass
ihn
rein

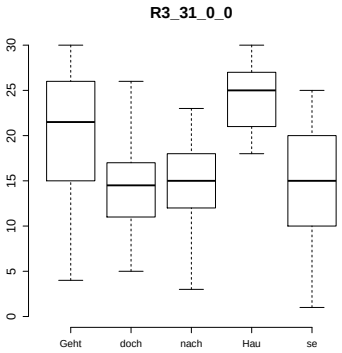

R2_Kont_2_1
0
50
100
150
200
250
300
Lass
ihn
rein

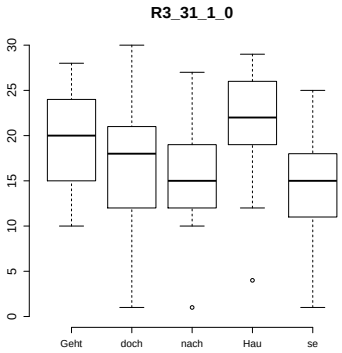

R3_4_0_0
R3_4_1_0
R3_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Geht
doch
nach
Hau
se

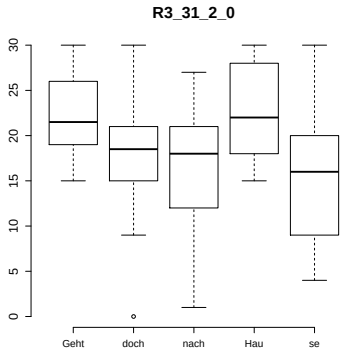

R3_4_0_1
R3_4_1_1
R3_4_2_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Geht
doch
nach
Hau
se

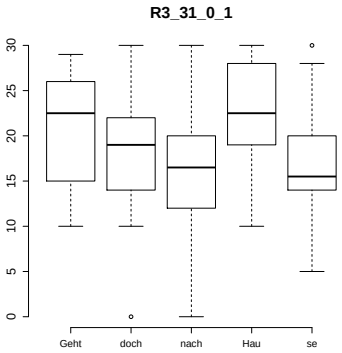

R3_11_0_0
R3_11_1_0
R3_11_2_0
0
2
4
6
8
10
Geht
doch
nach
Hau
se

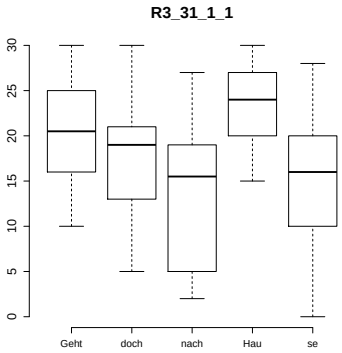

R3_11_0_1
R3_11_1_1
R3_11_2_1
0
2
4
6
8
10
Geht
doch
nach
Hau
se

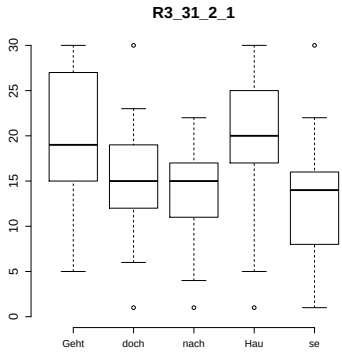

R3_31_0_0
0
5
10
15
20
25
30
Geht
doch
nach
Hau
se

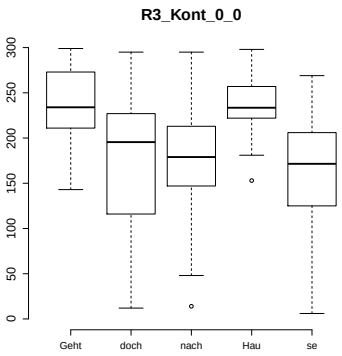

R3_31_1_0
0
5
10
15
20
25
30
Geht
doch
nach
Hau
se

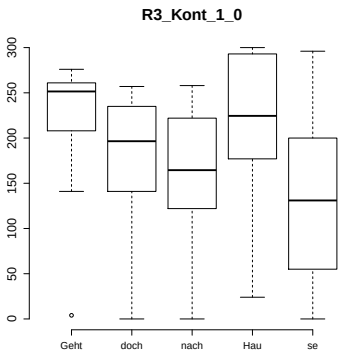

R3_31_2_0
0
5
10
15
20
25
30
Geht
doch
nach
Hau
se

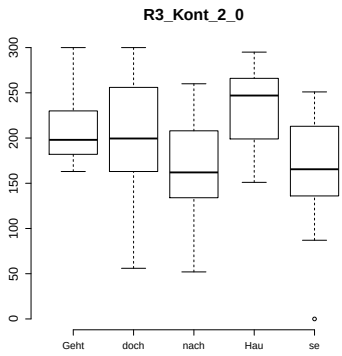

R3_31_0_1
0
5
10
15
20
25
30
Geht
doch
nach
Hau
se

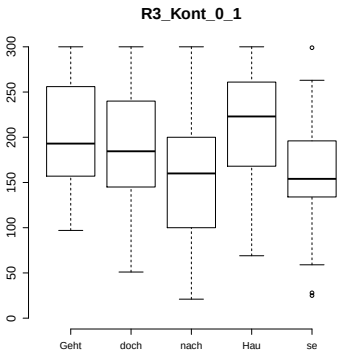

R3_31_1_1
0
5
10
15
20
25
30
Geht
doch
nach
Hau
se

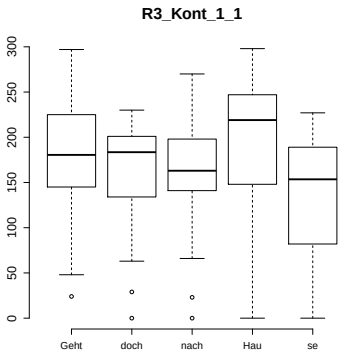

R3_31_2_1
0
5
10
15
20
25
30
Geht
doch
nach
Hau
se

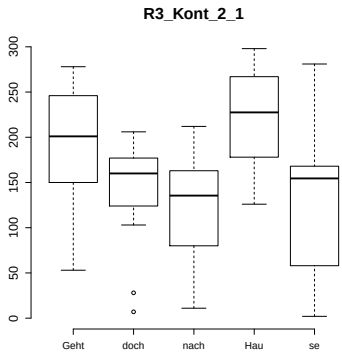

R3_Kont_0_0
0
50
100
150
200
250
300
Geht
doch
nach
Hau
se

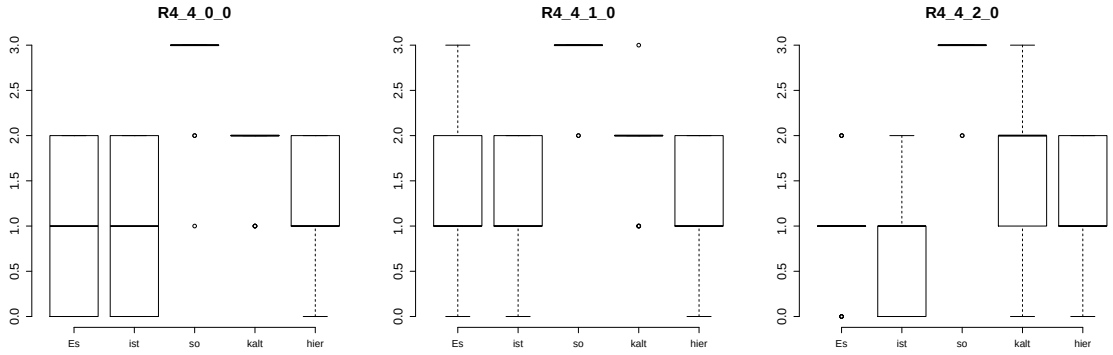

R3_Kont_1_0
0
50
100
150
200
250
300
Geht
doch
nach
Hau
se

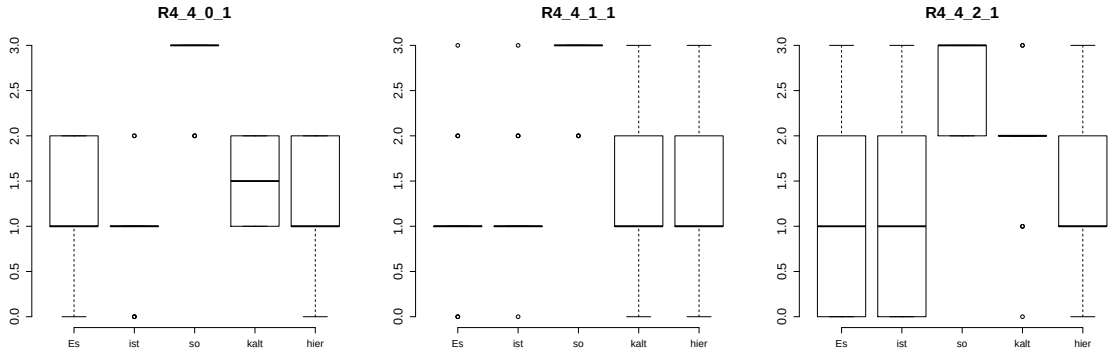

R3_Kont_2_0
0
50
100
150
200
250
300
Geht
doch
nach
Hau
se

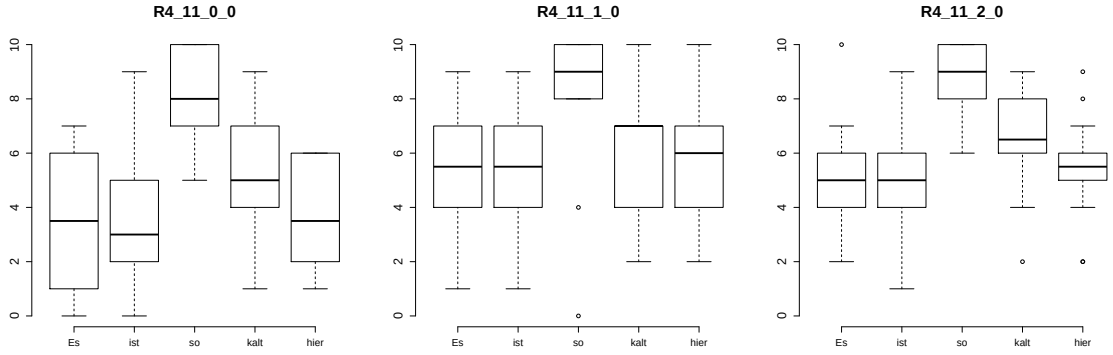

R3_Kont_0_1
0
50
100
150
200
250
300
Geht
doch
nach
Hau
se

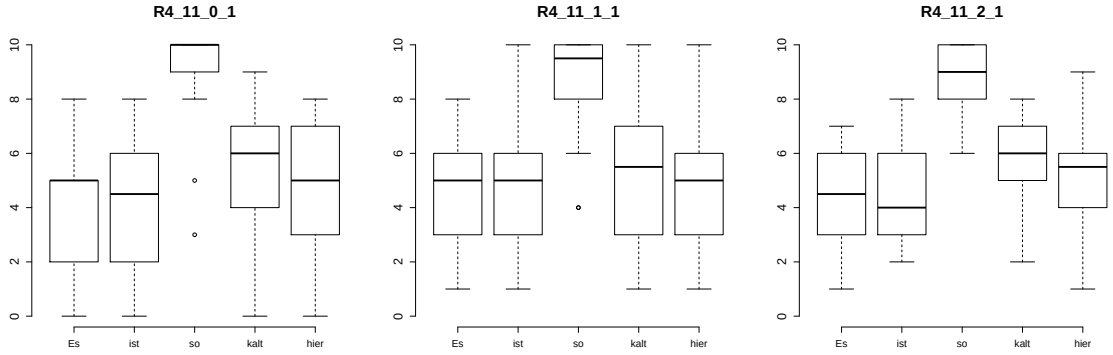

R3_Kont_1_1
0
50
100
150
200
250
300
Geht
doch
nach
Hau
se

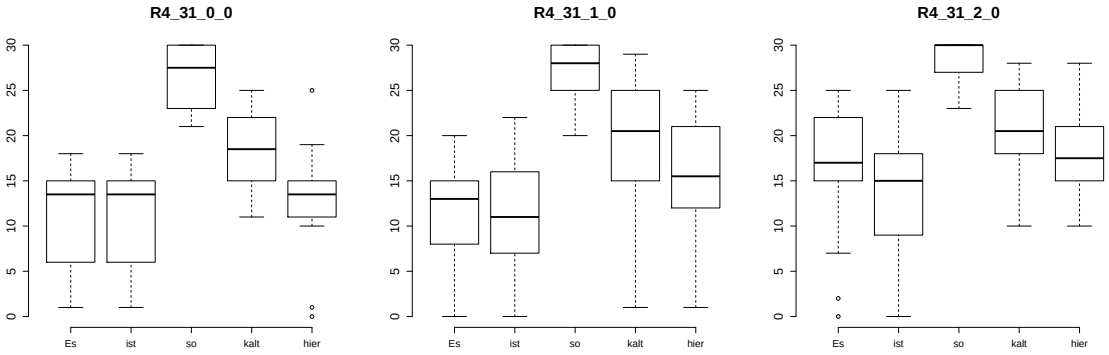

R3_Kont_2_1
0
50
100
150
200
250
300
Geht
doch
nach
Hau
se

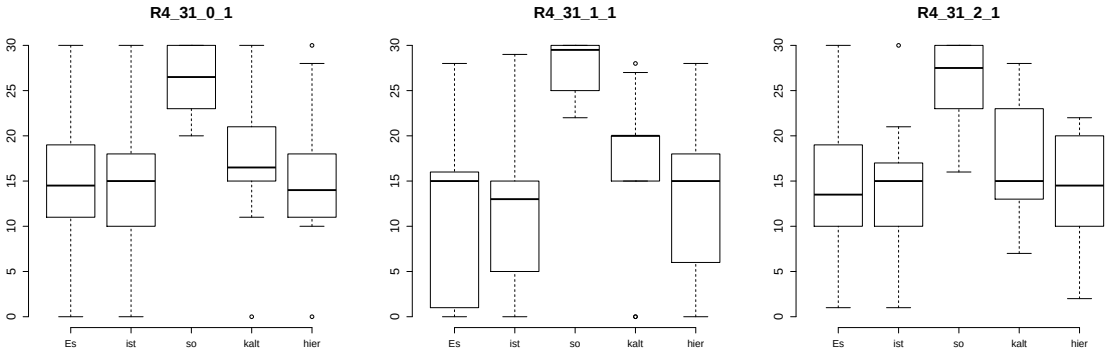

R4_4_0_0
R4_4_1_0
R4_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Es
ist
so
kalt
hier

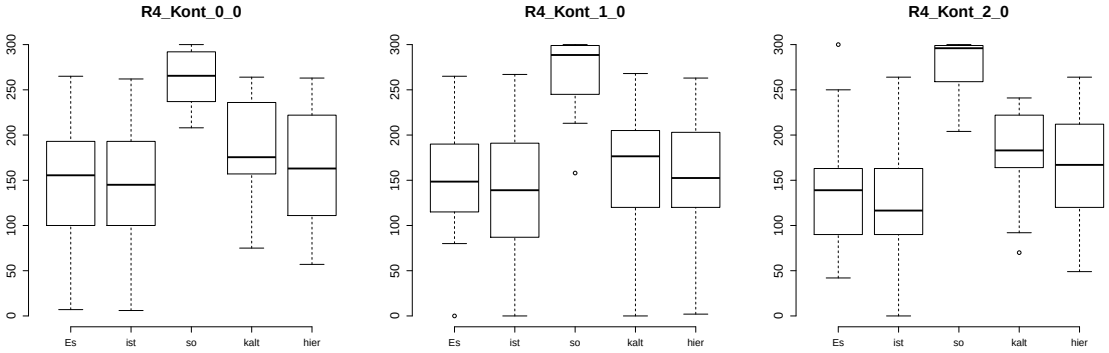

R4_4_0_1
R4_4_1_1
R4_4_2_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Es
ist
so
kalt
hier

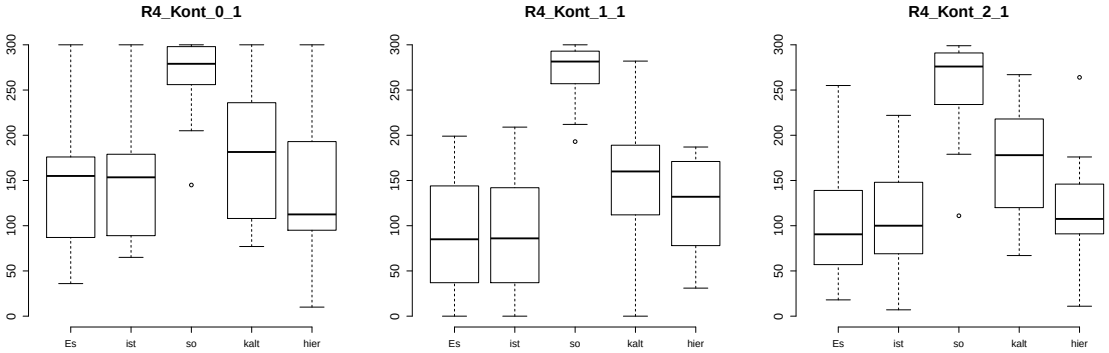

R4_11_0_0
R4_11_1_0
R4_11_2_0
0
2
4
6
8
10
Es
ist
so
kalt
hier

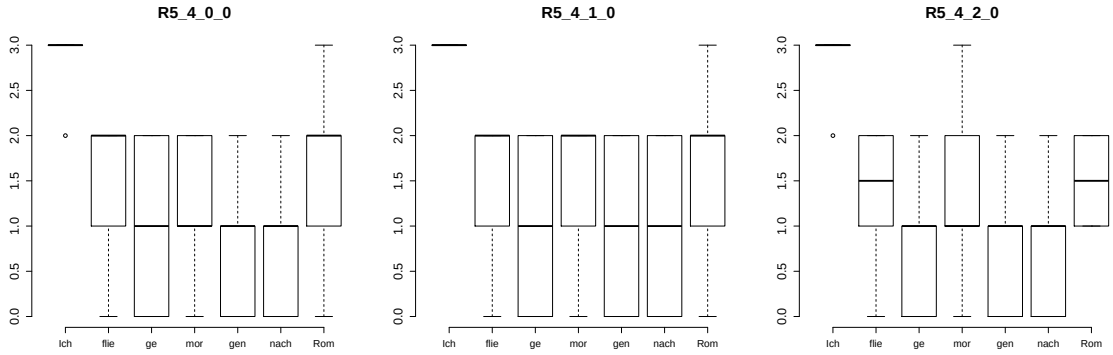

R4_11_0_1
R4_11_1_1
R4_11_2_1
0
2
4
6
8
10
Es
ist
so
kalt
hier

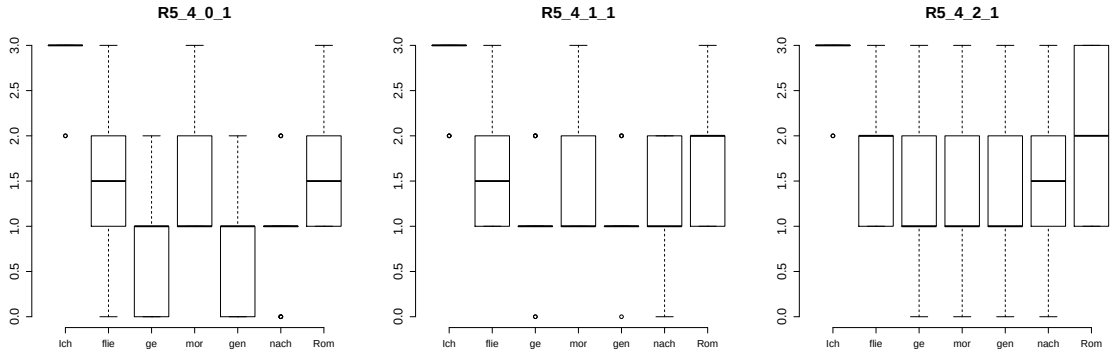

R4_31_0_0
R4_31_1_0
R4_31_2_0
0 5 10 15 20 25 30
Es ist so kalt hier

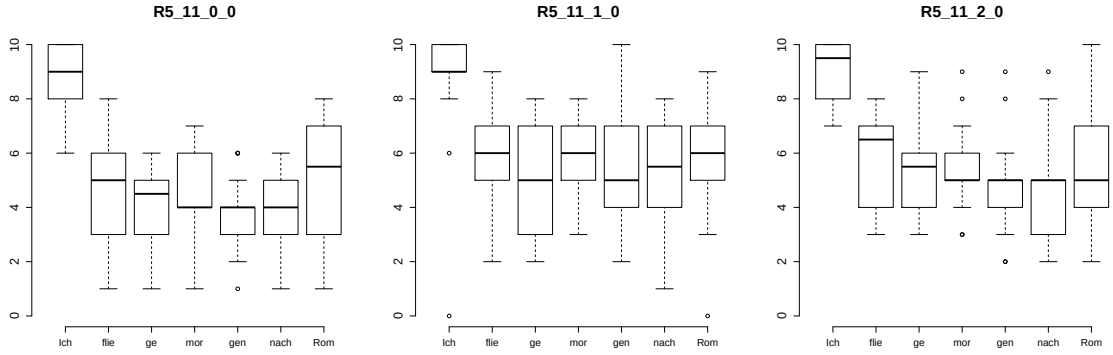

R4_31_0_1
R4_31_1_1
R4_31_2_1
0 5 10 15 20 25 30
Es ist so kalt hier

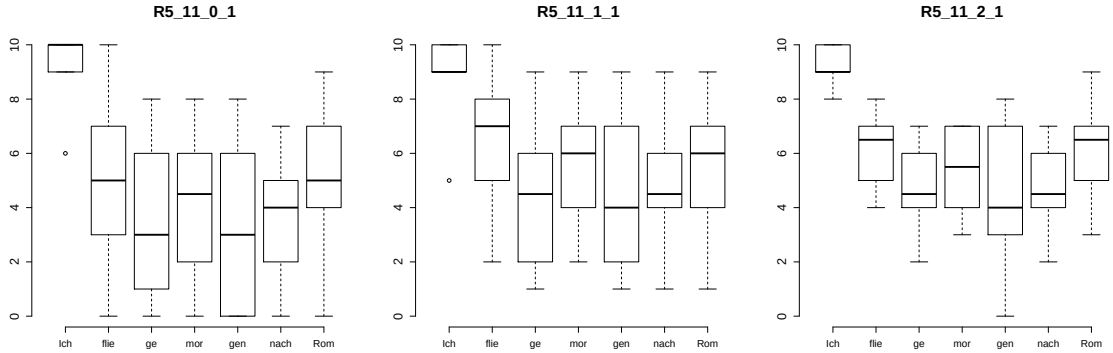

R4_Kont_0_0
R4_Kont_1_0
R4_Kont_2_0
0 50 100 150 200 250 300
Es ist so kalt hier

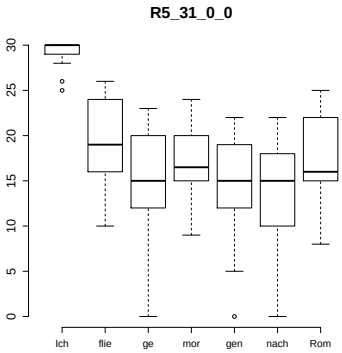

R4_Kont_0_1
R4_Kont_1_1
R4_Kont_2_1
0 50 100 150 200 250 300
Es ist so kalt hier

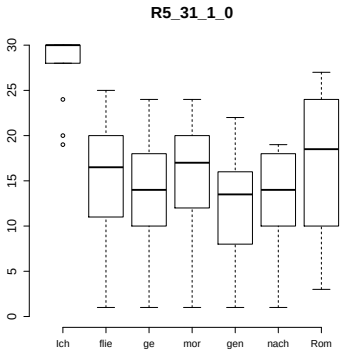

R5_4_0_0
R5_4_1_0
R5_4_2_0
Ich
flie
ge
mor
gen
nach
Rom

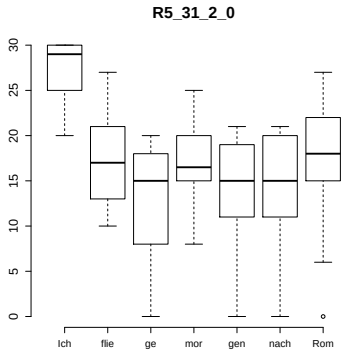

R5_4_0_1
R5_4_1_1
R5_4_2_1
Ich
flie
ge
mor
gen
nach
Rom

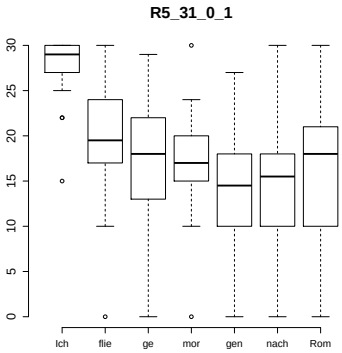

R5_11_0_0
R5_11_1_0
R5_11_2_0
Ich
flie
ge
mor
gen
nach
Rom

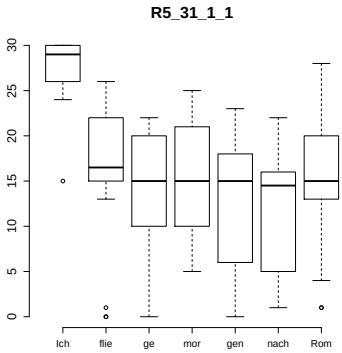

R5_11_0_1
R5_11_1_1
R5_11_2_1
Ich
flie
ge
mor
gen
nach
Rom

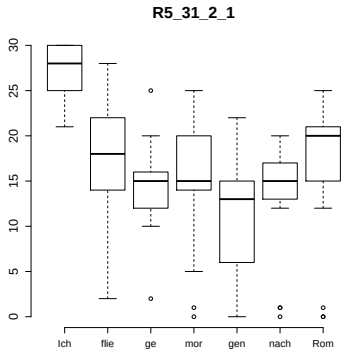

R5_31_0_0
0
5
10
15
20
25
30
Ich
flie
ge
mor
gen
nach
Rom

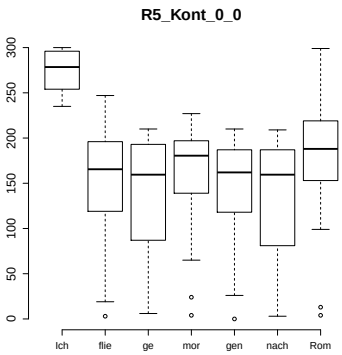

R5_31_1_0
0
5
10
15
20
25
30
Ich
flie
ge
mor
gen
nach
Rom

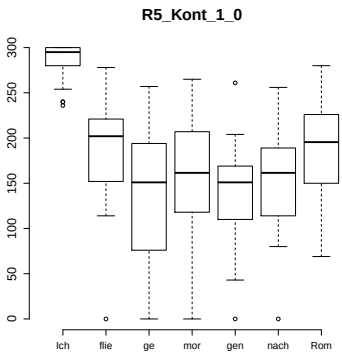

R5_31_2_0
0
5
10
15
20
25
30
Ich
flie
ge
mor
gen
nach
Rom

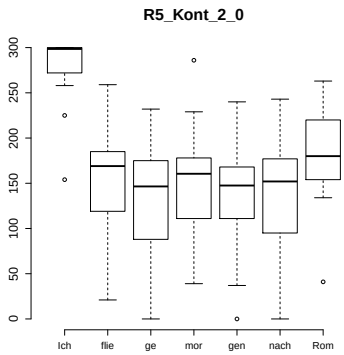

R5_31_0_1
0
5
10
15
20
25
30
Ich
flie
ge
mor
gen
nach
Rom

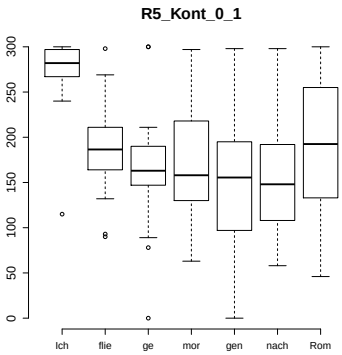

R5_31_1_1
0
5
10
15
20
25
30
Ich
flie
ge
mor
gen
nach
Rom

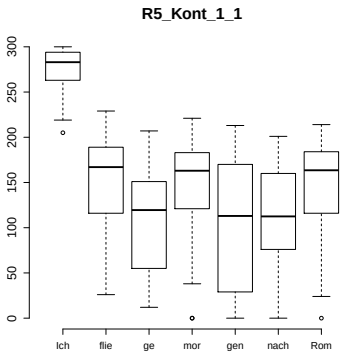

R5_31_2_1
0
5
10
15
20
25
30
Ich
flie
ge
mor
gen
nach
Rom

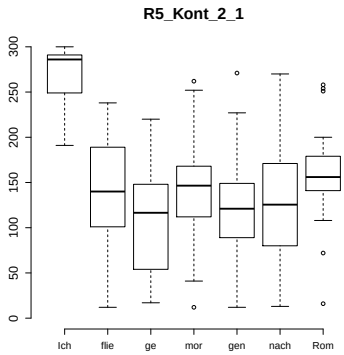

R5_Kont_0_0
0
50
100
150
200
250
300
Ich
flie
ge
mor
gen
nach
Rom

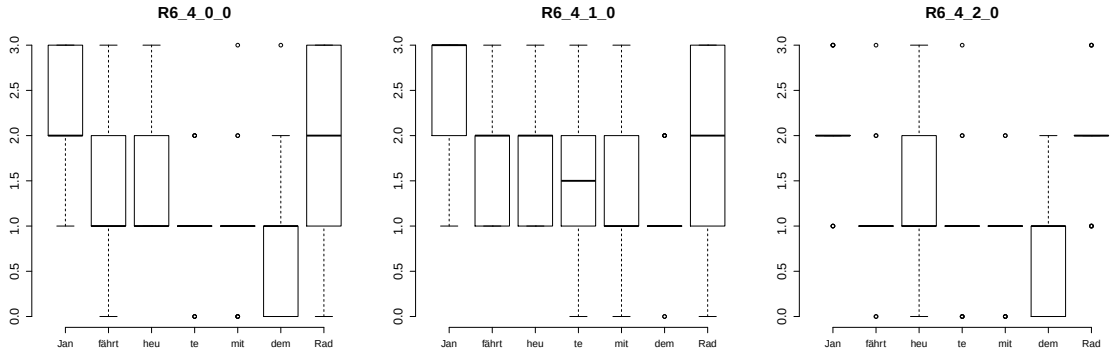

R5_Kont_1_0
0
50
100
150
200
250
300
Ich
flie
ge
mor
gen
nach
Rom

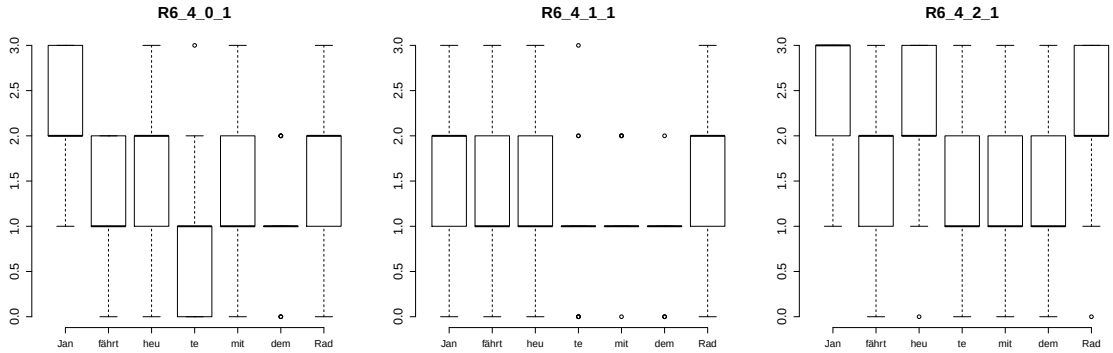

R5_Kont_2_0
0
50
100
150
200
250
300
Ich
flie
ge
mor
gen
nach
Rom

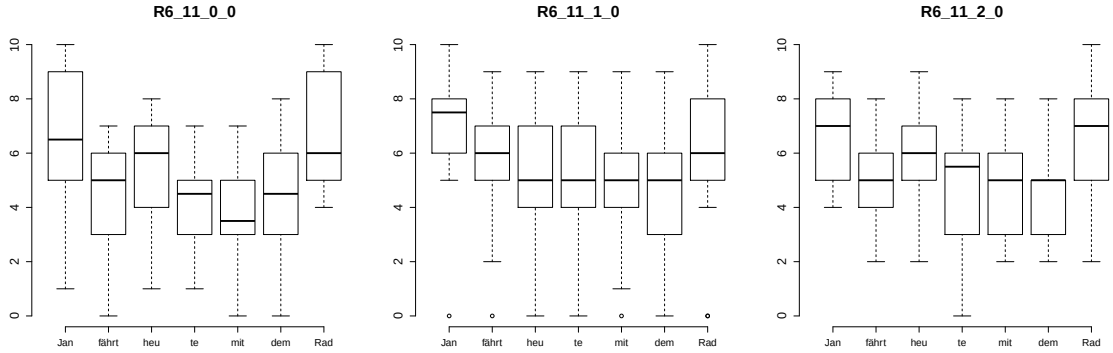

R5_Kont_0_1
0
50
100
150
200
250
300
Ich
flie
ge
mor
gen
nach
Rom

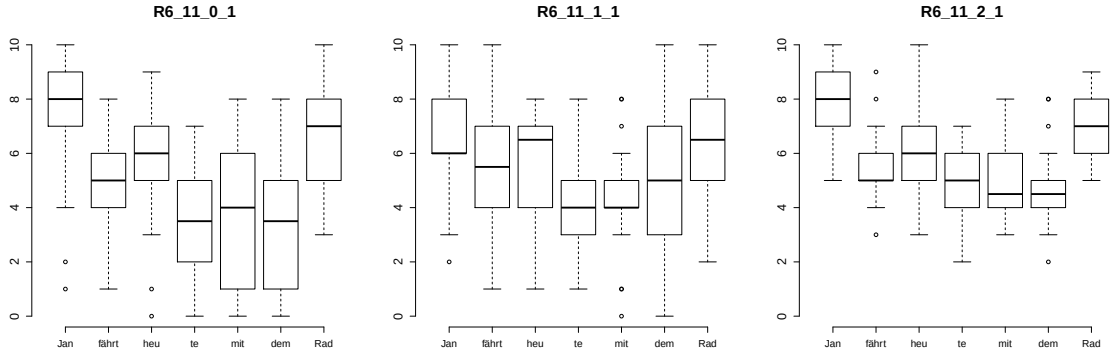

R5_Kont_1_1
0
50
100
150
200
250
300
Ich
flie
ge
mor
gen
nach
Rom

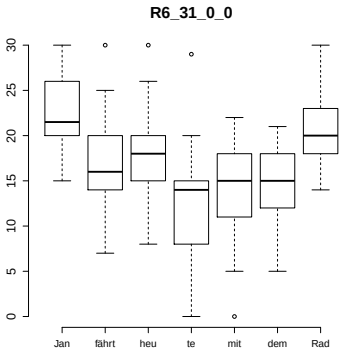

R5_Kont_2_1
0
50
100
150
200
250
300
Ich
flie
ge
mor
gen
nach
Rom

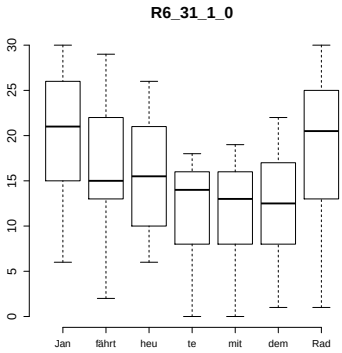

R6_4_0_0
R6_4_1_0
R6_4_2_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Jan fährt heu te mit dem Rad

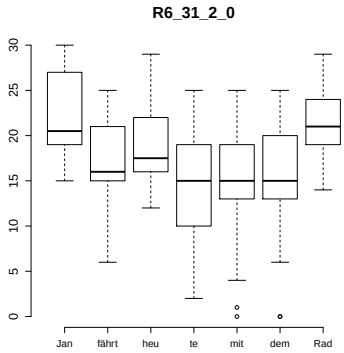

R6_4_0_1
R6_4_1_1
R6_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Jan fährt heu te mit dem Rad

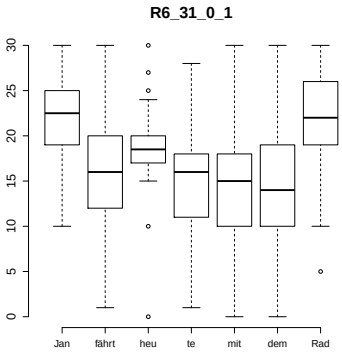

R6_11_0_0
R6_11_1_0
R6_11_2_0
0 2 4 6 8 10
Jan fährt heu te mit dem Rad

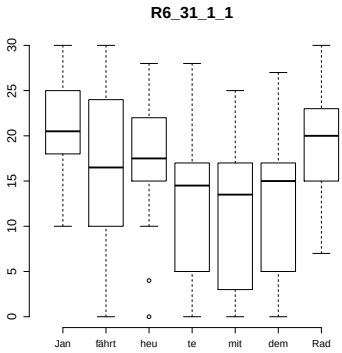

R6_11_0_1
R6_11_1_1
R6_11_2_1
0 2 4 6 8 10
Jan fährt heu te mit dem Rad

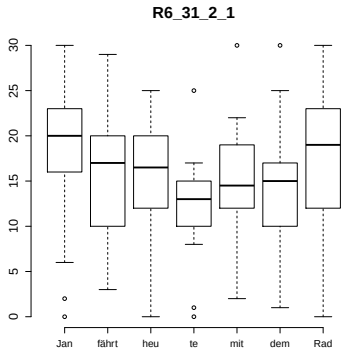

R6_31_0_0
0 5 10 15 20 25 30
Jan fährt heu te mit dem Rad

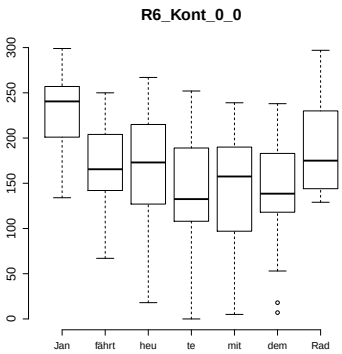

R6_31_1_0
0 5 10 15 20 25 30
Jan fährt heu te mit dem Rad

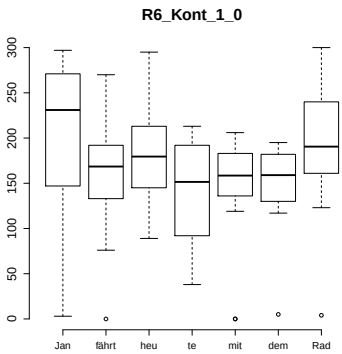

R6_31_2_0
0 5 10 15 20 25 30
Jan fährt heu te mit dem Rad

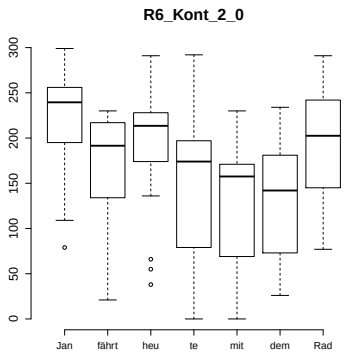

R6_31_0_1
0 5 10 15 20 25 30
Jan fährt heu te mit dem Rad

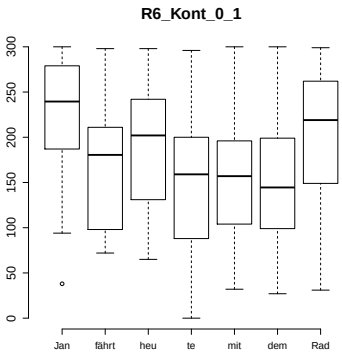

R6_31_1_1
0 5 10 15 20 25 30
Jan fährt heu te mit dem Rad

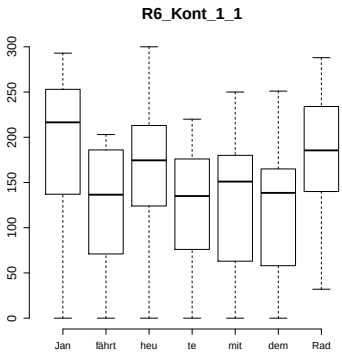

R6_31_2_1
0 5 10 15 20 25 30
Jan fährt heu te mit dem Rad

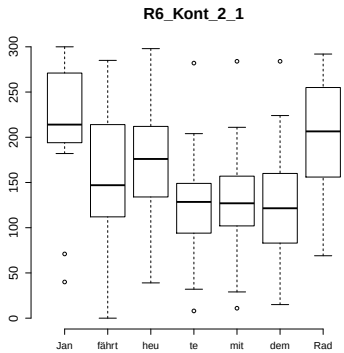

R6_Kont_0_0
0 50 100 150 200 250 300
Jan fährt heu te mit dem Rad

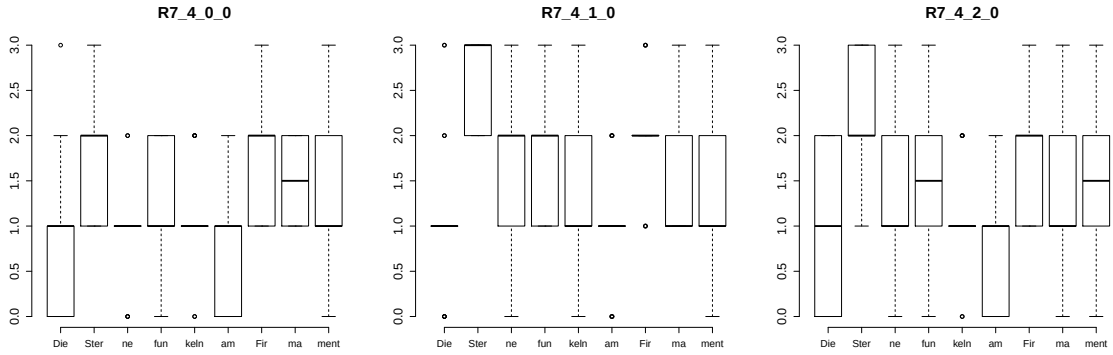

R6_Kont_1_0
0 50 100 150 200 250 300
Jan fährt heu te mit dem Rad

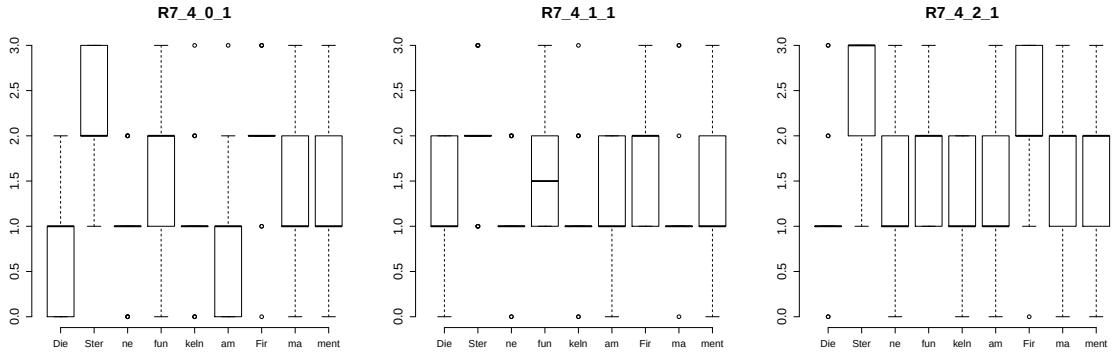

R6_Kont_2_0
0 50 100 150 200 250 300
Jan fährt heu te mit dem Rad

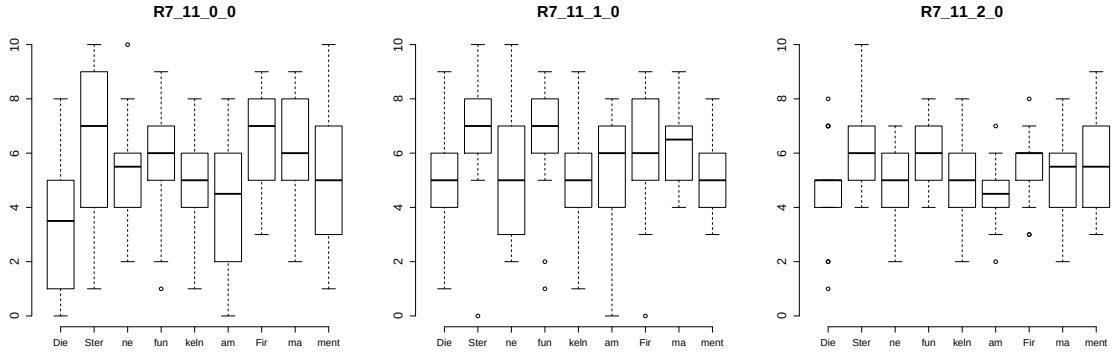

R6_Kont_0_1
0 50 100 150 200 250 300
Jan fährt heu te mit dem Rad

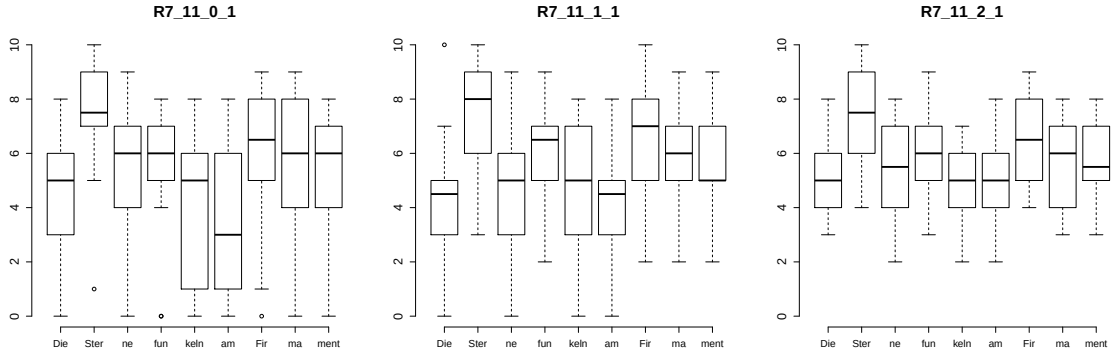

R6_Kont_1_1
0 50 100 150 200 250 300
Jan fährt heu te mit dem Rad

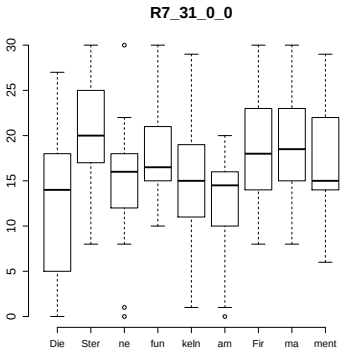

R6_Kont_2_1
0 50 100 150 200 250 300
Jan fährt heu te mit dem Rad

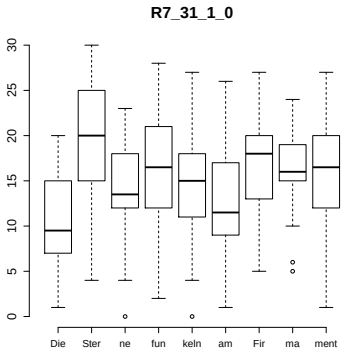

R7_4_0_0
R7_4_1_0
R7_4_2_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Die Ster ne fun keln am Fir ma ment

