universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Analyse von Konzernabschlüssen mithilfe von Lesbarkeitsindizes“

verfasst von / submitted by
Maximilian Wallisch, Bsc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 915
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Betriebswirtschaft
Betreut von / Supervisor:	PD Dr. Karina Sopp

Abstract:

In der folgenden Arbeit wird der Zusammenhang zwischen der Komplexität der Sprache in Konzernabschlüssen der letzten fünf Jahre und der wirtschaftlichen Leistung der Unternehmen überprüft. Die Komplexität der Sprache wird anhand von drei Lesbarkeitsindizes, FOG-, SMOG-Index und Flesch Reading Ease Formula, abgebildet. Die wirtschaftliche Leistung jedes Unternehmens in der beobachteten Periode wird mit den Earnings per Share (kurz: EPS) quantifiziert. Anhand von deskriptiven Analysen wird der lineare Zusammenhang der Variablen dargestellt und in der Folge mithilfe von einem linearen Regressionsmodell manifestiert. Die statistische Analyse hat eine signifikante positive Korrelation für Lesbarkeitsindizes FOG und SMOG und den EPS der Unternehmen bestätigt. Der Flesch-Ease Lesbarkeitsindex weist eine negative Korrelation mit den EPS der Unternehmen auf. Das Resultat der Analyse zeigt auf, dass mit ansteigender EPS eines Unternehemens auch die Komplexität der Konzernabschlüsse zunimmt.

Inhalt

1. Motivation

Ziel der Arbeit ist die Untersuchung vom Zusammenhang der Earnings per Share eines Unternehmens und der Komplexität der Sprache eines Unternehmens. Für die Analyse werden die Konzernabschlüsse der Unternehmen des prime markets der Wiener Börse gewählt. Die Quantifizierung der Komplexität der Sprache wird mittels Lesbarkeitsindizes dargestellt. Mithilfe von statistischen Methoden wird der Zusammenhang zwischen Earnings per Share und der Ergebnisse der Lesbarkeitsindizes bestimmt. Im Anschluss werden zukünftige Anwendungsgebiete aufgezählt und detailliert beschrieben. Am Ende der Arbeit werden die Resultate analysiert und graphisch dargestellt.

2. Lesbarkeitsindizes

2.1. Definition von Lesbarkeitsindizes

Lesbarkeitsindizes sind Indikatoren, welche die Komplexität eines Textes quantifizieren und dadurch Texte vergleichbar machen. Um einen Index zu berechnen, werden meist die Länge der Sätze, die Anzahl der Nebensätze oder etwa die Vielfältigkeit des Vokabulars in Betracht gezogen. Zur Quantifizierung der Vielfältigkeit des Vokabulars durch ein Computerprogramm, werden Wörterlisten verwendet, die ähnliche Wörter identifizieren. (Duffy T. M., 1985)

Ein Beispiel dafür ist die Dale-Chall Liste in der englischen Sprache. Diese Liste beinhaltet 3000 Wörter, von denen mindestens 80 Prozent der Schüler der 5. Schulstufe die Bedeutung kennen (Sprache, 1953). Hierbei handelt es sich nicht um die 3000 im Alltagsgebrauch am häufigsten verwendeten Wörter. Im Gegenteil eher unwichtig erscheinende Wörter wurden inkludiert und wichtige wiederum ausgeschlossen. Die Wörter wurden so gewählt, dass wenn das Auftreten der Wörter in einem beliebigen Text verglichen wird, so zeigt das Ergebnis eine signifikante Korrelation mit der Lesbarkeit des Textes.

Wenn all diese Informationen in die jeweiligen Formeln zur Berechnung des Lesbarkeitsindizes eingesetzt werden, ergeben diese einen Wert, der die Lesbarkeit kennzeichnet. Jener Wert, der aus der Formel resultiert, wird in eine Tabelle eingesetzt und anhand dieser wird die Lesbarkeit des Textes festgestellt. Die Einteilung wie gut lesbar ein Text ist, wird nach Schulstufen klassifiziert. Je höher der errechnete Wert ist, desto schwieriger ist der Text zu lesen und zu verstehen. In der Folge ist eine höhere Schulbildung erforderlich, um den Text zu verstehen.

Lesbarkeitsindizes erlangten zum ersten Mal nach Pressman (1979) in der amerikanischen Politik bedeutende Aufmerksamkeit. Der damalige Präsident Carter hatte einen Beschluss verabschiedet, der besagt, dass Formulare und Verträge, welche vom Staat veröffentlicht werden im sogenannten „plain english“ verfasst werden sollen (Pressman, 1979). Zusätzlich wurde dem Begriff „plain english“

Aufmerksamkeit geschenkt, da nach Duffy T. M. (1985) ein Gerichtsbeschluss erlassen wurde, welcher Hersteller dazu veranlasste, Produktbeschreibungen möglichst simpel zu formulieren, sodass die Produkte selbst sowie ihre Funktionsweisen für die gesamte Bevölkerung einfach zu verstehen sind. Diese Bewegung fand in der amerikanischen Bevölkerung hohen Anklang, da es sehr serviceorientiert sei, wenn etwaige Beschreibungen einfach verständlich verfasst sind.

Nun stellte sich jedoch die Frage, was unter dem Begriff „plain english“ zu verstehen ist und wie entschieden wird, ob ein Text einfach verständlich ist oder nicht. Da es zu dieser Zeit schon Lesbarkeitsindizes gab, die es möglich machten, die Verschachtelung von Texten zu bestimmen, wurde diese Methodik gewählt, um zu entscheiden, ob ein Text im sogenannten „plain english“ verfasst ist. Der große Vorteil gegenüber einer Bewertung von Menschen ohne Verwendung von Lesbarkeitsindizes, sei es nun ein Experte oder ein Richter, der entscheidet, ob ein Text nun einfach oder schwierig zu lesen sei, ist, dass der Beschluss frei von menschlichen Einflüssen ist und das Ergebnis quantifiziert werden kann. Der Ausschluss von der menschlichen Komponente und der Einführung von objektiven Bewertungsregel, wie in Lesbarkeitsindizes, sind entscheidende Vorteile, da dadurch ein Vergleich von mehreren Texten nach ihrer Lesbarkeit möglich ist. Die Lesbarkeit wird mit diesem Hilfsmittel in Form von Zahlen greifbar gemacht.

Die schlussendliche Entscheidung für Lesbarkeitsindizes und gegen eine menschliche Überprüfung bei der Bewertung von Formularen und Verträgen, welche vom Staat veröffentlicht werden, ist, dass bei der Berechnung von Formeln keine menschliche Komponente enthalten ist (Duffy T. M., 1985). Dies führt dazu, dass subjektive Empfindungen von einem Entscheidungsträger, sei er nun Experte oder Richter, nicht in die Bewertung einfließen. Aus diesem Grund kann das Bewertungssystem auch in vertraglichen Angelegenheiten verwendet werden, wie zum Beispiel zur Bewertung, ob eine Produktbeschreibung verständlich beschrieben ist.

Der zweite Grund, der für die Verwendung von Lesbarkeitsindizes spricht, ist, dass Lesbarkeitsindizes die Komplexität der Sätze quantifizierbar machen, somit ist ein direkter Vergleich der Komplexität von einer beliebiger Anzahl an Texten mithilfe des Index möglich. Folglich kann bei der Gegenüberstellung von verschiedenen Texten exakt festgestellt werden, welcher der beiden komplexer ist. Es macht keinen Unterschied, wer die Überprüfung durchführt, da das Ergebnis – die Ausprägung des Indizes– immer gleich ist. Voraussetzung ist nur, dass die richtigen Parameter in die Formel eingesetzt werden und die Berechnung richtig durchgeführt wird. Aus diesem Grund kann festgehalten werden, dass man auch bei mehrmaliger Berechnung des Lesbarkeitsindizes, immer exakt dasselbe Ergebnis erhält. Das Ergebnis sowie die Vorgangsweise sind standardisiert, das bedeutet, jede Person auf der Welt kann den Lesbarkeitsindex berechnen und wird bei richtiger Ausführung dasselbe Ergebnis erhalten.

Ein Kritikpunkt an Lesbarkeitsindizes ist genau diese strikte immergleiche Vorgehensweise, sie ist nicht dynamisch. In gewissen Fällen kann genau diese subjektive, aber dafür dynamische Herangehensweise eines Menschen, welche prinzipiell eine Schwachstelle der menschlichen Überprüfung darstellt, den ausschlaggebenden Unterschied machen. Weiters ist anzumerken, dass die Lesbarkeitsindizes menschlichen Ursprung haben, aus diesem Grund ist auch eine Lesbarkeitsformel nicht frei von jeglicher menschlichen Komponente. Das bedeutet, dass die Formel eventuell nicht alle Teile der Sprache abdecken kann und dies zur Missinterpretation der Komplexität der Sprach führen kann.

Das simple Einsetzen in eine Formel zur Berechnung eines Lesbarkeitsindizes führt dazu, dass der Vorgang auch von einem Computerprogramm übernommen werden kann. Es muss schlicht der Ablauf mit den Formeln programmiert werden und der Computer führt dies für beliebig viele Texte durch. So ist es möglich, dass die Berechnung auf beliebig viele Texte skaliert werden kann, sofern es genug Rechenkapazitäten gibt und alle Variablen korrekt in die Formel eingesetzt werden.