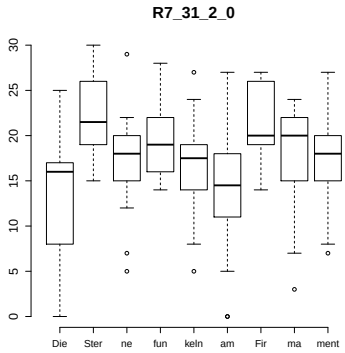

R7_4_0_1
R7_4_1_1
R7_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Die Ster ne fun keln am Fir ma ment

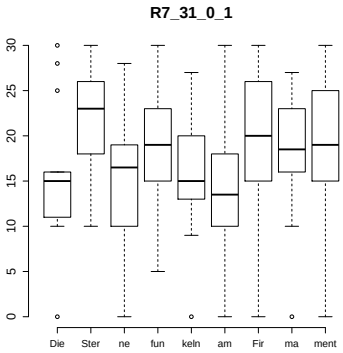

R7_11_0_0
R7_11_1_0
R7_11_2_0
0 2 4 6 8 10
Die Ster ne fun keln am Fir ma ment

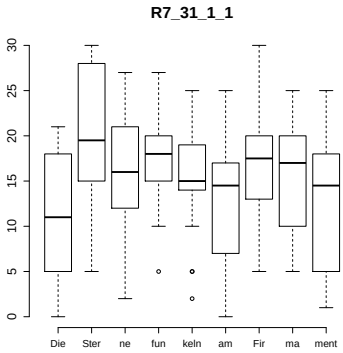

R7_11_0_1
R7_11_1_1
R7_11_2_1
0 2 4 6 8 10
Die Ster ne fun keln am Fir ma ment

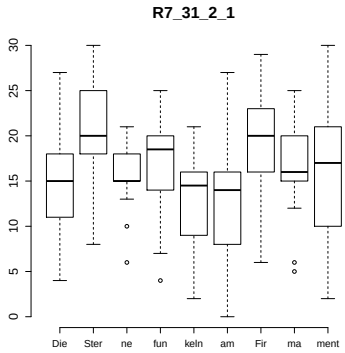

R7_31_0_0
0 5 10 15 20 25 30
Die Ster ne fun keln am Fir ma ment

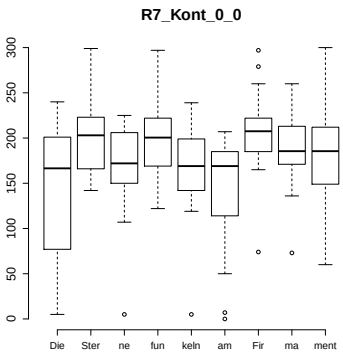

R7_31_1_0
0 5 10 15 20 25 30
Die Ster ne fun keln am Fir ma ment

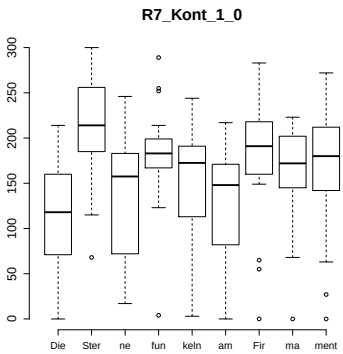

R7_31_2_0
0 5 10 15 20 25 30
Die Ster ne fun keln am Fir ma ment

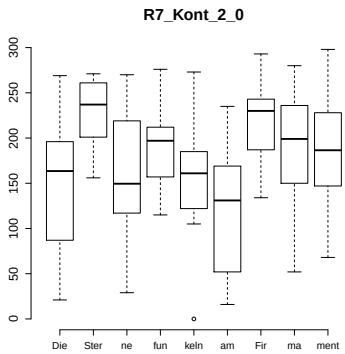

R7_31_0_1
0 5 10 15 20 25 30
Die Ster ne fun keln am Fir ma ment

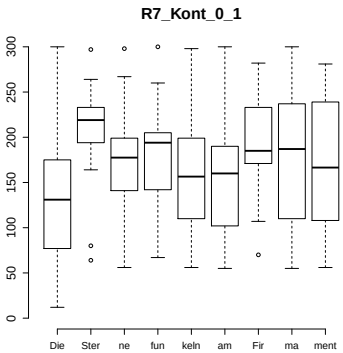

R7_31_1_1
0 5 10 15 20 25 30
Die Ster ne fun keln am Fir ma ment

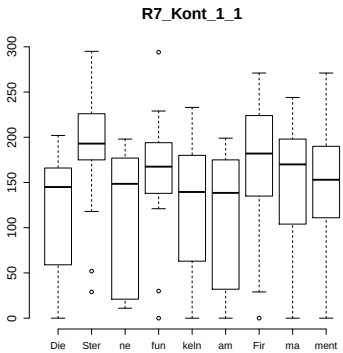

R7_31_2_1
0 5 10 15 20 25 30
Die Ster ne fun keln am Fir ma ment

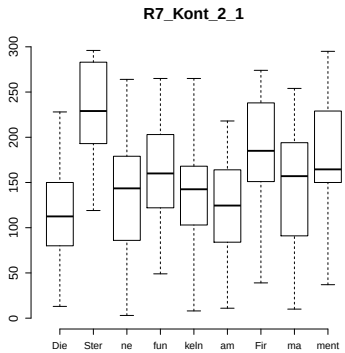

R7_Kont_0_0
0 50 100 150 200 250 300
Die Ster ne fun keln am Fir ma ment

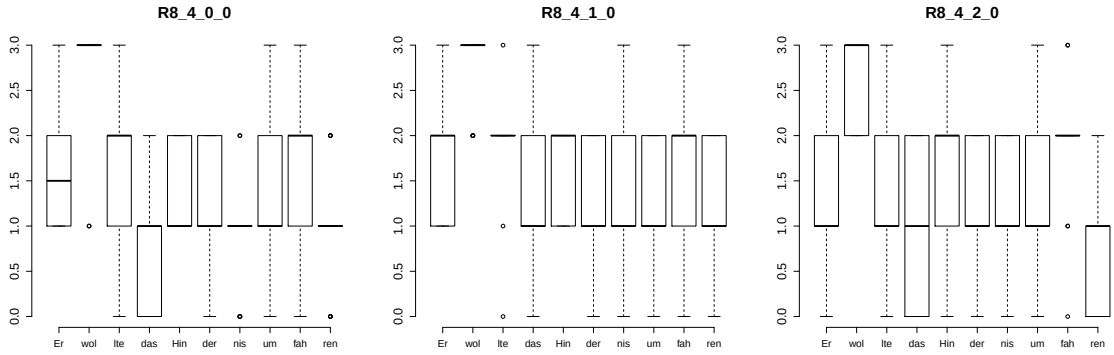

R7_Kont_1_0
0 50 100 150 200 250 300
Die Ster ne fun keln am Fir ma ment

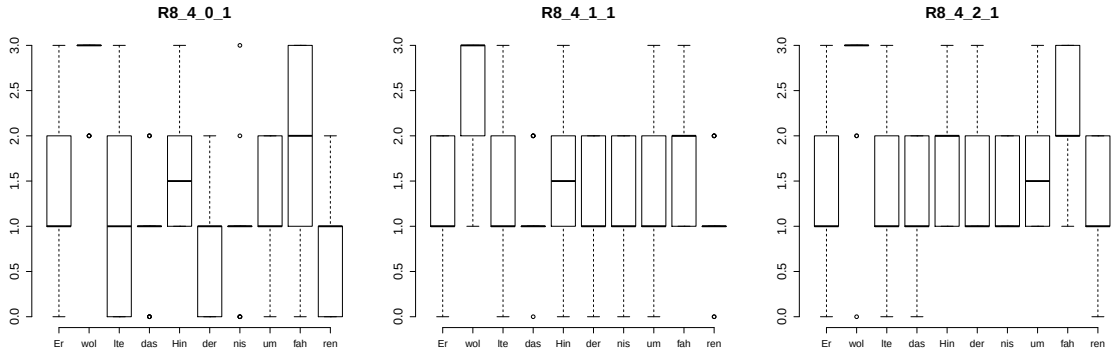

R7_Kont_2_0
0 50 100 150 200 250 300
Die Ster ne fun keln am Fir ma ment

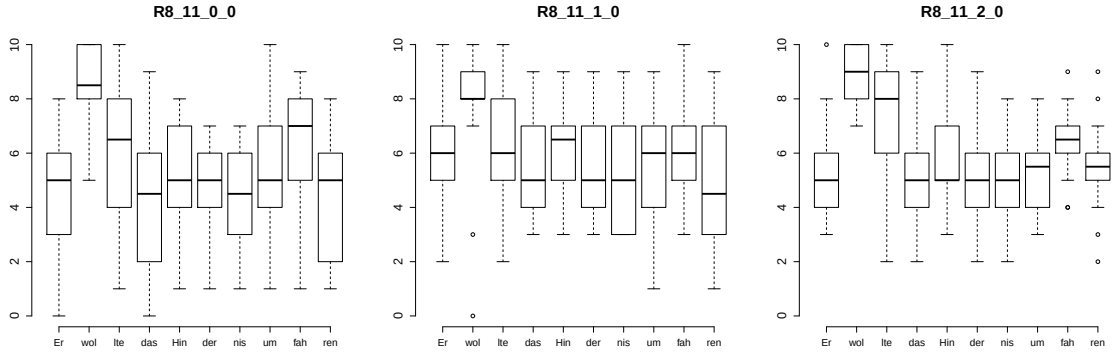

R7_Kont_0_1
0 50 100 150 200 250 300
Die Ster ne fun keln am Fir ma ment

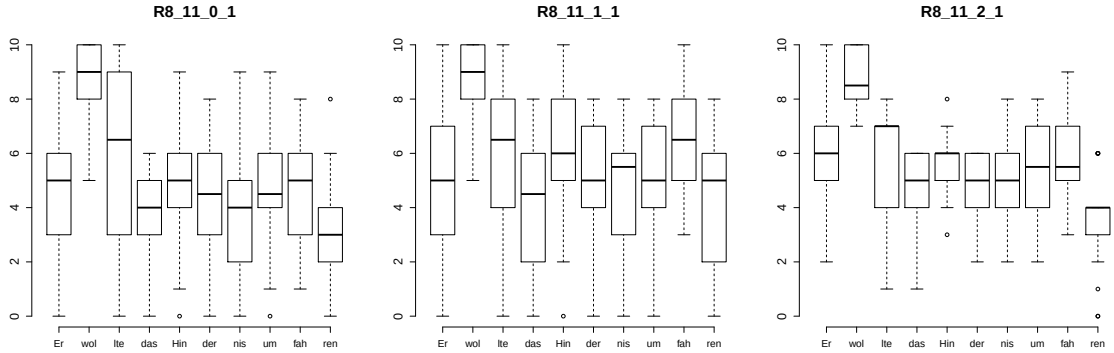

R7_Kont_1_1
0 50 100 150 200 250 300
Die Ster ne fun keln am Fir ma ment

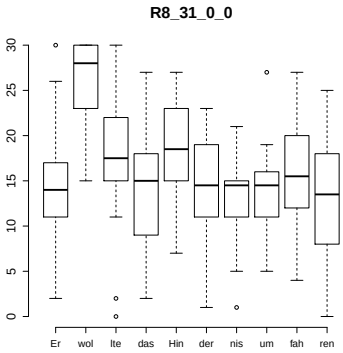

R7_Kont_2_1
0 50 100 150 200 250 300
Die Ster ne fun keln am Fir ma ment

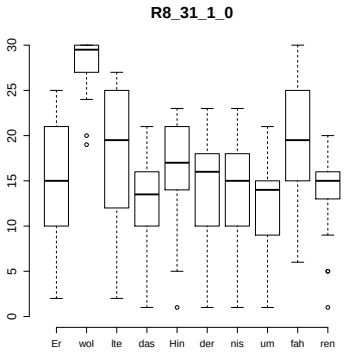

R8_4_0_0
R8_4_1_0
R8_4_2_0
Er wol lte das Hin der nis um fah ren

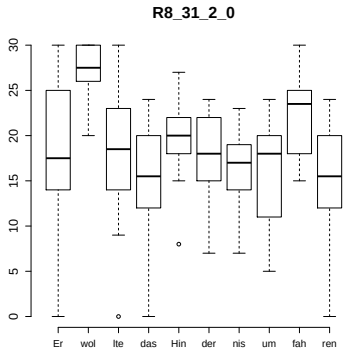

R8_4_0_1
R8_4_1_1
R8_4_2_1
Er wol lte das Hin der nis um fah ren

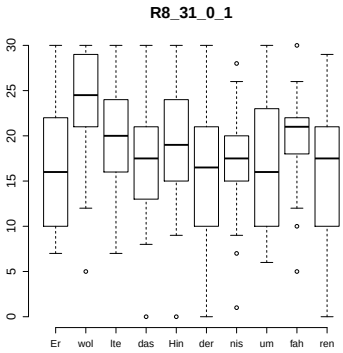

R8_11_0_0
R8_11_1_0
R8_11_2_0
Er wol lte das Hin der nis um fah ren

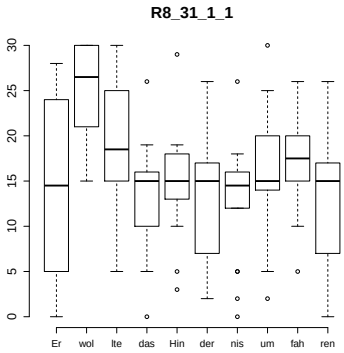

R8_11_0_1
R8_11_1_1
R8_11_2_1
Er wol lte das Hin der nis um fah ren

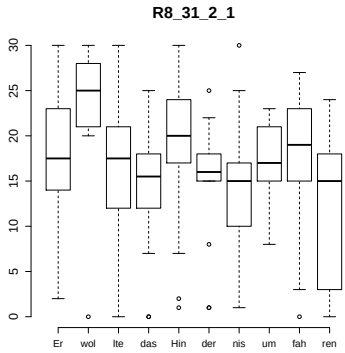

R8_31_0_0
0 5 10 15 20 25 30
Er wol lte das Hin der nis um fah ren

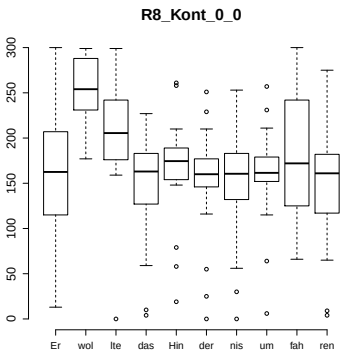

R8_31_1_0
0 5 10 15 20 25 30
Er wol lte das Hin der nis um fah ren

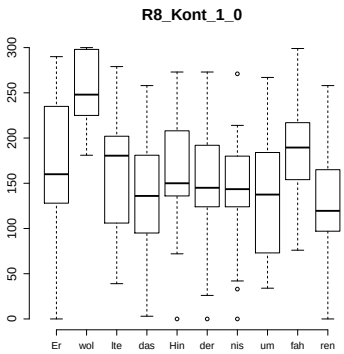

R8_31_2_0
0 5 10 15 20 25 30
Er wol lte das Hin der nis um fah ren

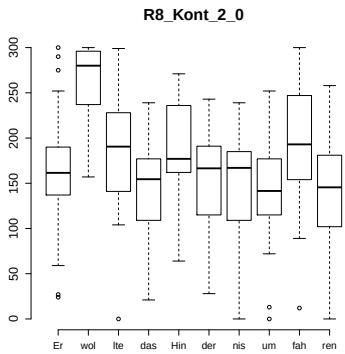

R8_31_0_1
0 5 10 15 20 25 30
Er wol lte das Hin der nis um fah ren

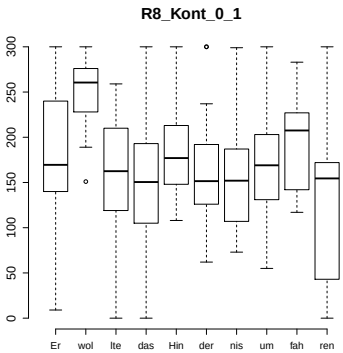

R8_31_1_1
0 5 10 15 20 25 30
Er wol lte das Hin der nis um fah ren

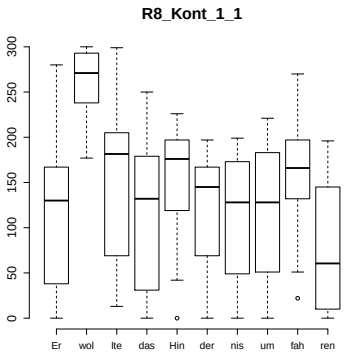

R8_31_2_1
0 5 10 15 20 25 30
Er wol lte das Hin der nis um fah ren

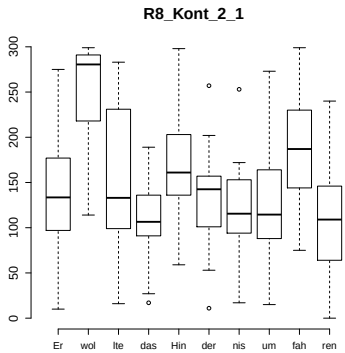

R8_Kont_0_0
0 50 100 150 200 250 300
Er wol lte das Hin der nis um fah ren

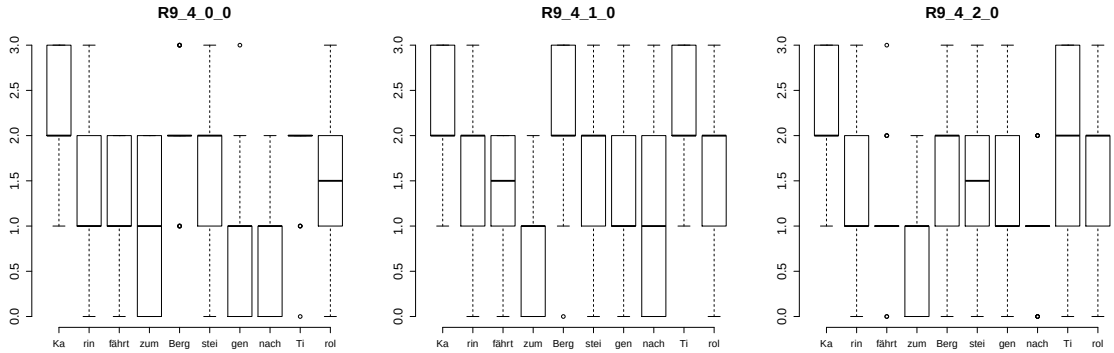

R8_Kont_1_0
0 50 100 150 200 250 300
Er wol lte das Hin der nis um fah ren

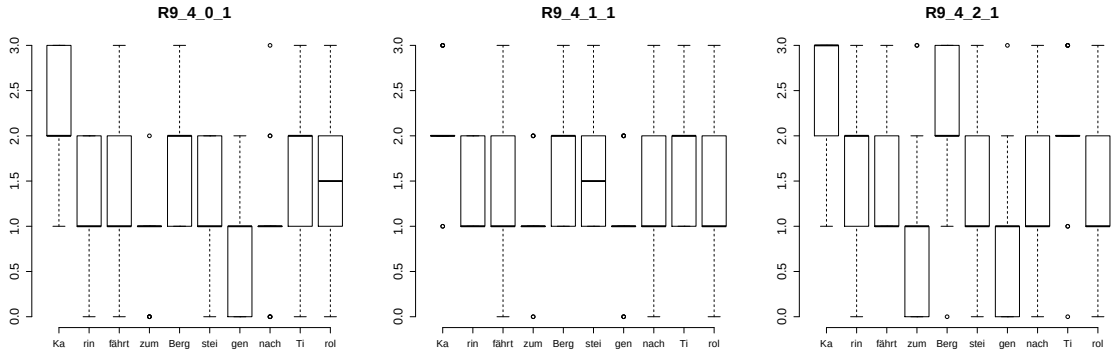

R8_Kont_2_0
0 50 100 150 200 250 300
Er wol lte das Hin der nis um fah ren

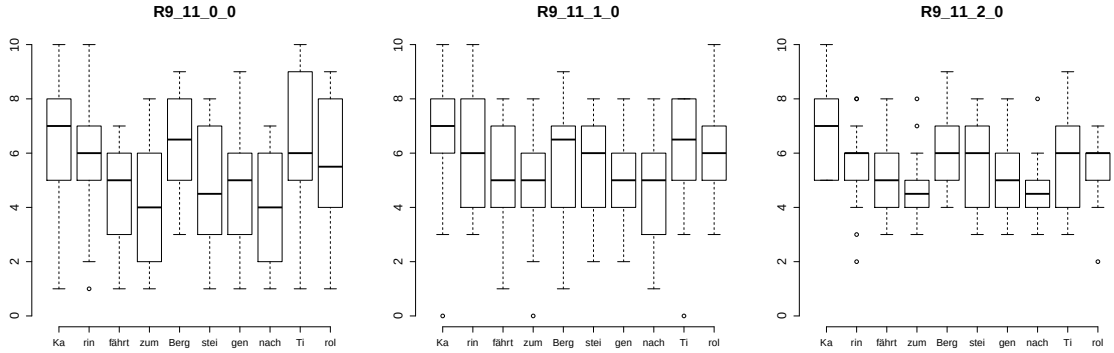

R8_Kont_0_1
0 50 100 150 200 250 300
Er wol lte das Hin der nis um fah ren

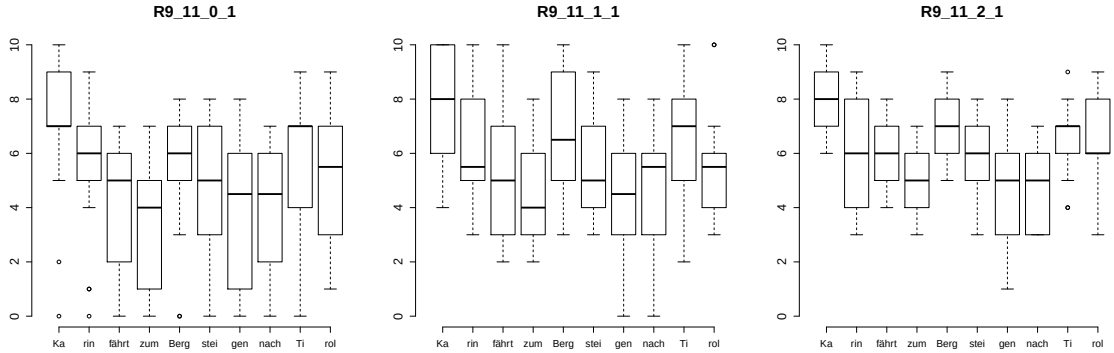

R8_Kont_1_1
0 50 100 150 200 250 300
Er wol lte das Hin der nis um fah ren

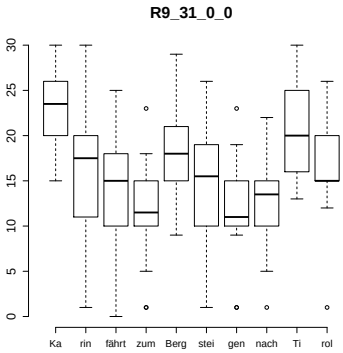

R8_Kont_2_1
0 50 100 150 200 250 300
Er wol lte das Hin der nis um fah ren

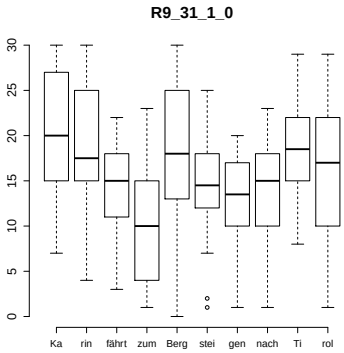

R9_4_0_0
R9_4_1_0
R9_4_2_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Ka rin fährt zum Berg stei gen nach Ti rol

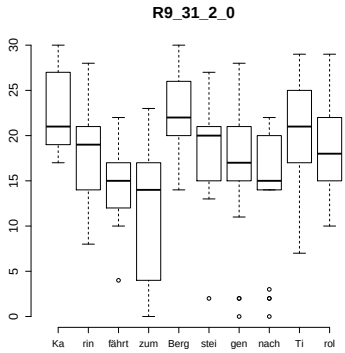

R9_4_0_1
R9_4_1_1
R9_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Ka rin fährt zum Berg stei gen nach Ti rol

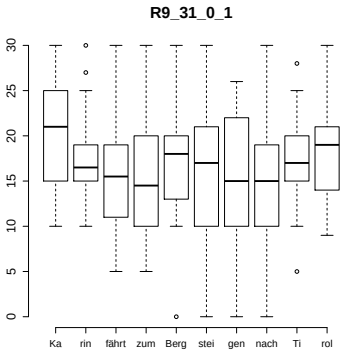

R9_11_0_0
R9_11_1_0
R9_11_2_0
0 2 4 6 8 10
Ka rin fährt zum Berg stei gen nach Ti rol

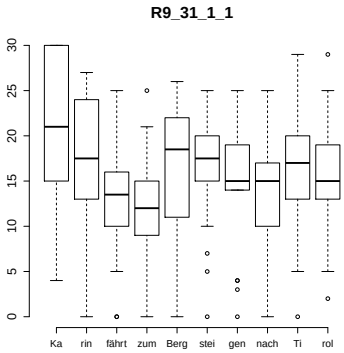

R9_11_0_1
R9_11_1_1
R9_11_2_1
0 2 4 6 8 10
Ka rin fährt zum Berg stei gen nach Ti rol

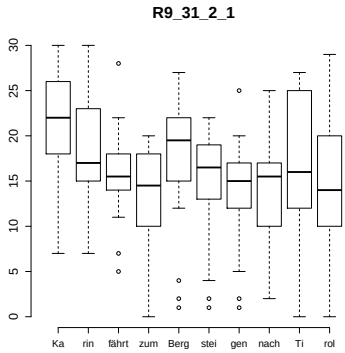

R9_31_0_0
0 5 10 15 20 25 30
Ka rin fährt zum Berg stei gen nach Ti rol

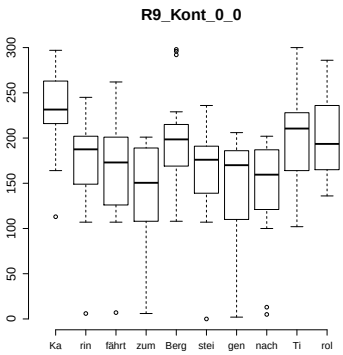

R9_31_1_0
0 5 10 15 20 25 30
Ka rin fährt zum Berg stei gen nach Ti rol

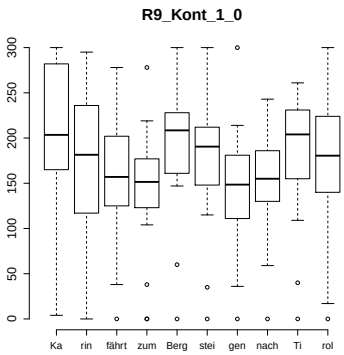

R9_31_2_0
0 5 10 15 20 25 30
Ka rin fährt zum Berg stei gen nach Ti rol

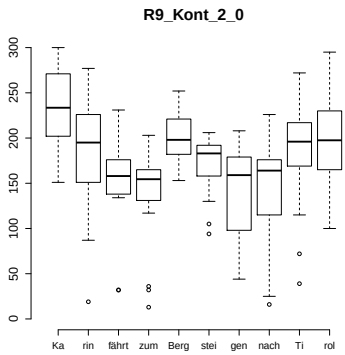

R9_31_0_1
0 5 10 15 20 25 30
Ka rin fährt zum Berg stei gen nach Ti rol

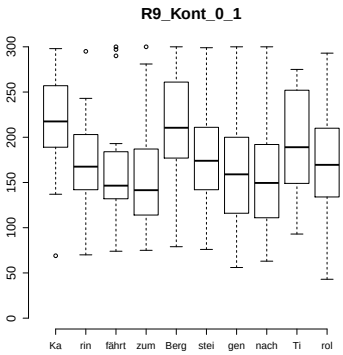

R9_31_1_1
0 5 10 15 20 25 30
Ka rin fährt zum Berg stei gen nach Ti rol

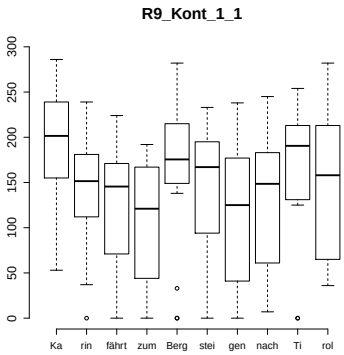

R9_31_2_1
0 5 10 15 20 25 30
Ka rin fährt zum Berg stei gen nach Ti rol

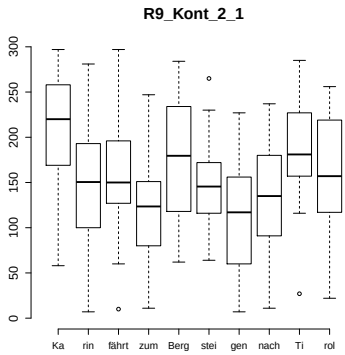

R9_Kont_0_0
0 50 100 150 200 250 300
Ka rin fährt zum Berg stei gen nach Ti rol

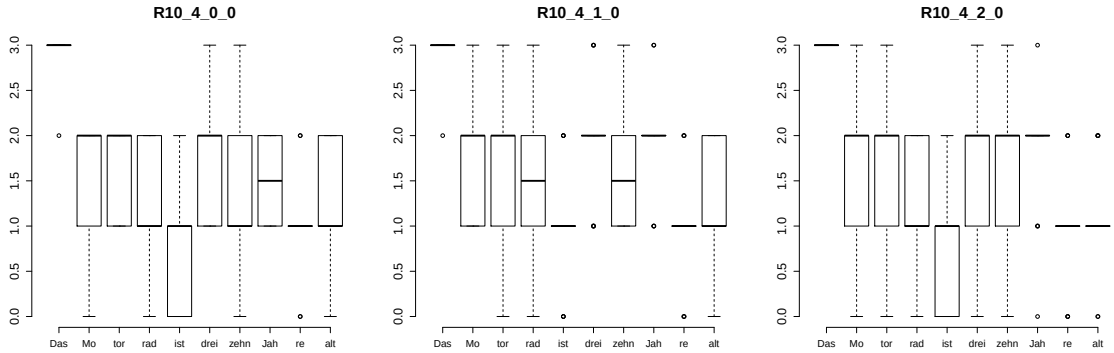

R9_Kont_1_0
0 50 100 150 200 250 300
Ka rin fährt zum Berg stei gen nach Ti rol

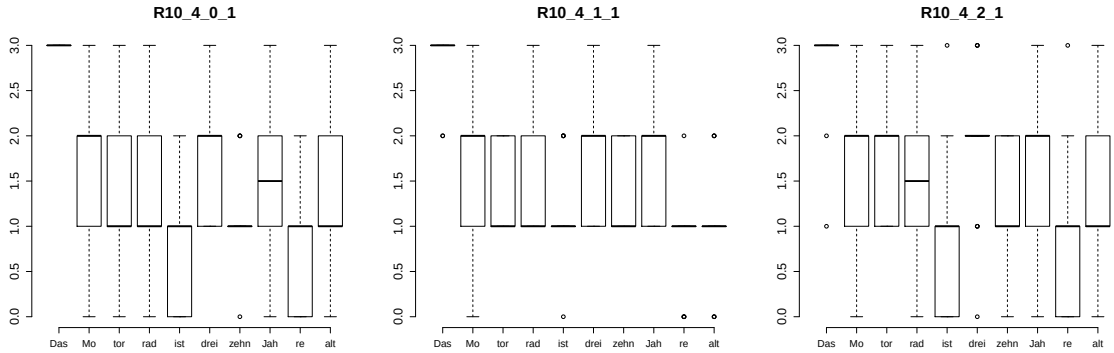

R9_Kont_2_0
0 50 100 150 200 250 300
Ka rin fährt zum Berg stei gen nach Ti rol

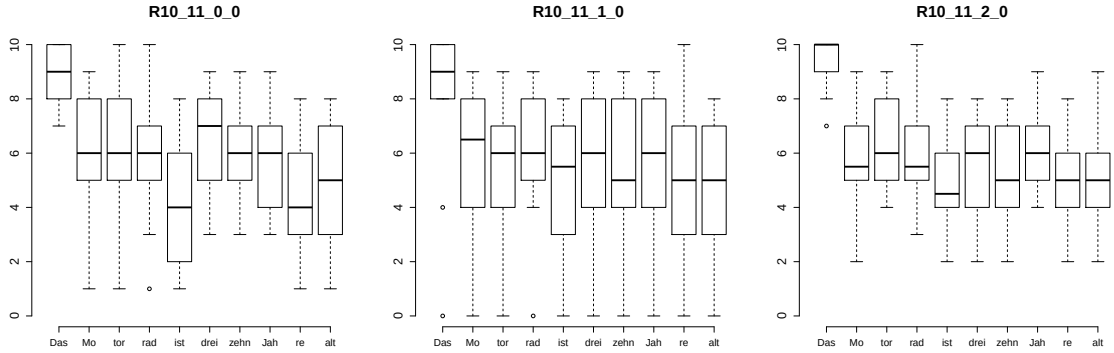

R9_Kont_0_1
0 50 100 150 200 250 300
Ka rin fährt zum Berg stei gen nach Ti rol

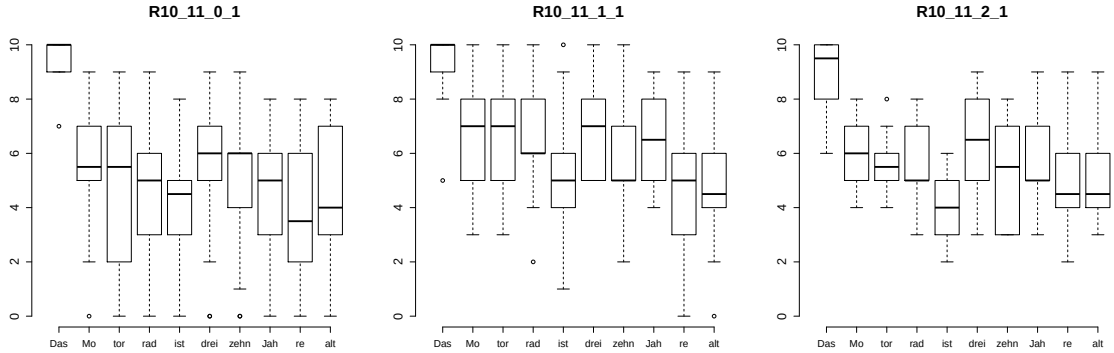

R9_Kont_1_1
0 50 100 150 200 250 300
Ka rin fährt zum Berg stei gen nach Ti rol

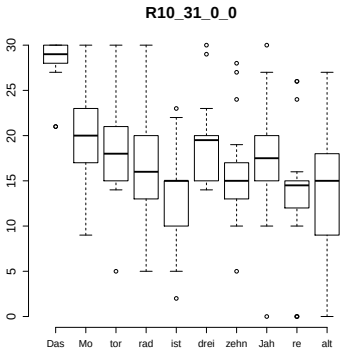

R9_Kont_2_1
0 50 100 150 200 250 300
Ka rin fährt zum Berg stei gen nach Ti rol

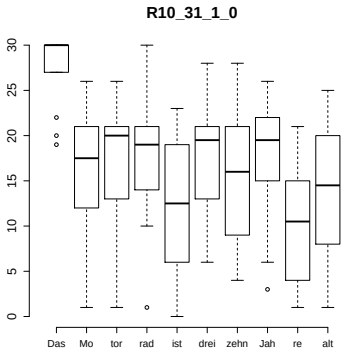

R10_4_0_0
R10_4_1_0
R10_4_2_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das Mo tor rad ist drei zehn Jah re alt

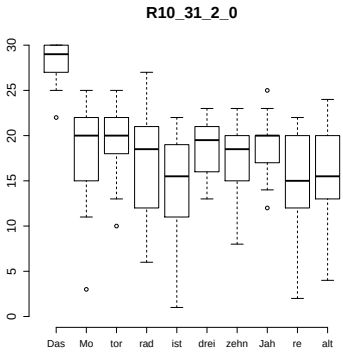

R10_4_0_1
R10_4_1_1
R10_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das Mo tor rad ist drei zehn Jah re alt

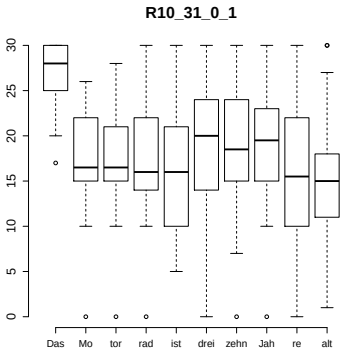

R10_11_0_0
R10_11_1_0
R10_11_2_0
0 2 4 6 8 10
Das Mo tor rad ist drei zehn Jah re alt

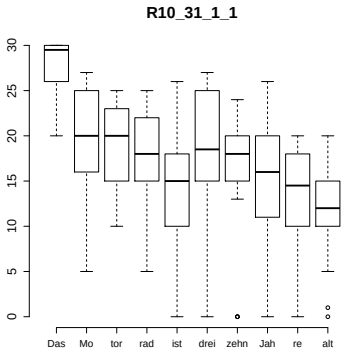

R10_11_0_1
R10_11_1_1
R10_11_2_1
0 2 4 6 8 10
Das Mo tor rad ist drei zehn Jah re alt