Interessant ist, dass besonders das amerikanische Militär Lesbarkeitsindizes große Aufmerksamkeit geschenkt hat. Nach Duffy T. M. (1982) ist das Militär eine geschlossene sogenannte Mini-Gesellschaft, welche eine Reihe standardisierter Texte benötigt. Diese Texte werden von einer Vielzahl von unterschiedlichen Menschen aus allen Gesellschaftsgruppen verwendet. Aus diesem Grund ist es notwendig, die Texte so zu formulieren, dass sie für jede Person verständlich sind. Sie umfassen nicht nur Beschreibungen von technischen Geräten, sondern auch Verträge, Ausbildungstexte, Trainingspläne und die Beschreibung von Arbeitsprozessen, welche möglichst einfach gestaltet werden sollen.

Der Anspruch simple Texte bereitzustellen, ist deshalb sehr hoch, da das amerikanische Militär auf der ganzen Welt Niederlassungen hat und Texte von allen diesen unterschiedlichen Personen auch gut verständlich sein müssen, sodass zum Beispiel Geräte an jedem Ort richtig verwendet werden können (Duffy T. M., 1985). In diesem Kontext haben sich Lesbarkeitsformeln als zu erfüllender Standard in Produktbeschreibungen durchgesetzt. Mithilfe der Formeln kann bestimmt werden, ob der ausgewählte Text den Ansprüchen gerecht wird und eingesetzt werden kann.

2.2. Quantifizierung der Komplexität eines Textes

Klare G. R. (1976) besagt, dass das Ziel einer Lesbarkeitsformel die Berechnung eines Scorewertes sein sollte, mithilfe derer die richtige Vorhersage der Komplexität für eine größere Anzahl an Textpassagen getroffen werden kann. Genauer gesagt, soll es mit einem Scorewert, der mithilfe einer Lesbarkeitsformel aus dem Text errechnet wird, möglich sein, mehrere Texte in Hinblick auf deren Komplexität direkt zu vergleichen. Eine Lesbarkeitsformel soll vorhersagen, ob ein beliebiger Beispieltext A für dasselbe Individuum schwieriger zu lesen sein wird als ein Beispieltext B. Die Fähigkeiten des Individuums, einen Text korrekt zu verstehen, sind zuerst nebensächlich, da es in diesem Fall nur wichtig ist, dass davon ausgegangen wird, dass beide Texte vom selben Individuum gelesen werden.

Um den Scorewert des Lesbarkeitsindizes zu bestimmen, ist schlussendlich die genauere Betrachtung der schulischen Ausbildung notwendig. Um die Schulbildung allgemein zu quantifizieren, hat man sich darauf geeinigt, dass die Anzahl der absolvierten Schuljahre eines Individuums den Scorewert angeben. Anhand der Schulbildung des Individuums wird entschieden, welche Ausprägung des Lesbarkeitsindizes ein Texten aufweist.

Wenn ein Text beispielsweise einen Scorewert von 5 beim FOG-Index aufweist, dann sind durchschnittlich 5 Jahre Schulbildung notwendig, damit ein durchschnittliches Individuum den Text sinnerfassend lesen und verstehen kann. Diese Methode lässt sich allgemein sehr gut verstehen, da jeder anhand seiner eigenen Erfahrungen die Skala ansatzweise durchblicken und interpretieren kann. Wie jeder aus eigener Erfahrung jedoch weiß, ist nicht jeder, auch wenn er exakt dieselbe Ausbildung genossen hat, nach einer gewissen Schulstufe auf demselben Leselevel. Dieser Sachverhalt wurde „literacy gap“ benannt und in wissenschaftlichen Texten, wie jenen von Duffy T. M. (1978), Mockovak (1974), Carver (1974), Caylor (1973) erklärt. Es wurde festgestellt, dass es eine signifikante Anzahl an Individuen gibt, welche nach etwa 5 Jahren Schulausbildung nicht die notwendigen Lesefertigkeiten besitzen, welche sie anhand des Readability-Scorewertes beherrschen sollten. Die

Literatur ist auf dieses „literacy gap“ aufmerksam geworden, als sie einige Probanden mit der gleichen Schulausbildung einen Text lesen ließen, der mithilfe eines Lesbarkeitsindizes adäquat zu ihrer schulischen Ausbildung passen sollte (Duffy T. M., 1985). Die Probanden waren jedoch nicht in der Lage die Texte zur Gänze sinnerfassend zu lesen, was zu dem Schluss führte, dass der Scorewert der Lesbarkeitsindizes die Lesefähigkeiten etwas überschätzt. Dies bedeutet, dass manche Texte im amerikanischen Militär, die mithilfe der Lesbarkeitsindizes erstellt wurden, zu komplex gestalten wurden. Folglich können die Texte nicht so effizient genützt werden, wie es von ihren Erstellern erwünscht wurde. In der Regel sind durchschnittlich mehr Jahre Schulausbildung notwendig, als ein Lesbarkeitsindex errechnet. In diesen wissenschaftlichen Texten wurde jedoch noch nicht festgehalten, wie groß dieser Bias für den jeweiligen Lesbarkeitsindex ist, sodass eine Anpassung des errechneten Scorewertes das Problem lösen könnte.

Kern (1980) hat festgestellt, dass Lesbarkeitsindizes speziell bei kürzeren Texten der Fehler größer wird und dadurch die Anwendung bei kürzeren Texten nicht zu empfehlen ist. Seine Erkenntnisse werden jedoch stark kritisiert, da sie sich nur auf sehr kurze Textpassagen beziehen. Weiters bemerkt Kern, dass bei einem größeren Stichprobenwert die Einschätzung sehr viel genauer wird und die Fehlerwahrscheinlichkeit bedeutend abnimmt. Daraus kann abgeleitet werden, dass für längere Textpassagen Lesbarkeitsindizes sehr wohl gut geeignet sind.

In der vorliegenden Arbeit wurden Konzernabschlüsse von Unternehmen nach IFRS herangezogen, die durchschnittlich mehr als 100 Seiten aufweisen, also als lange Texte deklariert werden können. Daher ist die Verwendung von Lesbarkeitsindizes zur Analyse der Komplexität des sprachlichen Gebrauchs in den vorliegenden Fällen durchaus legitim. Es muss jedoch beachtet werden, dass jeder Leser ein Individuum ist und somit einzigartig. Aus diesem Grund kann nicht genau gesagt werden, wie viele Jahre Schulausbildung eine Person absolviert haben muss, um einen gewissen Text sinnerfassend lesen und verstehen zu können. Der Scorewert eines Lesbarkeitsindex stellt lediglich einen Durchschnittwert dar, es ist ein Schätzwert, der es möglich macht, Schlüsse auf den Durchschnitt der Bevölkerung zu ziehen.

2.3. Wie wird eine Lesbarkeitsformel entwickelt?

Dieses Kapitel widmet sich der Zusammensetzung einer Lesbarkeitsformel, sowie der anschließenden Berechnung eines Lesbarkeitsindex. Bei der Entwicklung einer Lesbarkeitsformel wird nach Caylor, (1973) und Kincaid J. P. (1975) zuerst eine Auswahl an Textpassagen von verschiedenen Gruppen von Probanden getestet. Die Gruppen werden so zusammengesetzt, dass in einer Gruppe nur Probanden sind, in etwa dieselben Lesefähigkeiten besitzen. Basierend auf dem Verständnis der Textpassage der unterschiedlichen Gruppen wird ein Scorewert bestimmt.

Wird eine Textpassage nur von einigen der Gruppen verstanden, dann bestimmt die Anzahl der absolvierten Schuljahre der schlechtesten Gruppe, welche den Text sinnerfassend verstanden hat, den Scorewert.

Als nächstes wird eine Auswahl von Informationen jeder Textpassage gesammelt, wie zum Beispiel Anzahl der Präpositionen, Anzahl der Buchstaben pro Wort, Anzahl der Silben pro Wort, Wörter pro Satz, Anzahl der Nomen. Für jeden Lesbarkeitsindex gibt es eine eigene Methodik, wie sich die Formel zusammensetzt. Zum Beispiel werden in die Gunning Fog Lesbarkeitsformel nach Robert Gunning andere Informationen als in die SMOG-Lesbarkeitsformel und in die Flesch Reading Ease - Lesbarkeitsformel mitaufgenommen. Diese drei Lesbarkeitsindizes werden im Kapitel 2.4 ausführlich behandelt. Eine andere Methodik kann aber auch eine andere Gewichtung der Informationsvariablen bedeuten. Somit unterscheidet sich jede Lesbarkeitsformel von der anderen und wird folglich auch einen anderen Scorewert aufweisen. Aus diesem Grund werden in dieser Ausarbeitung drei Lesbarkeitsindizes für jeden Konzernabschluss meiner Analyse berechnet. Eine Methode, die entscheidenden Informationen für einen Lesbarkeitsindex zu erhalten, ist eine Regressionsanalyse der Informationen. Mithilfe einer Regressionsanalyse wird kurz gesagt die beste lineare Prognose der Daten produziert. Aus dieser Prognose kann gesagt werden welche Informationen ausschlaggebend sind und welche

Informationen den stärksten Zusammenhang mit der Komplexität der Textpassage aufweist.

Nach Entin (1978) sind die entscheidendsten Bestandteile einer Textpassage ein sogenannter Wortfaktor (z.B.: Anzahl der Silben pro Wort) in Kombination mit einem Satzfaktor (z.B.: Anzahl der Wörter pro Satz). Eine Kombination dieser beiden Elemente liefert den besten Vorhersagewert für das Leseverständnis. Dies führt dazu, dass die meisten Lesbarkeitsformeln die folgende Form aufweisen:

$$Scorewert = a + b * (Wortfaktor) + c * (Satzfaktor)$$

Wobei a eine Konstante ist, zu der die gewichteten Elemente, Satz- und Wortfaktor, addiert werden.

2.4. Warum werden die Lesbarkeitsformeln immer weiterentwickelt?

Es gibt eine Vielzahl von verschiedenen Lesbarkeitsformeln. Meistens wird von einem Autor eine neue Lesbarkeitsformel entwickelt, weil ein Forscher das Gefühl hat, dass speziell für diesen Fall noch keine anwendbare Formel existiert oder die bestehenden Formeln unzureichend für den anwendbaren Fall sind. In vielen Fällen werden Lesbarkeitsformeln zur Analyse eines Textes in Bezug auf einen speziellen Parameter verwendet. Genauer gesagt, soll die Lesbarkeitsformel analysieren, ob eine positive oder negative Nachricht im Text übermittelt wird, oder etwa der Tonfall der Ausdrucksweise erkannt werden. Zum Beispiel werden Pressemitteilungen mithilfe von Algorithmen analysiert und in computergesteuerte Portfoliomanager implementiert. Wenn etwa eine Nachricht zu einem bestimmten Unternehmen erfasst wird, dann deutet der Algorithmus die Nachricht und bestimmt ob diese einen positiven oder negativen Einfluss auf den Aktienkurs hat (Gomber, 2015). Aus dieser Information wird dann eine Kauf- oder Verkaufsentscheidung getroffen. Da jede wissenschaftliche Disziplin geprägt ist von einem spezifischen Fachvokabular, muss für jeden Fachbereich eine eigene Lesbarkeitsformel mit eigenen Wörterlisten entworfen werden (Klare G. , 1979).

Anfangs waren Lesbarkeitsformeln nur zur Analyse von Texten für Kinder und Jugendliche erstellt worden. Als Lesbarkeitsformeln später auch zur Analyse von Texten für das amerikanische Militär verwendet wurden, führte dies dazu, dass die Parameter nicht die richtige Schulstufe angegeben haben. Dadurch entstand eine Missinterpretation.

Kincaid J. P. (1975) hatte dieses Problem erkannt und festgestellt, dass die Formeln die Lesbarkeit der Texte überschätzten.

Diese Beobachtung hat zwei Gründe, erstens werden längere, technische Begriffe von Lesbarkeitsformeln als schwierig zu lesende Wörter erkannt, obwohl diese

Wörter für einen Leser mit Kenntnissen im jeweiligen Bereich sehr leicht zu verstehen sind. Dieses Problem kann gelöst werden, indem die Wortfaktorgewichtung *b* aus der Formel von oben verkleinert wird. Zweitens haben Erwachsene ein umfangreicheres Vokabular. Diese größere Kenntnis im Vergleich zu Jugendlichen führt dazu, dass der Intercept *a* in der Formel zu groß wird. Der Fehler wird beseitigt, indem der Intercept *a* verkleinert wird.

Wenn das Ziel der Analyse lediglich das Festlegen der richtigen Reihenfolge verschiedener Texte nach dessen Komplexität ist, dann können Lesbarkeitsformeln für Jugendliche ohne Bedenken verwendet werden, da der Fehler nur ein linearer Effekt ist. Dies bedeutet, dass Ergebnisse von Formeln für Erwachsenen sehr stark korreliert sind mit jenen für Jugendliche (Caylor, 1973).

2.5. Allgemeines zu Lesbarkeitsindizes

Die Analyse in dieser Arbeit bedient sich an den drei am weitest verbreiteten Lesbarkeitsindizes. Besonders in Publikationen zu wirtschaftlichen Themen findet der FOG-Index vermehrt Anwendung und wird von einigen Autoren zur Analyse verwendet, dieser Sachverhalt wird im Kapitel Literaturrecherche näher beschrieben. Zur Bestätigung und Überprüfung der Scorewerte zur Lesbarkeit eines Konzernabschlusses, wurden 3 verschiedene Lesbarkeitsindizes der analysierten Konzernabschlüssen berechnet. Weiterführend wird im Zuge einer linearen Regression überprüft, ob ein Zusammenhang zwischen den einzelnen Lesbarkeitsindizes besteht. Weisen die Lesbarkeitsindizes einen starken linearen Zusammenhang auf, kann gesagt werden, dass die Berechnung der Lesbarkeitsindizes richtig durchgeführt wurde und repräsentativ ist. Im folgenden Abschnitt wird auf die drei Lesbarkeitsindizes eingegangen und ihre Berechnung kurz beschrieben, sodass der Leser einen Einblick in die Vorgehensweise der Berechnung der Indizes bekommt. Danach folgen eine Interpretation sowie eine Deutung der Ergebnisse.

2.5.1. FOG-Index:

2.5.1.1. Beschreibung des FOG-Index

Ein Beispiel zur Analyse von Texten ist der FOG-Index, entwickelt von Robert Gunning. Nach seiner universitären Ausbildung war Robert Gunning ein Autor für die Scripps-Howard and Knight Zeitung. Sein größtes Anliegen war es, dass Firmen und Angestellte von Wirtschaftszeitungen und Magazinen Texte formulieren, welche leicht verständlich sind. In Gunning (1969) hat Robert Gunning zwanzig Jahre nach der Entwicklung des FOG-Index ein Resumé (Gunning, 1952) über den von ihm selbst erstellten Lesbarkeitsindex gemacht. Er stellte fest, dass der FOG-Index besonders im wirtschaftlichen Kontext Verbreitung fand. Der FOG-Index dient nicht als Hilfsmittel, sondern viel eher als Warnhinweis für Autoren, um Texte nicht zu komplex zu gestalten.

Sein Beweggrund zur Erstellung des Lesbarkeitsindex war, dass vorhergehende Lesbarkeitsindizes, seiner Meinung nach zu sehr auf die Wortvielfalt eines Textes und nicht auf die Komposition der Sätze, wie zum Beispiel die Satzlänge, eingehen (Gunning, 1969). Dies bedeutet, dass ein Text mit einer großen Wortvielfalt automatisch komplex ist. Außerdem waren bis dahin die Lesbarkeitsformeln und die Berechnung der Formel sehr komplex. Er wollte eine Lesbarkeitsformel entwickeln, welche einfach zu verwenden ist und gleichzeitig durch Anwendung der Formel die Textqualität verbessert. Der FOG-Index wurde nicht nur von dem amerikanischen Militär, der Navy und der Air Force verwendet, sondern auch in wirtschaftlichen Publikationen. Bei der Verwendung des FOG-Index werden besonders die Effektivität und die Einfachheit der Formel hervorgehoben (Gunning, 1969).

Auch wenn die Ausbildung der Menschen im Laufe der Jahre immer besser geworden ist, bedeutet dies nicht, dass die Komplexität der Sätze zwingend steigen muss. Mithilfe von Lesbarkeitsindizes, wie dem FOG-Index, haben es Autoren und Editoren geschafft, nicht nur die Komplexität der geschriebenen Sprache, etwa in Zeitungsartikeln, zu verkleinern, sondern auch einfachere sowie verständlichere Texte zu erstellen.

Der Effekt von Lesbarkeitsindizes hat vor allem auf Tageszeitungen einen deutlichen Einfluss genommen. Viele Herausgeber von Tageszeitungen verwenden heutzutage Lesbarkeitsindizes, um ihre Artikel verständlich zu machen (Gunning, 1969). Es ist jedoch anzumerken, dass ein gutes Verständnis der Texte, nicht gleichbedeutend mit der Qualität des Inhalts ist. Robert Gunning hat in den 1940er Jahren Studien von Zeitschriften, unter anderem in New York, Philadelphia, Washington, San Francisco, et cetera durchgeführt, um Lesegewohnheiten besser verstehen zu können (Gunning, 1969). In der Regel sind Leser in der Lage Texte zu verstehen, die sehr komplex formuliert sind, aber Sie bevorzugen oft einfach formulierte Texte. Anhand dieses Ergebnisses haben sich die Zeitungen dazu entschieden, mit Robert Gunning und seinem Team zu arbeiten, um ihre Texte in den Zeitschriften verbessern zu können. Es wurden den Zeitungsautoren Empfehlungen aber ebenso Warnungen für ihre Texte gegeben. Nur durch das Weglassen verschachtelter Sätze und unnützer

Wörter, konnte quasi über Nacht die Komplexität der Texte um etwa 10% gesenkt werden (Gunning, 1969). Für die Autoren wurde der FOG-Index als Warnsystem gegen verschleiertes und komplexes Schreiben im Arbeitsprozess implementiert. Besondere Erfolge konnten die Experten rund um Robert Gunning mit ihren Maßnahmen in der Fachpresse von Industrie und Wirtschaft erreichen. In fachspezifischen Texten verwenden Autoren meistens sehr spezielle Ausdrücke, anstatt diese sehr komplizierten Vorgehensweisen einfach und verständlich zu erklären. Nach Gunning (1969) geschieht dies, weil Autoren meistens mit ihrer Wortvielfalt beeindrucken wollen. Hinzukommend werden solche Fachzeitschriften meistens nur von Personen gelesen, welche über ausreichende Fachkenntnisse verfügen, sodass sie diese komplexen Ausdrücke verstehen, weil sie ihnen geläufig sind. Die potentiellen Leser sind daher eine sehr eingeschränkte Gruppe, wobei es auch hier besser wäre, wenn Texte simpler formuliert werden.