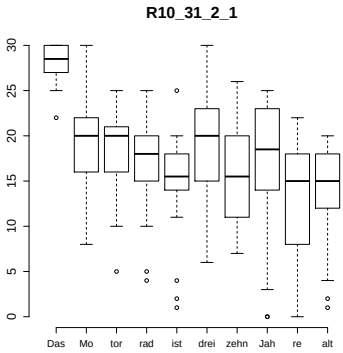

R10_31_0_0
0 5 10 15 20 25 30
Das Mo tor rad ist drei zehn Jah re alt

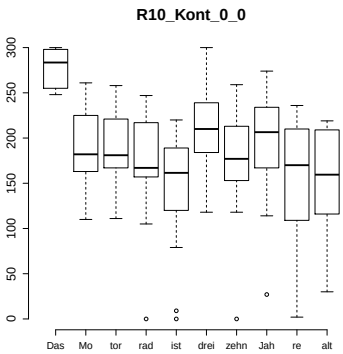

R10_31_1_0
0 5 10 15 20 25 30
Das Mo tor rad ist drei zehn Jah re alt

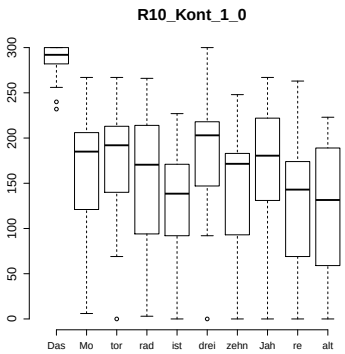

R10_31_2_0
0 5 10 15 20 25 30
Das Mo tor rad ist drei zehn Jah re alt

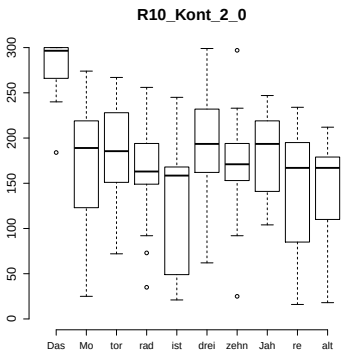

R10_31_0_1
0 5 10 15 20 25 30
Das Mo tor rad ist drei zehn Jah re alt

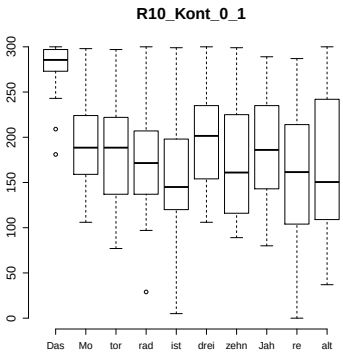

R10_31_1_1
0 5 10 15 20 25 30
Das Mo tor rad ist drei zehn Jah re alt

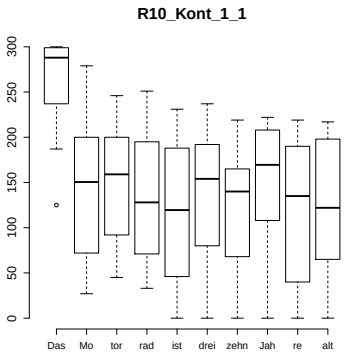

R10_31_2_1
0 5 10 15 20 25 30
Das Mo tor rad ist drei zehn Jah re alt

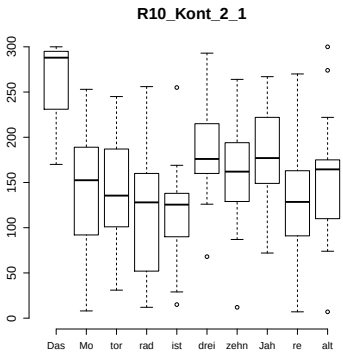

R10_Kont_0_0
0 50 100 150 200 250 300
Das Mo tor rad ist drei zehn Jah re alt

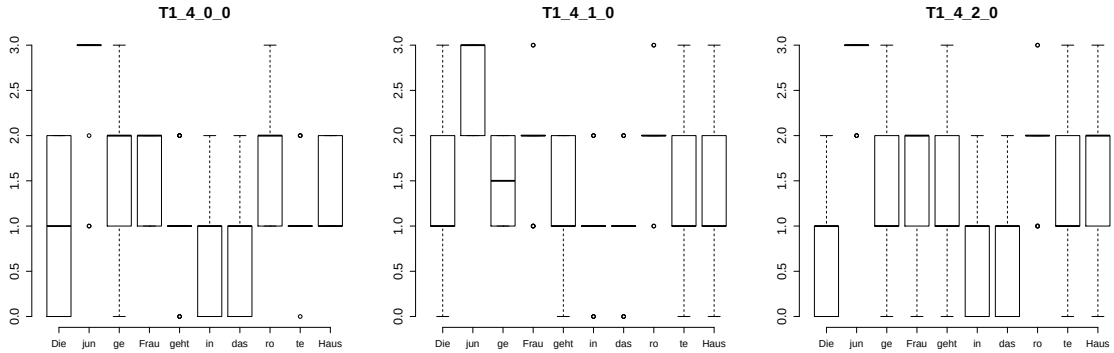

R10_Kont_1_0
0 50 100 150 200 250 300
Das Mo tor rad ist drei zehn Jah re alt

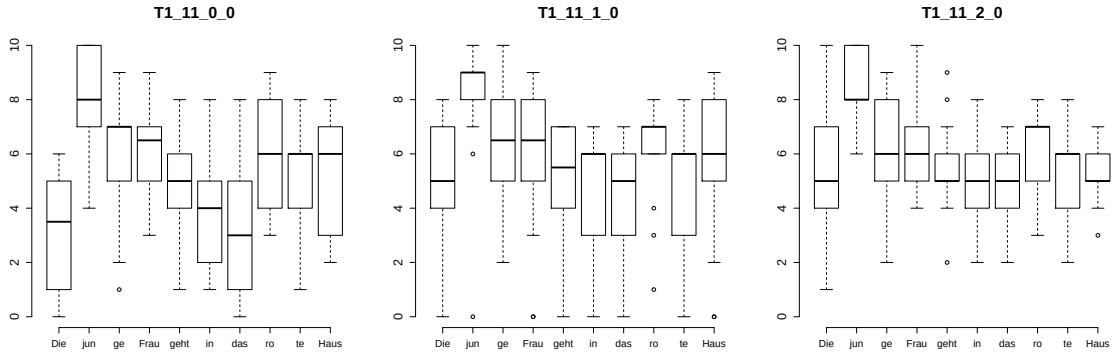

R10_Kont_2_0
0 50 100 150 200 250 300
Das Mo tor rad ist drei zehn Jah re alt

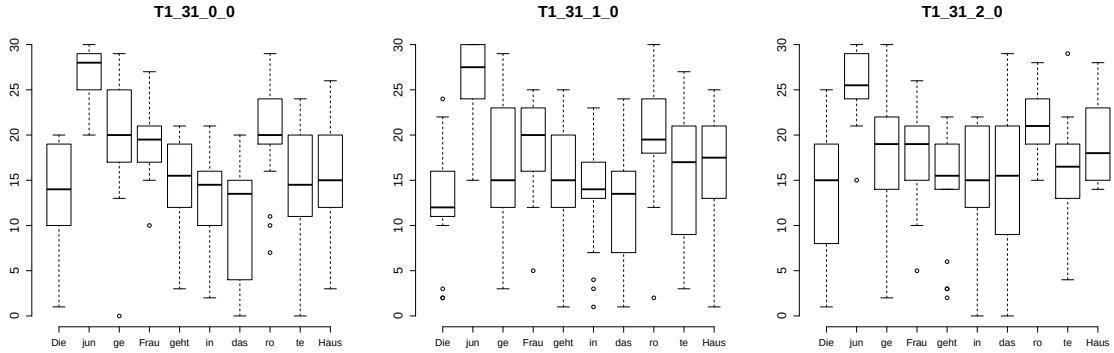

R10_Kont_0_1
0 50 100 150 200 250 300
Das Mo tor rad ist drei zehn Jah re alt

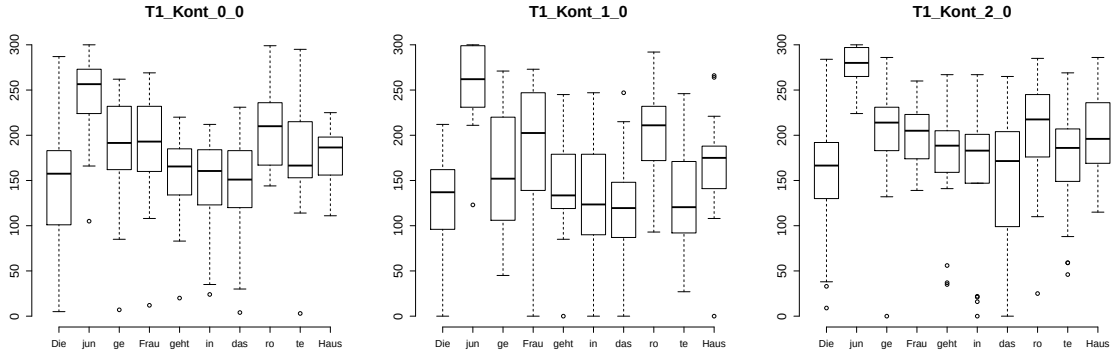

R10_Kont_1_1
0 50 100 150 200 250 300
Das Mo tor rad ist drei zehn Jah re alt

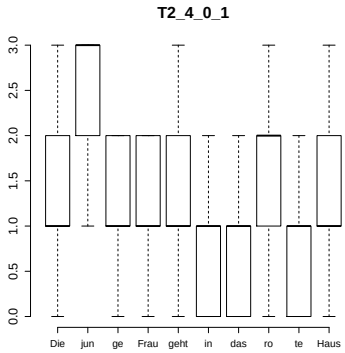

R10_Kont_2_1
0 50 100 150 200 250 300
Das Mo tor rad ist drei zehn Jah re alt

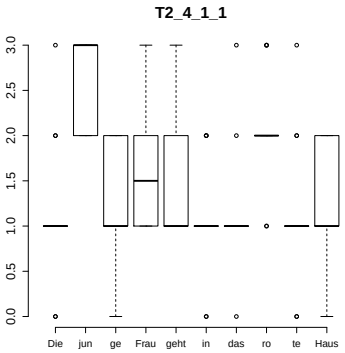

T1_4_0_0
T1_4_1_0
T1_4_2_0
Die jun ge Frau geht in das ro te Haus

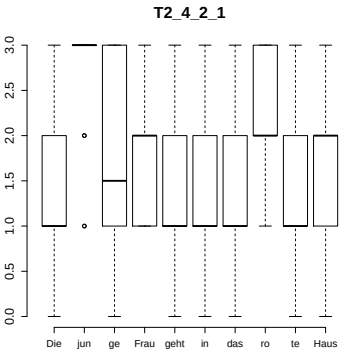

T1_11_0_0
T1_11_1_0
T1_11_2_0
Die jun ge Frau geht in das ro te Haus

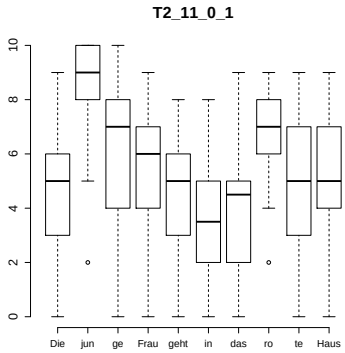

T1_31_0_0
T1_31_1_0
T1_31_2_0
Die jun ge Frau geht in das ro te Haus

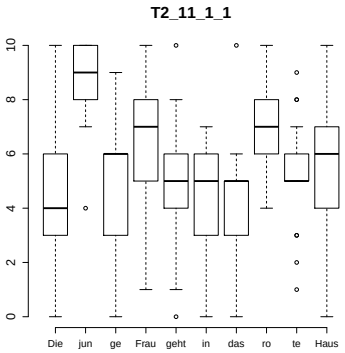

T1_Kont_0_0
T1_Kont_1_0
T1_Kont_2_0
Die jun ge Frau geht in das ro te Haus

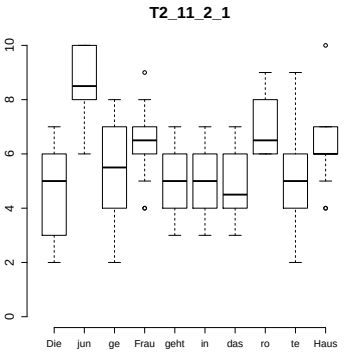

T2_4_0_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Die jun ge Frau geht in das ro te Haus

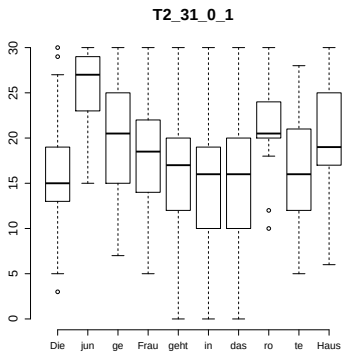

T2_4_1_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Die jun ge Frau geht in das ro te Haus

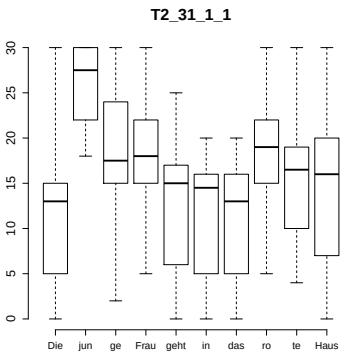

T2_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Die jun ge Frau geht in das ro te Haus

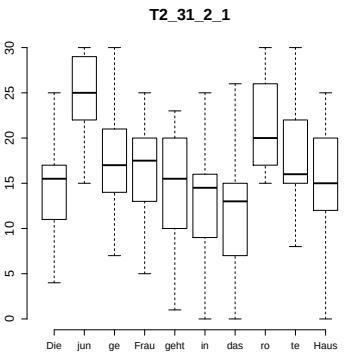

T2_11_0_1
0 2 4 6 8 10
Die jun ge Frau geht in das ro te Haus

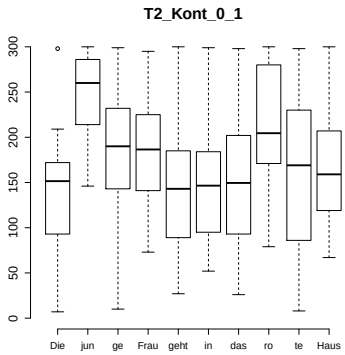

T2_11_1_1
0 2 4 6 8 10
Die jun ge Frau geht in das ro te Haus

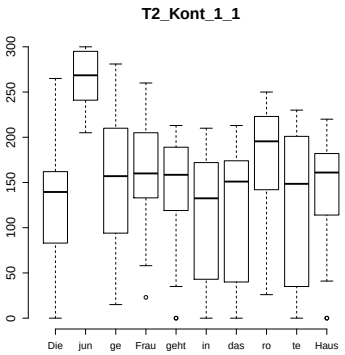

T2_11_2_1
0 2 4 6 8 10
Die jun ge Frau geht in das ro te Haus

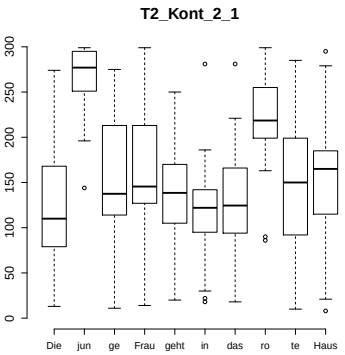

T2_31_0_1
0 5 10 15 20 25 30
Die jun ge Frau geht in das ro te Haus

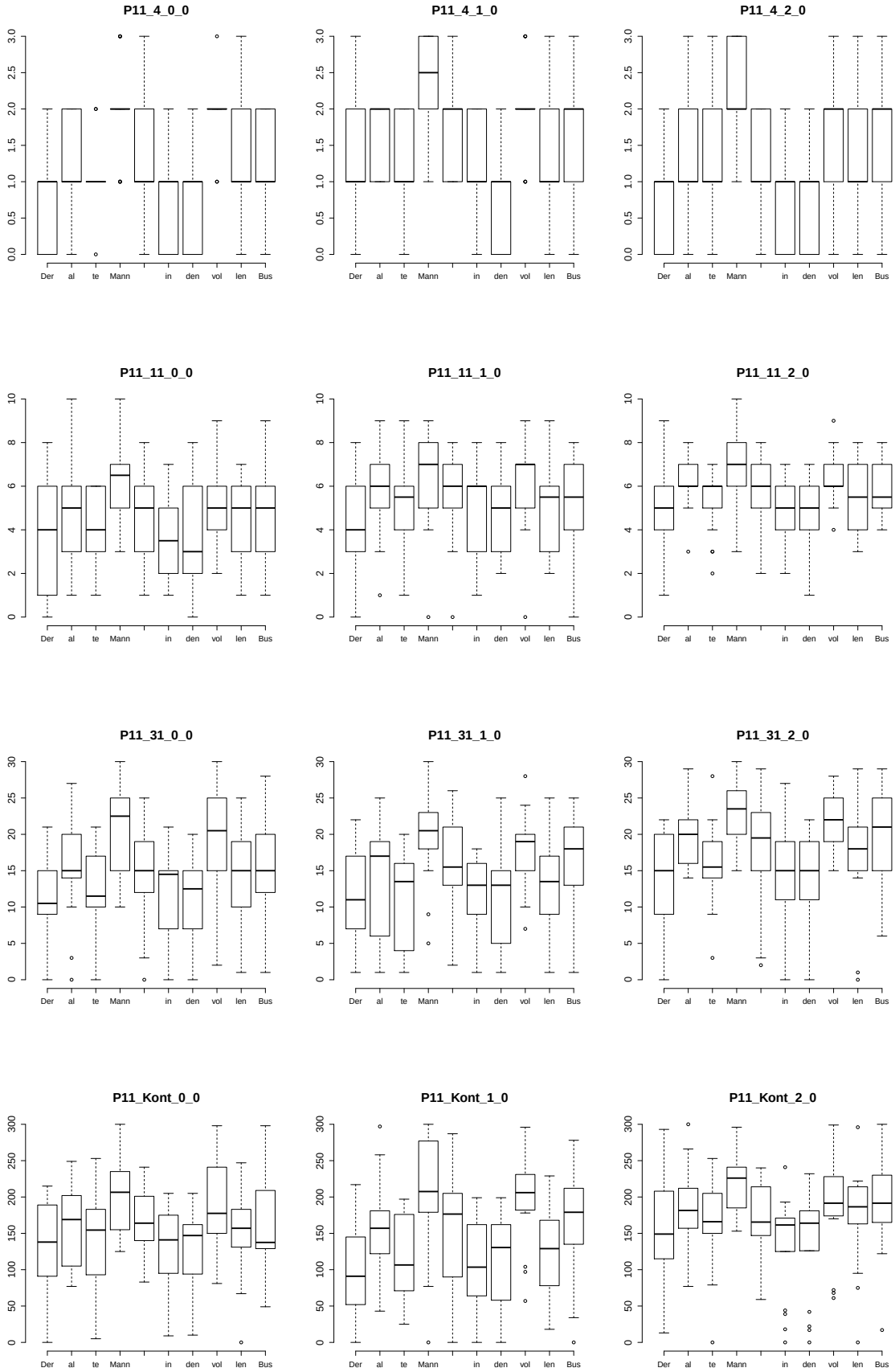

T2_31_1_1
0 5 10 15 20 25 30
Die jun ge Frau geht in das ro te Haus

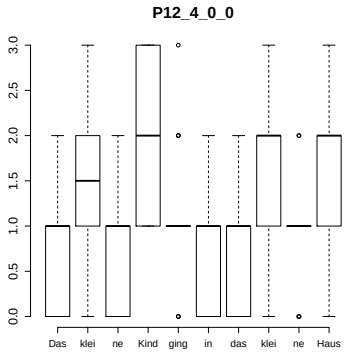

T2_31_2_1
0 5 10 15 20 25 30
Die jun ge Frau geht in das ro te Haus

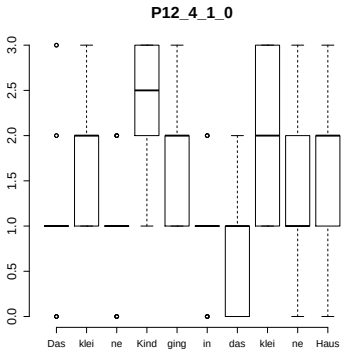

T2_Kont_0_1
0 50 100 150 200 250 300
Die jun ge Frau geht in das ro te Haus

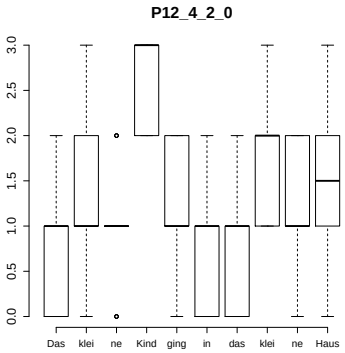

T2_Kont_1_1
0 50 100 150 200 250 300
Die jun ge Frau geht in das ro te Haus

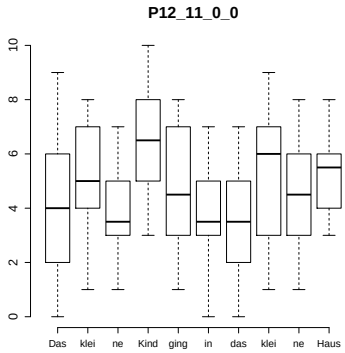

T2_Kont_2_1
0 50 100 150 200 250 300
Die jun ge Frau geht in das ro te Haus

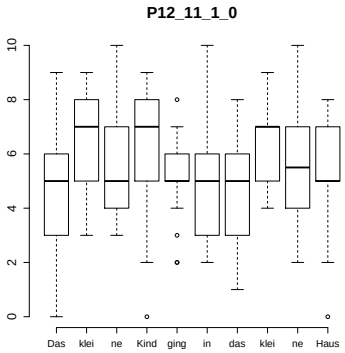

P11_4_0_0
P11_4_1_0
P11_4_2_0
P11_11_0_0
P11_11_1_0
P11_11_2_0
P11_31_0_0
P11_31_1_0
P11_31_2_0
P11_Kont_0_0
P11_Kont_1_0
P11_Kont_2_0
Der al te Mann in den vol len Bus

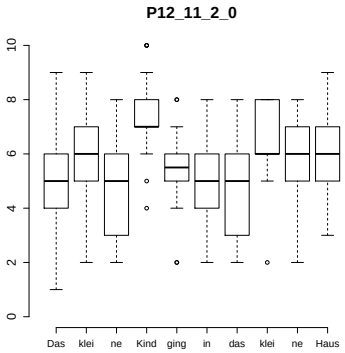

P12_4_0_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das klei ne Kind ging in das klei ne Haus

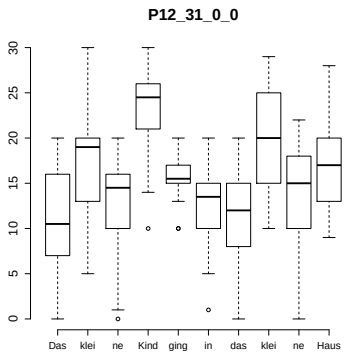

P12_4_1_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das klei ne Kind ging in das klei ne Haus

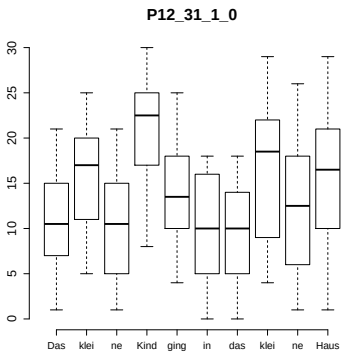

P12_4_2_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das klei ne Kind ging in das klei ne Haus

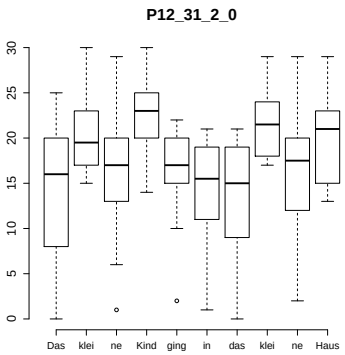

P12_11_0_0
0 2 4 6 8 10
Das klei ne Kind ging in das klei ne Haus

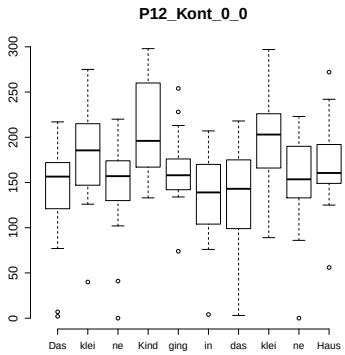

P12_11_1_0
0 2 4 6 8 10
Das klei ne Kind ging in das klei ne Haus

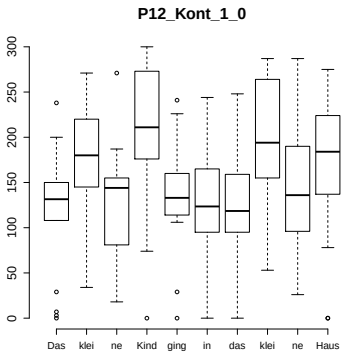

P12_11_2_0
0 2 4 6 8 10
Das klei ne Kind ging in das klei ne Haus

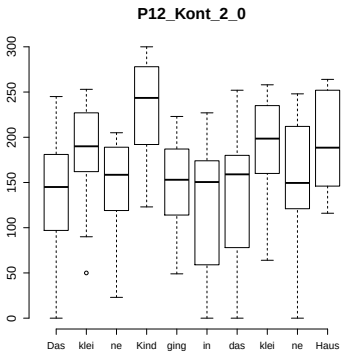

P12_31_0_0
0 5 10 15 20 25 30
Das klei ne Kind ging in das klei ne Haus

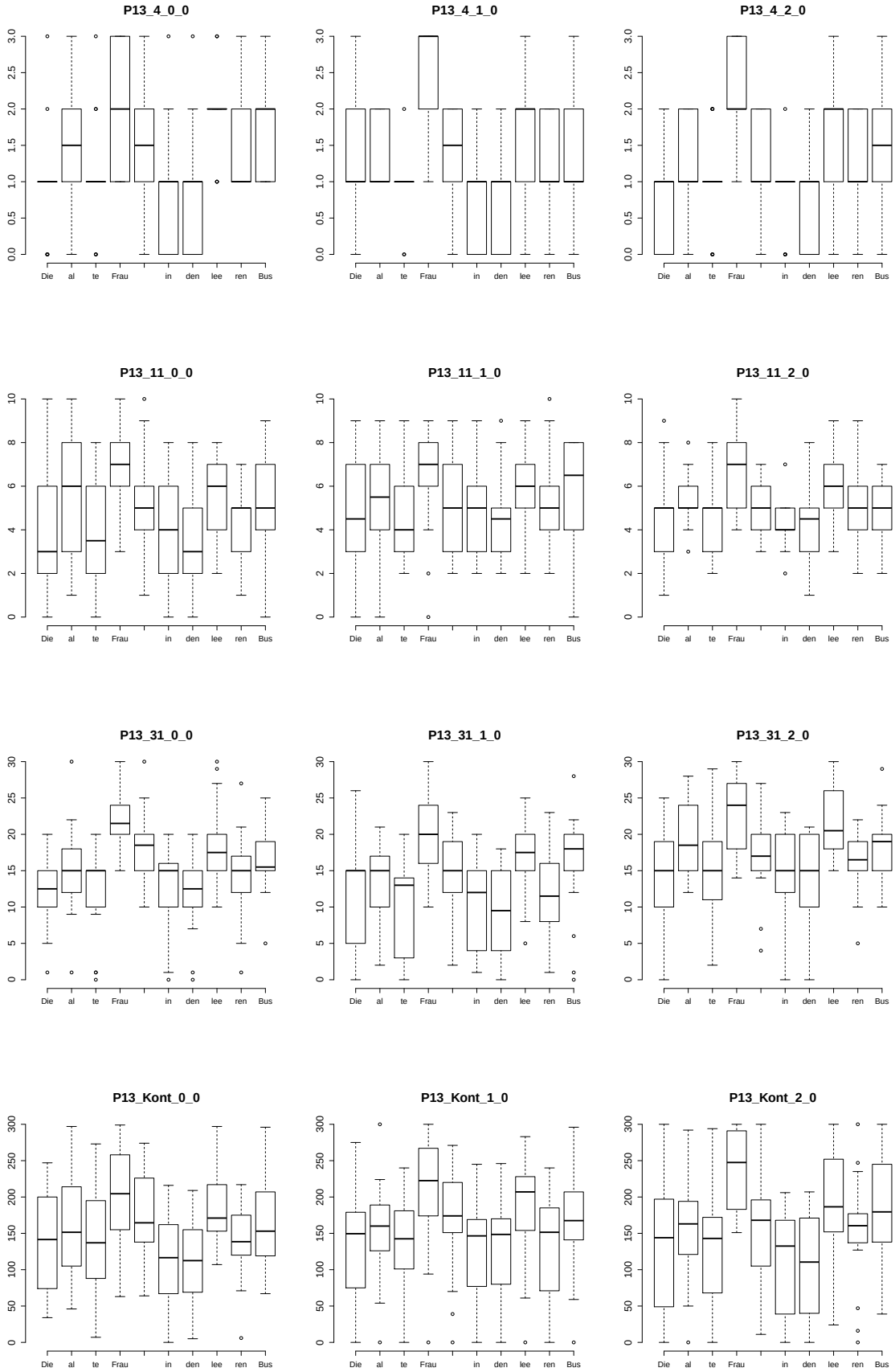

P12_31_1_0
0 5 10 15 20 25 30
Das klei ne Kind ging in das klei ne Haus

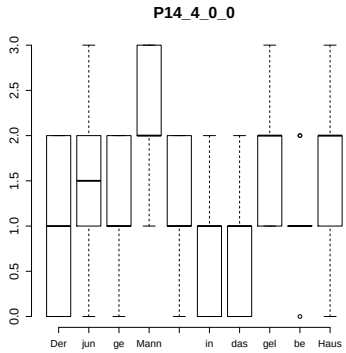

P12_31_2_0
0 5 10 15 20 25 30
Das klei ne Kind ging in das klei ne Haus

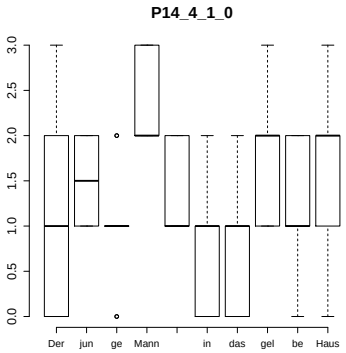

P12_Kont_0_0
0 50 100 150 200 250 300
Das klei ne Kind ging in das klei ne Haus

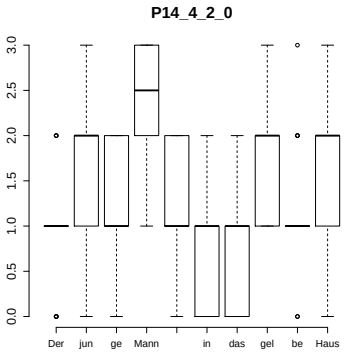

P12_Kont_1_0
0 50 100 150 200 250 300
Das klei ne Kind ging in das klei ne Haus

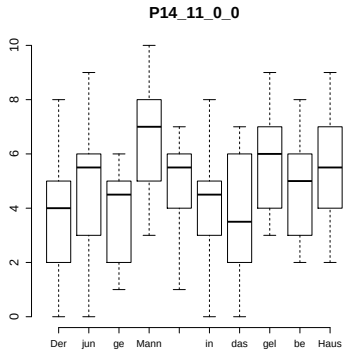

P12_Kont_2_0
0 50 100 150 200 250 300
Das klei ne Kind ging in das klei ne Haus

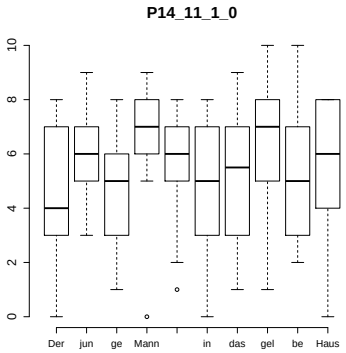

P13_4_0_0
P13_4_1_0
P13_4_2_0
P13_11_0_0
P13_11_1_0
P13_11_2_0
P13_31_0_0
P13_31_1_0
P13_31_2_0
P13_Kont_0_0
P13_Kont_1_0
P13_Kont_2_0
Die al te Frau in den lee ren Bus

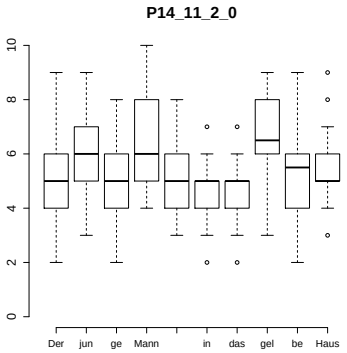

P14_4_0_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Der jun ge Mann in das gel be Haus

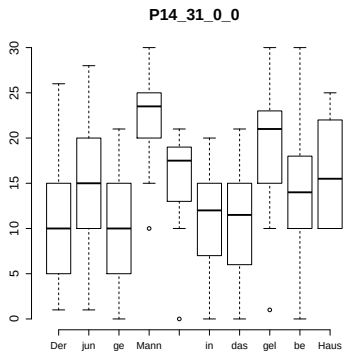

P14_4_1_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Der jun ge Mann in das gel be Haus

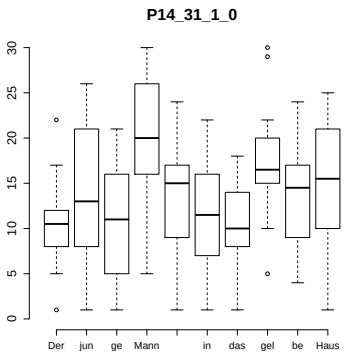

P14_4_2_0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Der jun ge Mann in das gel be Haus

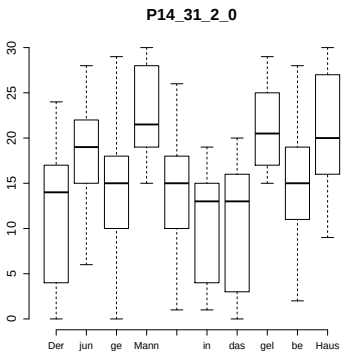

P14_11_0_0
0 2 4 6 8 10
Der jun ge Mann in das gel be Haus

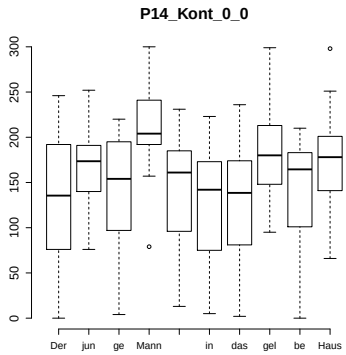

P14_11_1_0
0 2 4 6 8 10
Der jun ge Mann in das gel be Haus

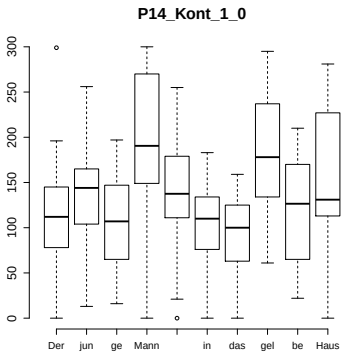

P14_11_2_0
0 2 4 6 8 10
Der jun ge Mann in das gel be Haus

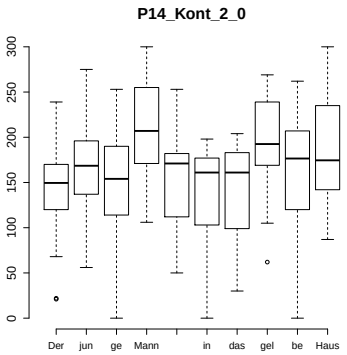

P14_31_0_0
0 5 10 15 20 25 30
Der jun ge Mann in das gel be Haus

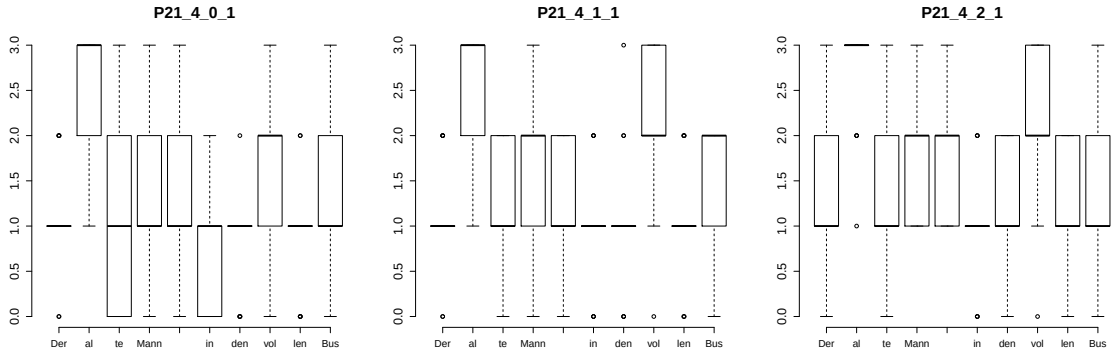

P14_31_1_0
0 5 10 15 20 25 30
Der jun ge Mann in das gel be Haus

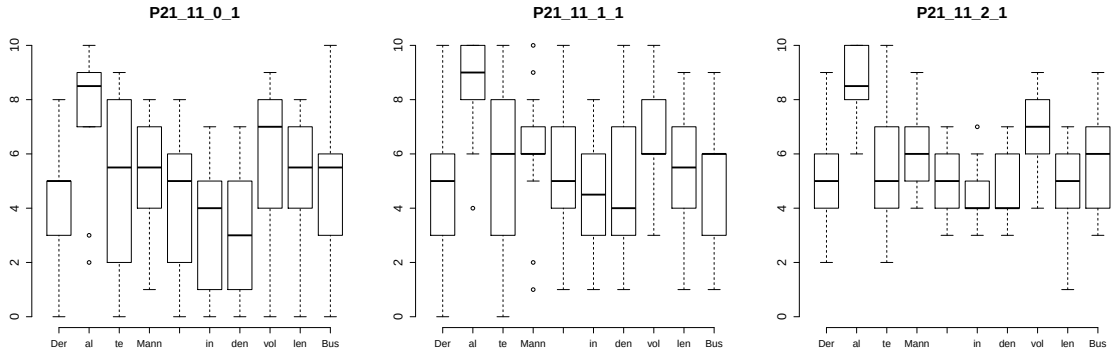

P14_31_2_0
0 5 10 15 20 25 30
Der jun ge Mann in das gel be Haus

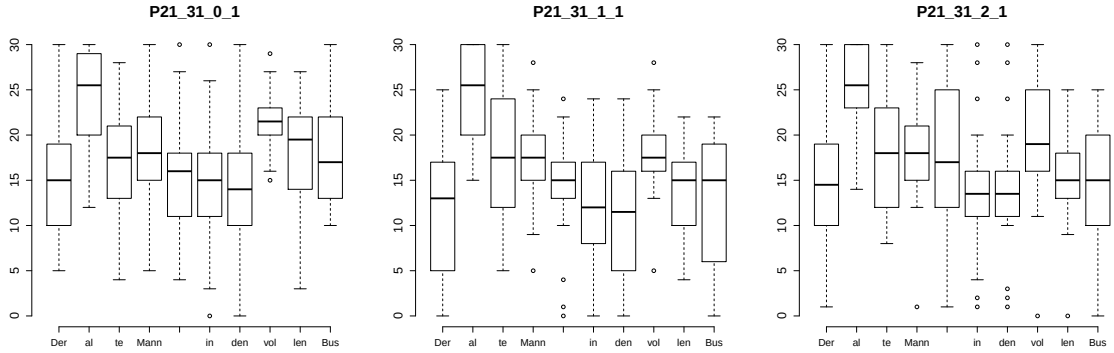

P14_Kont_0_0
0 50 100 150 200 250 300
Der jun ge Mann in das gel be Haus

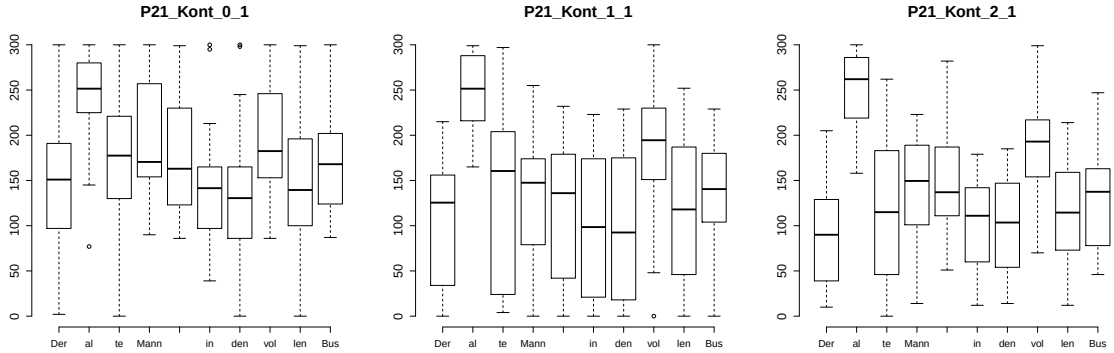

P14_Kont_1_0
0 50 100 150 200 250 300
Der jun ge Mann in das gel be Haus

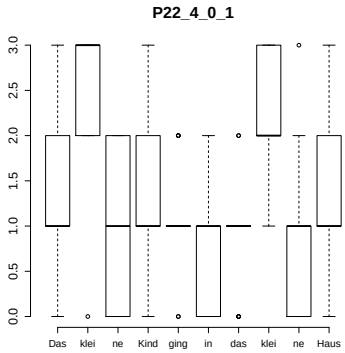

P14_Kont_2_0
0 50 100 150 200 250 300
Der jun ge Mann in das gel be Haus

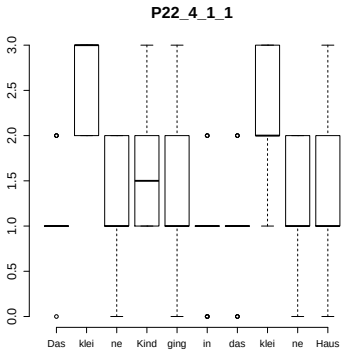

P21_4_0_1
P21_4_1_1
P21_4_2_1
Der al te Mann in den vol len Bus

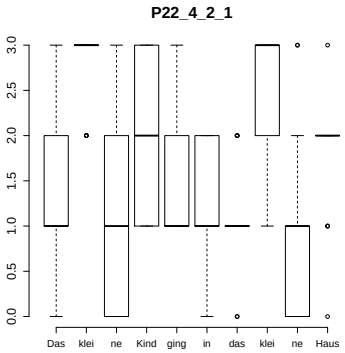

P21_11_0_1
P21_11_1_1
P21_11_2_1
Der al te Mann in den vol len Bus

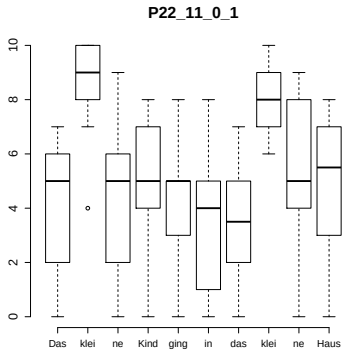

P21_31_0_1
P21_31_1_1
P21_31_2_1
Der al te Mann in den vol len Bus

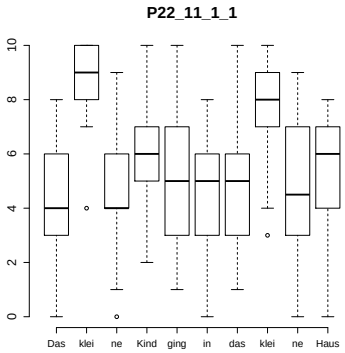

P21_Kont_0_1
P21_Kont_1_1
P21_Kont_2_1
Der al te Mann in den vol len Bus

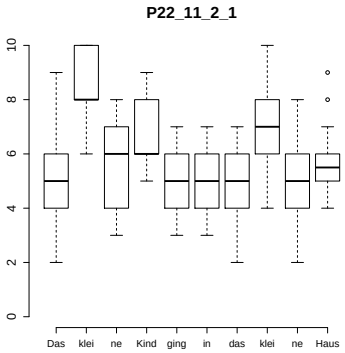

P22_4_0_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das klei ne Kind ging in das klei ne Haus

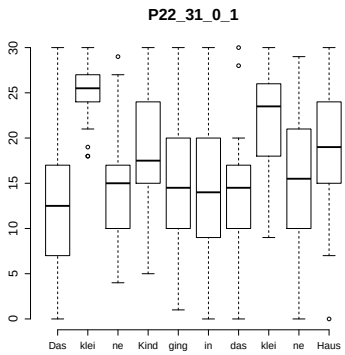

P22_4_1_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das klei ne Kind ging in das klei ne Haus

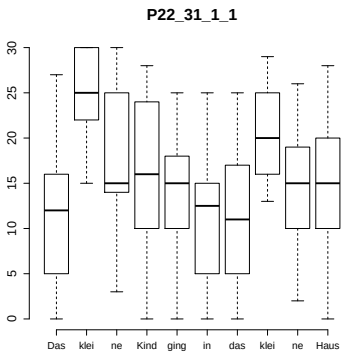

P22_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Das klei ne Kind ging in das klei ne Haus

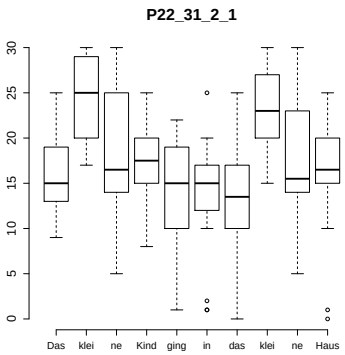

P22_11_0_1
0 2 4 6 8 10
Das klei ne Kind ging in das klei ne Haus

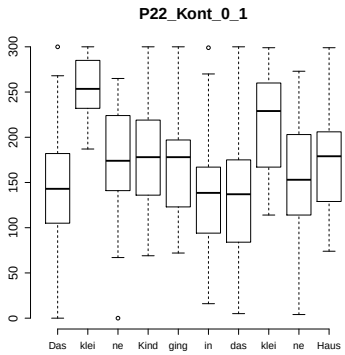

P22_11_1_1
0 2 4 6 8 10
Das klei ne Kind ging in das klei ne Haus

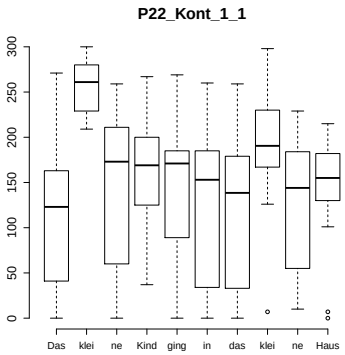

P22_11_2_1
0 2 4 6 8 10
Das klei ne Kind ging in das klei ne Haus

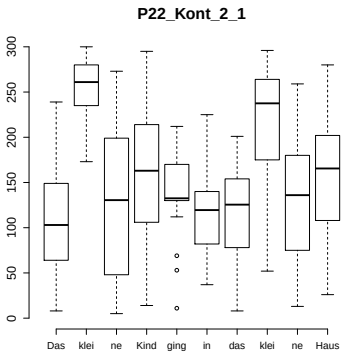

P22_31_0_1
0 5 10 15 20 25 30
Das klei ne Kind ging in das klei ne Haus

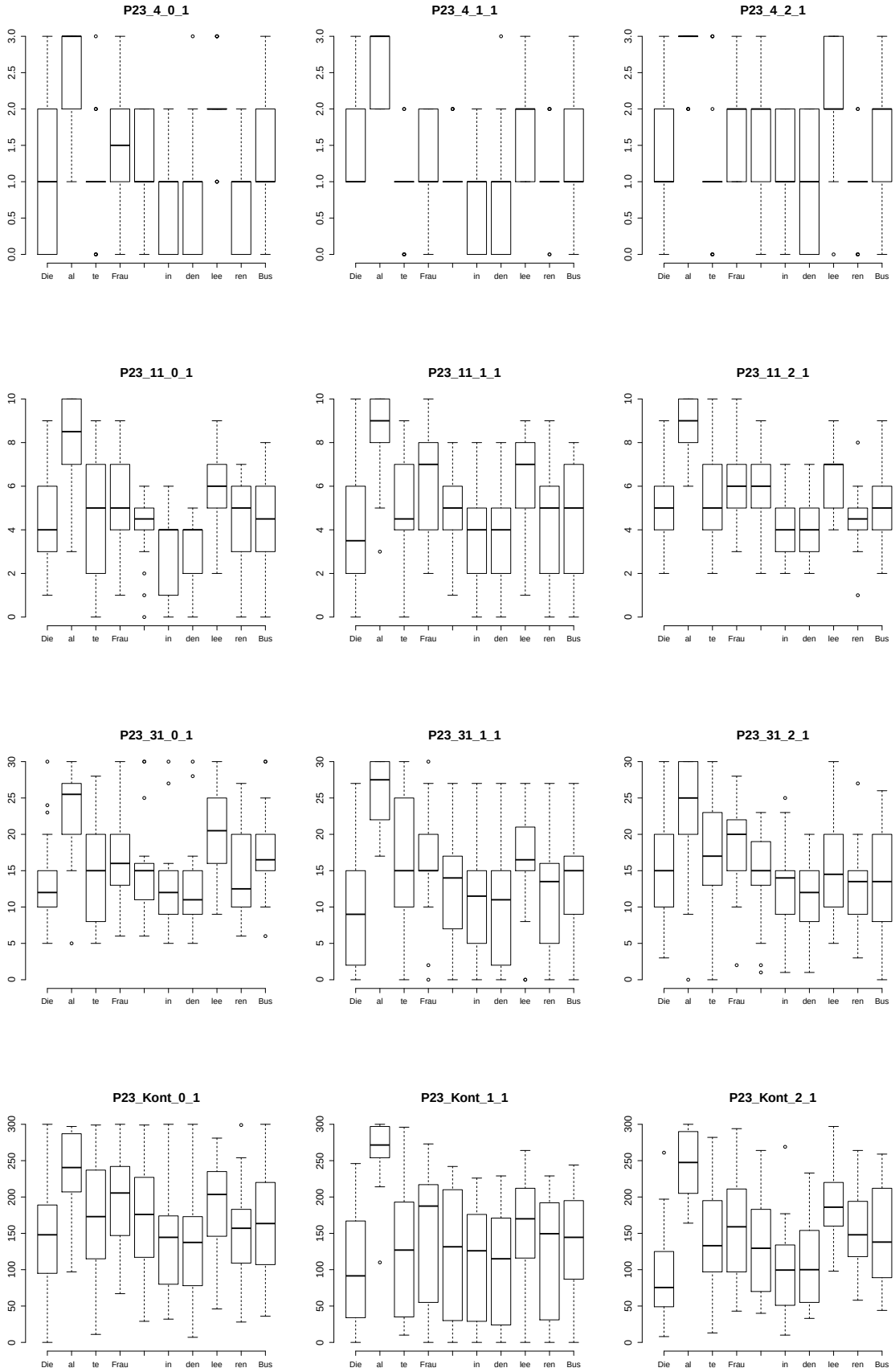

P22_31_1_1
0 5 10 15 20 25 30
Das klei ne Kind ging in das klei ne Haus

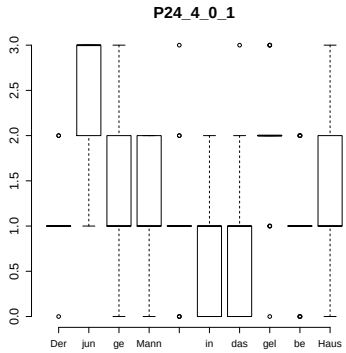

P22_31_2_1
0 5 10 15 20 25 30
Das klei ne Kind ging in das klei ne Haus

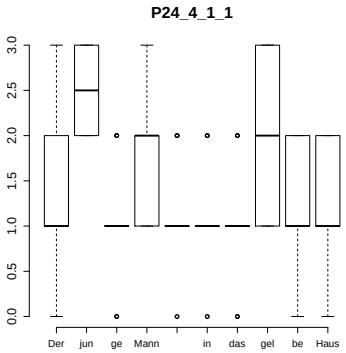

P22_Kont_0_1
0 50 100 150 200 250 300
Das klei ne Kind ging in das klei ne Haus

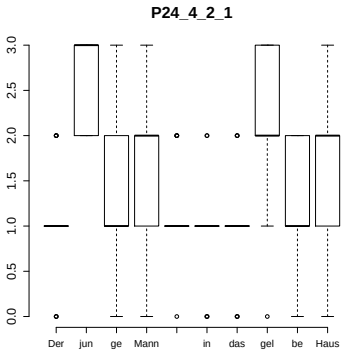

P22_Kont_1_1
0 50 100 150 200 250 300
Das klei ne Kind ging in das klei ne Haus

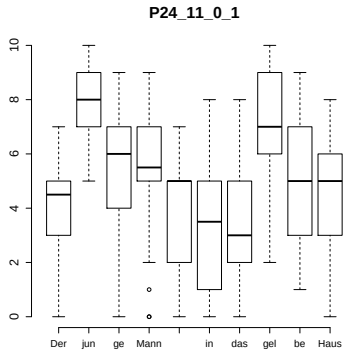

P22_Kont_2_1
0 50 100 150 200 250 300
Das klei ne Kind ging in das klei ne Haus

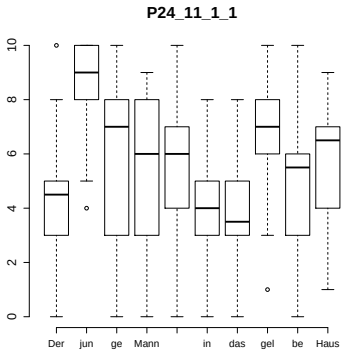

P23_4_0_1
P23_4_1_1
P23_4_2_1
P23_11_0_1
P23_11_1_1
P23_11_2_1
P23_31_0_1
P23_31_1_1
P23_31_2_1
P23_Kont_0_1
P23_Kont_1_1
P23_Kont_2_1
Die al te Frau in den lee ren Bus

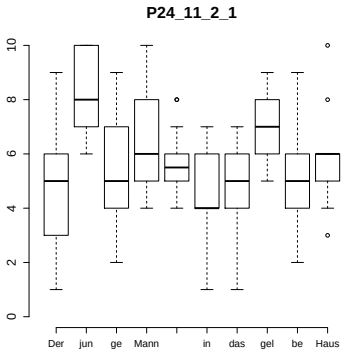

P24_4_0_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Der jun ge Mann in das gel be Haus

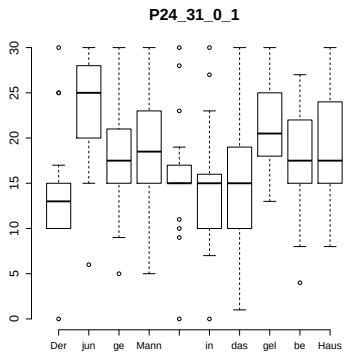

P24_4_1_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Der jun ge Mann in das gel be Haus

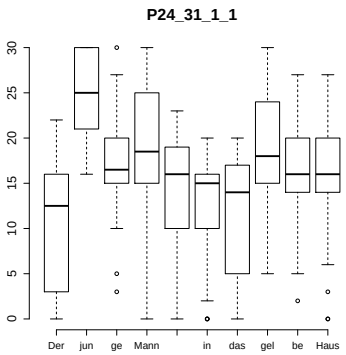

P24_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Der jun ge Mann in das gel be Haus

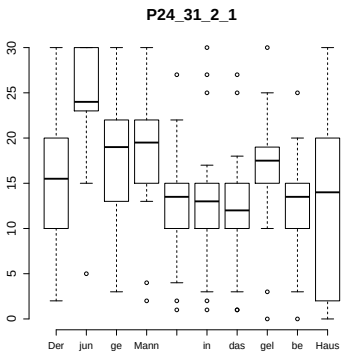

P24_11_0_1
0 2 4 6 8 10
Der jun ge Mann in das gel be Haus

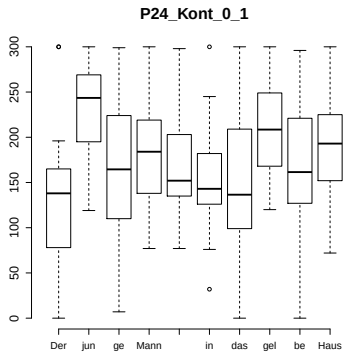

P24_11_1_1
0 2 4 6 8 10
Der jun ge Mann in das gel be Haus

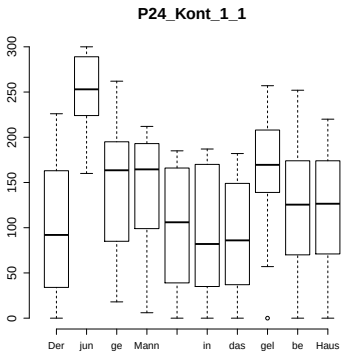

P24_11_2_1
0 2 4 6 8 10
Der jun ge Mann in das gel be Haus

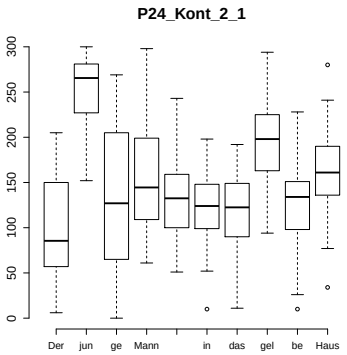

P24_31_0_1
0 5 10 15 20 25 30
Der jun ge Mann in das gel be Haus


P24_31_1_1
0 5 10 15 20 25 30
Der jun ge Mann in das gel be Haus


P24_31_2_1
0 5 10 15 20 25 30
Der jun ge Mann in das gel be Haus

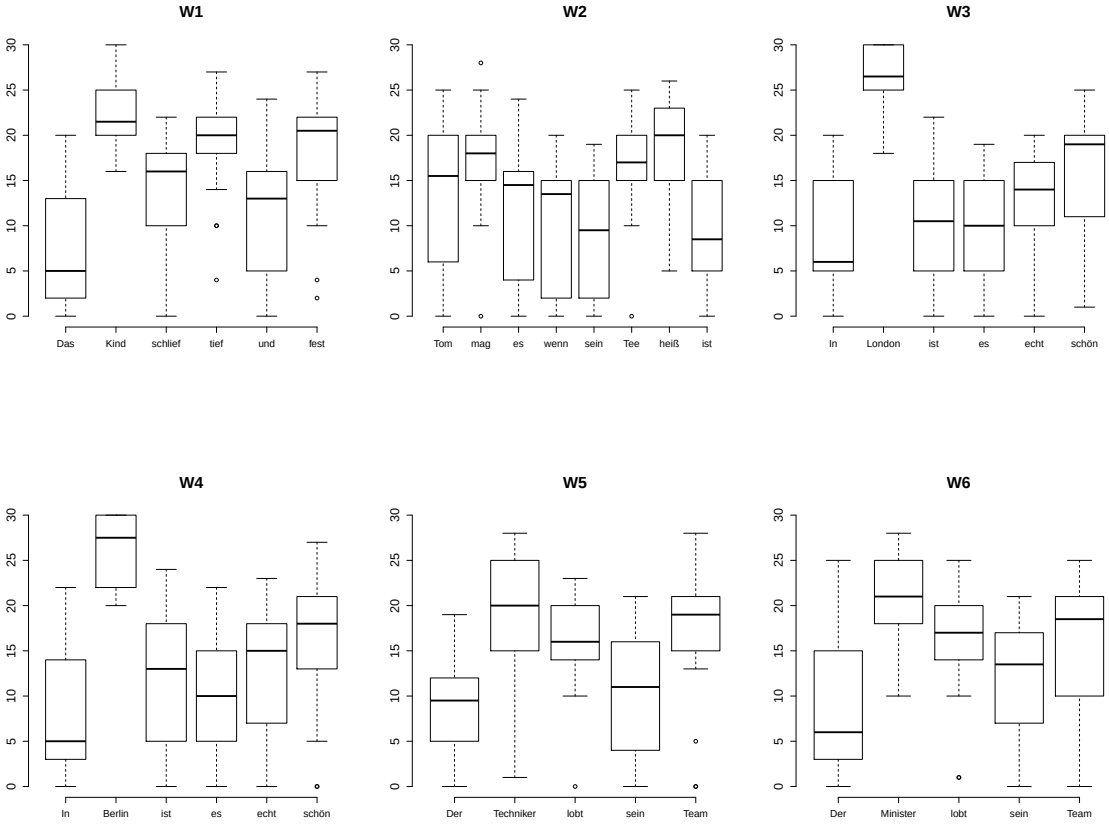

P24_Kont_0_1
0 50 100 150 200 250 300
Der jun ge Mann in das gel be Haus

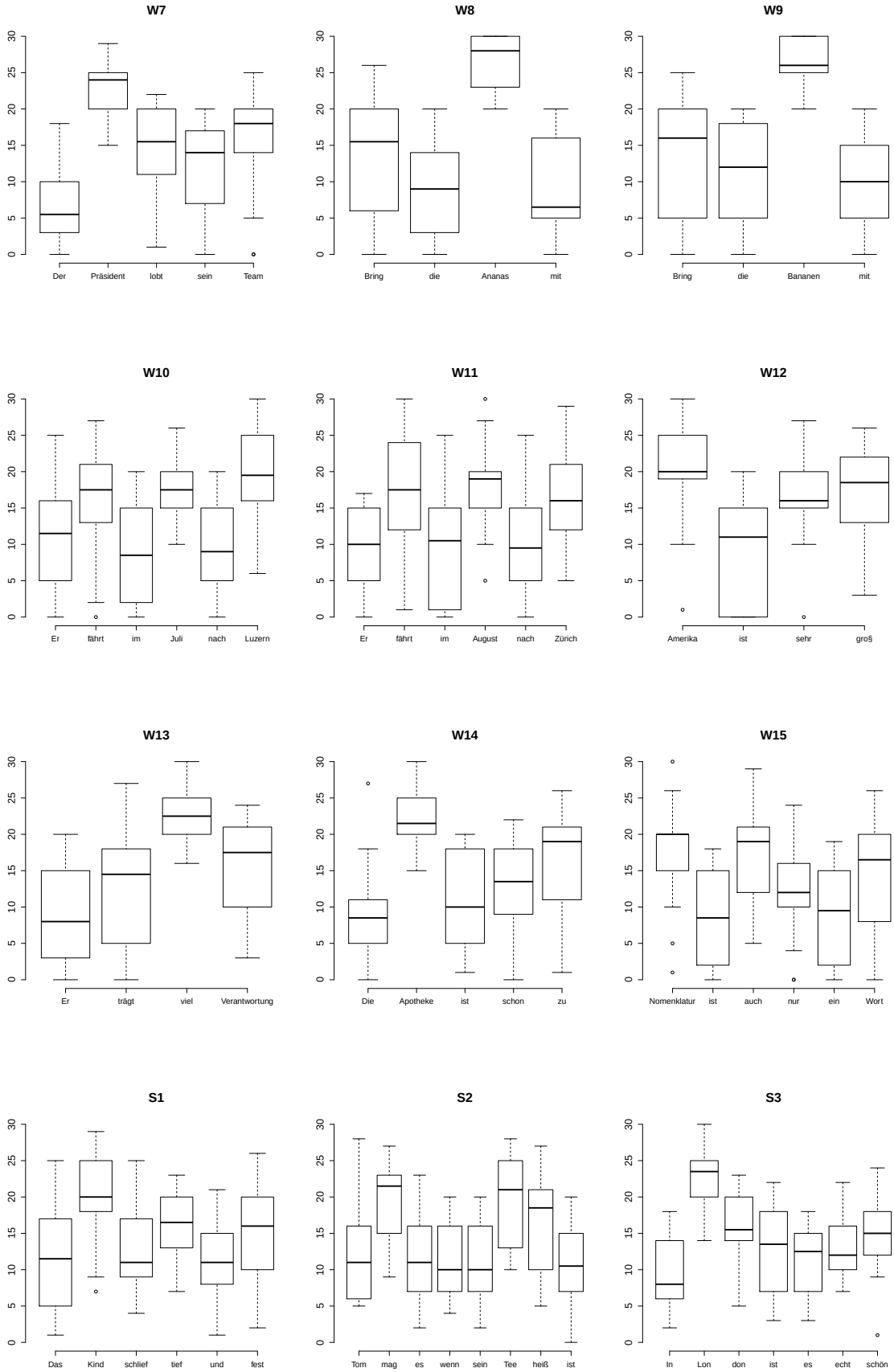

P24_Kont_1_1
0 50 100 150 200 250 300
Der jun ge Mann in das gel be Haus

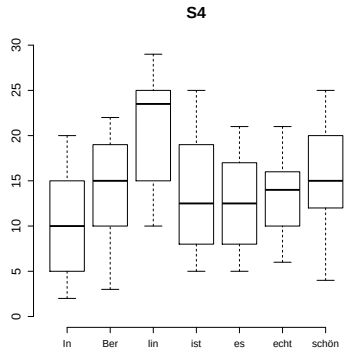

P24_Kont_2_1
0 50 100 150 200 250 300
Der jun ge Mann in das gel be Haus

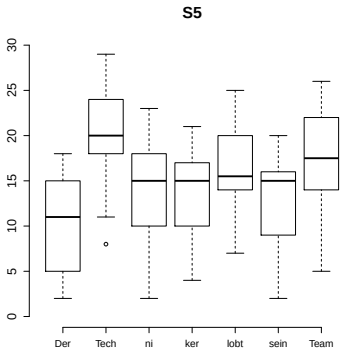
Anhang D. Boxplots Experiment 2

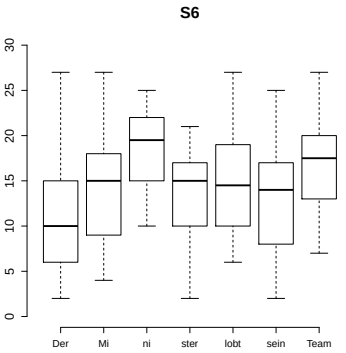
Auf den folgenden Seiten werden die Ratings aller Sätze durch die Probanden des zweiten Experiments in Boxplots dargestellt. Hierbei setzt sich der Name der Abbildung wie folgt zusammen: GruppeSatznummer. W steht für die Bewertung auf Wortebene und S für die Bewertung auf Silbenebene.

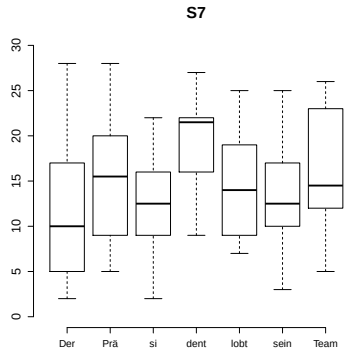


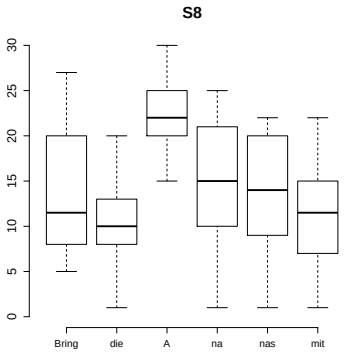

W7
Der Präsident lobt sein Team
W8
Bring die Ananas mit
W9
Bring die Bananen mit
W10
Er fährt im Juli nach Luzern
W11
Er fährt im August nach Zürich
W12
Amerika ist sehr groß
W13
Er trägt viel Verantwortung
W14
Die Apotheke ist schon zu
W15
Nomenklatur ist auch nur ein Wort
S1
Das Kind schlief tief und fest
S2
Tom mag es wenn sein Tee heiß ist
S3
In Lon don ist es echt schön

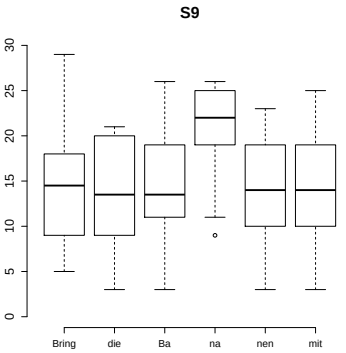

S4
0
5
10
15
20
25
30
In
Ber
lin
ist
es
echt
schön

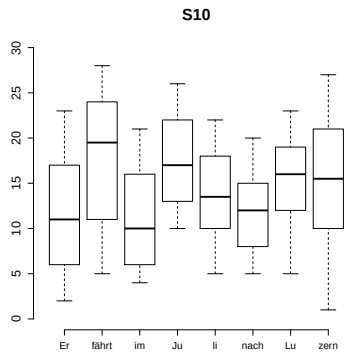

S5
0
5
10
15
20
25
30
Der
Tech
ni
ker
lobt
sein
Team

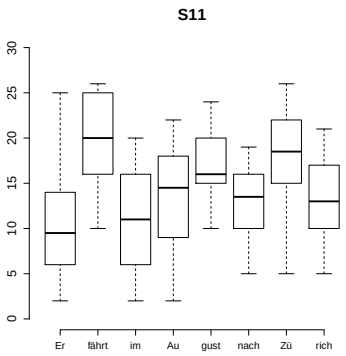

S6
0
5
10
15
20
25
30
Der
Mi
ni
ster
lobt
sein
Team

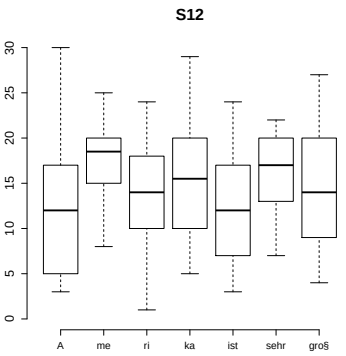

S7
0
5
10
15
20
25
30
Der
Prä
si
dent
lobt
sein
Team

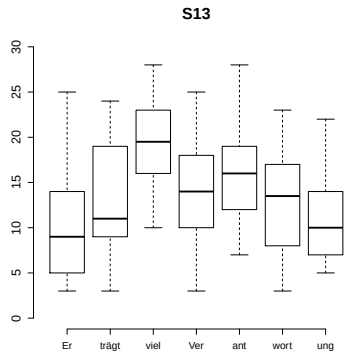

S8
0
5
10
15
20
25
30
Bring
die
A
na
nas
mit

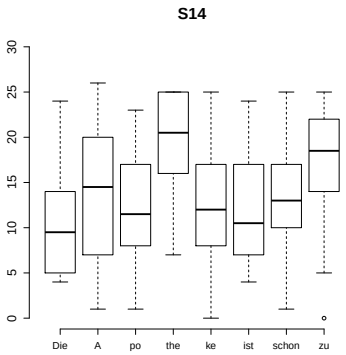

S9
0
5
10
15
20
25
30
Bring
die
Ba
na
nen
mit

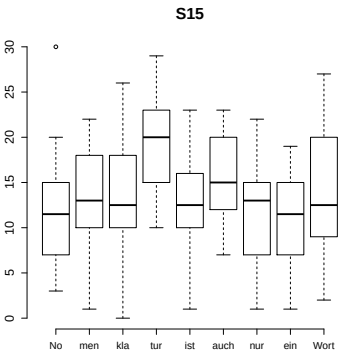

S10
0
5
10
15
20
25
30
Er
fährt
im
Ju
li
nach
Lu
zern


S11
0
5
10
15
20
25
30
Er
fährt
im
Au
gust
nach
Zü
rich


S12
0
5
10
15
20
25
30
A
me
ri
ka
ist
sehr
groß

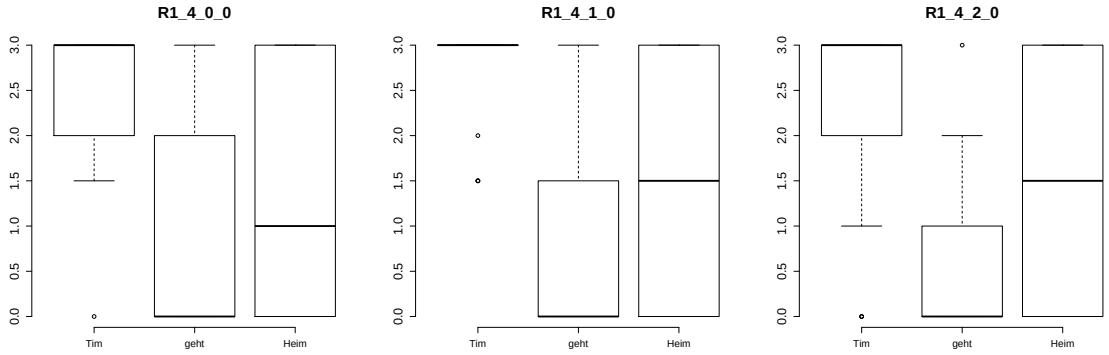

S13
0
5
10
15
20
25
30
Er
trägt
viel
Ver
ant
wort
ung

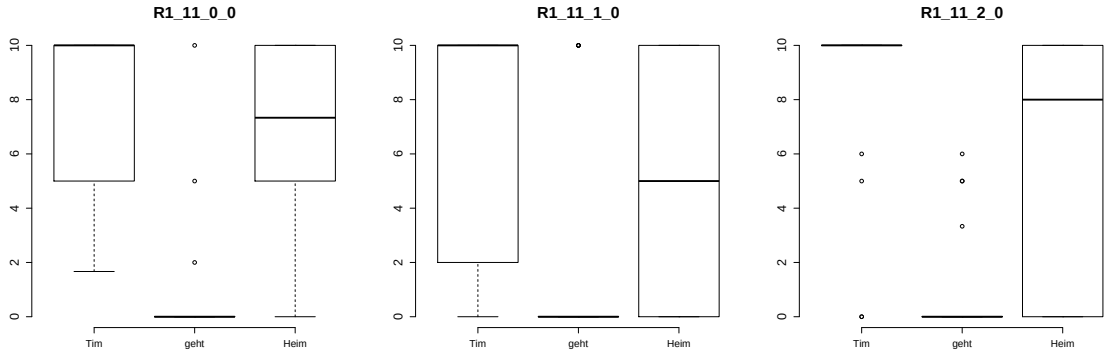

S14
0
5
10
15
20
25
30
Die
A
po
the
ke
ist
schon
zu

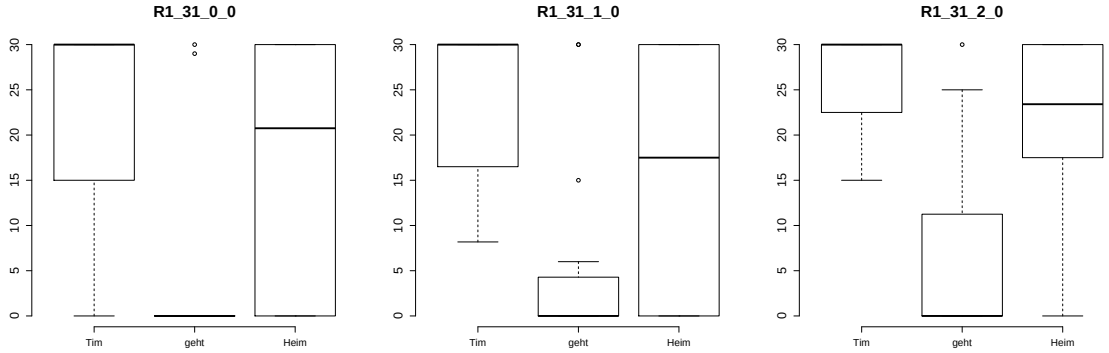

S15
0
5
10
15
20
25
30
No
men
kla
tur
ist
auch
nur
ein
Wort

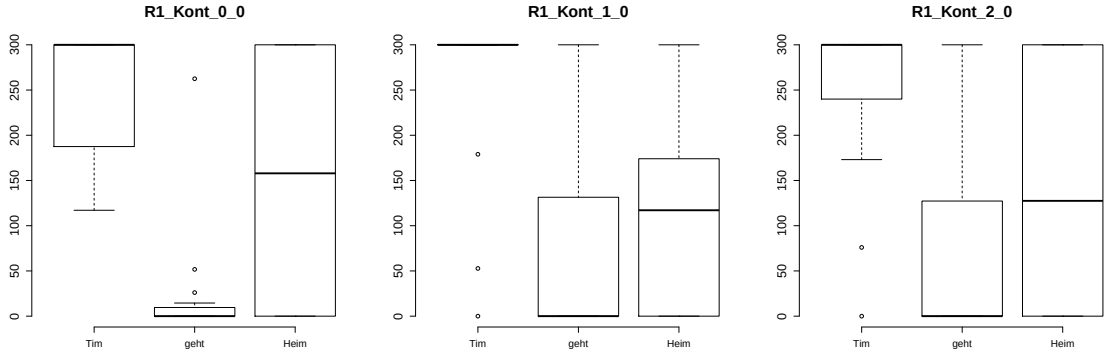
Anhang E. Boxplots Normalisierung Eriksson

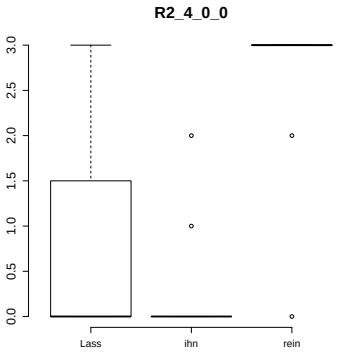
Auf den folgenden Seiten werden die Ratings der Sätze R1 - R10 durch die Probanden des ersten Experiments nach der linearen Transformation in Boxplots dargestellt. Hierbei setzt sich der Name der Abbildung wie folgt zusammen: Satzreferenz_Skala_Akkuratheitsbedingung_Priminggruppe. Boxplot R10_4_0_0 zeigt also die Bewertungen des Satzes R10 mit der 4-Punkt Skala unter Akkuratheitsbedingung 0 und Priminggruppe 0.

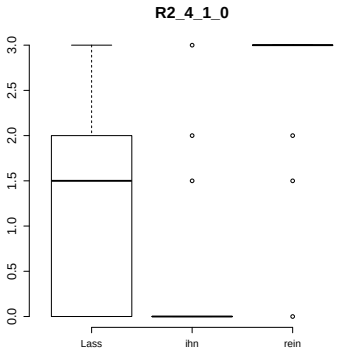

R1_4_0_0
R1_4_1_0
R1_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Tim
geht
Heim

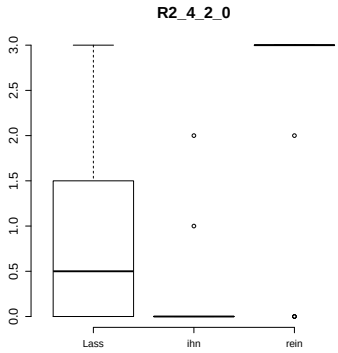

R1_11_0_0
R1_11_1_0
R1_11_2_0
0
2
4
6
8
10
Tim
geht
Heim

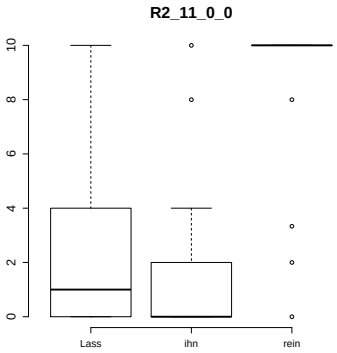

R1_31_0_0
R1_31_1_0
R1_31_2_0
0
5
10
15
20
25
30
Tim
geht
Heim

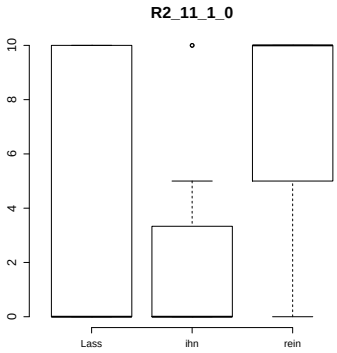

R1_Kont_0_0
R1_Kont_1_0
R1_Kont_2_0
0
50
100
150
200
250
300
Tim
geht
Heim

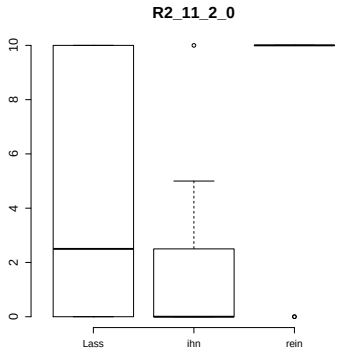

R2_4_0_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lass
ihn
rein

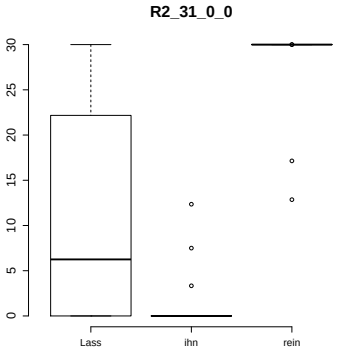

R2_4_1_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lass
ihn
rein

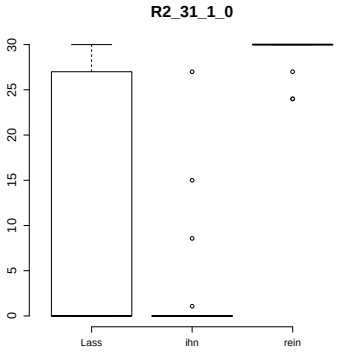

R2_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lass
ihn
rein

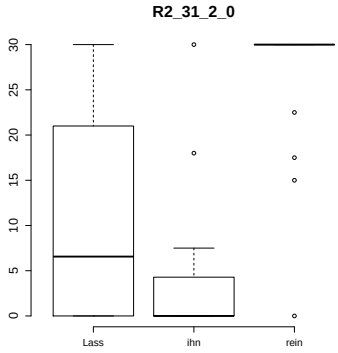

R2_11_0_0
0
2
4
6
8
10
Lass
ihn
rein

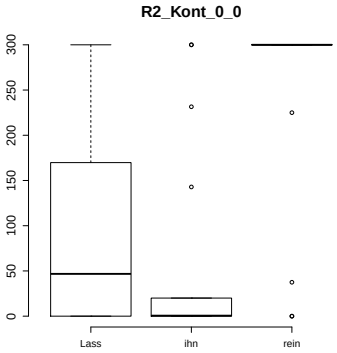

R2_11_1_0
0
2
4
6
8
10
Lass
ihn
rein

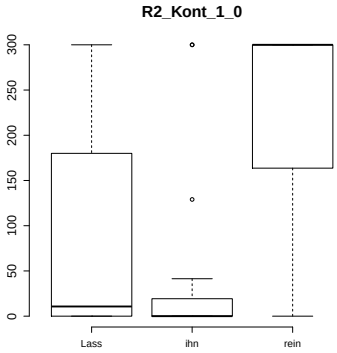

R2_11_2_0
0
2
4
6
8
10
Lass
ihn
rein

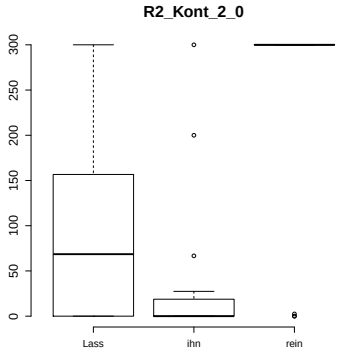

R2_31_0_0
0
5
10
15
20
25
30
Lass
ihn
rein

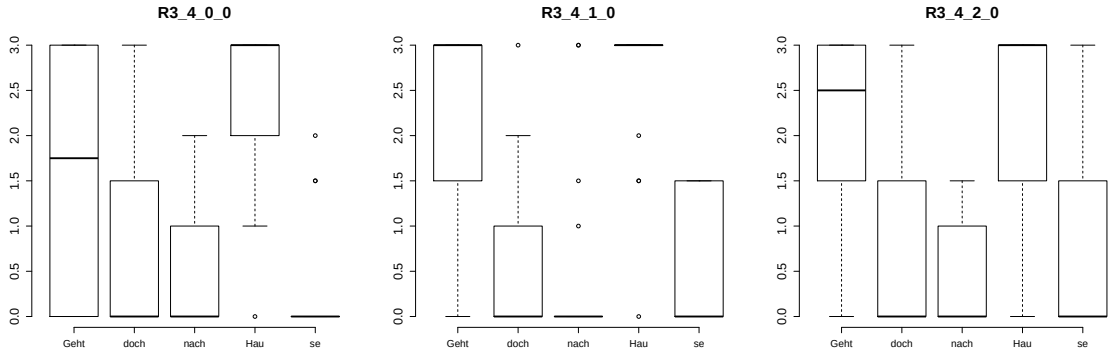

R2_31_1_0
0
5
10
15
20
25
30
Lass
ihn
rein

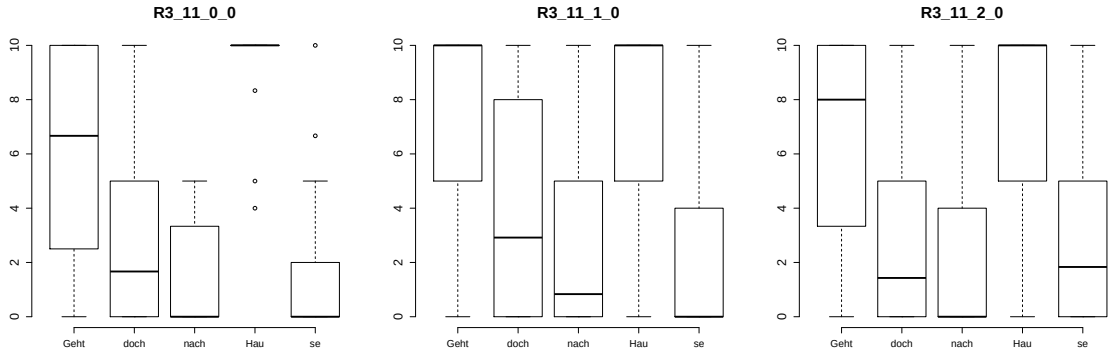

R2_31_2_0
0
5
10
15
20
25
30
Lass
ihn
rein

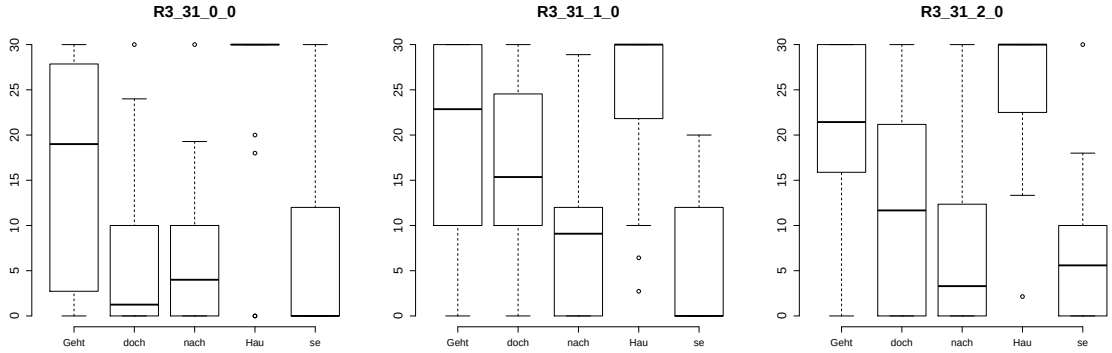

R2_Kont_0_0
0
50
100
150
200
250
300
Lass
ihn
rein

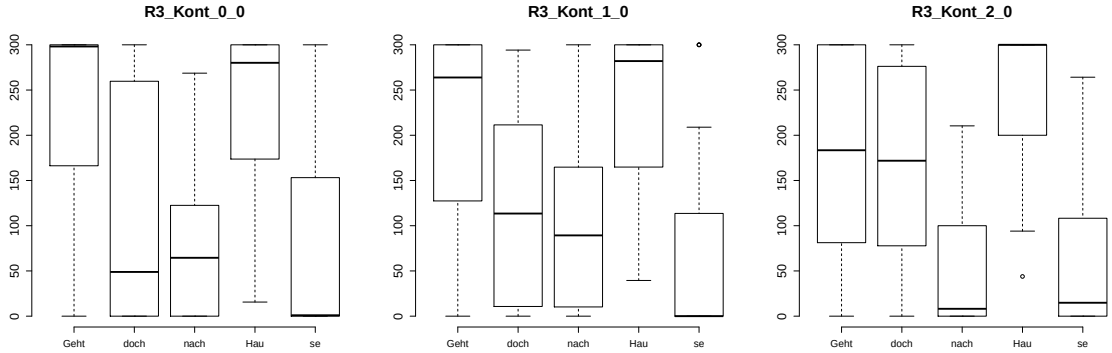

R2_Kont_1_0
0
50
100
150
200
250
300
Lass
ihn
rein

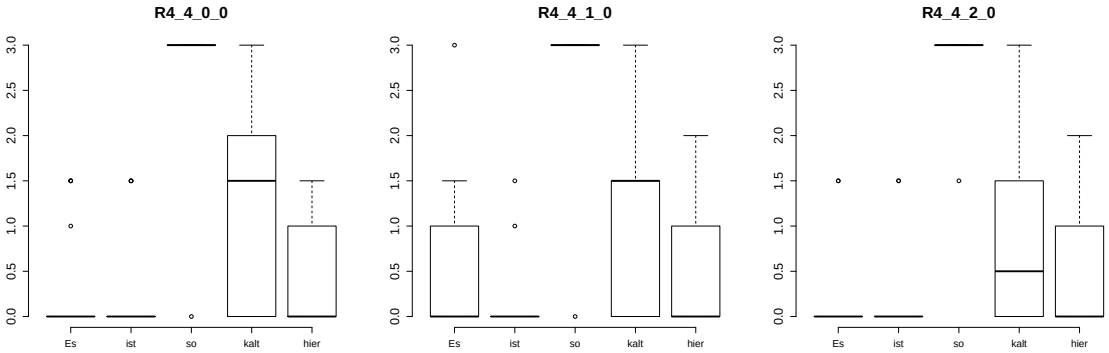

R2_Kont_2_0
0
50
100
150
200
250
300
Lass
ihn
rein

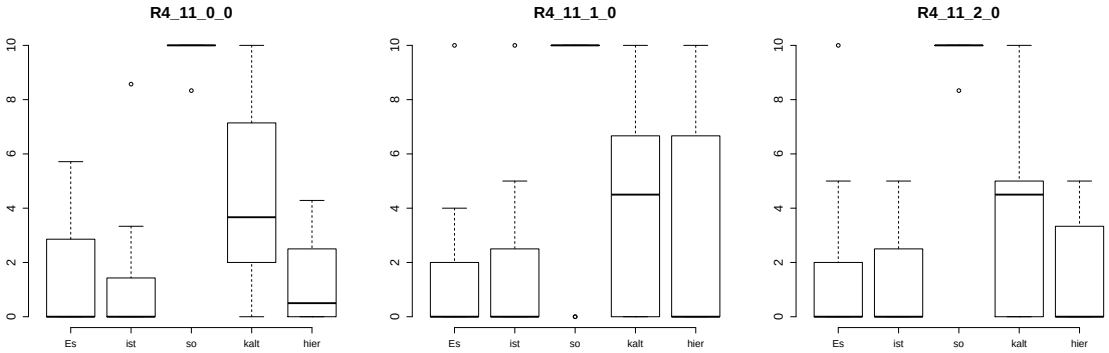

R3_4_0_0
R3_4_1_0
R3_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Geht
doch
nach
Hau
se

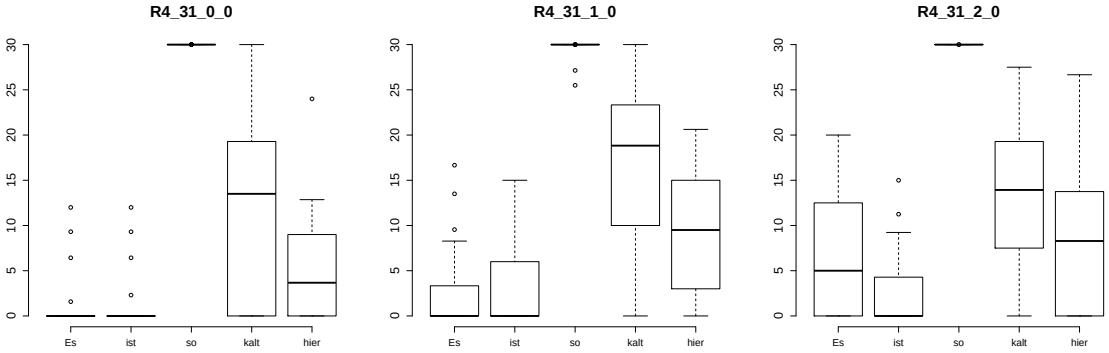

R3_11_0_0
R3_11_1_0
R3_11_2_0
0
2
4
6
8
10
Geht
doch
nach
Hau
se

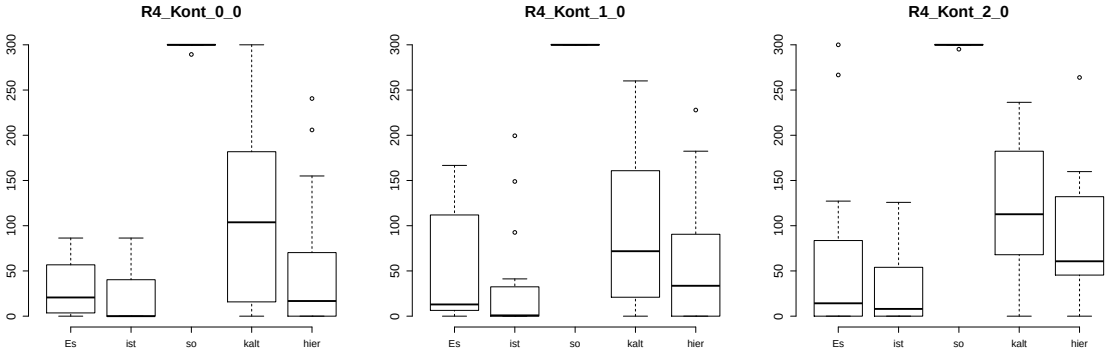

R3_31_0_0
R3_31_1_0
R3_31_2_0
0
5
10
15
20
25
30
Geht
doch
nach
Hau
se

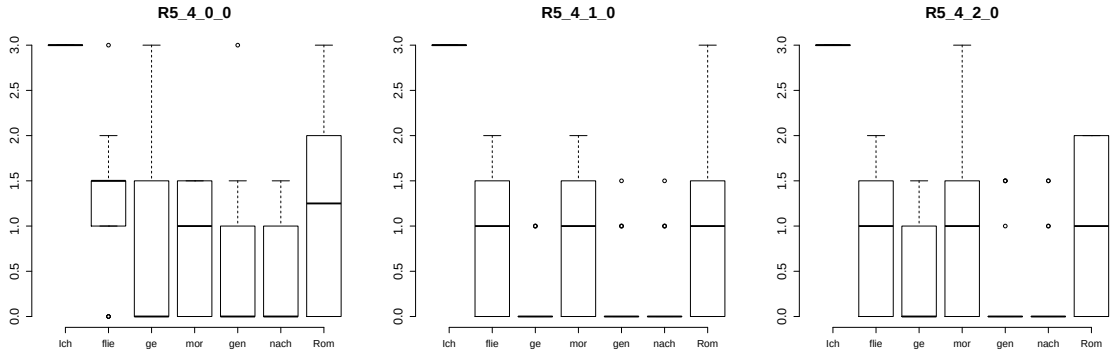

R3_Kont_0_0
R3_Kont_1_0
R3_Kont_2_0
0
50
100
150
200
250
300
Geht
doch
nach
Hau
se