Auch wenn der FOG-Index als ein sehr einfach zu berechnendes Warnsystem konzipiert wurde, ist es trotzdem empfehlenswert ein Expertenteam zu engagieren, welches unterstützend bei Anpassungen zur Seite steht, da die Effekte sich sonst negativ auf den Text auswirken können. Nach einer Informationsphase vom Expertenteam, kann der FOG-Index als Selbstüberprüfungs- bzw. Warnsystem für eigene Texte fungieren, sodass die Texte nicht zu komplex gestaltet werden.

Im Gegensatz dazu ist der FOG-Index kein Maßstab für die Qualität eines Textes, etwa wie gut ein Inhalt recherchiert wird. Ein qualitativ schlechter Inhalt, der nur einfacher formuliert wird, ist und bleibt schlecht qualitativ minderwertig.

2.5.1.2. *Berechnung des FOG-Index nach Gunning (1952):*

Schritt 1)

Im ersten Schritt muss die exakte Anzahl an Wörtern und Sätzen der gewählten Textpassage gezählt werden.

Schritt 2)

Anschließend wird die Anzahl der Wörter durch die Anzahl der Sätze dividiert, um die durchschnittliche Satzlänge der Textpassage zu erhalten.

Schritt 3)

Als nächstes werden alle Wörter mit drei oder mehr Silben zusammengezählt. Ausgenommen: Nomen, zusammengesetzte Wörter oder zweisilbige Wörter, welche aus mehreren zweisilbigen Wörtern zusammengesetzt sind.

Schritt 4)

Nun wird die Anzahl der mehrsilbigen Wörter gemäß Schritt 3) durch die Anzahl aller Wörter der Textpassage dividiert.

Schritt 5)

Zu dem Ergebnis aus Schritt 4) wird die durchschnittliche Satzlänge aus Schritt 2) addiert.

Schritt 6) Multipliziere das Ergebnis aus Schritt 5) mit dem Faktor 0,4.

$$FOG - Index = \left(\frac{Anzahl\ der\ Wörter}{Anzahl\ der\ Sätze} + Anzahl\ aller\ Wörter \right) * 0{,}4$$

Das Ergebnis des FOG-Index gibt an, wie viele Jahre Schulausbildung ein Mensch genossen haben muss, um einen Text beim ersten Durchlesen verstehen zu können. Ein FOG-Index von 12 bedeutet zum Beispiel, dass eine Person die Schulausbildung bis zur 8.Klasse Gymnasium erfolgreich abgeschlossen haben müsste, um den Text verstehen zu können. Wenn ein Text für den Großteil der Bevölkerung verständlich sein soll, dann sollte der FOG-Index des Textes kleiner als 12 sein. Wenn der Text für jeden verständlich sein soll, ist ein FOG-Index kleiner als 8 notwendig, im österreichischen Schulsystem wäre das mit dem Pflichtschulabschluss erreicht. Fachtexte besitzen durch ihr komplexes Fachvokabular meist einen FOG-Index von etwa 18 und sind daher sehr schwer zu verstehen. In diesem Fall sind schon ein

einschlägiger Studienabschluss sowie Berufserfahrung notwendig, um so einen Text sinnerfassend verstehen zu können.

Mindestalter	Schulstufe	Indexwert (z.B.: FOG-Index
6	Volksschule	1
7		2
8		3
9		4
10	AHS Unterstufe bzw. neue Mittelschule	5
11		6
12		7
13		8
14	AHS Oberstufe, BHS, BMS usw.	9
15		10
16		11
17		12
ab 18	Universität, Fachhochschule oder berufliche Erfahrung	höher als 13

Quelle: International Standard Classification of Education Fields of Education and Training 2015, http://uis.unesco.org/en/topic/international-standard-classification-education-isced (27.08.2018)

2.5.2. SMOG-Index:

2.5.2.1. Beschreibung des SMOG-Index:

Mc Laughlin (1969) hat im Jahre 1969 eine neue Lesbarkeitsformel präsentiert, die sogenannte SMOG-Formel. Diese verspricht schneller und leichter anwendbar zu sein, und exaktere Ergebnisse zu produzieren als vorhergehende Formeln. Mc Laughlin (1969) geht davon aus, dass in der englischen Sprache nicht nur die Länge der Wörter entscheidend ist, sondern auch die Satzlänge. Diese beiden Faktoren hat er als Erster in diesem Gebiet miteinander multipliziert anstatt diese einfach aufzuaddieren. Vorhergehende Studien ignorieren seines Erachtens, dass ein kleiner Unterschied in der Satz- bzw. Wortlänge zwischen zwei Satzpassagen nicht darauf hindeutet, dass der Absatz schwieriger oder einfacher zu lesen ist. Aus diesem Grund hat er sich gegen die Addition und für die Multiplikation der beiden Faktoren Satzlänge und Wortlänge entschieden.

Mathematisch ausgedrückt haben vorhergehende Formeln folgende Struktur:

$$Lesbarkeit = 3 + \sqrt{Anzahl\ an\ mehrsilbigen\ Wörter}$$

Das Ergebnis spiegelt die Anzahl der Jahre an schul- bzw. universitärer Ausbildung wieder, welche der Leser absolviert haben muss, um den Text beim ersten Lesen verstehen zu können. Weiters weist er darauf hin, dass seine Formel einfacher zu berechnen ist, da sie eine Konstante weniger benötigt. Wie auch schon im FOG-Index von Robert Gunning werden im SMOG-Index die mehrsilbigen Wörter gezählt und analysiert. Die Leseschwierigkeit wird gleichfalls mit den Jahren der benötigten Schulbildung angegeben. Ein Wert zwischen 13 und 16 deutet darauf hin, dass der Leser Maturaniveau erlangen muss. Werte zwischen 17 und 18 benötigen einen Bachelor und Werte über 19 deuten darauf hin, dass man eine weitere akademische oder berufliche Ausbildung benötigt.

2.5.2.2. Berechnung des SMOG-Index nach (Mc Laughlin, 1969)

Schritt 1)

Im ersten Schritt müssen jeweils zehn zusammenhängende Sätze am Anfang, in der Mitte und am Ende der zu bearbeitenden Textpassage ausgewählt werden.

Schritt 2)

Danach muss jedes Wort mit drei oder mehr Silben der Gruppe der auserwählten Sätze gezählt werden. In diesem Schritt werden nur Wörter mitgezählt, welche öfter als einmal auftreten.

Schritt 3)

Als nächstes muss die Quadratwurzel vom Ergebnis aus Schritt 2) berechnet und zur nächsten Zehnerstelle auf- bzw. abgerundet werden.

Schritt 4)

Zuletzt wird drei zu dem Ergebnis aus Schritt 3) addiert, um den SMOG-Index zu erhalten. Die Zahl drei wird addiert, damit die Parametrisierung korrekt ist. Genauer gesagt, wird so sichergestellt, dass der errechnete Scorewert die richtige Schulstufe angibt.

2.5.3. Flesch Reading Ease Formel:

2.5.3.1. Beschreibung der Flesch Reading Ease Formel:

Der dritte und letzte Lesbarkeitsindex, der in dieser statistischen Analyse von Konzernabschlüssen berechnet wird, ist der Flesch-Ease Formula nach (Flesch, 1948). Unter Flesch-Kincaid Lesbarkeitstests sind zwei verschiedene Tests bekannt, welche auf denselben Prinzipien beruhen, aber eine unterschiedliche Gewichtung der Wortlänge und auch der Satzlänge aufweisen. Wie schon der FOG-Index, hatte die Flesch-Ease Formula seine ersten Anwendungsgebiete im amerikanischen Militär zur Überprüfung der Verschachtelung von Beschreibungen für technische Geräte. Die Formel wurde 1979 zum United States Military Standard ausgewählt (Kincaid J. P., 1975). Diese Formel wurde im Anschluss als Hilfsmittel zur Vereinfachung von Polizzen bei Automobilversicherungen angewendet und dient nun in vielen amerikanischen Staaten als rechtliche Voraussetzung für Dokumente, wie zum Beispiel Verträge (McClure, 1987).

Im Gegensatz zu den ersten beiden Lesbarkeitsformeln, gibt das Ergebnis dieser Lesbarkeitsformel nicht an, wie viele Jahre Schulausbildung der Leser absolviert haben muss, um den Text verstehen zu können. Das Intervall der Ergebnisse beginnt bei 100, was einen sehr einfach zu lesenden Text beschreibt und je kleiner das Ergebnis wird, desto schwieriger ist der Text zu verstehen.

Die untere Schranke liegt bei null, was bedeutet, dass der Text am besten von Universitätsabsolventen verstanden werden kann. Ein Wert bei zirka 50 bedeutet, dass der Text etwa Maturaniveau dem Leser abverlangt. Im Gegensatz zu SMOG- und FOG-Index bedeutet bei der Flesch Reading Ease Formula ein höherer Wert eine leichtere Lesbarkeit des ausgewählten Textes.