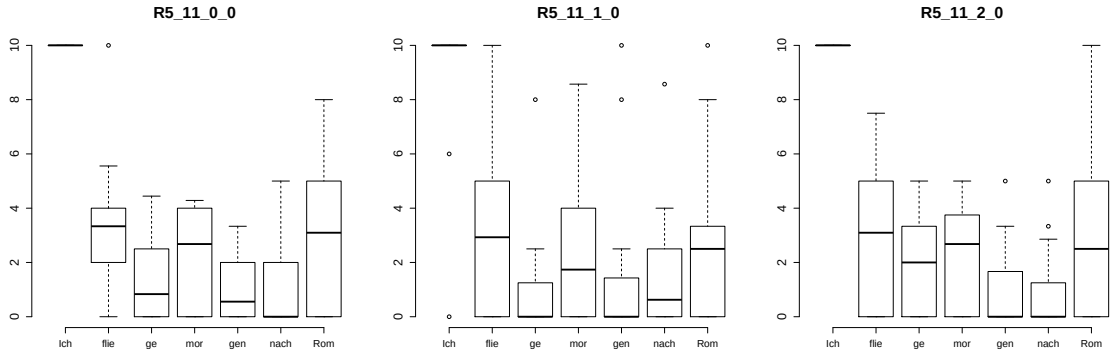

R4_4_0_0
R4_4_1_0
R4_4_2_0
Es
ist
so
kalt
hier

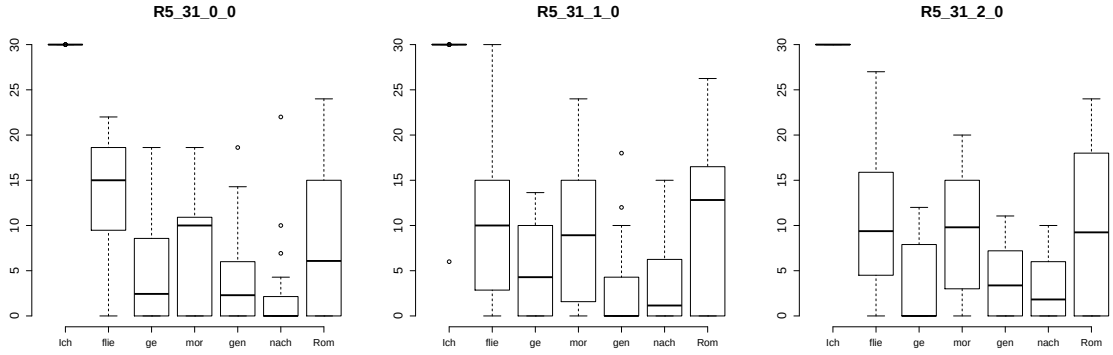

R4_11_0_0
R4_11_1_0
R4_11_2_0
Es
ist
so
kalt
hier

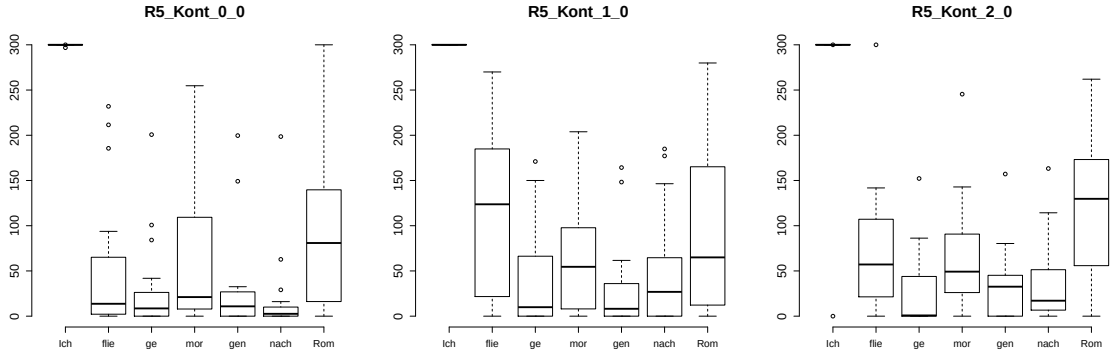

R4_31_0_0
R4_31_1_0
R4_31_2_0
Es
ist
so
kalt
hier

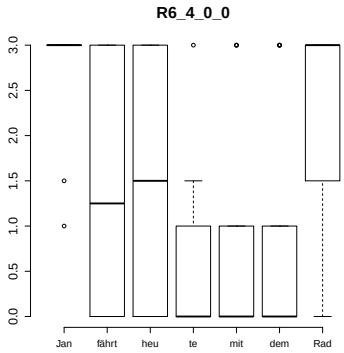

R4_Kont_0_0
R4_Kont_1_0
R4_Kont_2_0
Es
ist
so
kalt
hier

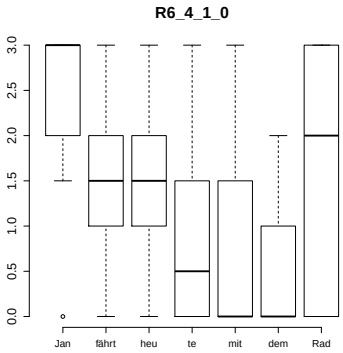

R5_4_0_0
R5_4_1_0
R5_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Ich
flie
ge
mor
gen
nach
Rom

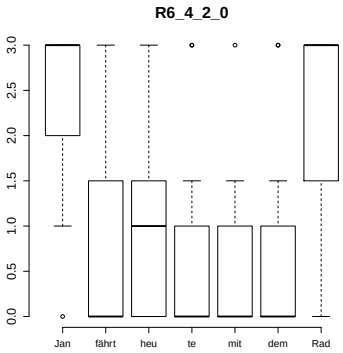

R5_11_0_0
R5_11_1_0
R5_11_2_0
0
2
4
6
8
10
Ich
flie
ge
mor
gen
nach
Rom

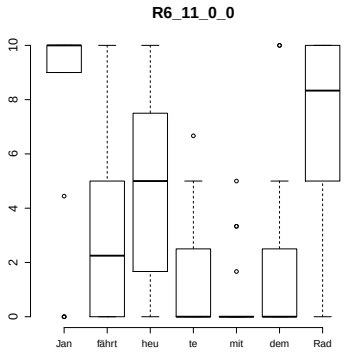

R5_31_0_0
R5_31_1_0
R5_31_2_0
0
5
10
15
20
25
30
Ich
flie
ge
mor
gen
nach
Rom

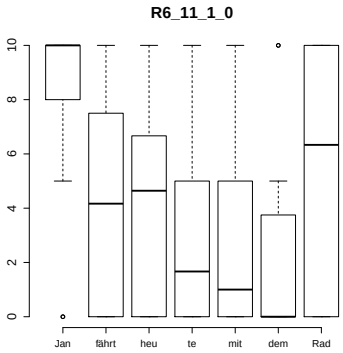

R5_Kont_0_0
R5_Kont_1_0
R5_Kont_2_0
0
50
100
150
200
250
300
Ich
flie
ge
mor
gen
nach
Rom

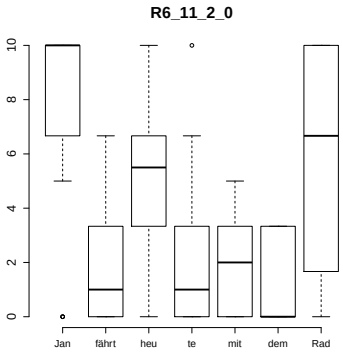

R6_4_0_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Jan
fährt
heu
te
mit
dem
Rad

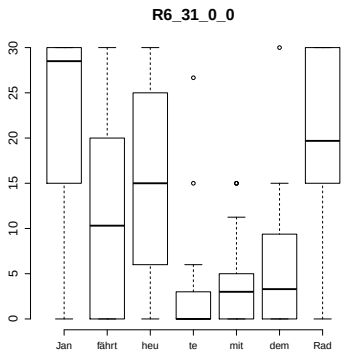

R6_4_1_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Jan
fährt
heu
te
mit
dem
Rad

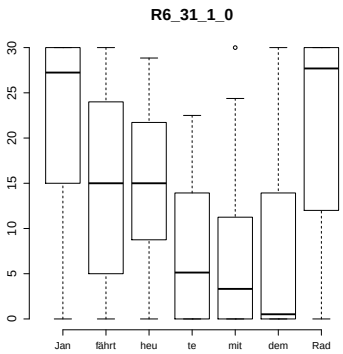

R6_4_2_0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Jan
fährt
heu
te
mit
dem
Rad

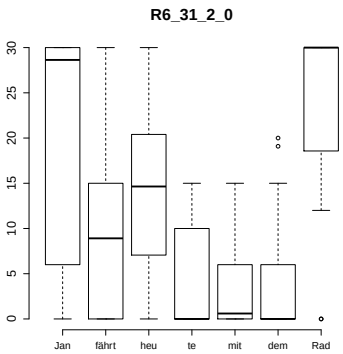

R6_11_0_0
0
2
4
6
8
10
Jan
fährt
heu
te
mit
dem
Rad

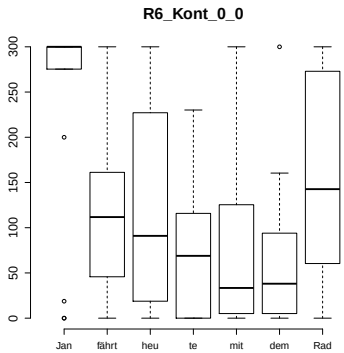

R6_11_1_0
0
2
4
6
8
10
Jan
fährt
heu
te
mit
dem
Rad

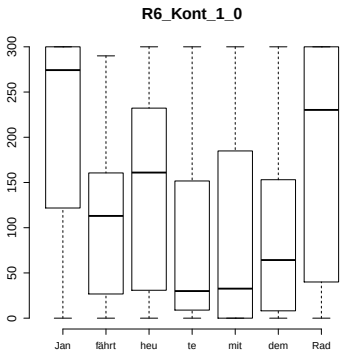

R6_11_2_0
0
2
4
6
8
10
Jan
fährt
heu
te
mit
dem
Rad

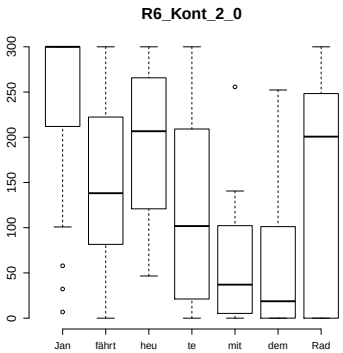

R6_31_0_0
0
5
10
15
20
25
30
Jan
fährt
heu
te
mit
dem
Rad

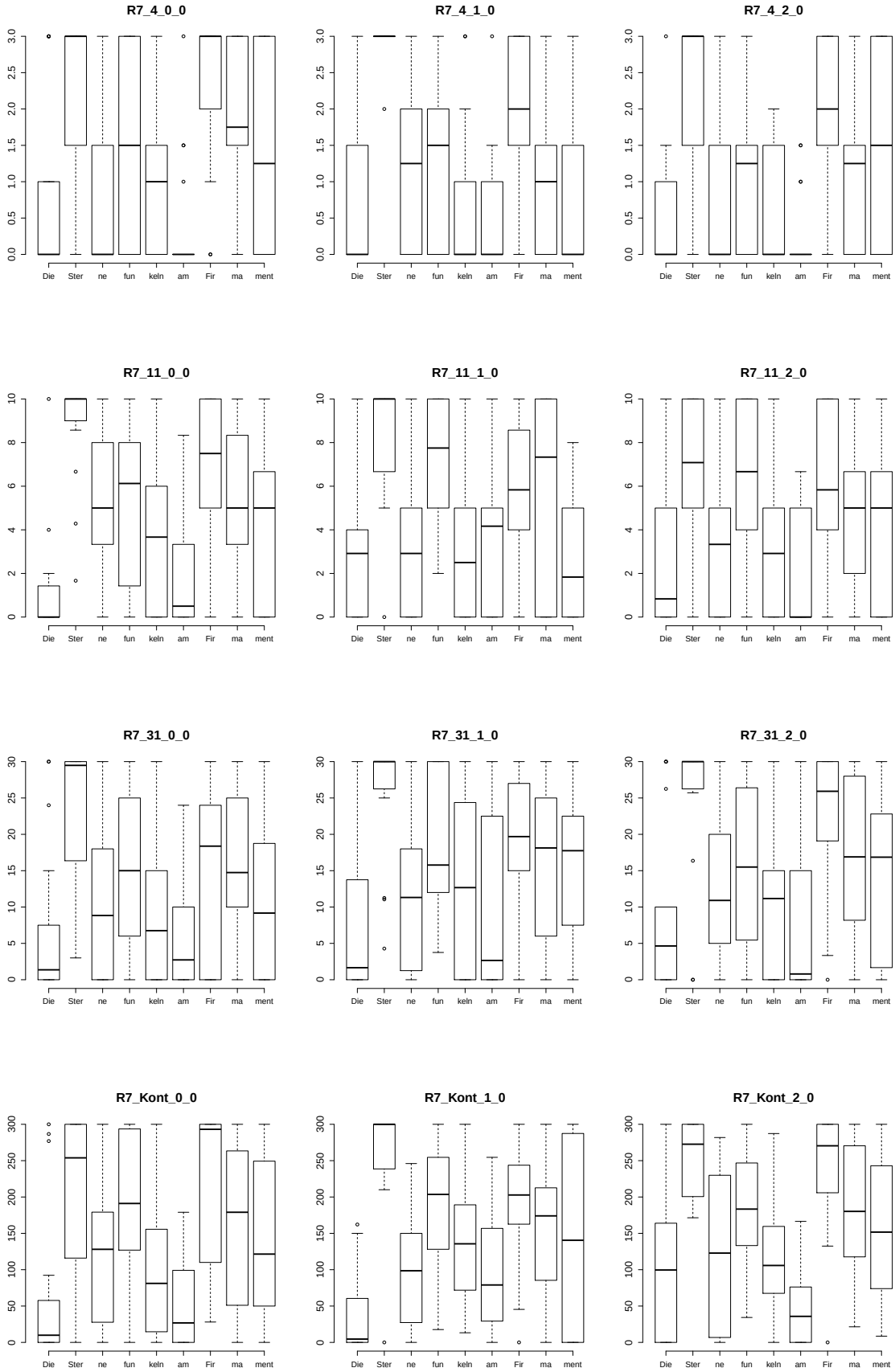

R6_31_1_0
0
5
10
15
20
25
30
Jan
fährt
heu
te
mit
dem
Rad

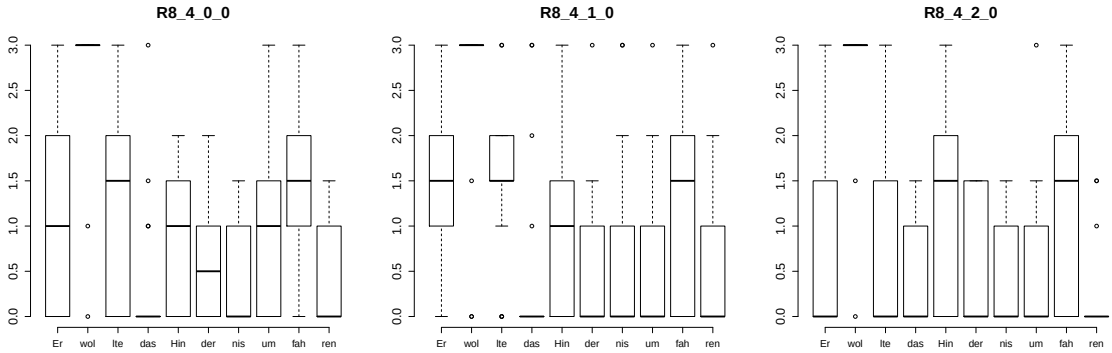

R6_31_2_0
0
5
10
15
20
25
30
Jan
fährt
heu
te
mit
dem
Rad

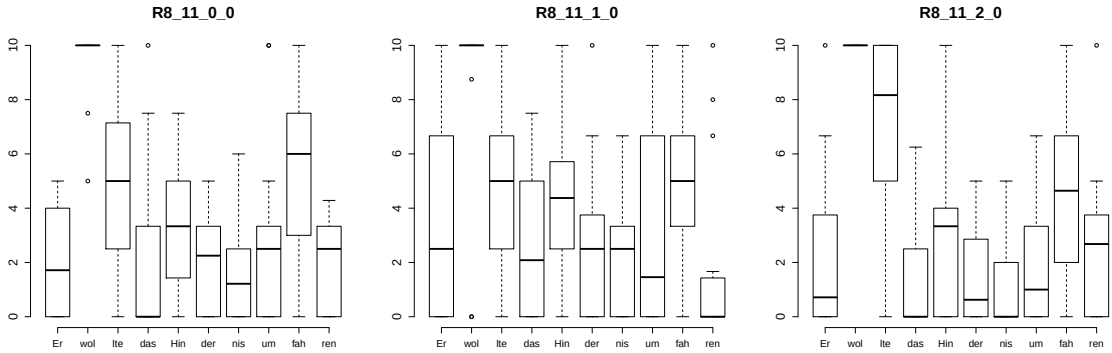

R6_Kont_0_0
0
50
100
150
200
250
300
Jan
fährt
heu
te
mit
dem
Rad

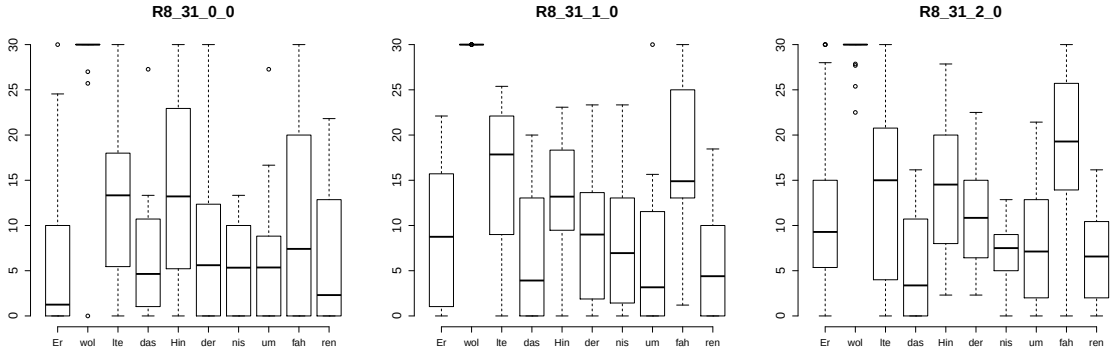

R6_Kont_1_0
0
50
100
150
200
250
300
Jan
fährt
heu
te
mit
dem
Rad

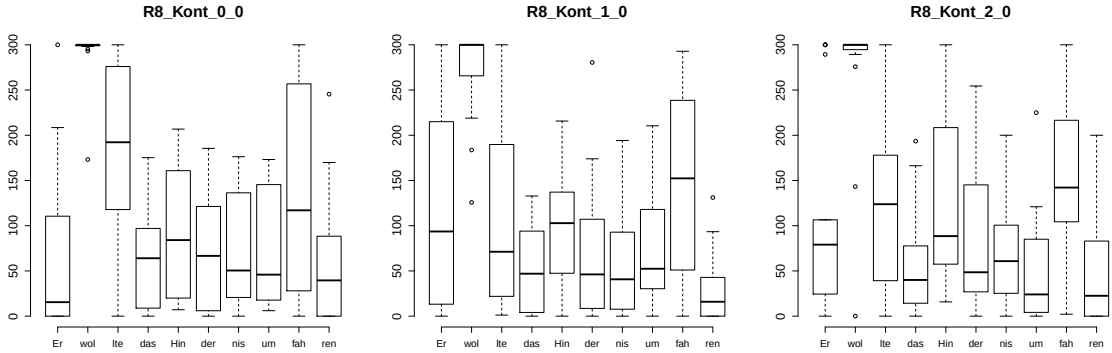

R6_Kont_2_0
0
50
100
150
200
250
300
Jan
fährt
heu
te
mit
dem
Rad

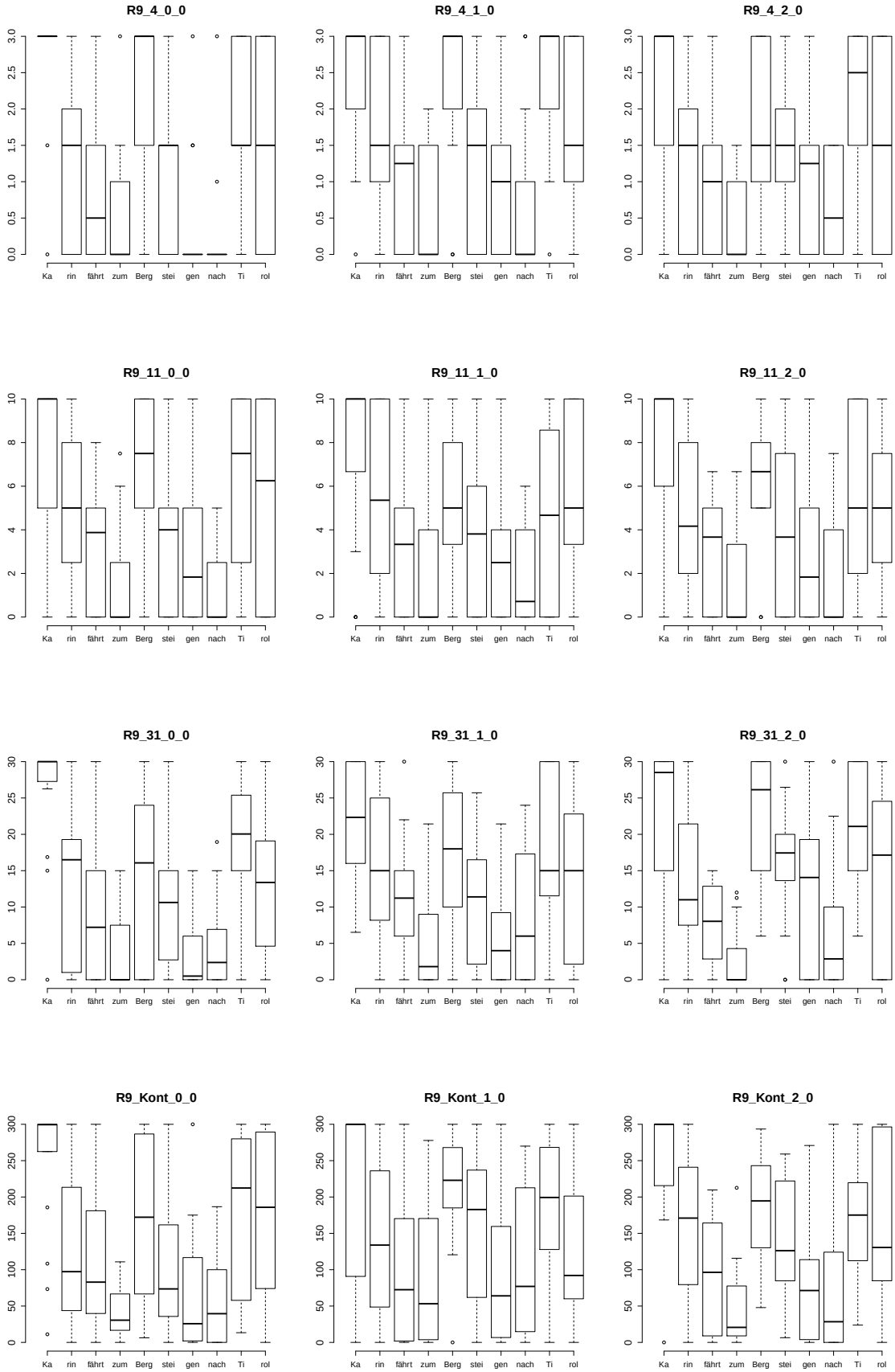

R7_4_0_0
R7_4_1_0
R7_4_2_0
R7_11_0_0
R7_11_1_0
R7_11_2_0
R7_31_0_0
R7_31_1_0
R7_31_2_0
R7_Kont_0_0
R7_Kont_1_0
R7_Kont_2_0
Die Ster ne fun keln am Fir ma ment

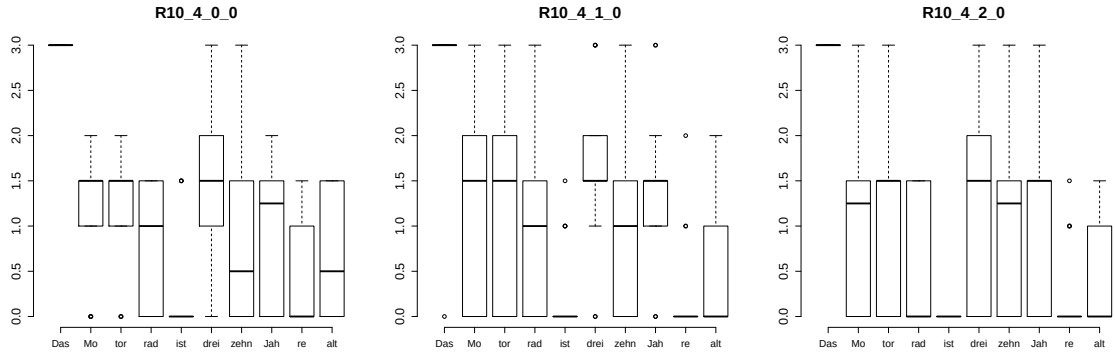

R8_4_0_0
R8_4_1_0
R8_4_2_0
Er wol lte das Hin der nis um fah ren

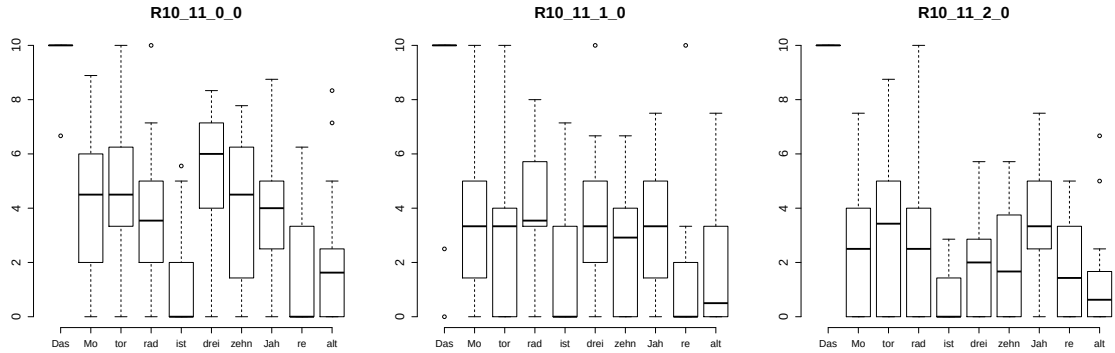

R8_11_0_0
R8_11_1_0
R8_11_2_0
Er wol lte das Hin der nis um fah ren

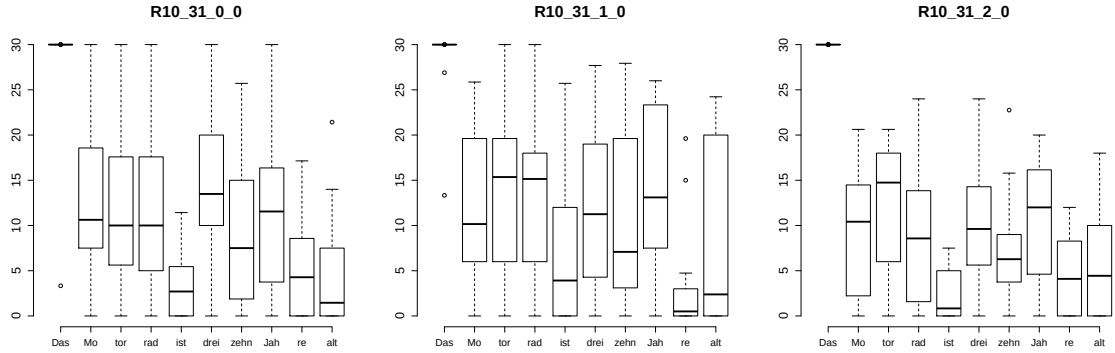

R8_31_0_0
R8_31_1_0
R8_31_2_0
Er wol lte das Hin der nis um fah ren

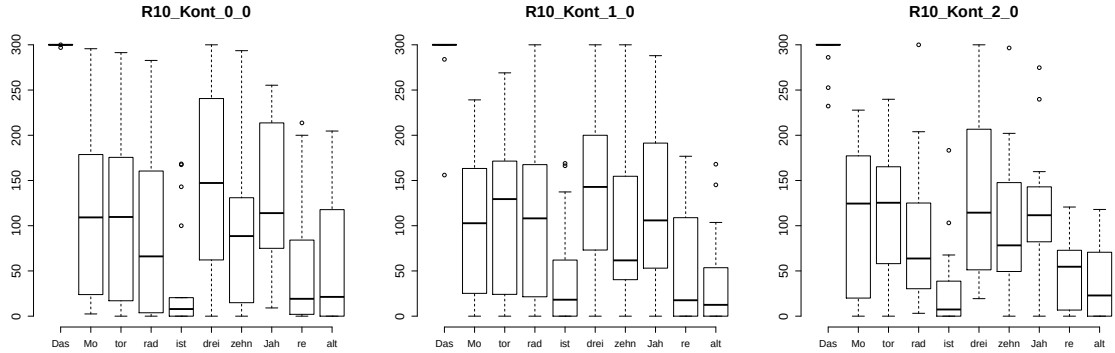

R8_Kont_0_0
R8_Kont_1_0
R8_Kont_2_0
Er wol lte das Hin der nis um fah ren

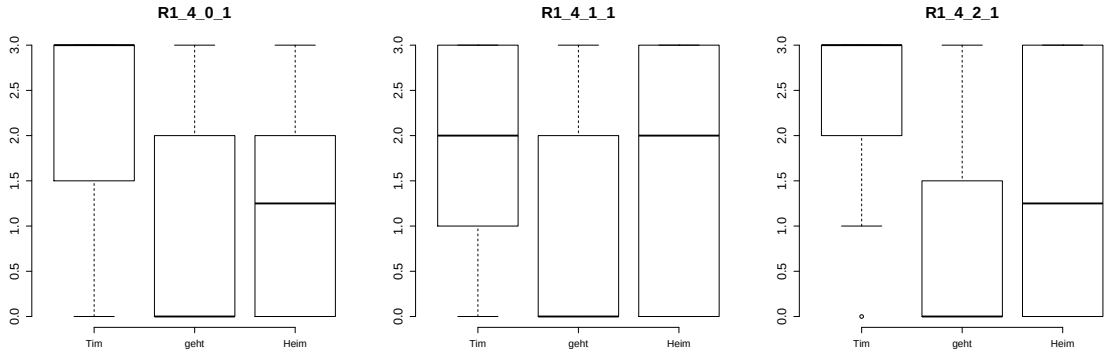

R9_4_0_0
R9_4_1_0
R9_4_2_0
R9_11_0_0
R9_11_1_0
R9_11_2_0
R9_31_0_0
R9_31_1_0
R9_31_2_0
R9_Kont_0_0
R9_Kont_1_0
R9_Kont_2_0
Ka rin fährt zum Berg stei gen nach Ti rol

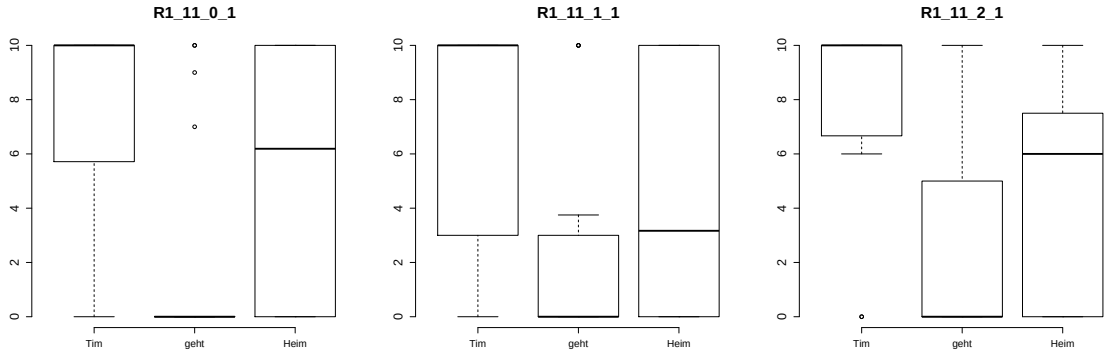

R10_4_0_0
R10_4_1_0
R10_4_2_0
Das Mo tor rad ist drei zehn Jah re alt

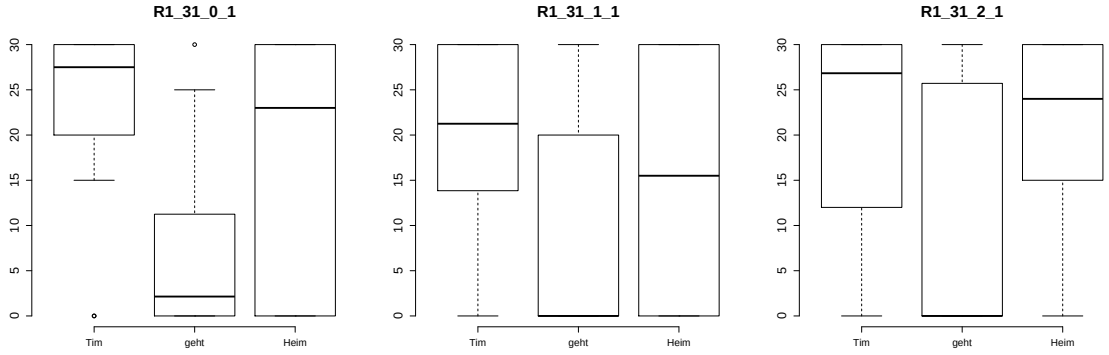

R10_11_0_0
R10_11_1_0
R10_11_2_0
Das Mo tor rad ist drei zehn Jah re alt

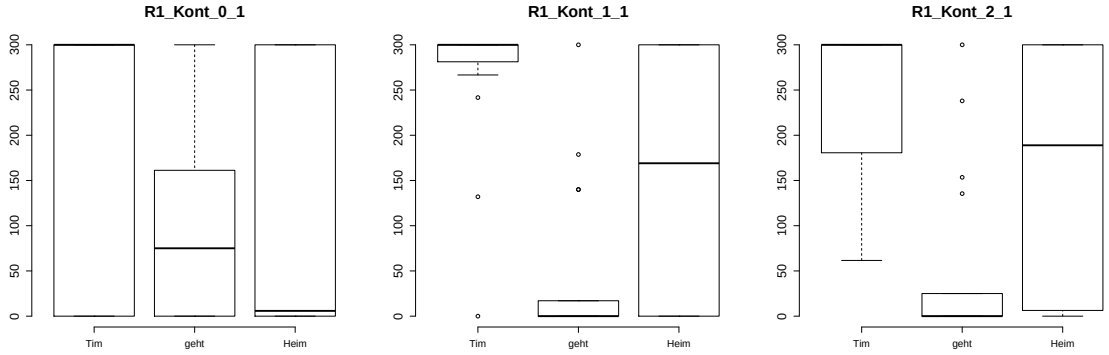

R10_31_0_0
R10_31_1_0
R10_31_2_0
Das Mo tor rad ist drei zehn Jah re alt

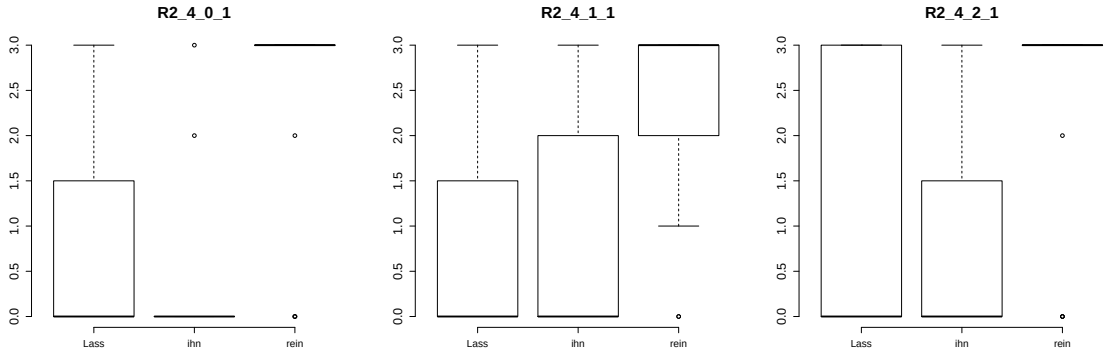

R10_Kont_0_0
R10_Kont_1_0
R10_Kont_2_0
Das Mo tor rad ist drei zehn Jah re alt

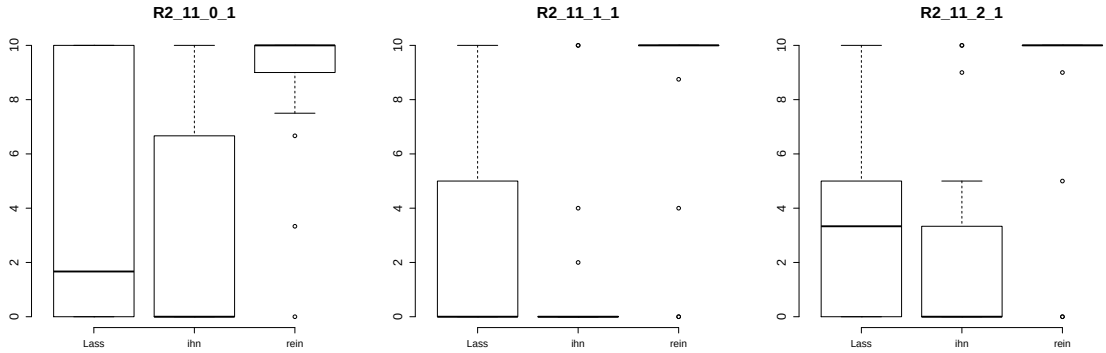

R1_4_0_1
R1_4_1_1
R1_4_2_1
Tim
geht
Heim

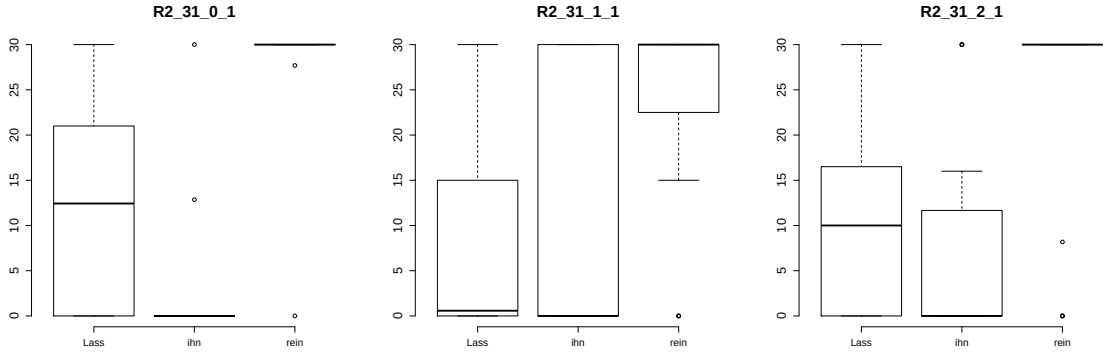

R1_11_0_1
R1_11_1_1
R1_11_2_1
Tim
geht
Heim

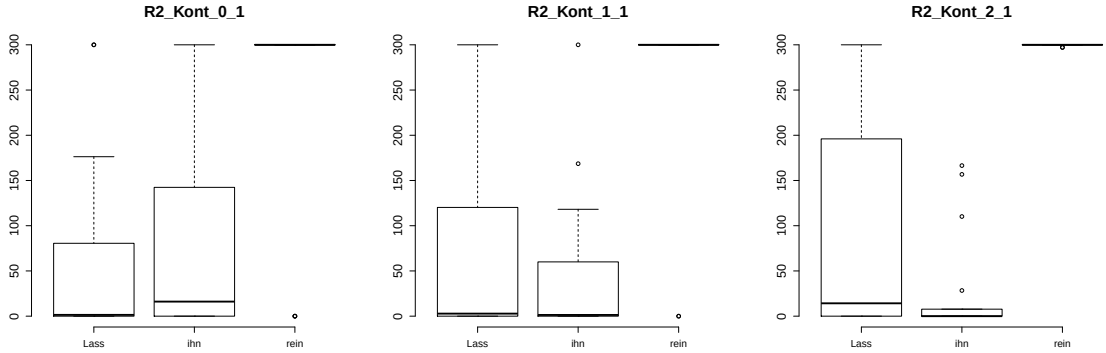

R1_31_0_1
R1_31_1_1
R1_31_2_1
Tim
geht
Heim

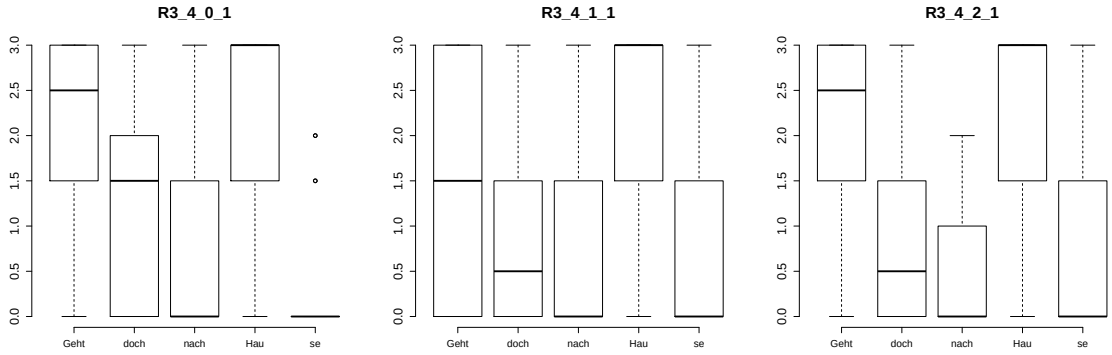

R1_Kont_0_1
R1_Kont_1_1
R1_Kont_2_1
Tim
geht
Heim

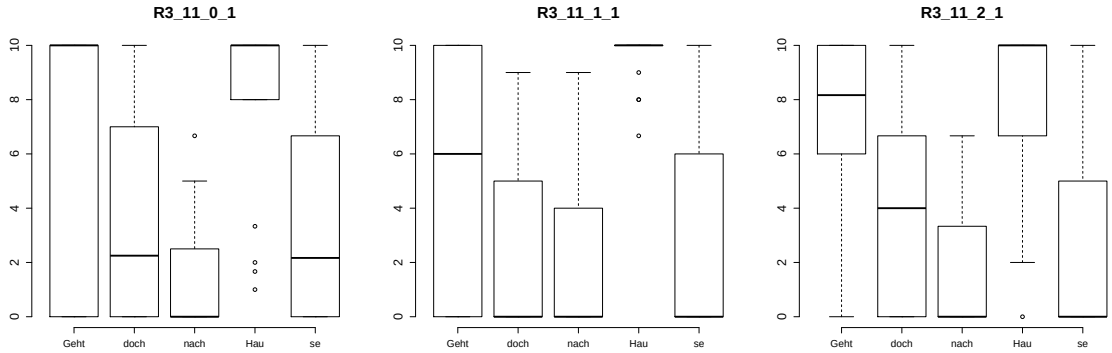

R2_4_0_1
R2_4_1_1
R2_4_2_1
Lass
ihn
rein

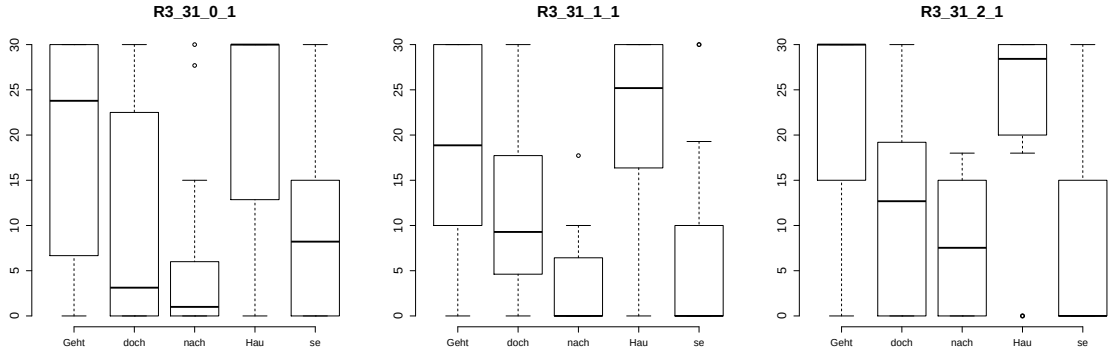

R2_11_0_1
R2_11_1_1
R2_11_2_1
Lass
ihn
rein

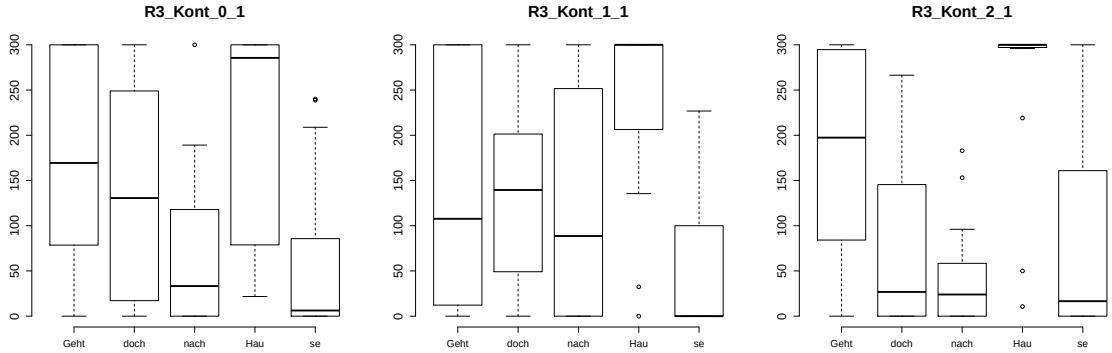

R2_31_0_1
R2_31_1_1
R2_31_2_1
Lass
ihn
rein

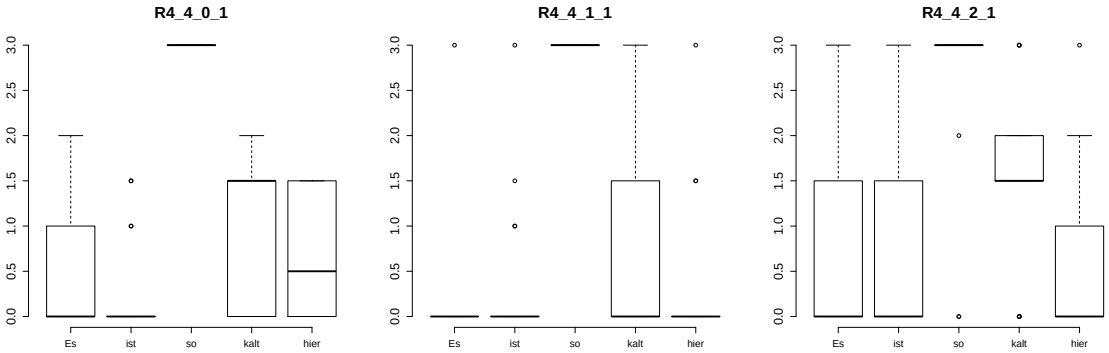

R2_Kont_0_1
R2_Kont_1_1
R2_Kont_2_1
Lass
ihn
rein

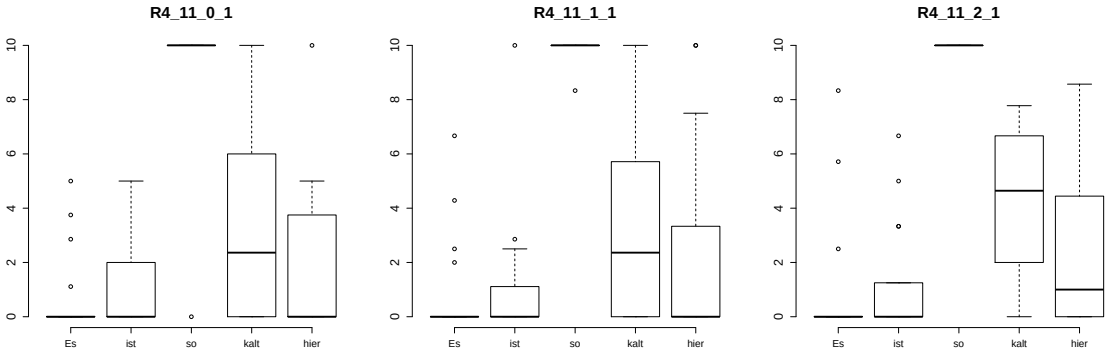

R3_4_0_1
R3_4_1_1
R3_4_2_1
Geht
doch
nach
Hau
se

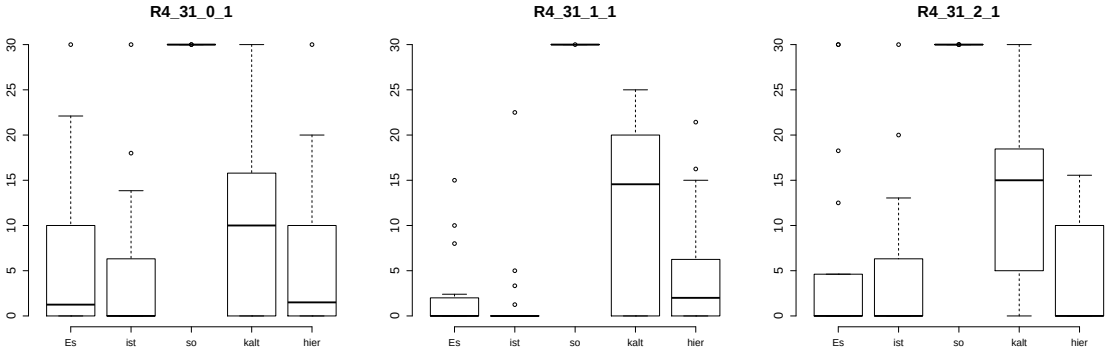

R3_11_0_1
R3_11_1_1
R3_11_2_1
Geht
doch
nach
Hau
se

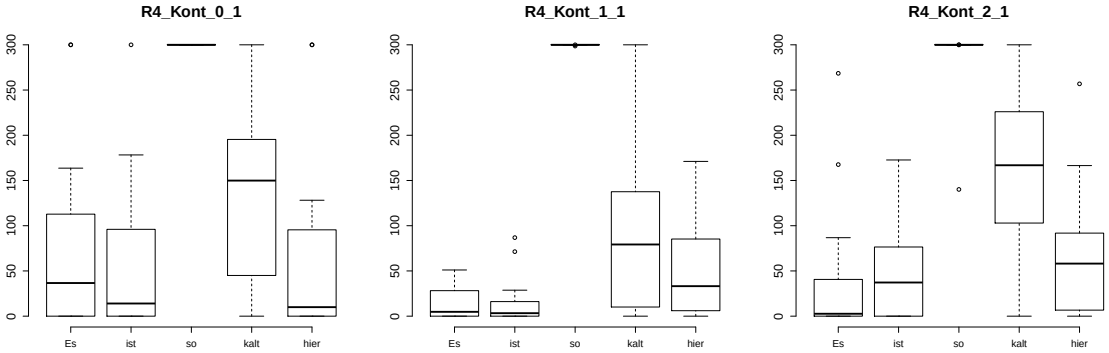

R3_31_0_1
R3_31_1_1
R3_31_2_1
Geht
doch
nach
Hau
se

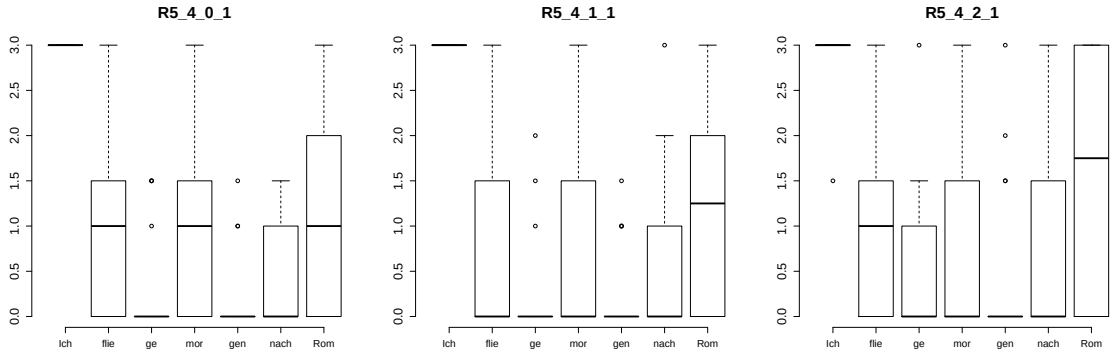

R3_Kont_0_1
R3_Kont_1_1
R3_Kont_2_1
Geht
doch
nach
Hau
se

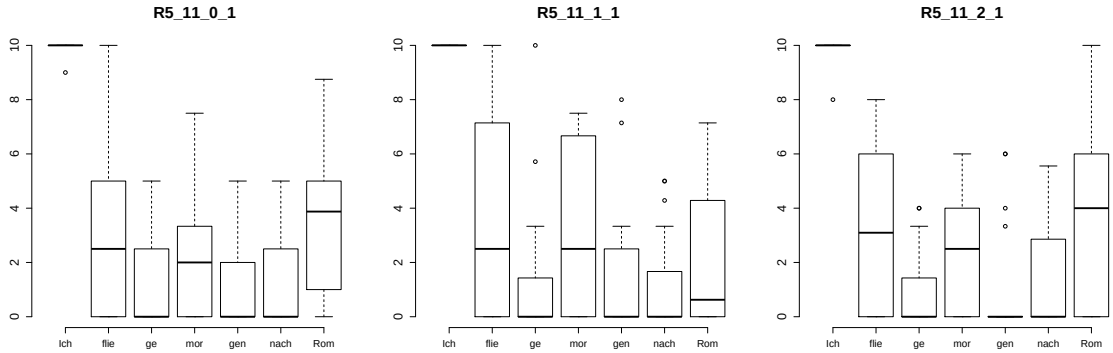

R4_4_0_1
R4_4_1_1
R4_4_2_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Es
ist
so
kalt
hier

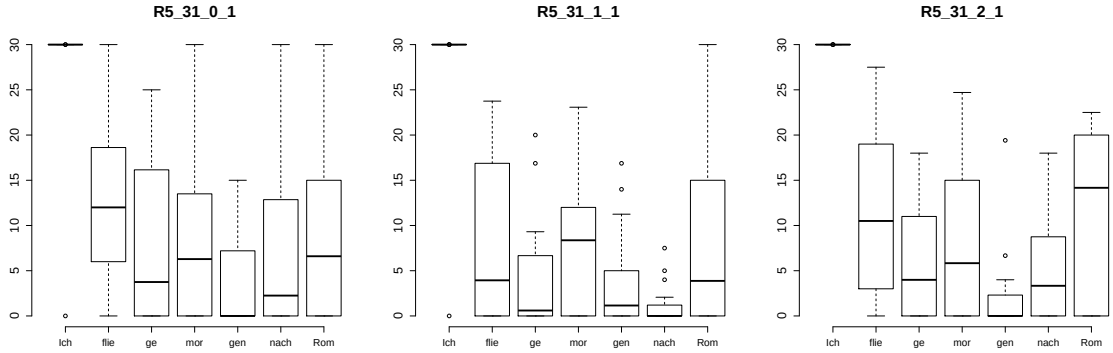

R4_11_0_1
R4_11_1_1
R4_11_2_1
0
2
4
6
8
10
Es
ist
so
kalt
hier

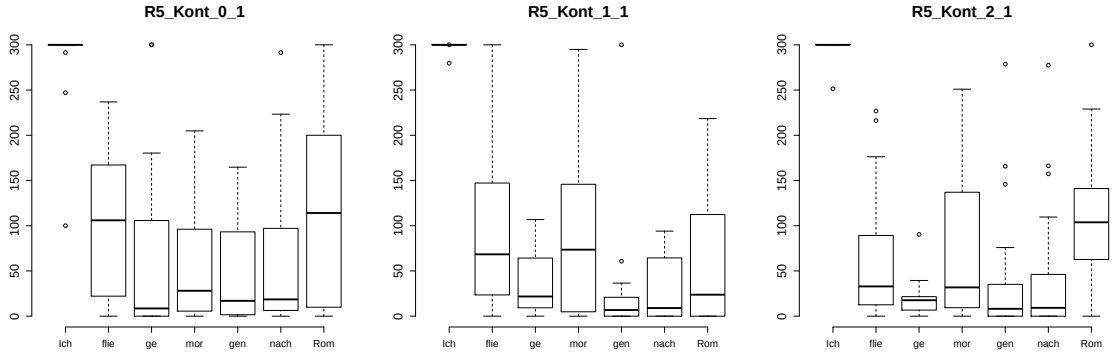

R4_31_0_1
R4_31_1_1
R4_31_2_1
0
5
10
15
20
25
30
Es
ist
so
kalt
hier

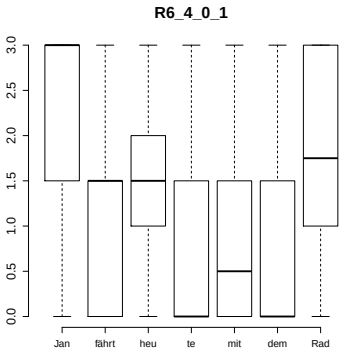

R4_Kont_0_1
R4_Kont_1_1
R4_Kont_2_1
0
50
100
150
200
250
300
Es
ist
so
kalt
hier

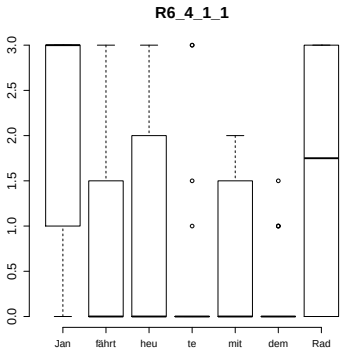

R5_4_0_1
R5_4_1_1
R5_4_2_1
Ich
flie
ge
mor
gen
nach
Rom

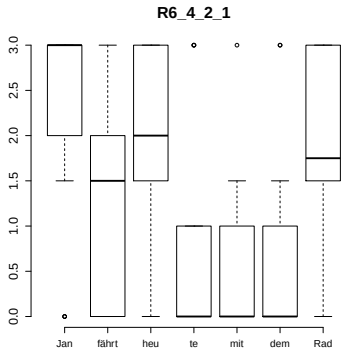

R5_11_0_1
R5_11_1_1
R5_11_2_1
Ich
flie
ge
mor
gen
nach
Rom

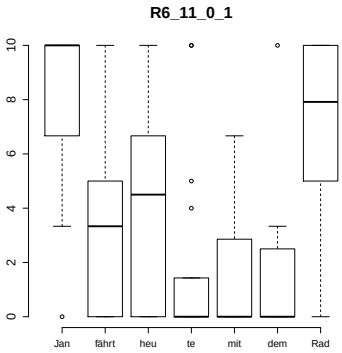

R5_31_0_1
R5_31_1_1
R5_31_2_1
Ich
flie
ge
mor
gen
nach
Rom

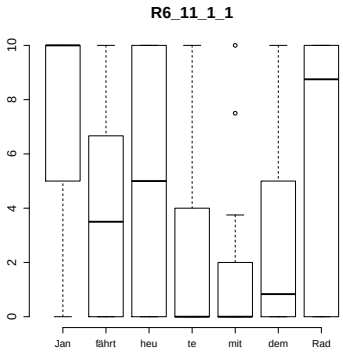

R5_Kont_0_1
R5_Kont_1_1
R5_Kont_2_1
Ich
flie
ge
mor
gen
nach
Rom

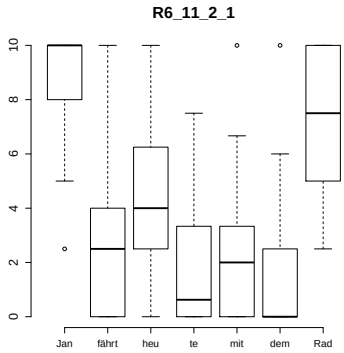

R6_4_0_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Jan
fährt
heu
te
mit
dem
Rad

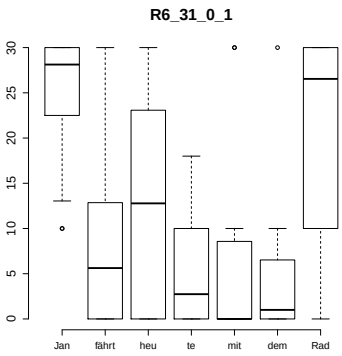

R6_4_1_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Jan
fährt
heu
te
mit
dem
Rad

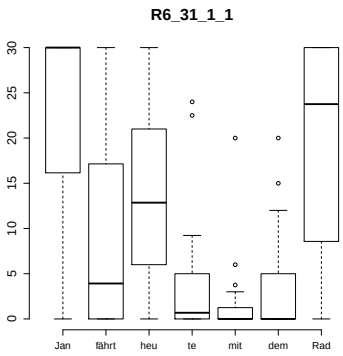

R6_4_2_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Jan
fährt
heu
te
mit
dem
Rad

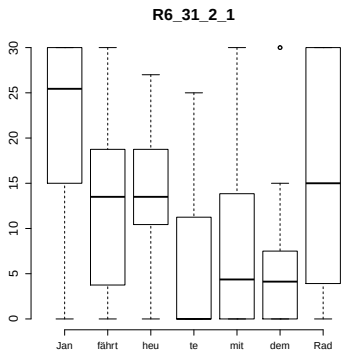

R6_11_0_1
0
2
4
6
8
10
Jan
fährt
heu
te
mit
dem
Rad

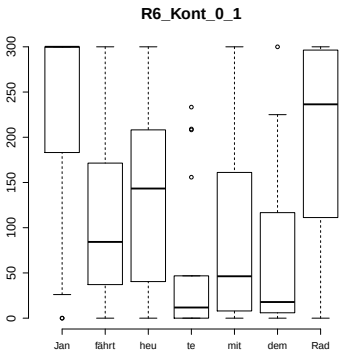

R6_11_1_1
0
2
4
6
8
10
Jan
fährt
heu
te
mit
dem
Rad

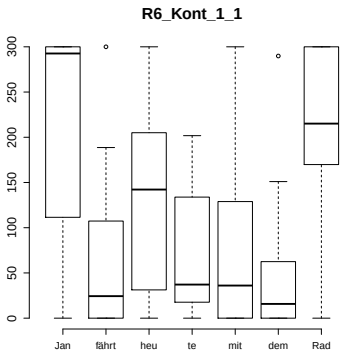

R6_11_2_1
0
2
4
6
8
10
Jan
fährt
heu
te
mit
dem
Rad

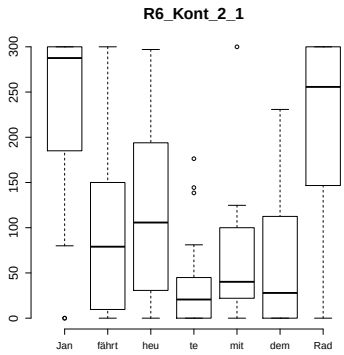

R6_31_0_1
0
5
10
15
20
25
30
Jan
fährt
heu
te
mit
dem
Rad

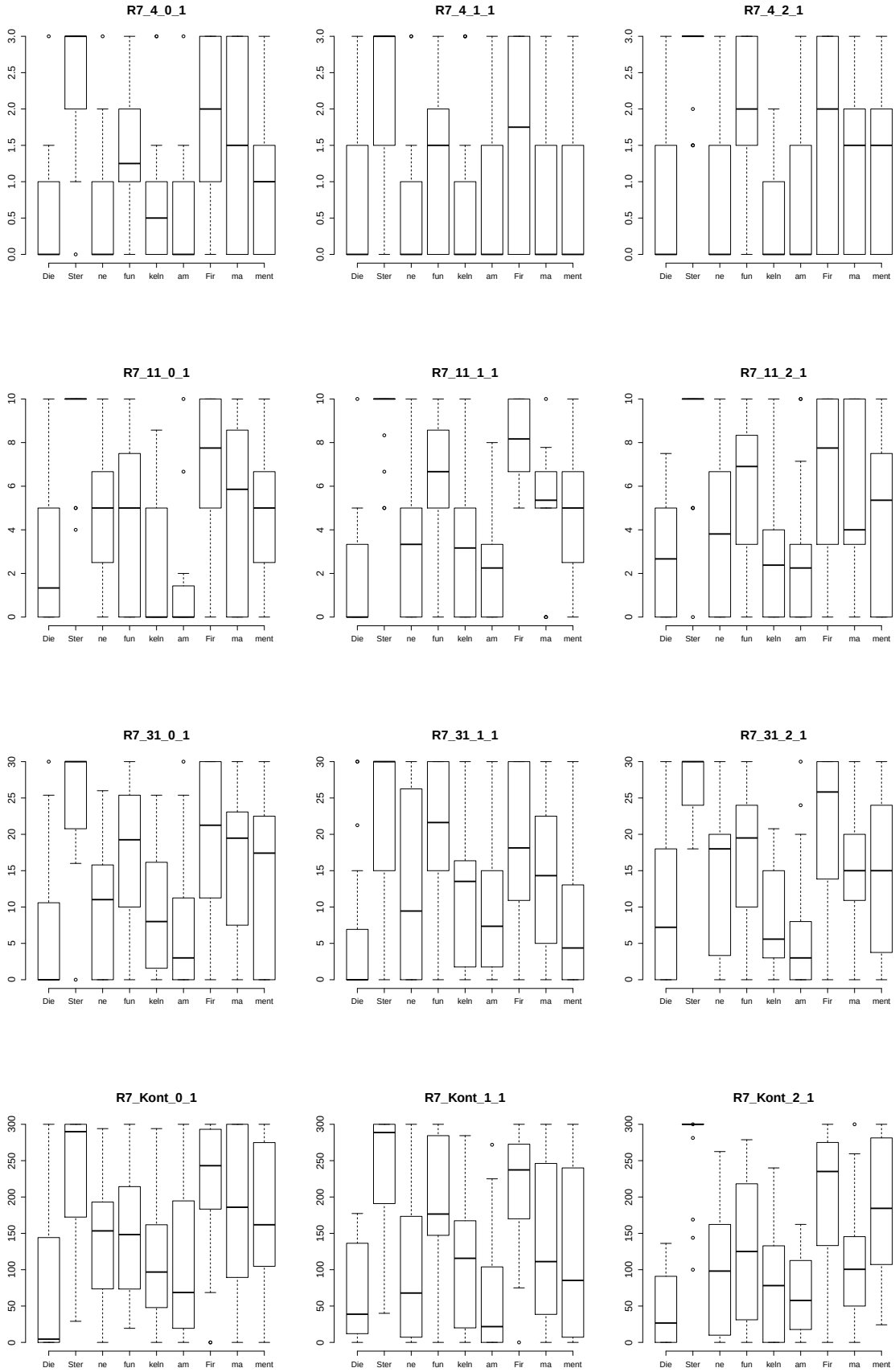

R6_31_1_1
0
5
10
15
20
25
30
Jan
fährt
heu
te
mit
dem
Rad

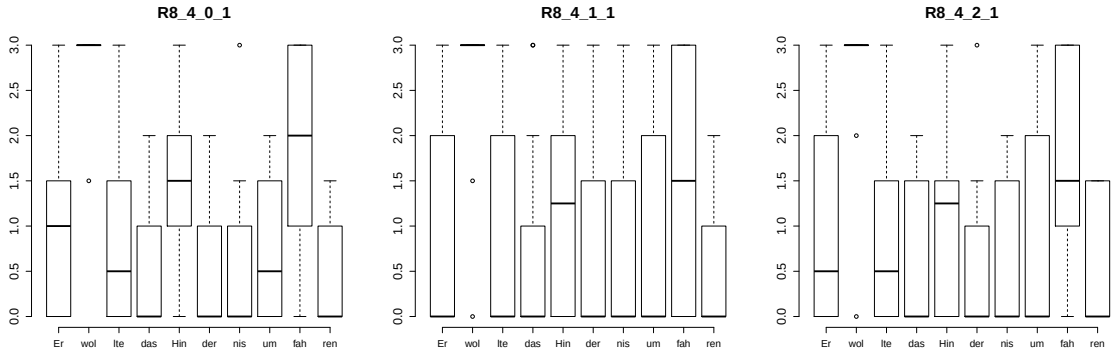

R6_31_2_1
0
5
10
15
20
25
30
Jan
fährt
heu
te
mit
dem
Rad

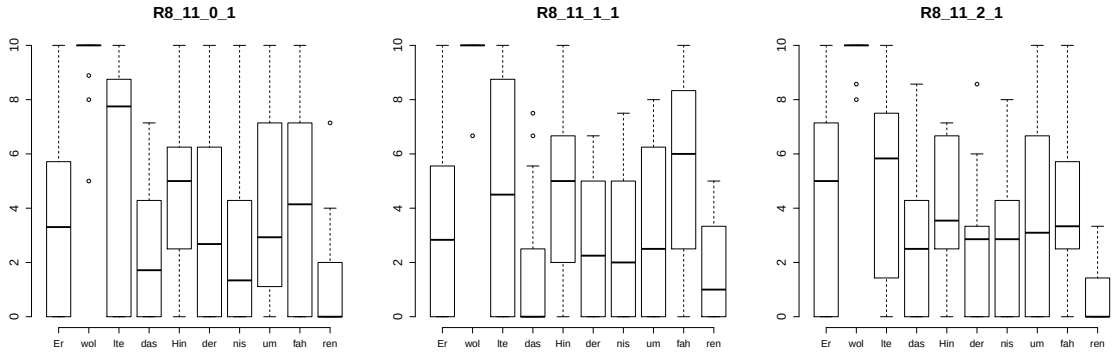

R6_Kont_0_1
0
50
100
150
200
250
300
Jan
fährt
heu
te
mit
dem
Rad

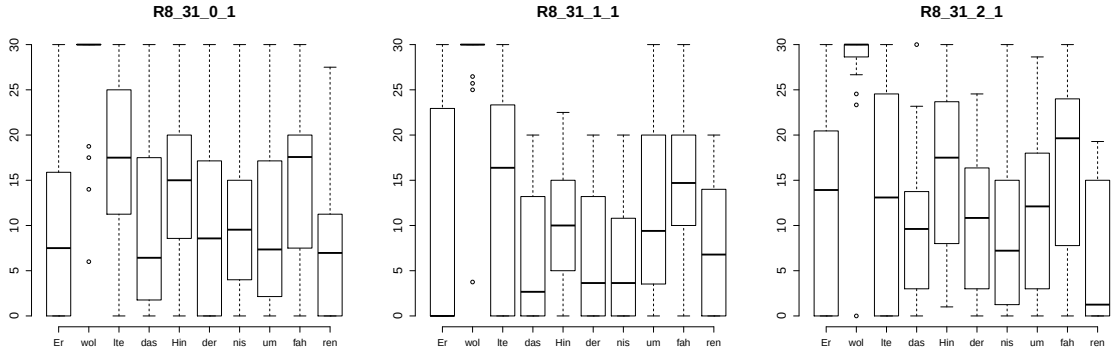

R6_Kont_1_1
0
50
100
150
200
250
300
Jan
fährt
heu
te
mit
dem
Rad

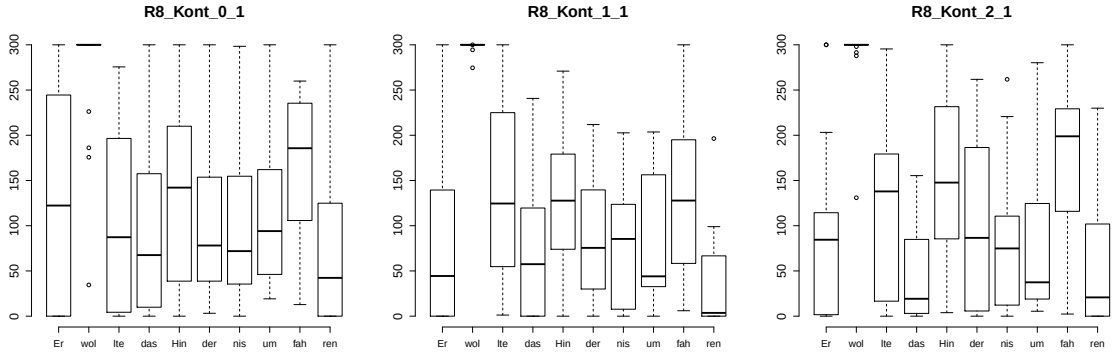

R6_Kont_2_1
0
50
100
150
200
250
300
Jan
fährt
heu
te
mit
dem
Rad

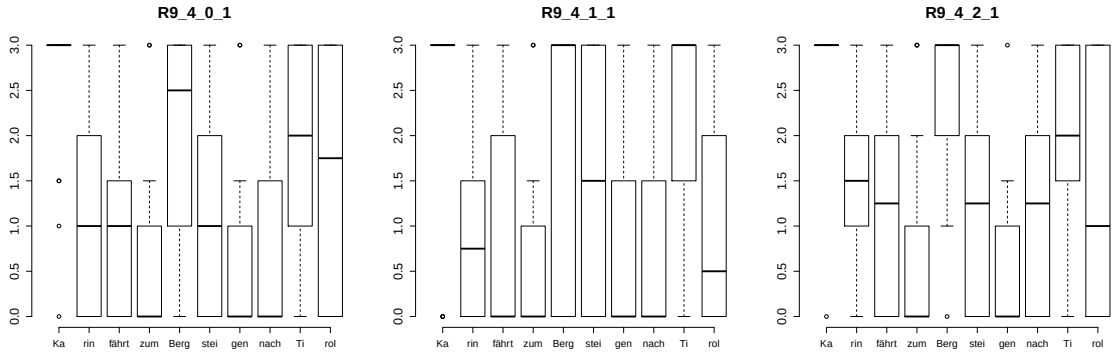

R7_4_0_1
R7_4_1_1
R7_4_2_1
R7_11_0_1
R7_11_1_1
R7_11_2_1
R7_31_0_1
R7_31_1_1
R7_31_2_1
R7_Kont_0_1
R7_Kont_1_1
R7_Kont_2_1
Die Ster ne fun keln am Fir ma ment

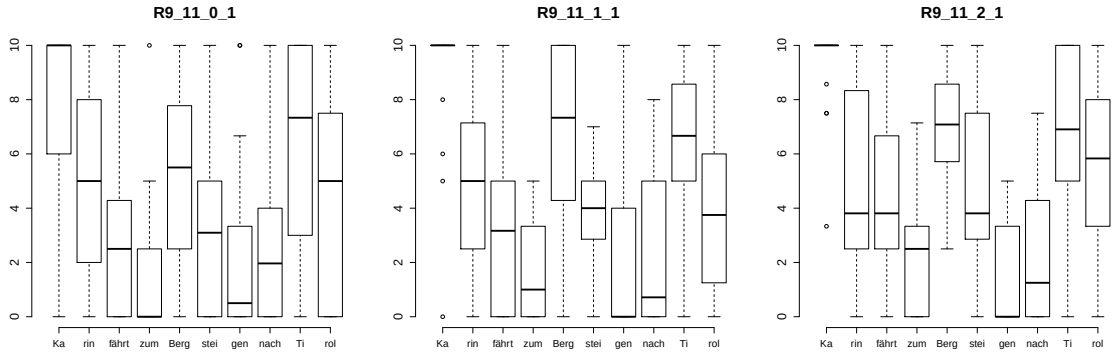

R8_4_0_1
R8_4_1_1
R8_4_2_1
Er wol lte das Hin der nis um fah ren

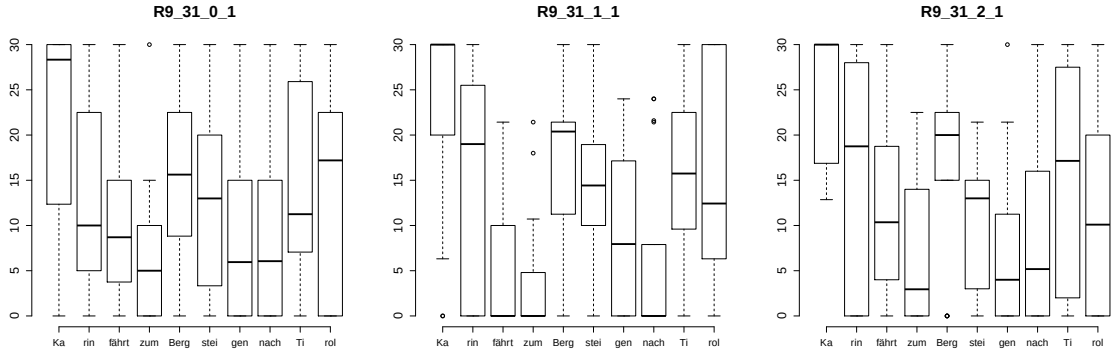

R8_11_0_1
R8_11_1_1
R8_11_2_1
Er wol lte das Hin der nis um fah ren

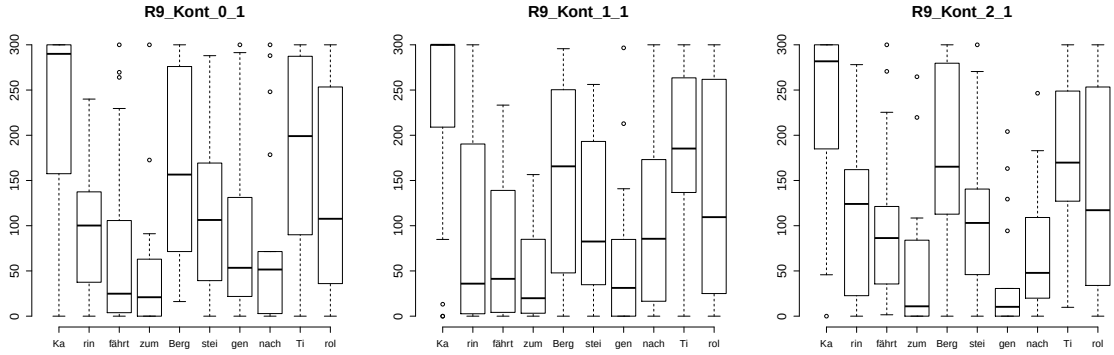

R8_31_0_1
R8_31_1_1
R8_31_2_1
Er wol lte das Hin der nis um fah ren

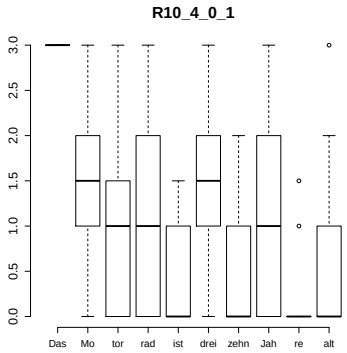

R8_Kont_0_1
R8_Kont_1_1
R8_Kont_2_1
Er wol lte das Hin der nis um fah ren

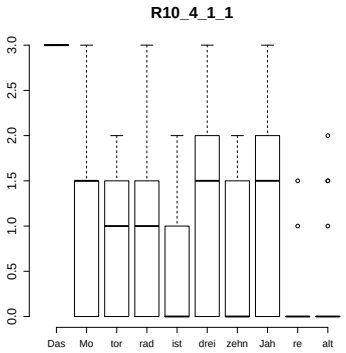

R9_4_0_1
R9_4_1_1
R9_4_2_1
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Ka rin fährt zum Berg stei gen nach Ti rol

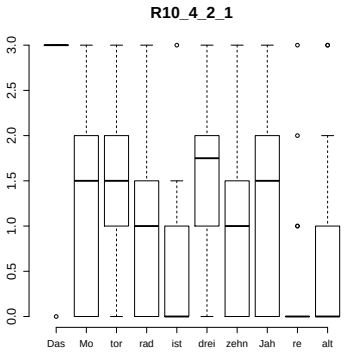

R9_11_0_1
R9_11_1_1
R9_11_2_1
0 2 4 6 8 10
Ka rin fährt zum Berg stei gen nach Ti rol

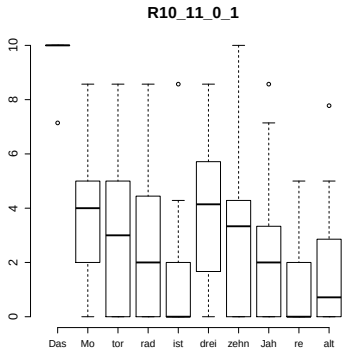

R9_31_0_1
R9_31_1_1
R9_31_2_1
0 5 10 15 20 25 30
Ka rin fährt zum Berg stei gen nach Ti rol

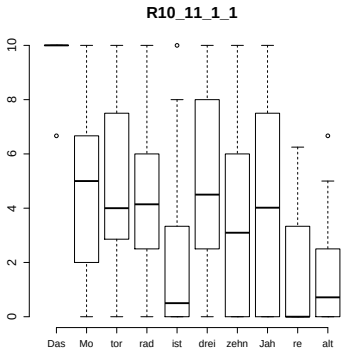

R9_Kont_0_1
R9_Kont_1_1
R9_Kont_2_1
0 50 100 150 200 250 300
Ka rin fährt zum Berg stei gen nach Ti rol

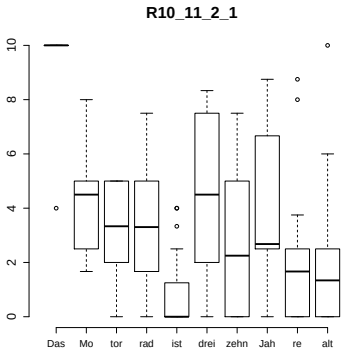

R10_4_0_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

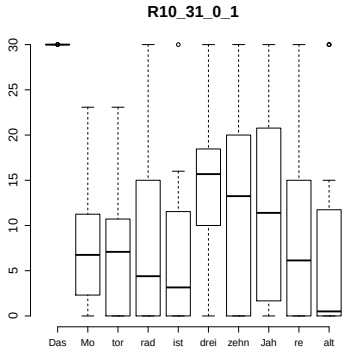

R10_4_1_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

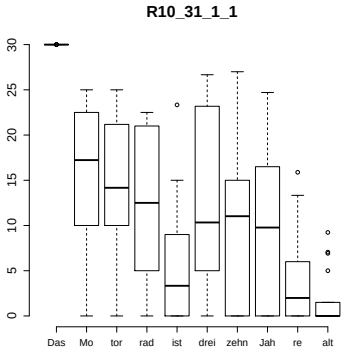

R10_4_2_1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

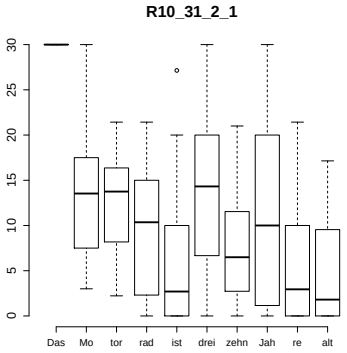

R10_11_0_1
0
2
4
6
8
10
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

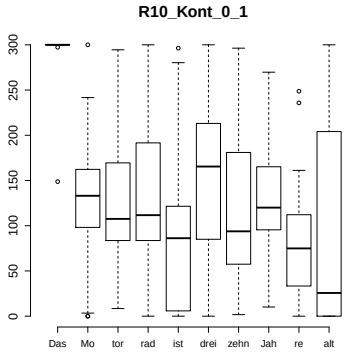

R10_11_1_1
0
2
4
6
8
10
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

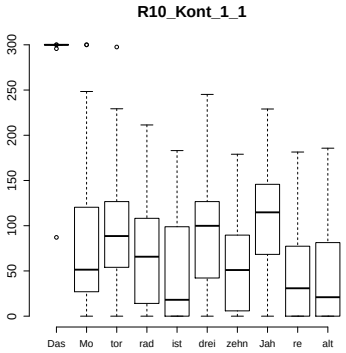

R10_11_2_1
0
2
4
6
8
10
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

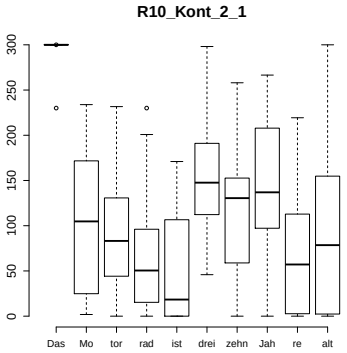

R10_31_0_1
0
5
10
15
20
25
30
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt


R10_31_1_1
0
5
10
15
20
25
30
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt


R10_31_2_1
0
5
10
15
20
25
30
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

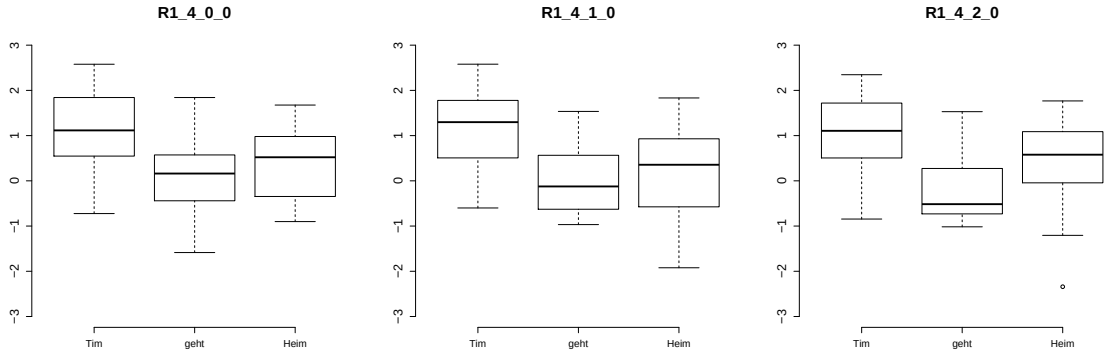

R10_Kont_0_1
0
50
100
150
200
250
300
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

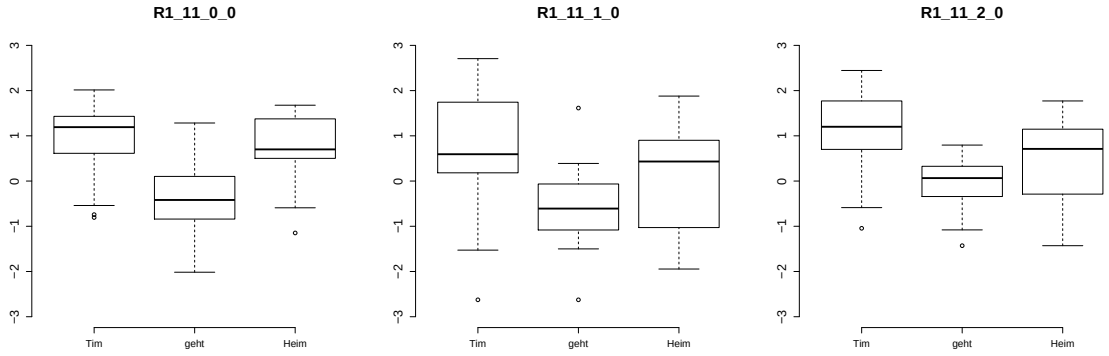

R10_Kont_1_1
0
50
100
150
200
250
300
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

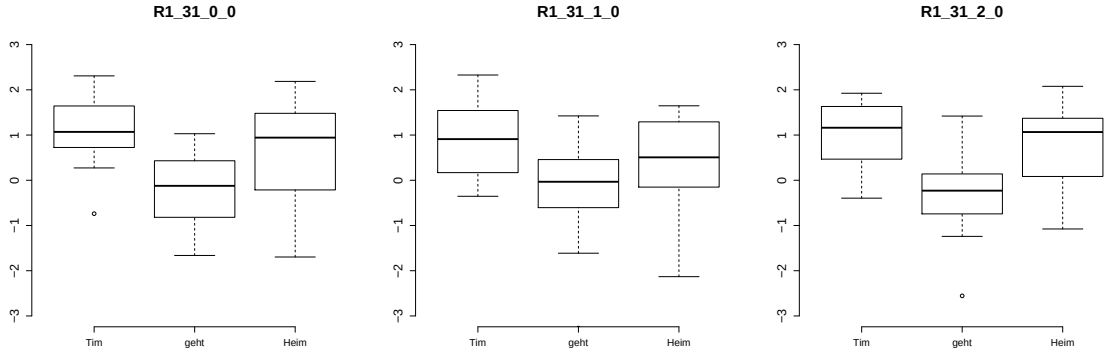

R10_Kont_2_1
0
50
100
150
200
250
300
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

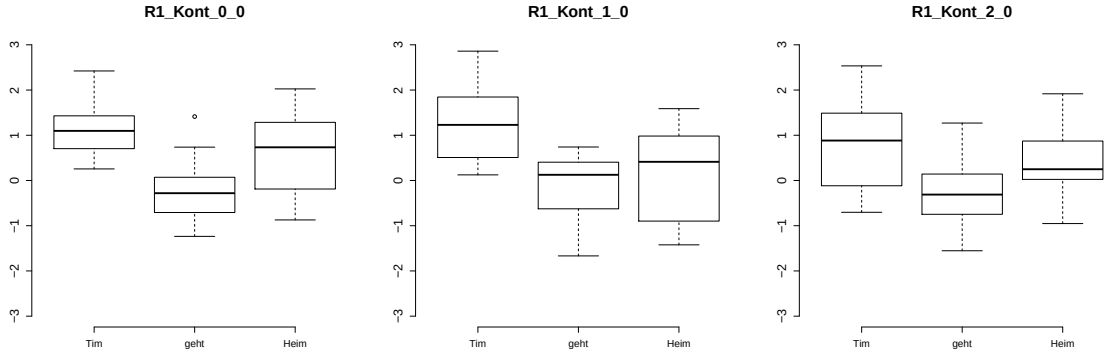
Anhang F. Boxplots Normalisierung z-Transformation

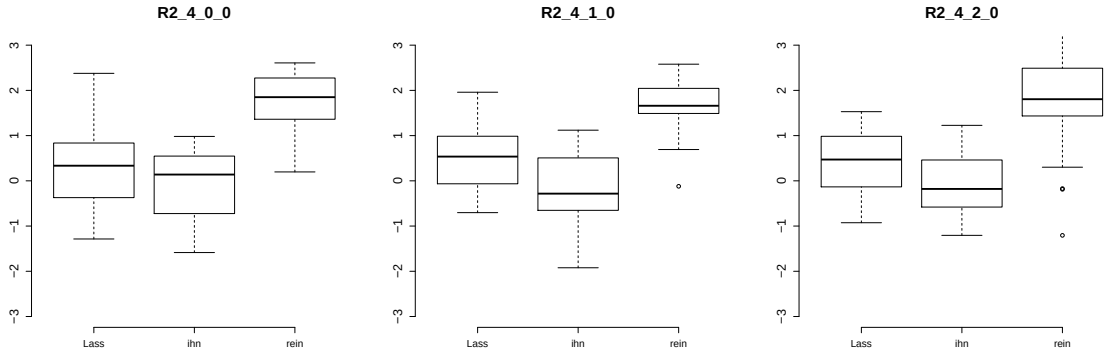
Auf den folgenden Seiten werden die Ratings der Sätze R1 - R10 durch die Probanden des ersten Experiments nach der z-Transformation dargestellt. Hierbei setzt sich der Name der Abbildung wie auch schon in den anderen Anhängen wie folgt zusammen: Satzreferenz_Skala_Akkuratheitsbedingung_Priminggruppe. Die Abbildung Boxplot R10_4_0_0 zeigt also die Bewertungen des Satzes R10 mit der 4-Punkt Skala unter Akkuratheitsbedingung 0 und Priminggruppe 0.