2.5.3.2. Berechnung des Flesch Reading Ease Formula -Index nach (Flesch, 1948):

Schritt 1)

Wie bereits in der Formel für den FOG-Index muss die durchschnittliche Satzlänge berechnet werden.

Schritt 2)

Hinzukommend muss für diese Formel auch noch die durchschnittliche Anzahl von Silben pro Wort berechnet werden.

Schritt 3)

Anschließend werden die Ergebnisse in die folgende Formel eingesetzt:

$$FK - Index = 206{,}835 - (1{,}015 * Durchschnittliche\ Satzlänge) - (84{,}6 * Durschnittliche\ Anzahl\ Silben\ pro\ Wort)$$

3. Anwendungsgebiete von Lesbarkeitsindizes

3.1. Verwendung von Lesbarkeitsindizes bei der Analyse von Konzernabschlüssen

Die Offenlegungspflicht für Konzernabschlüsse ermöglicht eine Analyse zum Zweck dieser Arbeit. Gesetzlich vorgeschrieben ist es, dass jedes börsennotierte Unternehmen sicherstellt, dass sein Jahresfinanzbericht veröffentlich wird und mindestens zehn Jahre lang öffentlich zugänglich bleibt. [1]

In Österreich umfasst das Aktiengesellschaften, Societas Europaea, GmbHs und Genossenschaften, aber auch sogenannte Mischformen, wie zum Beispiel die GmbH & Co KG und AG & Co KG, sofern sie keinen persönlich haftenden Gesellschafter mit Vertretungsbefugnis haben. Daher sind auch die Konzernabschlüsse der Unternehmen, die am Prime Market der Wiener Börse notiert sind, einfach auf der Unternehmenshomepage zu finden.

Laut dem IAS 1.15 True-and-Fair-View-Grundsatz der IFRS muss ein Jahres- bzw. Konzernabschluss nach International Financial Reporting Standards (kurz: IFRS) die tatsächlichen Verhältnisse wahrheitstreu wiederspiegeln. Dies bedeutet, dass alle Geschehnisse, aber auch zukünftige Gefahren für das Unternehmen dokumentiert sein müssen, sodass ein möglicher Inverstor einen ausreichenden Einblick in die aktuelle sowie zukünftige Lage des Unternehmens erhält.

Die Analyse dieser Arbeit wird jedoch nicht auf den Informationsgehalt des Textes selbst eingehen, sondern es wird die sprachliche Formulierung im Text analysiert. Unter Verwendung eines Lesbarkeitsindizes wird die Komplexität jedes Konzernabschlusses bestimmt. Der Grund, warum ein Lesbarkeitsindex für die Analyse gewählt wird ist, dass durch die Verwendung von Lesbarkeitsindizes die Komplexität eines Textes quantifiziert werden kann. In Folge können zwei Texte gegenübergestellt und mit Hilfe der Lesbarkeitsindizes eine objektive Aussage über

[1] *Quelle: §§ 277 UGB*

die Komplexität der Texte getroffen werden. Noch viel wichtiger ist, man erhält einen Zahlenwert, welcher die Komplexität des Textes quantifiziert. Mit diesem Zahlenwert kann im direkten Vergleich von zwei unterschiedlichen Texten bestimmt werden, welcher der beiden Texte komplexer formuliert worden ist.

3.2. Zukünftige Anwendungsbereiche der Analyse von Konzernabschlüssen mithilfe von Lesbarkeitsindizes

Im folgenden Kapitel werden die zukünftigen Anwendungsmöglichkeiten der Methodik diskutiert. Konzernabschlüsse von Unternehmen zeigen, wie sich das Unternehmen entwickelt hat und stellen die wirtschaftliche Performance des Unternehmens dar. Dies ist nur einer der vielen Gründe, warum es für Unternehmen vorteilhaft ist, sich besonders gut in ihrem Konzernabschluss zu präsentieren. Des Weiteren ist der Konzernabschluss von Unternehmen eine wichtige Grundlage für Investoren, ob sie in ein Unternehmen investieren oder nicht. Wenn Unternehmen über einen längeren Zeitraum keine ausreichenden wirtschaftlichen Leistungen erbringen, sich im Konzernabschluss jedoch besser als den Tatsachen entsprechend präsentieren, so kann dies dazu führen, dass sogar multinationale Unternehmen in finanzielle Schwierigkeiten gelangen können (Sadasivam, 2016). Die wahre wirtschaftliche Leistung eines Unternehmens zu erkennen ist eine große Herausforderung bei Analyse des Konzernabschlusses eines Unternehmens.

Aus diesem Grund ist es wichtig, dass Analysten diese sogenannten *management frauds* erkennen um die wahre wirtschaftliche Leistung des Unternehmens zu beurteilen. Wenn jedes Unternehmen einzeln analysiert werden muss, dann ist es für Analysten enorm zeit- und daher auch sehr kostenintensiv. Folglich ist es wichtig neuartige Möglichkeiten zu entwickeln, welche die wirtschaftliche Leistung des Unternehmens abbilden.

Optimal wäre ein automatisierter Vorgang, welcher jedes Unternehmen objektiv analysiert. Eine Analyse der sprachlichen Gestaltung eines Konzernabschlusses kann einen Teil einer Analyse darstellen. Die Analyse könnte verbessert werden, da diese quantitative Analyse eine objektive Herangehensweise ist und somit eine qualitative Analyse verbessern könnte. Es könnte hilfreich sein die Komplexität der Sprache als zusätzliche Kennzahl heranzuziehen um eine qualitative Analyse zu ergänzen.

Aus diesem Grund wird im Zuge dieser Arbeit ein Code programmiert, welcher die Konzernabschlüsse einliest, die Sprache der Texte analysiert und für jeden Konzernabschluss mehrere Lesbarkeitsindizes berechnet.

Für Investoren können die Ergebnisse der Lesbarkeitsindizes für Konzernabschlüsse als Überprüfung für den Börsenkurs eines Unternehmens dienen, sodass für ein Unternehmen nicht mehr bezahlt wird, als dieses eigentlich Wert ist.

Die Hypothese, welche in dieser Arbeit aufgestellt wird, besagt, dass ein Konzernabschluss mit einer höheren Komplexität eine schlechtere Performance des Unternehmens verschleiert.

Grundlage für diese These ist die Beobachtung, dass sobald die Zahlen in der Bilanz und Gewinn- und Verlustrechnung nicht zufriedenstellend oder besser als in der Vorperiode sind, sind Unternehmen gezwungen dieses Versäumnis genauer und vor allem detaillierter zu beschreiben. Zumal sich Unternehmen dennoch gut präsentieren möchten, ist die Theorie jene, dass die Komplexität der Sprache in ihrem Konzernabschluss höher ist als ihre tatsächliche wirtschaftliche Leistung. Diese Methodik soll eine einfache und schnelle Möglichkeit darstellen, Unternehmen objektiv zu analysieren, bevor überhaupt eine Analyse durch einen Menschen durchgeführt wurde. In Zukunft könnte es einen zusätzlichen Teil einer Unternehmensanalyse darstellen, welcher objektiver ist, da jegliche menschliche Komponente bei der Berechnung ausgeschlossen ist.

Eine wichtige Frage, welche geklärt werden muss ist, wie die wahre Performance des Unternehmens genauer gemessen werden kann. Hierfür wurde eine Kennzahl ausgewählt, welche für jedes Unternehmen verfügbar und einfach zu berechnen ist, die Earnings per Share.

Aus dem Jahresüberschuss wird die Dividende für die Anteilseigner bezahlt, was ein wichtiges Indiz für Investoren darstellt. Die Kennzahl Earnings per Share (kurz: EPS) wurde für jedes Unternehmen herausgesucht und ist in die Analyse eingeflossen. Die Formel für EPS ist wie folgt:

$$EPS = \frac{Net\ Income - \ Dividende\ an\ Vorteilsaktieninhaber}{durchschnittliche\ Anzahl\ an\ Aktien}$$

Die Kennzahl wurde ausgewählt, da sie direkt mit der wirtschaftlichen Leistung des Unternehmens zusammenhängt. EPS zeigt wie viel Jahresüberschuss das Unternehmen abzüglich Dividende, je Aktie erreicht hat ohne dadurch die zukünftige Leistung zu gefährden, besonders, da etwaige Investitionen in das eigene Unternehmen, wie zum Beispiel die Neuanschaffung einer Maschine oder eine andere Investition vorher abgezogen wurden. Es wird also auch die Entscheidung des Unternehmens, wie viel Dividende ausgeschüttet wird, miteinbezogen. Weiters werden diese Kennzahlen in vielen anderen Publikationen verwendet, wie im Abschnitt Literaturrecherche genauer beschrieben wird.

3.3. Limitationen der Analyse von Konzernabschlüssen mithilfe von Lesbarkeitsindizes

Zunächst sollte erwähnt werden, dass das Ergebnis der Lesbarkeitsindizes in Hinblick auf die Güte des Konzernabschlusses keinen Einfluss hat. Es wird lediglich die sprachliche Formulierung analysiert und mehrere Indizes werden berechnet, welche die Lesbarkeit quantifizieren. Es wird nicht analysiert, wie schön ein Konzernabschluss zusammengestellt und recherchiert wurde.