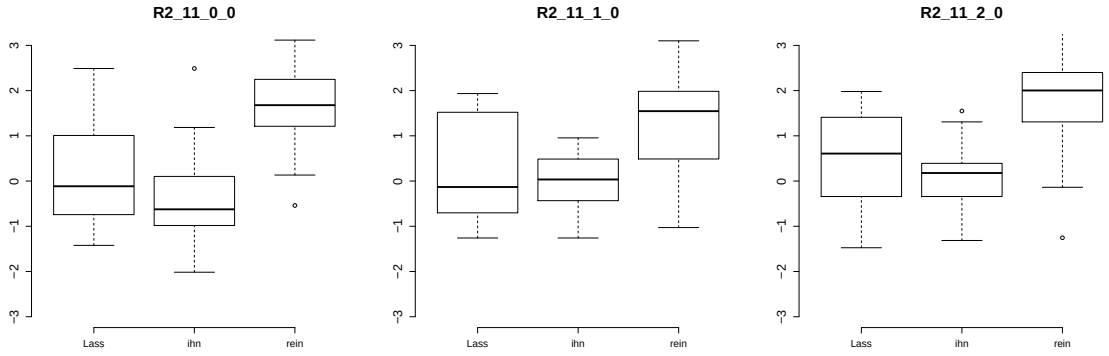

R1_4_0_0
R1_4_1_0
R1_4_2_0
Tim
geht
Heim

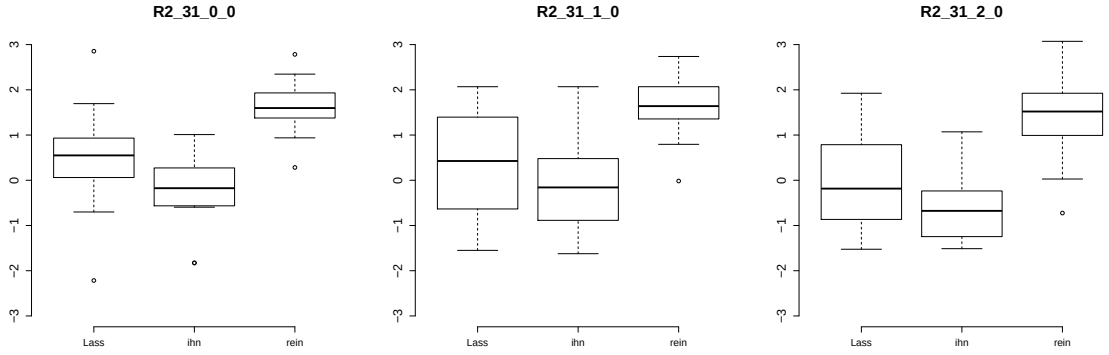

R1_11_0_0
R1_11_1_0
R1_11_2_0
Tim
geht
Heim

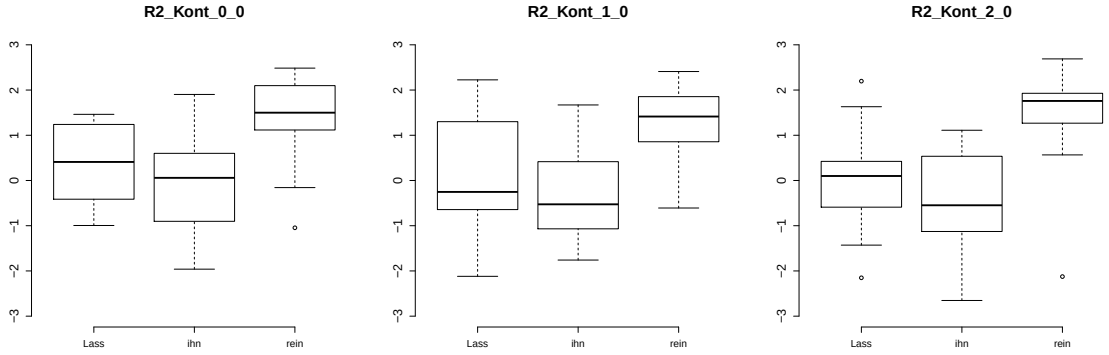

R1_31_0_0
R1_31_1_0
R1_31_2_0
Tim
geht
Heim

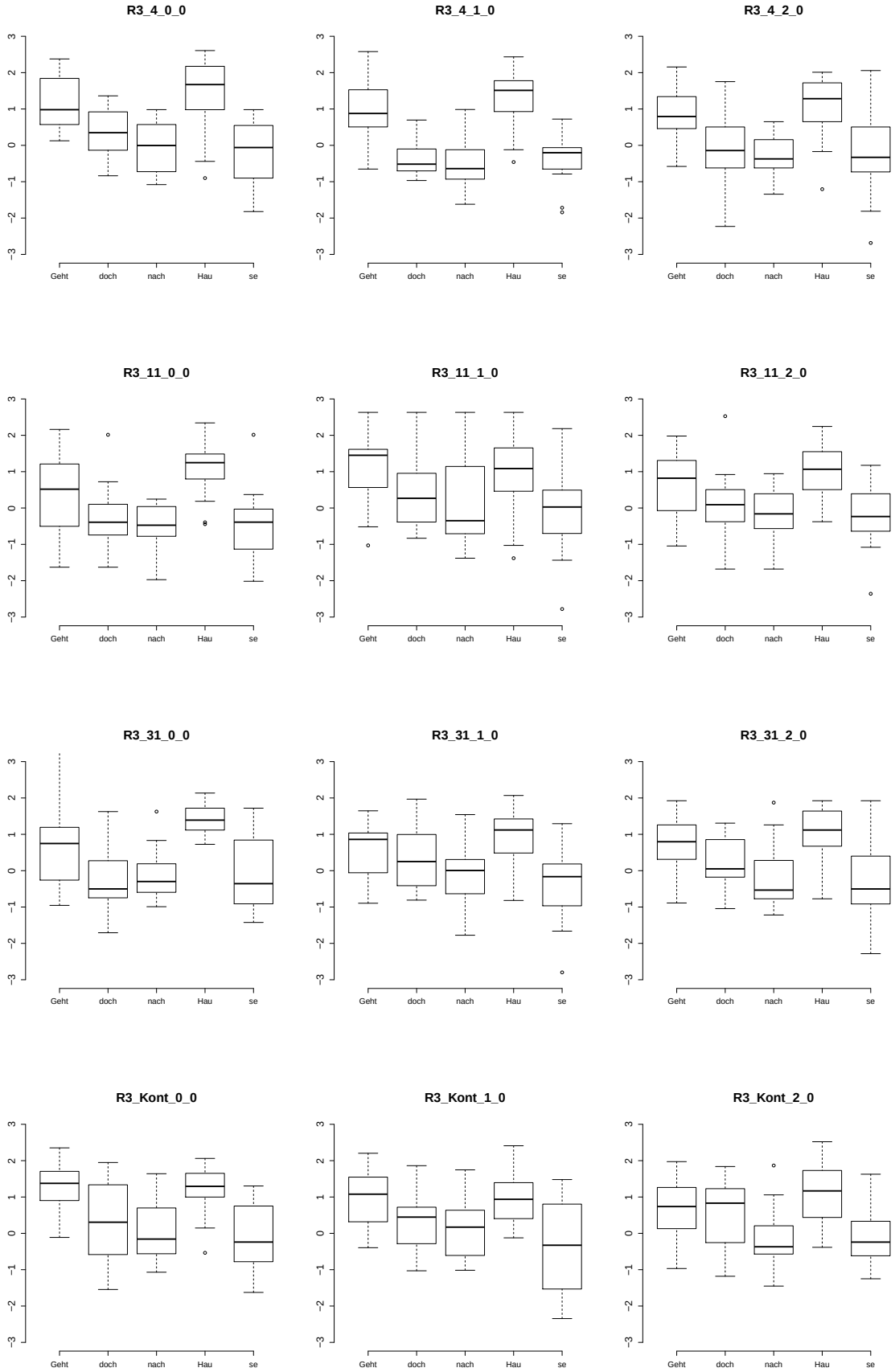

R1_Kont_0_0
R1_Kont_1_0
R1_Kont_2_0
Tim
geht
Heim

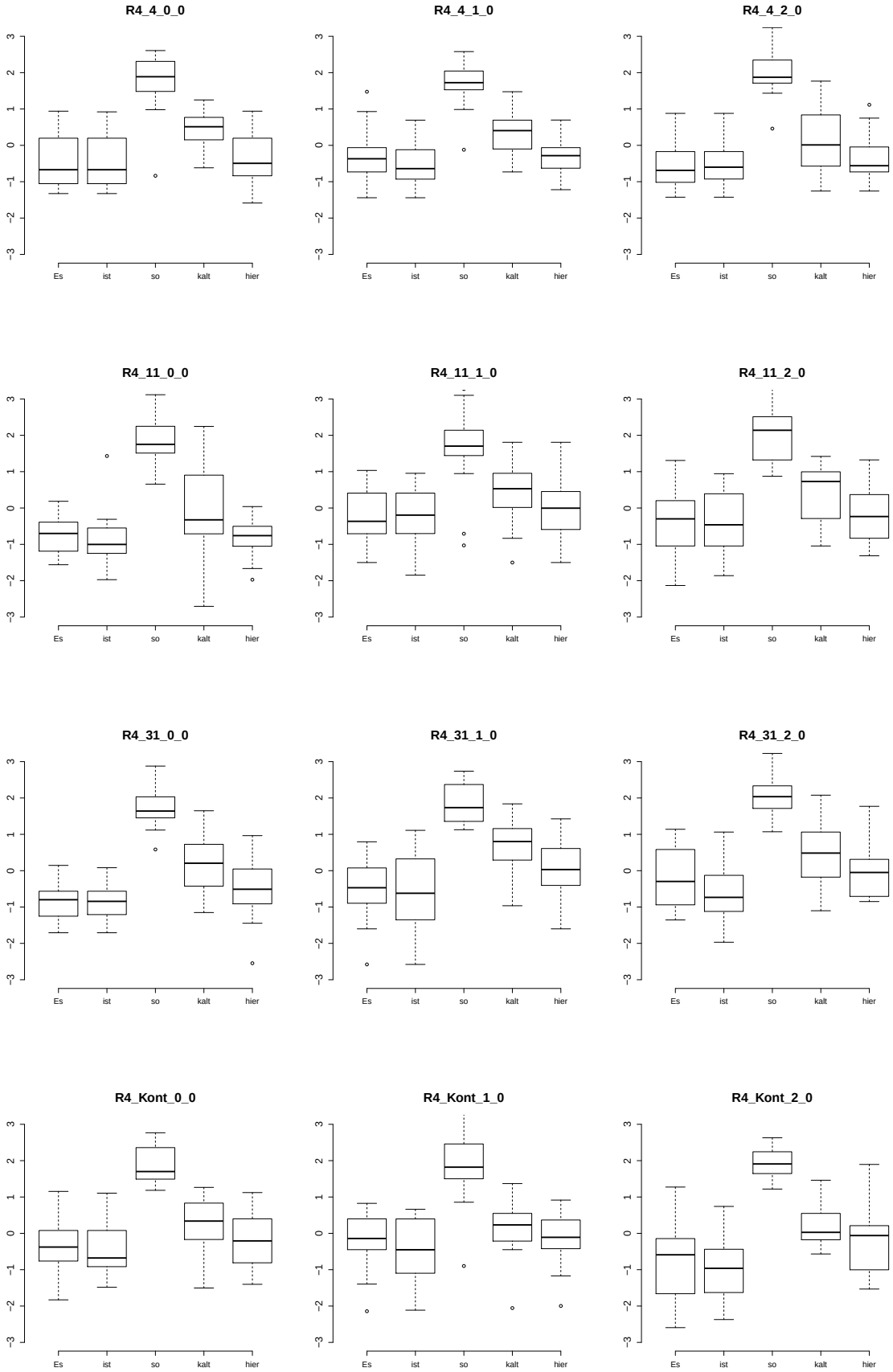

R2_4_0_0
R2_4_1_0
R2_4_2_0
Lass
ihn
rein

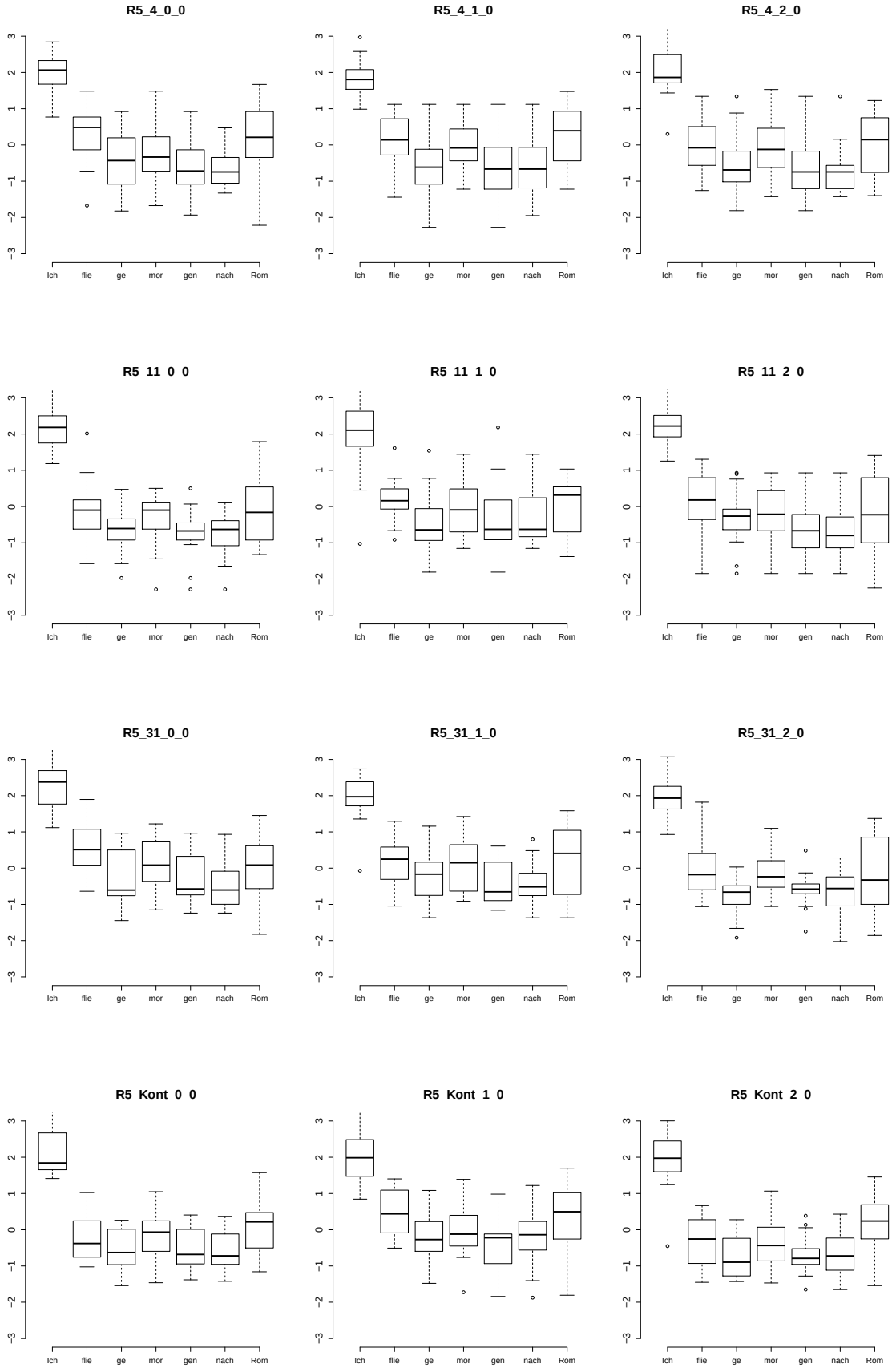

R2_11_0_0
R2_11_1_0
R2_11_2_0
Lass
ihn
rein

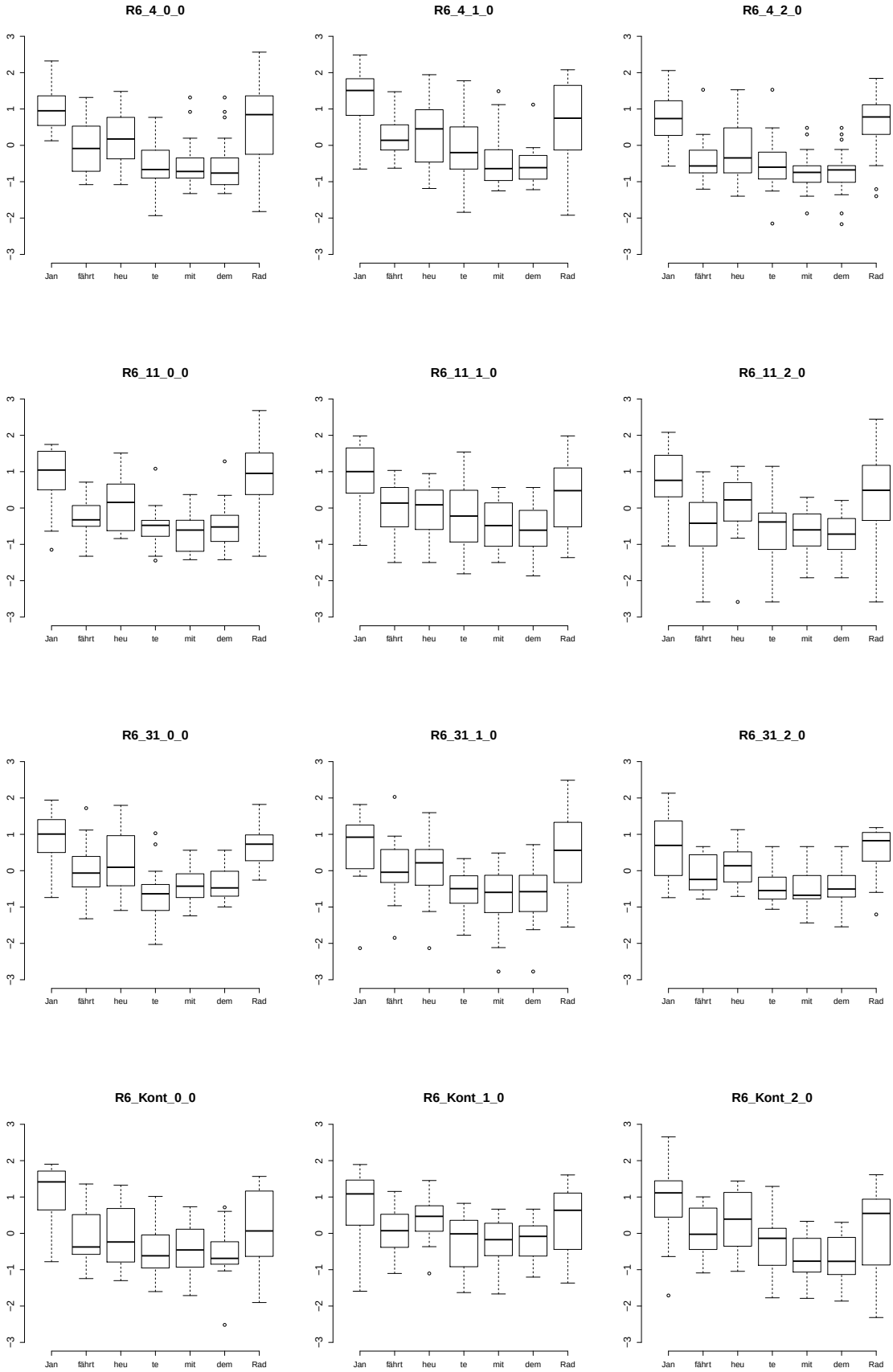

R2_31_0_0
R2_31_1_0
R2_31_2_0
Lass
ihn
rein

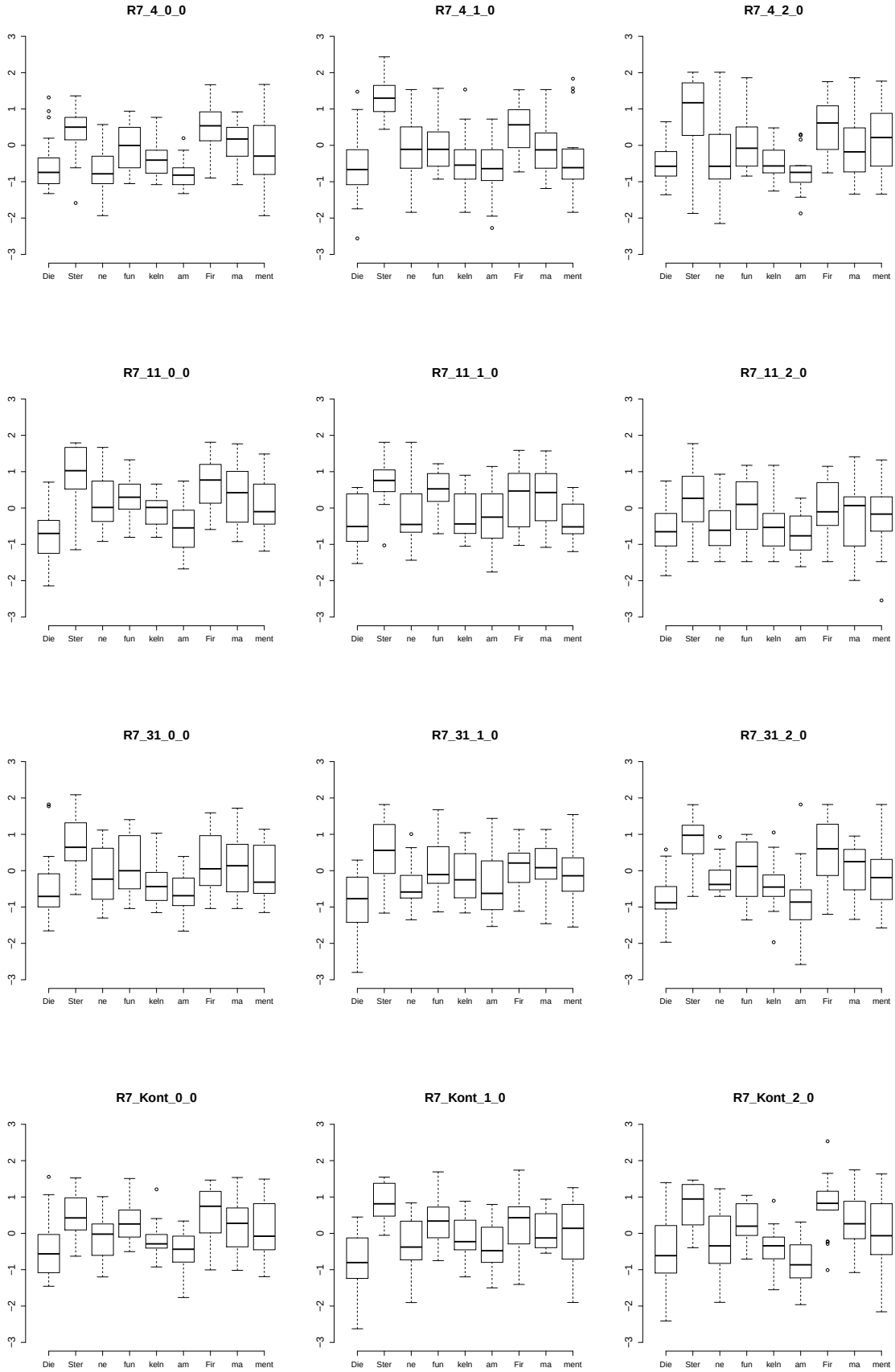

R2_Kont_0_0
R2_Kont_1_0
R2_Kont_2_0
Lass
ihn
rein

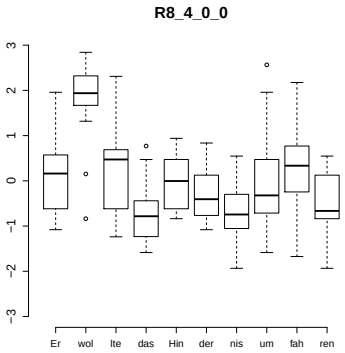

R3_4_0_0
R3_4_1_0
R3_4_2_0
R3_11_0_0
R3_11_1_0
R3_11_2_0
R3_31_0_0
R3_31_1_0
R3_31_2_0
R3_Kont_0_0
R3_Kont_1_0
R3_Kont_2_0
Geht
doch
nach
Hau
se

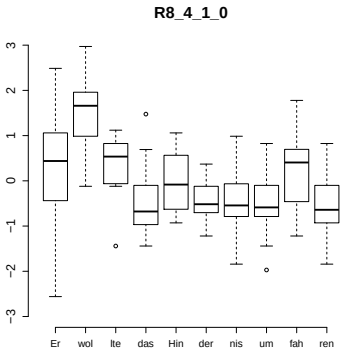

R4_4_0_0
R4_4_1_0
R4_4_2_0
R4_11_0_0
R4_11_1_0
R4_11_2_0
R4_31_0_0
R4_31_1_0
R4_31_2_0
R4_Kont_0_0
R4_Kont_1_0
R4_Kont_2_0
Es
ist
so
kalt
hier

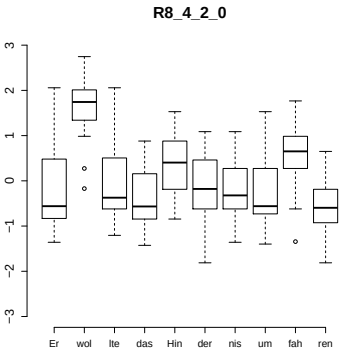

R5_4_0_0
R5_4_1_0
R5_4_2_0
R5_11_0_0
R5_11_1_0
R5_11_2_0
R5_31_0_0
R5_31_1_0
R5_31_2_0
R5_Kont_0_0
R5_Kont_1_0
R5_Kont_2_0
Ich
flie
ge
mor
gen
nach
Rom

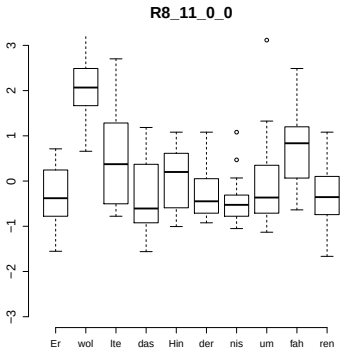

R6_4_0_0
R6_4_1_0
R6_4_2_0
R6_11_0_0
R6_11_1_0
R6_11_2_0
R6_31_0_0
R6_31_1_0
R6_31_2_0
R6_Kont_0_0
R6_Kont_1_0
R6_Kont_2_0
Jan
fährt
heu
te
mit
dem
Rad
−3
−2
−1
0
1
2
3

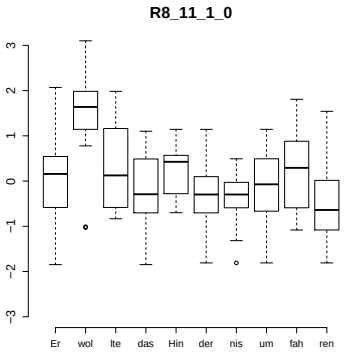

R7_4_0_0
R7_4_1_0
R7_4_2_0
R7_11_0_0
R7_11_1_0
R7_11_2_0
R7_31_0_0
R7_31_1_0
R7_31_2_0
R7_Kont_0_0
R7_Kont_1_0
R7_Kont_2_0
Die Ster ne fun keln am Fir ma ment
−3 −2 −1 0 1 2 3

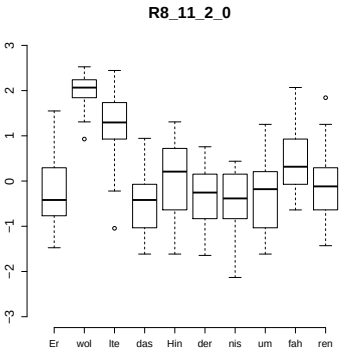

R8_4_0_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

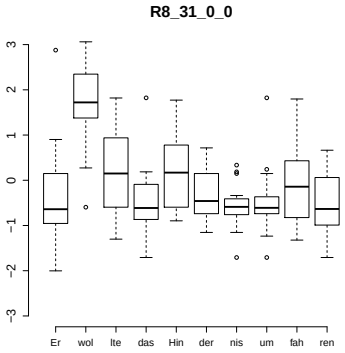

R8_4_1_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

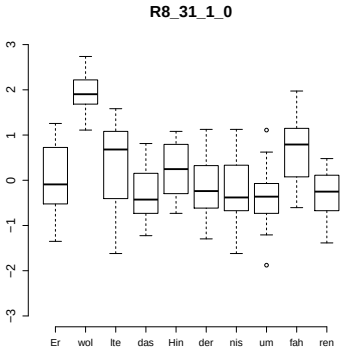

R8_4_2_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

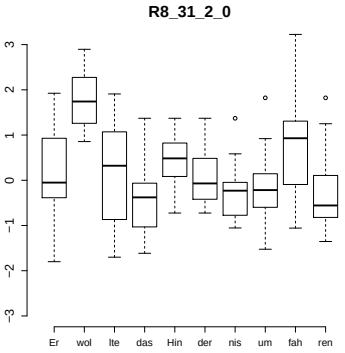

R8_11_0_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

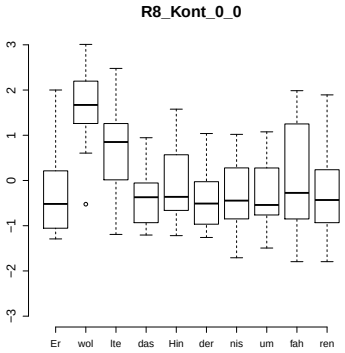

R8_11_1_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

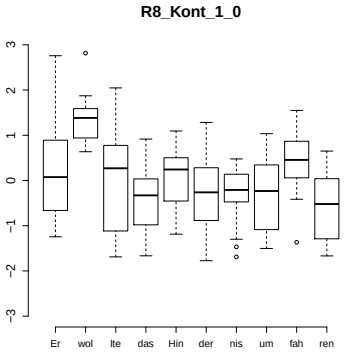

R8_11_2_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

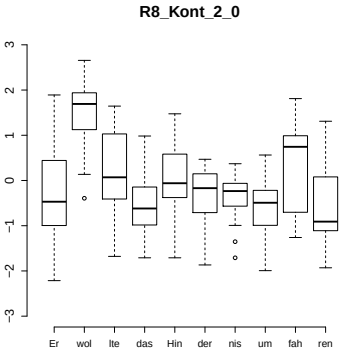

R8_31_0_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

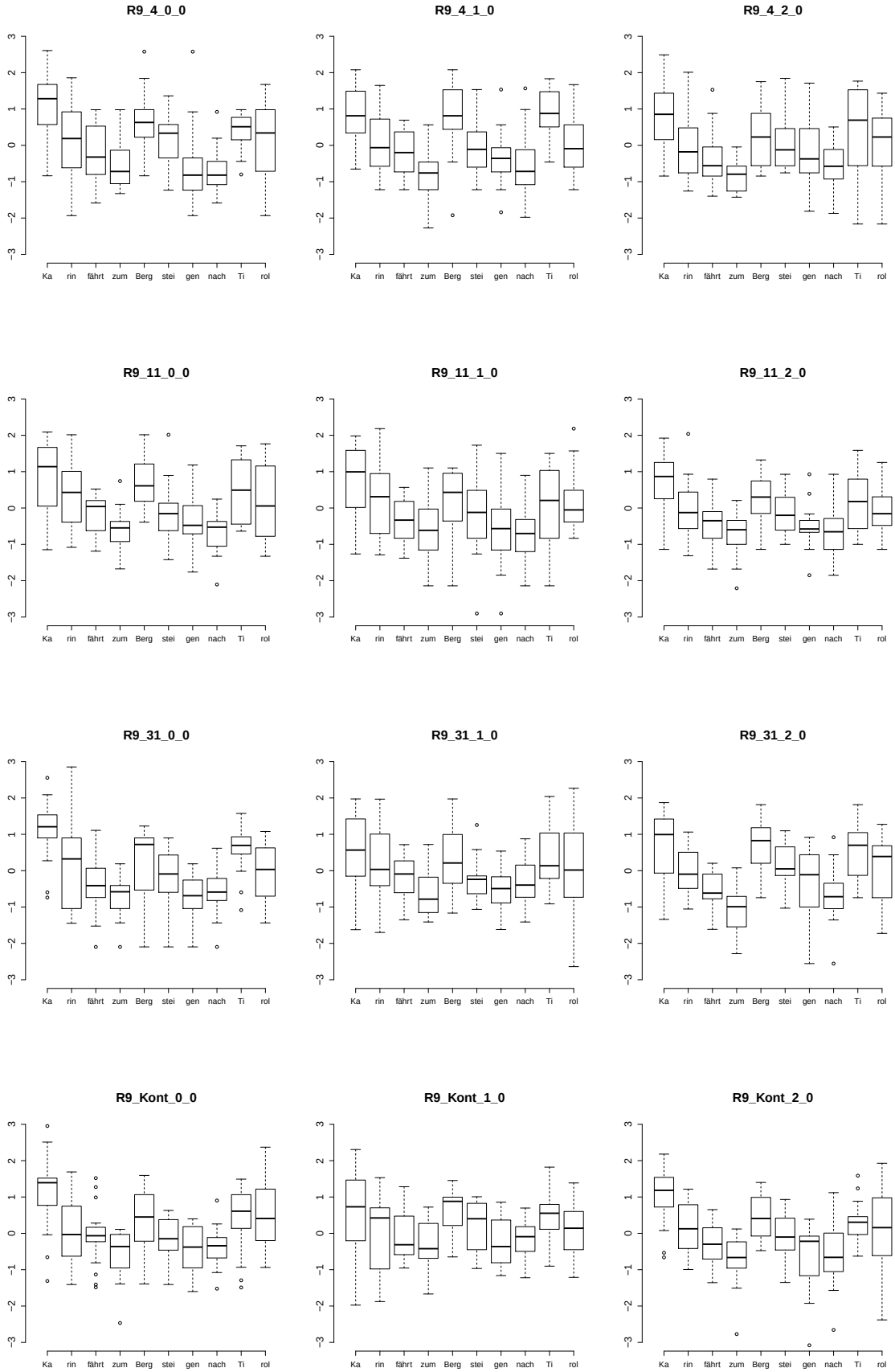

R8_31_1_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

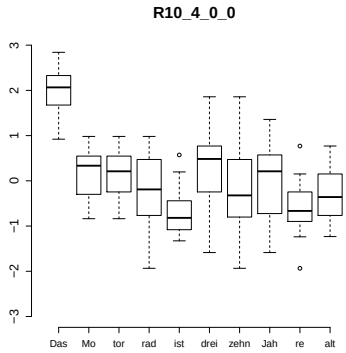

R8_31_2_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

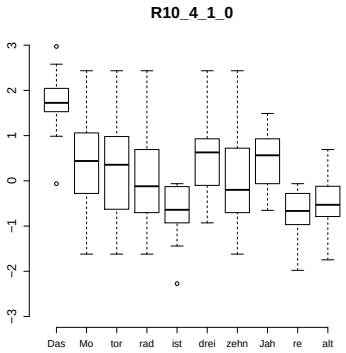

R8_Kont_0_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

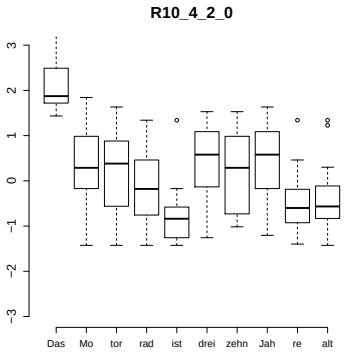

R8_Kont_1_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

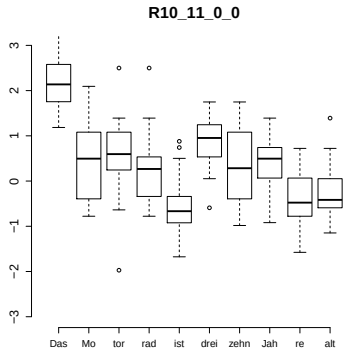

R8_Kont_2_0
3
2
1
0
−1
−2
−3
Er wol lte das Hin der nis um fah ren

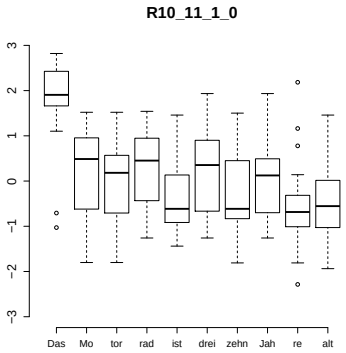

R9_4_0_0
R9_4_1_0
R9_4_2_0
R9_11_0_0
R9_11_1_0
R9_11_2_0
R9_31_0_0
R9_31_1_0
R9_31_2_0
R9_Kont_0_0
R9_Kont_1_0
R9_Kont_2_0
Ka rin fährt zum Berg stei gen nach Ti rol
−3 −2 −1 0 1 2 3