Anschließend muss festgehalten werden, dass die Lesbarkeitsindizes nicht für die Analyse von finanzwirtschaftlichen Texten konzipiert wurden. Wie in fachgleicher Literatur beschrieben wird, erstellen manche Autoren eigene Lesbarkeitsindizes speziell für finanzwirtschaftliche Texte (Loughran, 2011). Die meisten dieser neuen Methoden sind aber im Gegensatz zu den in dieser Arbeit verwendeten Lesbarkeitsindizes nicht frei verfügbar. Aus Gründen der Verfügbarkeit wurden daher die Lesbarkeitsindizes FOG, SMOG und Flesch Reading Ease Formula verwendet. Es ist unbestritten, dass Konzernabschlüsse ein komplexes Vokabular verwenden, was sich auch in den Ergebnissen dieser Arbeit wiederspiegeln wird. Da dieses Vokabular aber für alle Konzernabschlüsse etwa gleich komplex ist, wird der resultierende Fehler in den Ergebnissen nicht ausschlaggebend für die Analyse sein, weil er ein additiver Faktor ist und für jedes Unternehmen und jeden Konzernabschluss etwa gleich hoch ist. Aus diesem Grund kann der Fehler vernachlässigt werden und die Lesbarkeitsindizes können ohne Bedenken für die Analyse verwendet werden. Hinzukommend ist anzumerken, dass die Lesbarkeitsindizes in der Analyse nur relativ miteinander verglichen werden, sodass dieser Faktor keinen Einfluss auf die Resultate hat.

4. Programmierung

4.1. Einleitung in den Programmierungsprozess:

Im folgenden Kapitel wird der Programmierungsprozess dokumentiert und die Beschaffung der Daten genauer beschrieben. Des Weiteren werden auftretende Probleme behandelt und deren Lösungswege gezeigt.

Bevor die Daten analysiert werden können, müssen die Daten zunächst selbst beschafft werden. Auf den Homepages der Unternehmen, welche auf dem Prime Market der Wiener Börse notiert sind, wurden die konsolidierten Konzernabschlüsse der letzten fünf Jahre herausgesucht. Nach dem Herunterladen der .pdf-Dateien, mussten diese Dateien so umgewandelt werden, dass nur mehr der reine Text zur Berechnung der Lesbarkeitsindizes verfügbar ist.

Nach den ersten Testumwandlungen wurde der Entschluss gefasst, die Konvertierung mit dem Tool der Firma Smallpdf durchzuführen, weil mit diesem Tool die Umwandlung der .pdf-Dateien in .doc-Dateien fehlerfrei und schnell funktioniert. Das Problem von den anderen Konvertierungsprogrammen, wie etwa pdf2doc ist, dass mit einem Passwort geschütze .pdf-Dateien nicht umwandelbar sind. Mit dem Programm Smallpdf funktioniert dies jedoch fehlerfrei. Im Anschluss erfolgte die Umwandlung aller Konzernabschlüsse in .doc-Dateien ohne weitere Probleme.

Als nächsten Schritt habe ich mir überlegt, wie ich Bilder Grafiken und Tabellen der Konzernabschlüsse behandeln soll. Die meisten Grafiken enthalten einen Text, welcher bei der Umwandlung nicht verloren geht, daher habe ich die Entscheidung getroffen, alle Grafiken zu löschen und falls ein Text enthalten ist, wird dieser in die Analyse mitaufgenommen. Anschließend werden die Dateien von einer .doc in eine .txt-Datei umgespeichert. Dieser Vorgang bringt mehrere Vorteile mit sich, erstens werden Bilder, Grafiken und Tabellen, welche für die literarische Analyse des reinen Textes nicht gebraucht werden, aus dem Dokument entfernt. Somit ist das Problem mit den Grafiken und Tabellen gelöst. Zweitens werden dabei alle Formatierungen gelöscht, und nur der reine Text bleibt übrig. Dies ermöglicht es, den Text leichter

einzulesen, sodass keine Zeilen- bzw. Seitenumbrüche im Text enthalten sind, welche die Textanalyse beeinflussen würden.

Der dritte und wichtigste Punkt, der für das Umspeichern in ein .txt-Dokument spricht, ist, dass es im Statistikprogramm R eine Funktion gibt, welche .txt-Dateien problemlos einliest und somit anschließend weitere literarische Analysen ermöglicht.

Bei der Abspeicherung als .txt-Datei bietet Microsoft Word zwei weitere Optionen an. Erstens, die Option, dass Word selbstständig Zeilenumbrüche einfügt. Diese Möglichkeit wurde abgewählt, da nach eingehender Überprüfung, nur sinnfreie Zeilenumbrüche eingefügt werden und das wiederum beeinflusst deutlich die Lesbarkeitsindizes.

Diesen Sachverhalt kann man am Beispiel des Konzernabschlusses der Agrana AG beobachten. In diesem Beispiel wird die zuvor besprochene Option angewählt, Word fügt anschließend unerklärlich Zeilenumbrüche ein. Dies macht erstens den Text unlesbar und der Text entspricht nicht mehr der ursprünglichen Form. Am besten ist diese Veränderung an der Seitenanzahl zu beobachten. Die ursprüngliche Version des Konzernabschlusses besteht aus 209 Seiten, mit dieser Option schwillt der Konzernabschluss auf 655 Seiten an. Daher erscheint es am sinnvollsten diese Option abzuwählen. Etwaige Absätze oder die Anzahl der Absätze in einem Text sind unerheblich für die Analyse, da nur die Sätze selbst analysiert werden, wie aus den Formeln der Lesbarkeitsindizes erkennbar ist. Keiner der Indizes beinhaltet eine Variable, welche von Absätzen in einem Text beeinflusst wird.

Die zweite Option bietet an, dass Zeichen von Word ersetzt werden, falls bei der Umwandlung von .doc in .txt etwaige Fehler auftreten. Diese hilfreiche Funktion ermöglicht, dass bei der Umwandlung alle Sonderzeichen richtig mitumgewandelt und somit alle Wörter richtig übernommen werden können.
In Abbildung 1 ist dieses Optionsfenster ersichtlich.

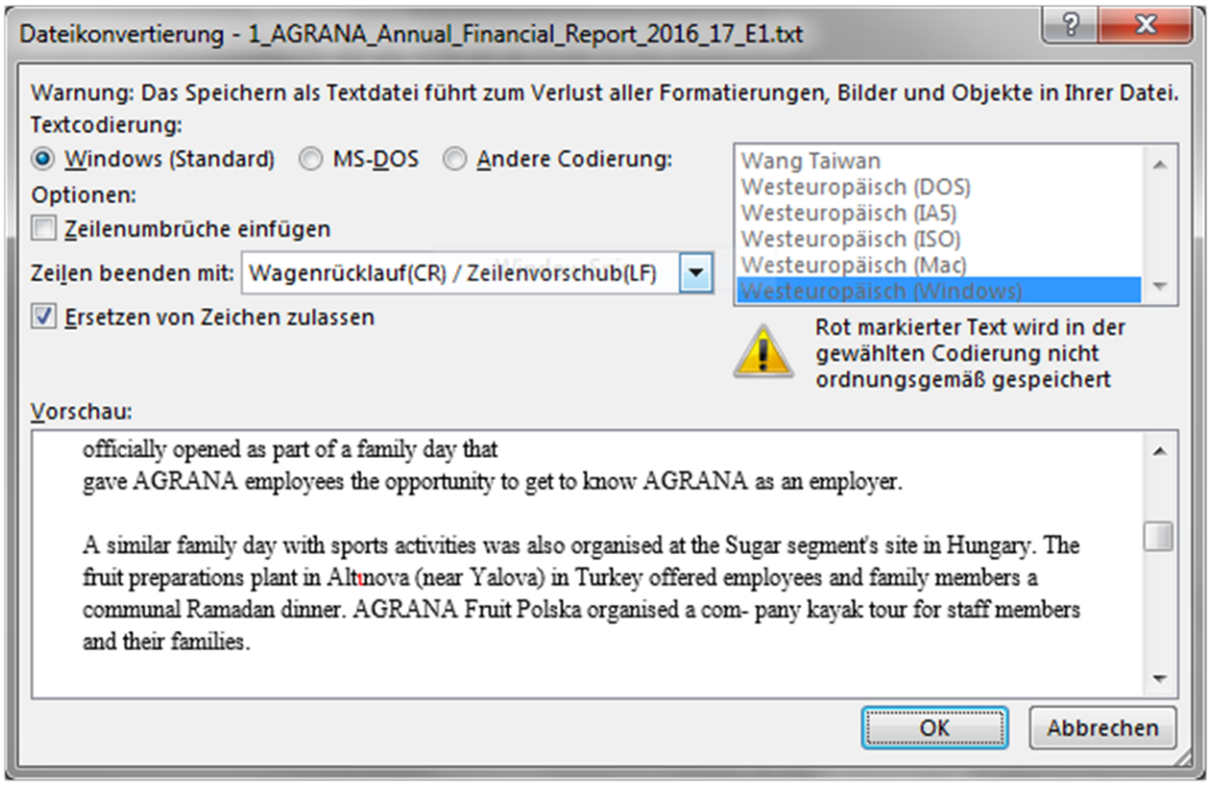


Abbildung 2 eigene Darstellung aus Microsoft Word

Um die Auswahl der Optionen in Microsoft Word zu untersuchen, wurden die relevanten Optionen durchgetestet und dessen Auswirkungen auf die Lesbarkeitsindizes überprüft.

In der folgenden Tabelle werden die Lesbarkeitsindizes für die gewählte Formatierung, „ohne Zeilenumbrüche einfügen“ und mit „Ersetzen von Zeichen“ in Vergleich gesetzt, zu der Version mit Option mit Zeilenumbrüche einfügen, „Ersetzen von Zeichen zulassen“ und zu der Version „ohne Zeilenumbrüche einfügen“, „ohne Ersetzen von Zeichen.“

Umwandlungsmethoden			FOG - Index	Flesch Reading Ease Formula	SMOG - Index
Versuch 1	*Ersetzen von Zeichen*	JA	14,45	50,82	13,44
	Zeilenumbrüche einfügen	NEIN			
Versuch 2	*Ersetzen von Zeichen*	JA	14,78	44,69	13,65
	Zeilenumbrüche einfügen	JA			
Versuch 3	*Ersetzen von Zeichen*	NEIN	14,5	45,31	13,43
	Zeilenumbrüche einfügen	NEIN			

Tabelle 1 Umwandlungsmethoden

Betrachtung des Textes bei der Option „Zeilenumbrüche einfügen“ ein nicht befriedigendes Ergebnis zustande, weshalb diese Option ebenfalls nicht gewählt wird. Die Option „Ersetzen von Zeichen“ zulassen scheint die optimalen und zuverlässigsten Resultate zu liefern, indem Sonderzeichen richtig übernommen, somit die Richtigkeit des Textes sichergestellt werden kann. Diese Erkenntnis liefern nicht nur die Werte in der Tabelle, sondern auch bei näherer Betrachtung der Text, liefert dieser Versuch das beste Ergebnis.

Nun beginnt die eigentliche Programmierung des Codes zur literarischen Analyse von Textstellen. Hierfür wird das Statistikprogramm R herangezogen.

Zu Beginn wird das Verzeichnis mit dem Befehl „path“ angegeben. Dieser Schritt gibt für das Programm den Speicherpfad der Konzernabschlüsse an. Von diesem Verzeichnis soll der Code nun die Konzernabschlüsse laden und anschließend die Analyse durchführen.

Der Befehl „library()“ lädt die gewünschte Library. Eine Library ist eine Sammlung von Funktionen, welche das Repertoire des Programms für die ausgewählte Sitzung erweitert. Libraries werden immer nur für die aktuelle Sitzung aus dem Internet geladen, sodass der Speicherplatz möglichst klein gehalten werden kann. Libraries werden nur bei Bedarf geladen werden, daher bleibt für die Anwendung des Codes genügend Speicherkapazität verfügbar. Dies ermöglicht eine schnellere und effizientere Berechnung des Codes, sodass nur die gerade benötigten Anwendungen geladen werden.

Es gibt Libraries für die verschiedensten Anwendungsbereiche, weil sie einfach von Anwendern selbst erstellt werden können. Die sogenannte R-Community ist sehr groß, besonders an der technischen Universität Wien und an der Wirtschaftsuniversität Wien, gibt es viele engagierte Wissenschaftler, die sich mit dem Programm R beschäftigen und immer wieder neue Libraries zur Verfügung stellen, die für jedermann frei verfügbar sind.

Nach Recherche im Internet erschien die „koRpus“-Library eine gute Wahl für die Analyse, weil sie speziell für literarische Analysen von Texten in R erstellt wurde. Sie enthält eine Vielzahl von Funktionen, um Texte in das Programm zu laden, zu bearbeiten und wieder neu zu speichern. Routinierte R-Anwender wissen, dass besonders das Einlesen von neuen Datensätzen, sei es nur ein Excel-file oder ein Textdokument, schwierig sein kann.

Das Problem ist meistens, dass das Programm die Daten unformatiert einliest, weshalb weitere Analysen schwierig durchführbar sind. Wenn Textdokumente eingelesen werden, ist dies nicht anders. Wie auch bei einer Exceldatei, ist es gängig, die Daten im Microsoft Office, in diesen Fall Microsoft Word, soweit zu bearbeiten, sodass nach dem Einlesen in R keine großen Veränderungen an der Datei vorgenommen werden müssen.

Aus Gründen der Einfachheit wurde entschieden, jede Datei in Word zu bearbeiten und anschließend die fertige .txt-Datei in R einzulesen.

Die nächste Funktion „list-files()“ betrachtet das durch die Funktion „path()“ angegebene Verzeichnis und gibt an, wie viele .txt-Dateien sich in diesem Ordner befinden. Dieser Befehl wird benötigt, damit eine Schleife programmiert werden kann, welche alle Dateien aus dem angegebenen Verzeichnis einliest. Bevor jedoch die Schleife programmiert werden kann, müssen noch Vektoren erstellt werden, in welchen anschließend die Lesebarkeitsindizes abgespeichert werden. Dabei werden die Lesbarkeitsindizes SMOG, Flesch Reading Ease und FOG berechnet und für jeden Index wird ein eigener Vektor erstellt. Abermals wird die Anzahl der .txt-Dateien aus dem bestimmten Verzeichnis entnommen, welche in diesem Fall die Länge der Vektoren angibt. Für jede Datei wird hier eine Zeile im Vektor angelegt.

Nun beginnt die Programmierung der Schleife. Klassischerweise beginnt die Schleife mit der Definition, welche Zahlen von eins bis zur Variable i durchgenommen werden sollen, i kann in diesen Fall jede ganze Zahl annehmen. Ähnlich wie vorher soll jede .txt-Datei bearbeitet werden, weshalb die Funktion „length(ll.files)“ verwendet wird, um jede Datei des Verzeichnisses zu inkludieren. Das englische Wort für Länge ist „length“, diese Funktion gibt die Anzahl der im Vektor gespeicherten Elemente wieder. In diesem Fall beschreibt es die Anzahl der Textdokumente, welche im angegebenen Verzeichnis enthalten sind.

Im nächsten Schritt wird die Funktion „tokenize()“ verwendet. Wie der Name der Funktion schon vermuten lässt, umfasst die Funktion „tokenize()“ alle Funktionen eines „Tokenizers“. Auf Deutsch nennt sich dieser Vorgang eine lexikalische Analyse. Ein sogenannter „Tokenizer“ (Grefenstette, 1994) liest zuerst den gewünschten Text ein. Der erste Bearbeitungsschritt ist die Entfernung sämtlicher Formatierungen. Ein Text in elektronischer Form enthält nicht nur den reinen Text, sondern meistens auch extra Leerzeichen oder andere Umbrüche, welche zu Formatierungszwecken eingesetzt werden. Obwohl diese Zeichen sehr wohl bedeutsam sind, genauer gesagt, sie helfen dem Leser den Text einfacher zu verstehen, werden sie zuerst entfernt bevor die lexikalische Analyse durch den „Tokenizer“ durchgeführt werden kann. Sie sind unbedeutend für die sprachliche Formulierung sowie für den Inhalt des Textes. Einige Formatierungen wurden schon in einem vorherigen Schritt,

nämlich bei der Umformatierung der Konzernabschlüsse, von einer Worddatei in eine Textdatei umgewandelt. Mithilfe des „Tokenizers“ werden nun die übrigen Formatierungen entfernt. Es müssen jedoch nicht nur diese Zeichen entfernt, sondern einige Stellen wieder zusammengesetzt werden. Diese Zusammensetzung muss an Stellen durchgeführt werden, wo einzelne Wörter in einem Worddokument etwa durch Zeilen- oder Seitenumbrüche in zwei Teile getrennt worden sind. Dies geschieht im nächsten Schritt mithilfe der Funktion „hyphen()“. Durch die Entfernung der Formatierung stehen die getrennten Wörter nur als einzelne Wortfetzen im Text und sind nicht lesbar. Aus diesem Grund müssen diese Teile wieder zusammengesetzt werden. Da der Ort der sogenannten Worttrennung durch die Formatierung des Textes im Worddokument, insbesondere durch den Zeileneinzug, welcher die Breite der Zeilen festlegt, hervorgerufen wird und nicht durch sprachliche Mittel, welche eine unterstützende Funktion haben und zur Übermittlung des Inhalts beitragen, können diese Wörter entfernt werden. Es ist nicht notwendig eine zusätzliche Dokumentation über diesen Sachverhalt anzulegen, weil die Stellen unbedeutend für den Inhalt des Textes sind.

Die Funktion erkennt die Worttrennungen zuerst anhand von kleingeschriebenen Buchstaben am Ende einer Zeile, gefolgt von einem Trennzeichen, anschließend ein Umbruchzeichen in Form von Tab, Leerzeichen oder Zeilenumbruch und zuletzt wieder ein kleingeschriebener Buchstabe. Es werden nur diese Wörter mit der Funktion zusammengesetzt und alle anderen Textbestandteile bleiben unverändert. Klarerweise könnten mit der Funktion auch Wörter entdeckt werden, welche auch ohne Zeilenumbruch einen Bindestich in der Mitte aufweisen (z.B.: ein französisch-deutsches Wörterbuch) oder etwa ein Bindestrich als Ergänzungsstrich, wie im Beispiel „Grün- und Buntspechte“. Bei diesem Beispiel würde das neue Wort „Grünund“ entstehen. Aus diesem Grund sind in der Funktion Wörterlisten enthalten, welche einzelne Wörter erkennen und im Anschluss diese nicht zusammen. Es wird die Annahme getroffen, dass diese Fälle nur sehr selten auftreten und daher bedeutsam sind.