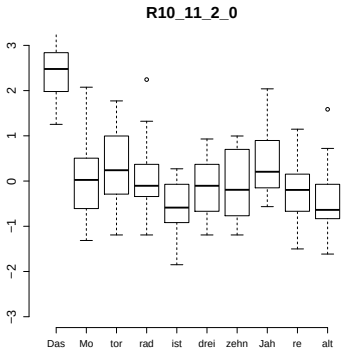

R10_4_0_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

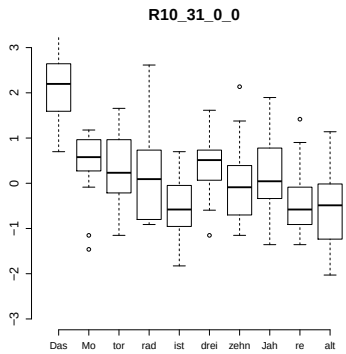

R10_4_1_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

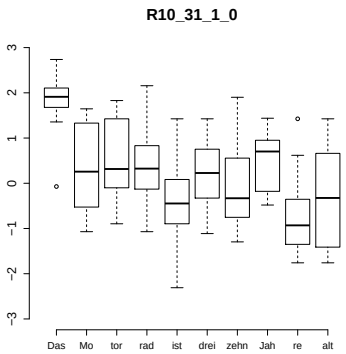

R10_4_2_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

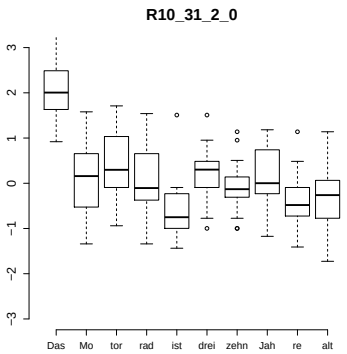

R10_11_0_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

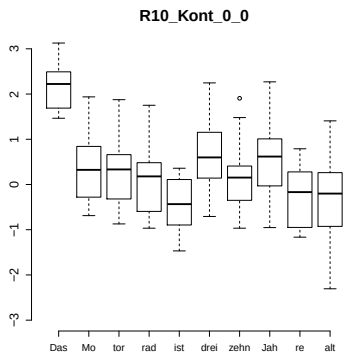

R10_11_1_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

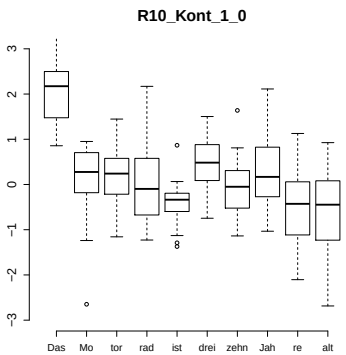

R10_11_2_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

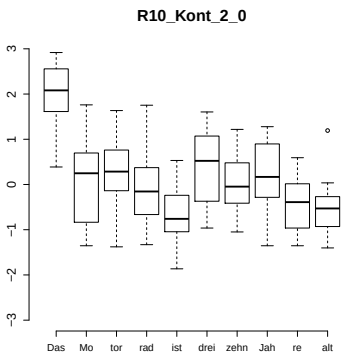

R10_31_0_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

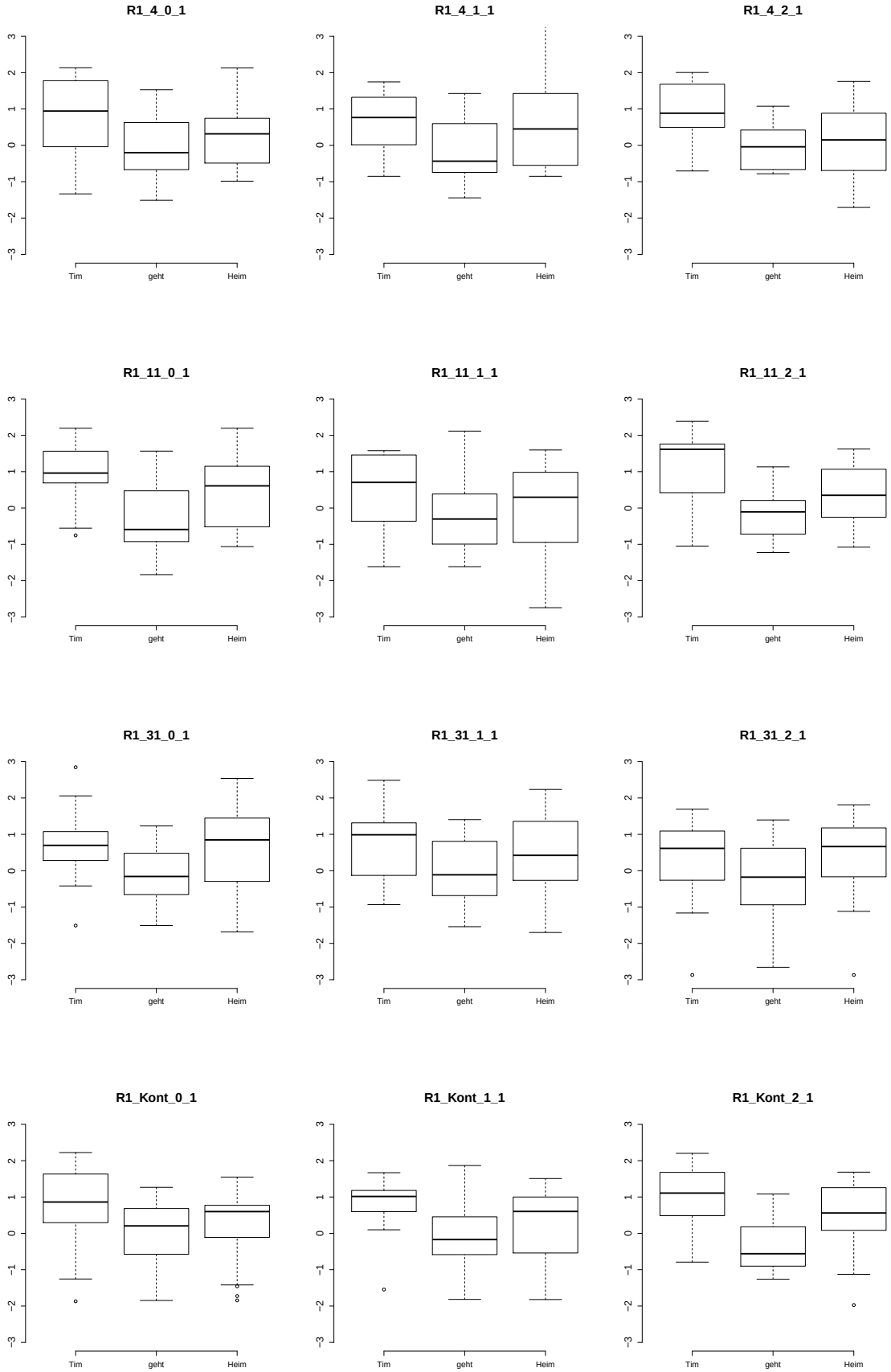

R10_31_1_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

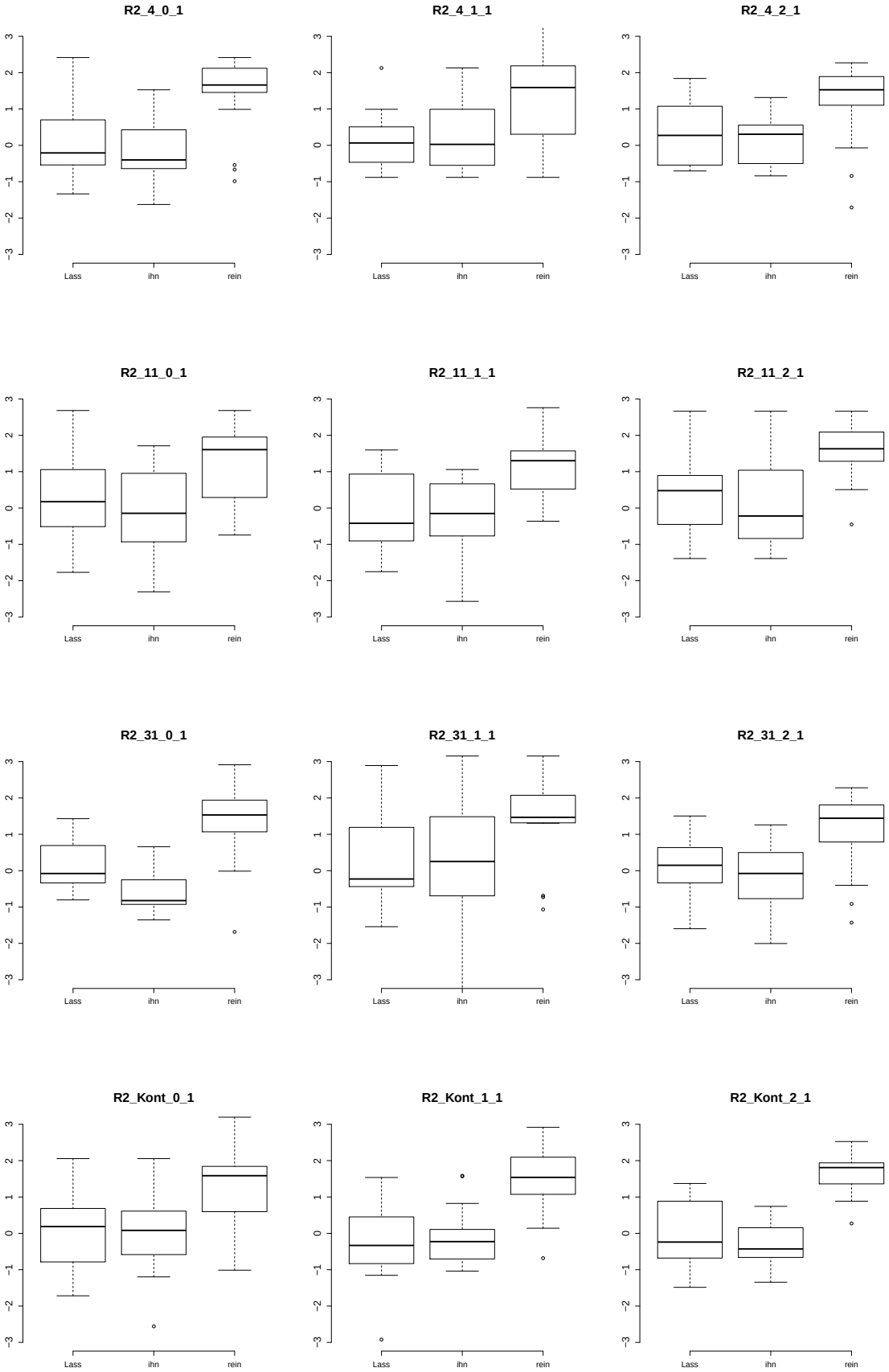

R10_31_2_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

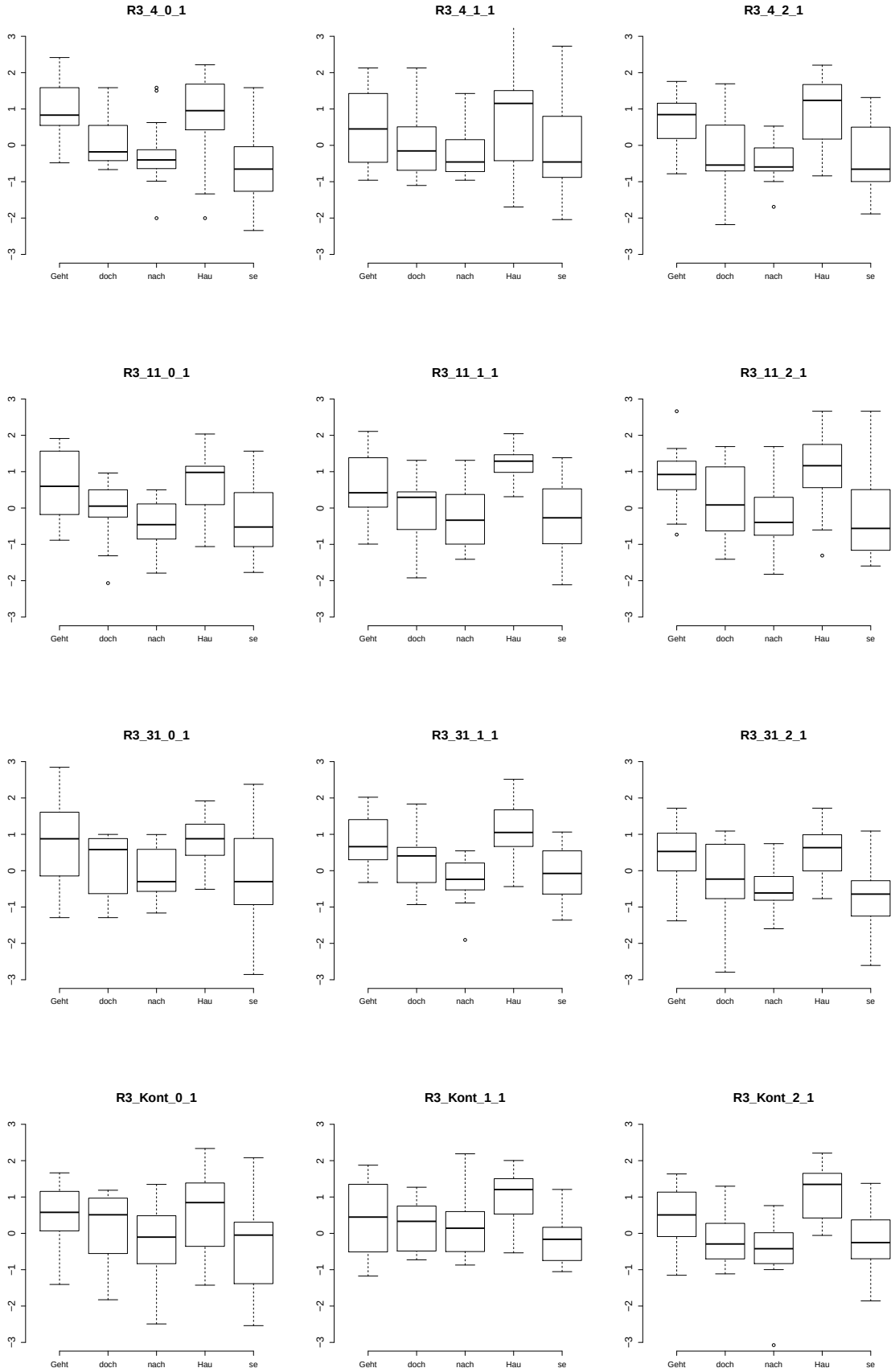

R10_Kont_0_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

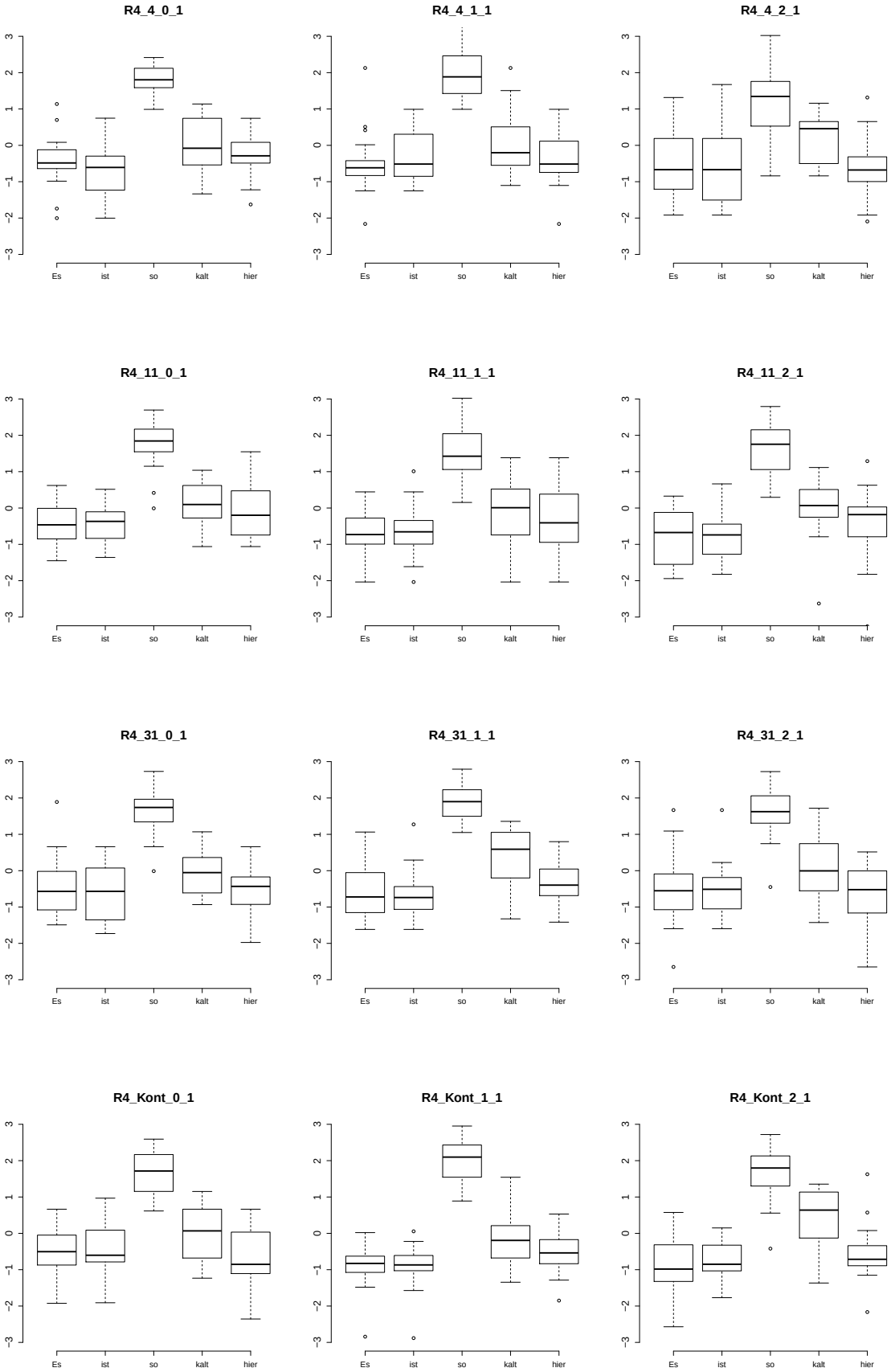

R10_Kont_1_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

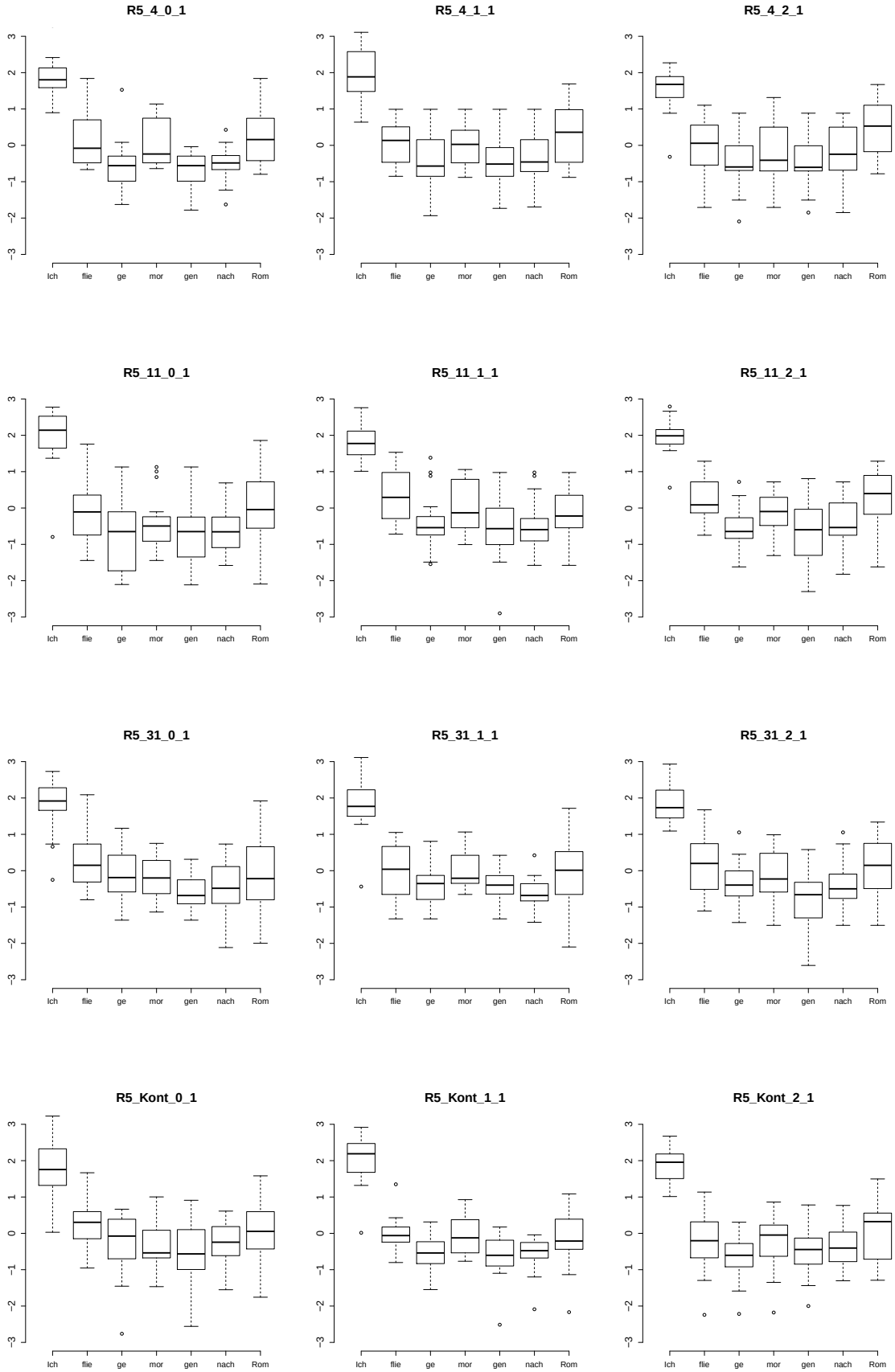

R10_Kont_2_0
3
2
1
0
−1
−2
−3
Das Mo tor rad ist drei zehn Jah re alt

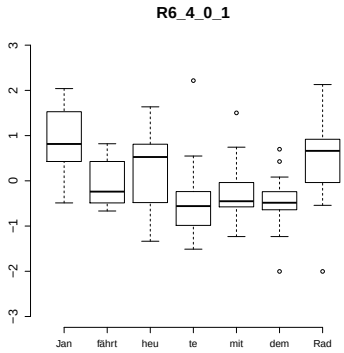

R1_4_0_1
R1_4_1_1
R1_4_2_1
R1_11_0_1
R1_11_1_1
R1_11_2_1
R1_31_0_1
R1_31_1_1
R1_31_2_1
R1_Kont_0_1
R1_Kont_1_1
R1_Kont_2_1
Tim
geht
Heim

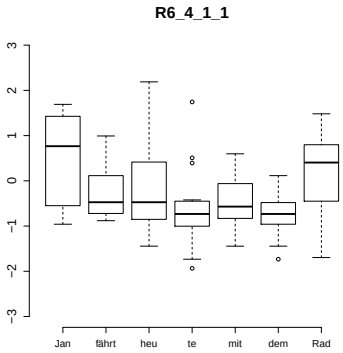

R2_4_0_1
R2_4_1_1
R2_4_2_1
R2_11_0_1
R2_11_1_1
R2_11_2_1
R2_31_0_1
R2_31_1_1
R2_31_2_1
R2_Kont_0_1
R2_Kont_1_1
R2_Kont_2_1
Lass
ihn
rein

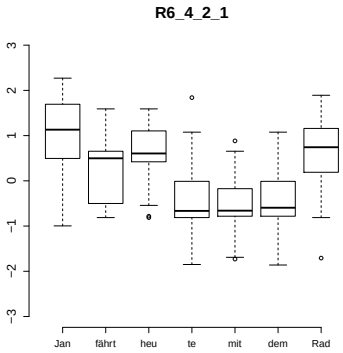

R3_4_0_1
R3_4_1_1
R3_4_2_1
R3_11_0_1
R3_11_1_1
R3_11_2_1
R3_31_0_1
R3_31_1_1
R3_31_2_1
R3_Kont_0_1
R3_Kont_1_1
R3_Kont_2_1
Geht
doch
nach
Hau
se
−3
−2
−1
0
1
2
3

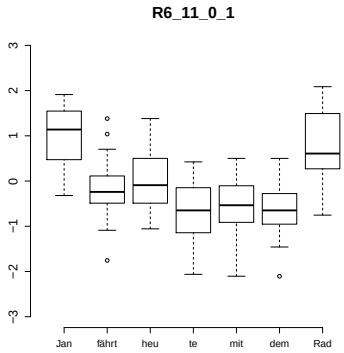

R4_4_0_1
R4_4_1_1
R4_4_2_1
R4_11_0_1
R4_11_1_1
R4_11_2_1
R4_31_0_1
R4_31_1_1
R4_31_2_1
R4_Kont_0_1
R4_Kont_1_1
R4_Kont_2_1
Es
ist
so
kalt
hier
−3
−2
−1
0
1
2
3

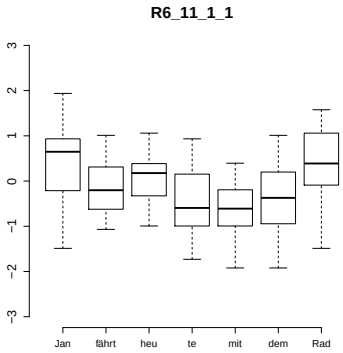

R5_4_0_1
R5_4_1_1
R5_4_2_1
R5_11_0_1
R5_11_1_1
R5_11_2_1
R5_31_0_1
R5_31_1_1
R5_31_2_1
R5_Kont_0_1
R5_Kont_1_1
R5_Kont_2_1
Ich
flie
ge
mor
gen
nach
Rom

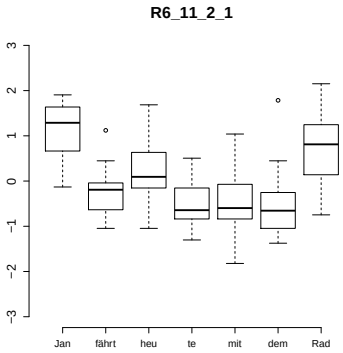

R6_4_0_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

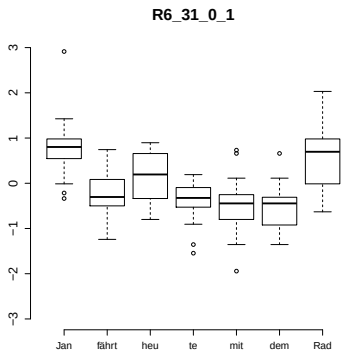

R6_4_1_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

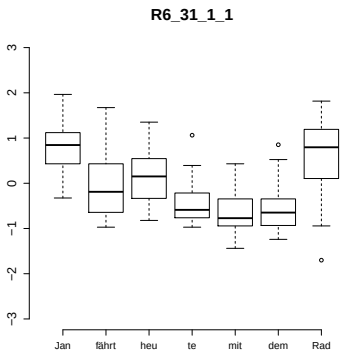

R6_4_2_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

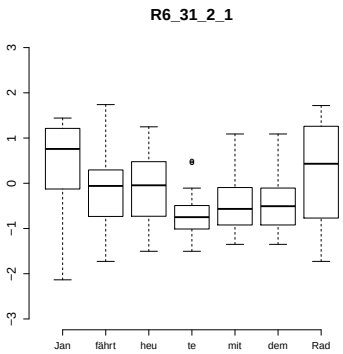

R6_11_0_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

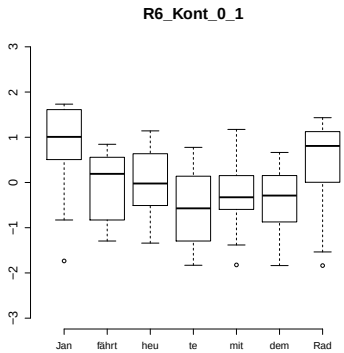

R6_11_1_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

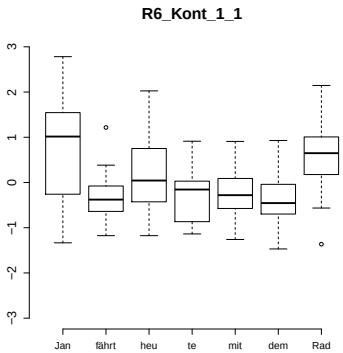

R6_11_2_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

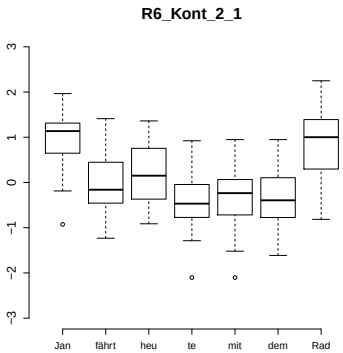

R6_31_0_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

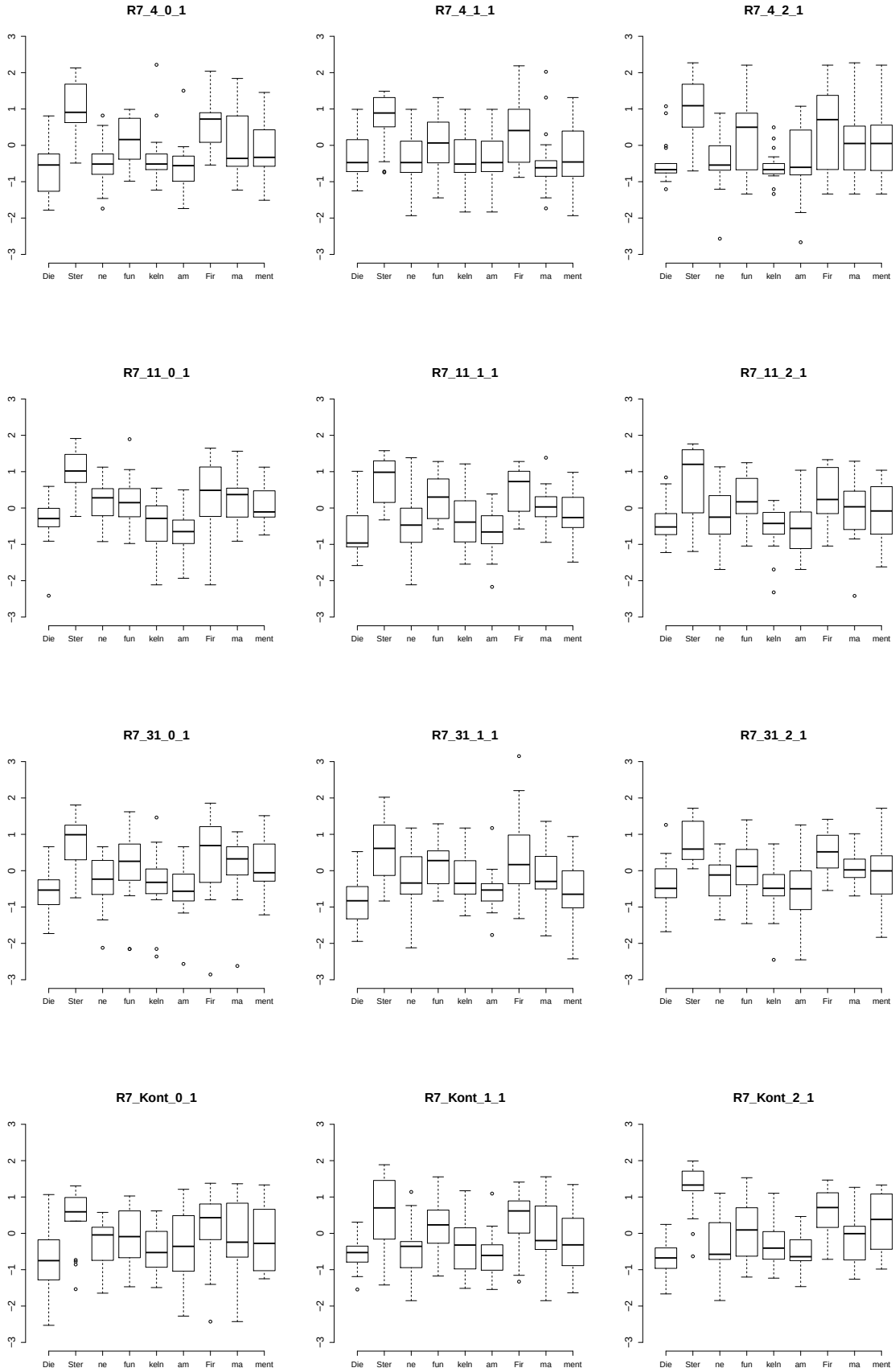

R6_31_1_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

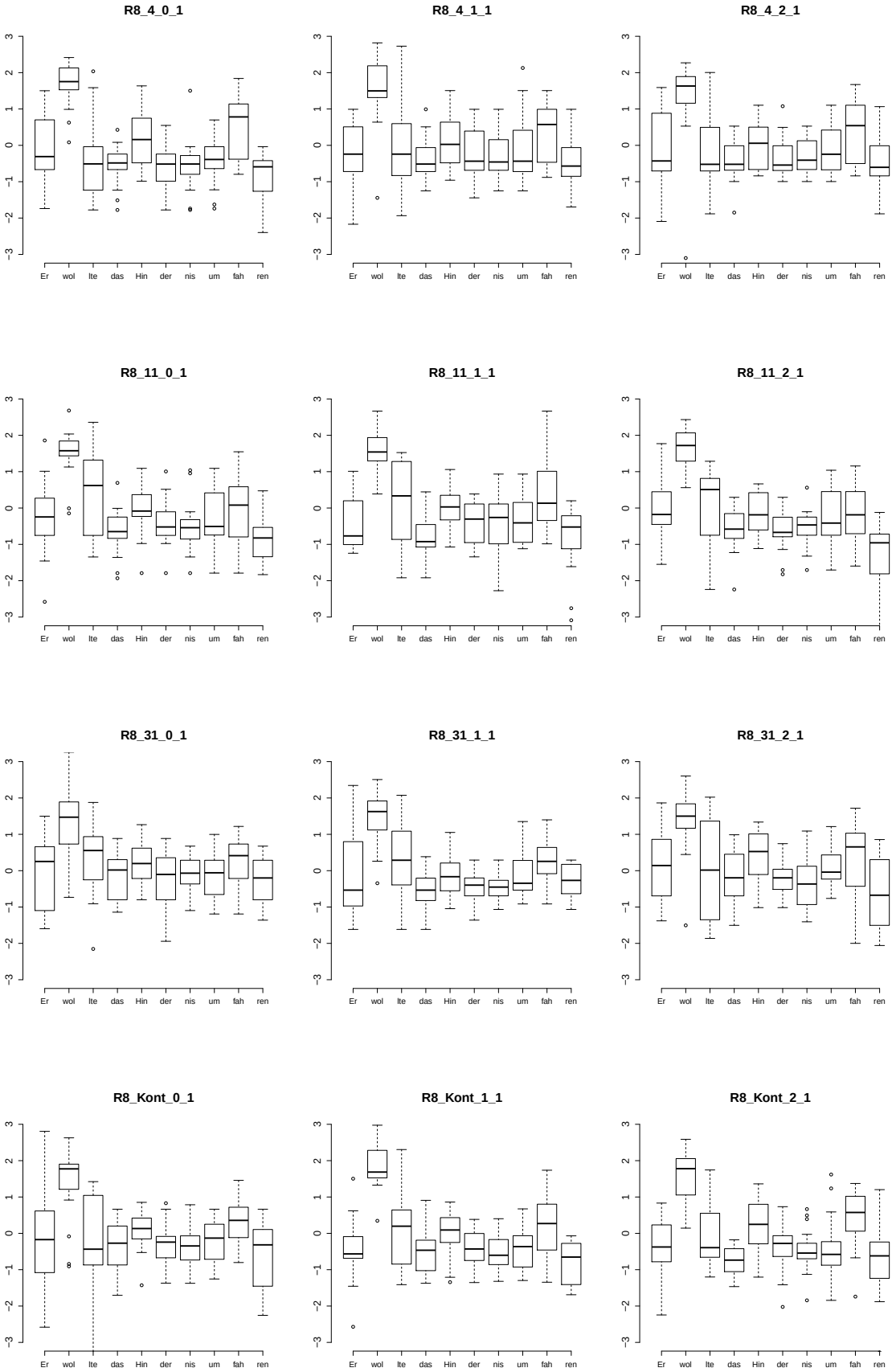

R6_31_2_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

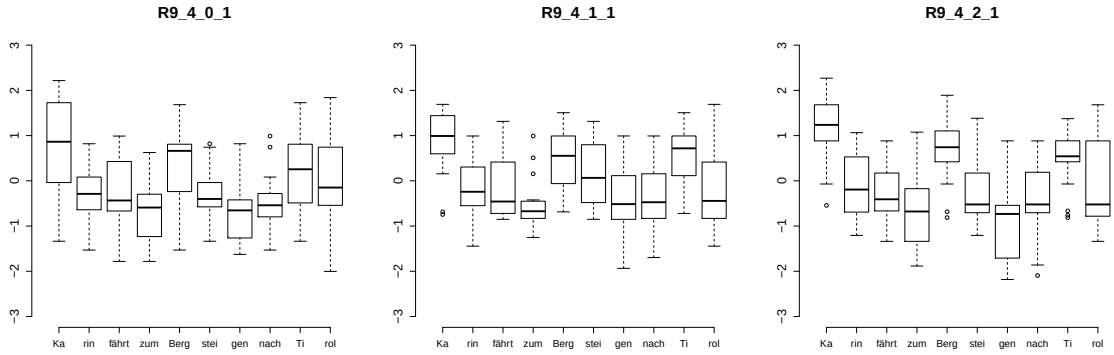

R6_Kont_0_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

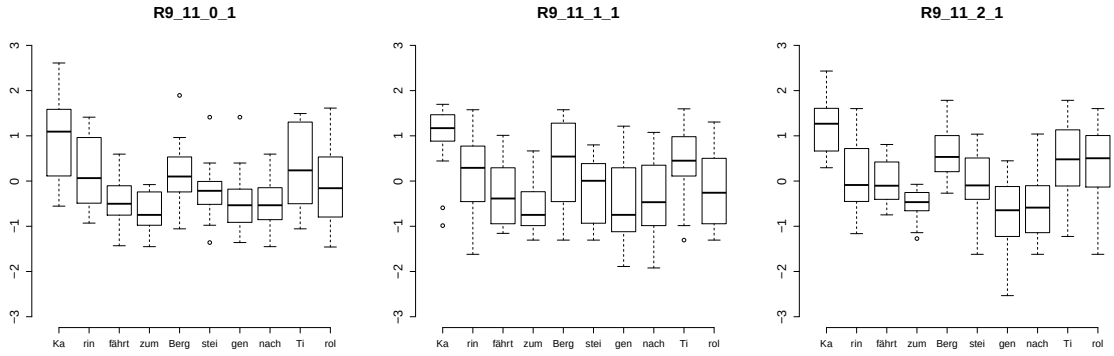

R6_Kont_1_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

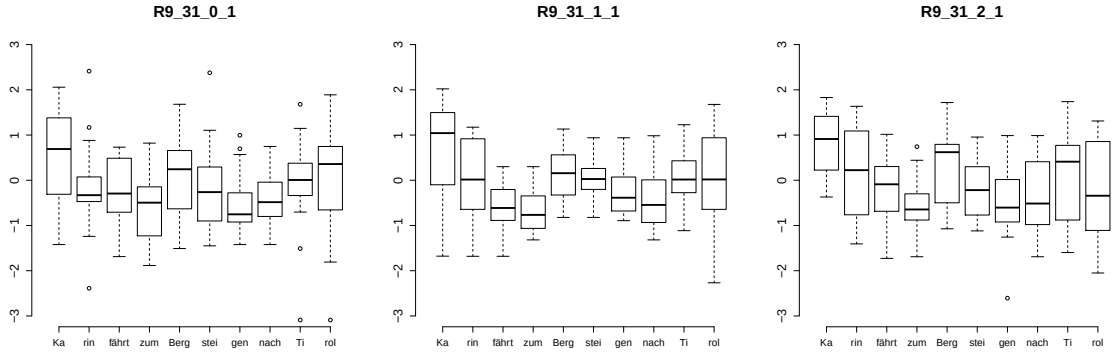

R6_Kont_2_1
3
2
1
0
−1
−2
−3
Jan
fährt
heu
te
mit
dem
Rad

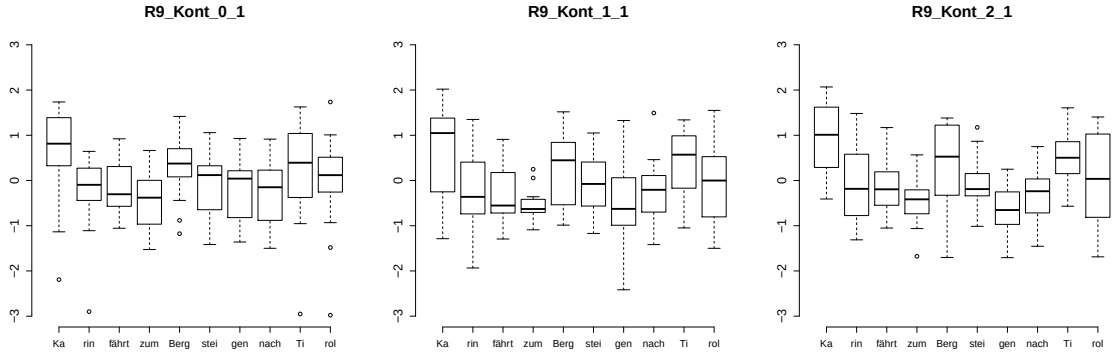

R7_4_0_1
R7_4_1_1
R7_4_2_1
R7_11_0_1
R7_11_1_1
R7_11_2_1
R7_31_0_1
R7_31_1_1
R7_31_2_1
R7_Kont_0_1
R7_Kont_1_1
R7_Kont_2_1
Die Ster ne fun keln am Fir ma ment
−3 −2 −1 0 1 2 3

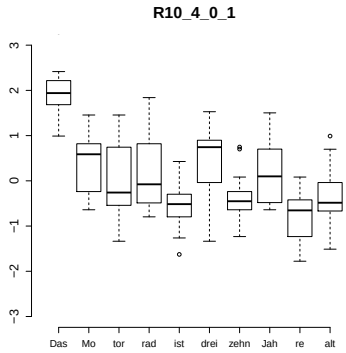

R8_4_0_1
R8_4_1_1
R8_4_2_1
R8_11_0_1
R8_11_1_1
R8_11_2_1
R8_31_0_1
R8_31_1_1
R8_31_2_1
R8_Kont_0_1
R8_Kont_1_1
R8_Kont_2_1
Er wol lte das Hin der nis um fah ren
−3 −2 −1 0 1 2 3

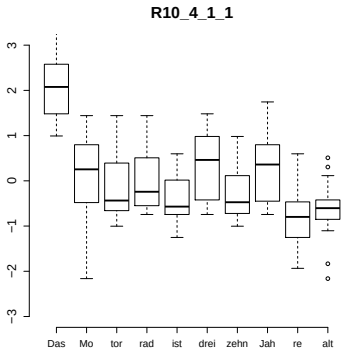

R9_4_0_1
R9_4_1_1
R9_4_2_1
Ka rin fährt zum Berg stei gen nach Ti rol

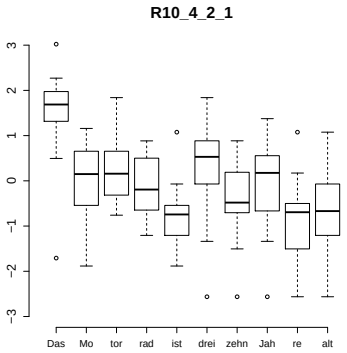

R9_11_0_1
R9_11_1_1
R9_11_2_1
Ka rin fährt zum Berg stei gen nach Ti rol

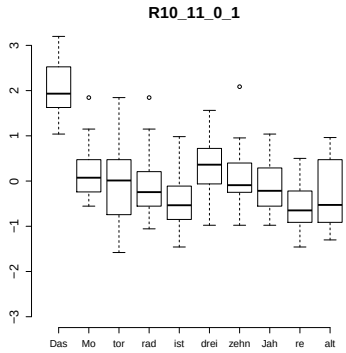

R9_31_0_1
R9_31_1_1
R9_31_2_1
Ka rin fährt zum Berg stei gen nach Ti rol

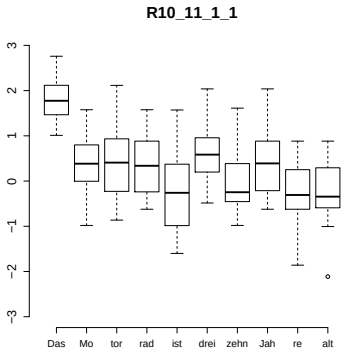

R9_Kont_0_1
R9_Kont_1_1
R9_Kont_2_1
Ka rin fährt zum Berg stei gen nach Ti rol

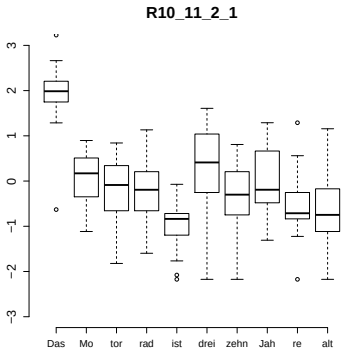

R10_4_0_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

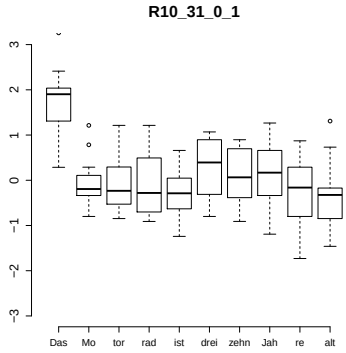

R10_4_1_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

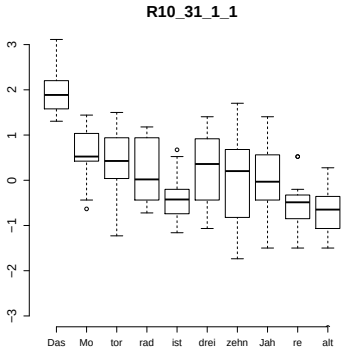

R10_4_2_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

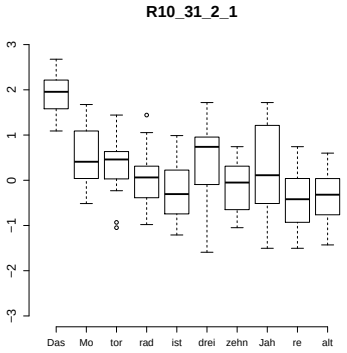

R10_11_0_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

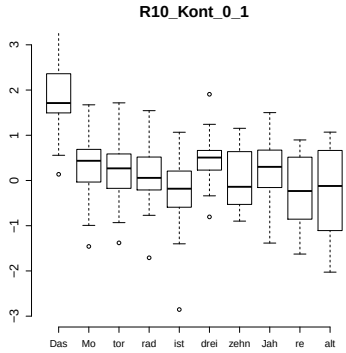

R10_11_1_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

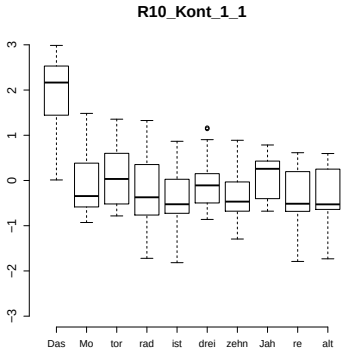

R10_11_2_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt

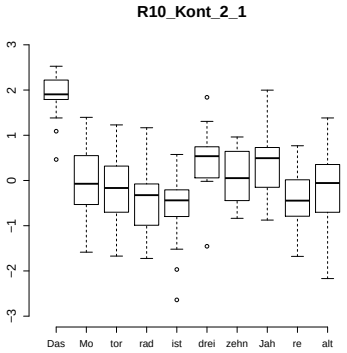

R10_31_0_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt


R10_31_1_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt


R10_31_2_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt


R10_Kont_0_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt


R10_Kont_1_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt


R10_Kont_2_1
3
2
1
0
−1
−2
−3
Das
Mo
tor
rad
ist
drei
zehn
Jah
re
alt