Sollte hingegen ein Fehler bei der Analyse auftreten, so wird dieses eine Wort nicht erkannt, was aber statistisch gesehen, vernachlässigt werden kann, zumal bei der Anzahl an Wörtern eines Konzernabschlusses dieses eine Wort keinen spürbaren Einfluss auf das Ergebnis hat. Francis (1982) hat den Brown Corpus Text von Francis (1979) mit einer Million Wörter analysiert. Dabei sind nur 615 Fehler entstanden.

Der Brown Corpus wurde im Zuge von Francis (1979) in den 1960er Jahren entwickelt. Diese Textsammlung soll die zeitgenössische englische Sprache wiederspiegeln und alle Wörter und Phrasen abdecken, welche im englischen Sprachgebrauch gängig sind.

Francis (1982) hat den Text nach Wörtern durchsucht, welche nicht im ursprünglichen Text aufgetreten sind. Dies ergibt eine Fehlerquote von 0,0515%, welche bedenkenlos vernachlässigt werden kann. Diese Vorgangsweise wird von Grefenstette (1994) ebenfalls unterstützt.

Zusammenfassend ist nach Grefenstette (1994) zu sagen, dass bei einer lexikalischen Analyse viele verschiedene Problemstellungen zu beachten sind. Bevor die eigentliche lexikalische Analyse beginnt, müssen eine Reihe von Maßnahmen getroffen werden, um den Text richtig einzulesen. Diese betreffen hauptsächlich Formatierungsmaßnahmen, insbesondere die Zusammenführung von getrennten Wörtern, die aufgrund von Zeilenumbrüchen entstehen, sowie die Entfernung von Formatierungsmaßnahmen, die vom Autor eingefügt worden sind, um dem Leser eine ansehnlichere Form des Textes zu bieten. Diese Formatierungen haben jedoch keinen Zusammenhang mit der sprachlichen Formulierung und werden daher entfernt.

Wenn die Vorformatierung abgeschlossen ist, kann nun der linguistische Prozess nach Grefenstette (1994) fortgesetzt werden. Bei diesem Vorgang wird jedes Wort des Textes analysiert und durch das Programm die zugehörige Syntaxklasse für dieses bestimmt. Geht man zum Beispiel vom Wort „Hund“ aus, wird dieses bestimmt als Nomen, singulär. Im tokenizations Prozess werden zum ersten Mal die

einzelnen Sätze eines Textes definiert. Zuvor ist der gesamte Text nur als ein einziger String für das Programm abgespeichert. Das bedeutet, dass ein Text für den Computer nur als ein Satz erkannt wird und die einzelnen Sätze, welche ein Konzernabschluss aufweist, nicht ersichtlich sind. Da aber die Erkennung einzelner Sätze für die lexikalische Analyse notwendig ist, muss diese Aufsplittung zuerst durchgeführt werden.

Um dies zu erreichen muss dem Programm erklärt werden, welche Eigenschaften ein Satz aufweist, sodass es für den Computer möglich ist Sätze und Wörter zu erkennen. Es stellt sich die Frage, was ein Wort ist und was einen Satz auszeichnet. Zu Beginn kann festgehalten werden, dass Sätze grundsätzlich mit einem Punkt, Ausrufezeichen oder Fragezeichen enden. Dieser Sachverhalt wurde von Francis (1982) abermals mit dem Brown Corpus Text überprüft, welcher 48885 Sätze enthält. Es wurde festgestellt, dass in einem von 14 Sätzen Sonderfälle auftreten, welche problematisch bei der Satzerkennung sind. Diese Fälle müssen gesondert betrachtet werden. Sie umfassen neben Datumsangaben, Zahlen mit Tausendertrennzeichen, auch Abkürzungen. Wenn solche Fälle erkannt werden, soll das Programm das Zeichen nicht als Satzende erkennen, damit der ganze Satz richtig erkannt werden kann. Kurz gesagt, es sollen die falsch definierten Satzenden gelöscht werden, sodass die lexikalische Analyse korrekt durchgeführt werden kann.

Nun beginnt die eigentliche Tokenization. Hauptbestandteil dieses Vorgangs ist die Erkennung von eigenständigen Wörtern und Sätzen und die anschließende Bestimmung des Syntaxes des Satzelements. Wenn dieser Vorgang beendet ist, ist die lexikalische Analyse im Grunde genommen erledigt und die ganze Aufmerksamkeit kann nun allfälligen Spezialfällen gewidmet werden. Diese Spezialfälle betreffen primär falsch erfasste Satzenden und inkorrekte Worttrennungen. Wie schon vorher erwähnt, ist die Anzahl an falschen Worttrennungen nicht signifikant und können bedenkenlos ignoriert werden.

Ein Tokenizer kann nach Grefenstette (1994) als automatischer Texterfassungsprozess bezeichnet werden, der einige Spezialfälle generiert, welche

extra behandelt werden müssen. Nichtsdestotrotz ist es ein legitimer Weg einen Text mit einem Computer zu erfassen, um anschließend einen Lesbarkeitsindex zu berechnen.

Die Funktion „readability()“ berechnet aus der koRpus-Library mithilfe der Informationen der Funktionen „tokenize()“ und „hyphen()“ zahlreiche Lesbarkeitsindizes. Nach Durchlaufen des Codes, fällt auf, dass mehr als 50 Warnungen auftreten. Warnungen sind bei einem Code nicht zwingend ein Problem, da der Code trotzdem weiter berechnet wird. In diesen Fall treten Warnungen auf, die signalisieren, dass zur Berechnung für einige Lesbarkeitsindizes Wörterlisten zur Verfügung gestellt werden müssen. Für die drei Lesbarkeitsindizes, welche berechnet werden, sind keine Wörterlisten notwendig, daher werden auch keine abgespeichert. Aus diesem Grund sind alle Warnungen zur Berechnung der Indizes FOG, SMOG und Flesch Reading Ease kein Hindernis und die Ergebnisse werden fehlerfrei berechnet.

Die nächste Zeile in der Schleife nimmt die berechneten Lesbarkeitsindizes SMOG, Flesch Reading Ease und FOG aus dem Output der „readability()“ – Funktion in die zuvor definierten Matrizen ab, sodass sie nach dem Ende des Ablaufs der Schleife weiter verwendet werden können.

Wenn alle Befehle der Schleife für alle Konzernabschlüsse berechnet worden sind, ist die Schleife beendet und der Code kann nach der geschlossenen Klammer „}“ normal fortgesetzt werden. Die folgenden Befehle führen die Matrizen der Lesbarkeitsindizes zusammen. Im nächsten Schritt werden die Matrizen sowie eine Matrix mit allen Lesbarkeitsindizes im csv-Format abgespeichert, sodass diese einfach im Excel aufgerufen werden können.

4.2. Anmerkungen zur Stichprobe und zur Auswahl der Unternehmen für die Analyse:

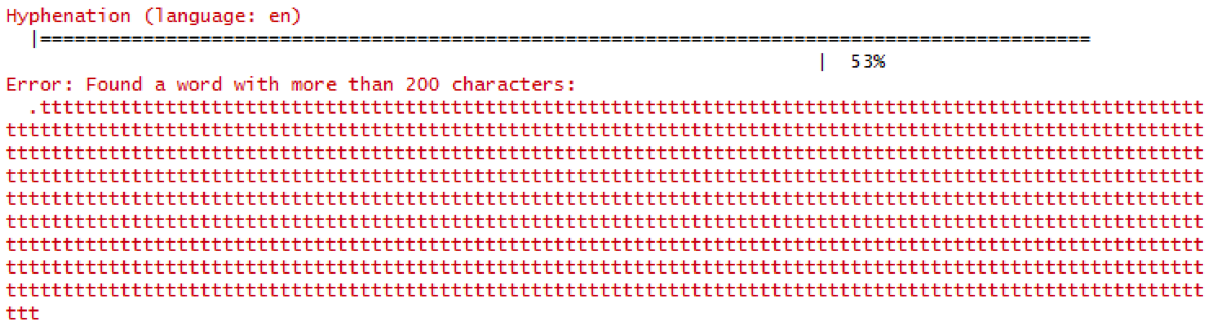
Die Konzernabschlüsse der BUWOG AG sind nur für die letzten 4 Perioden, also ab dem Fiskaljahr 2013/2014, verfügbar. Beziehungsweise die der FACC AG sind nur die Konzernabschlüsse der letzten 3 Perioden, ab 2014/2015, veröffentlicht worden, da die IPO der FACC AG erst seit 2014 am Prime Market der Wiener Börse notiert ist. FACC AG ist erst seit 25.6.2014 an der Wiener Börse notiert, bis dahin gab es laut Konzernabschluss 2014 nur die Aerospace Innovation Investment GmbH, die aus der Fusion der FACC AG und der Xi'an Aircraft Industry (Group) Company Ltd. („XAC") entstanden ist. Für dieses Unternehmen werden nur die Konzernabschlüsse ab 2014 in die Analyse einbezogen.

Bei der Immofinanz AG endet das Geschäftsjahr am 31.12.2016, im Vergleich dazu endeten die vorherigen Geschäftsjahre mit 30. April. Aus diesem Grund gibt es einen Sonderkonzernabschluss für das Geschäftsjahr 1.5.-31.12.2016. Davor endete das Fiskaljahr am 30.04..

Die RHI AG wurde am 25.10.2017 das letzte Mal an der Wiener Börse gehandelt. Die RHI AG hat mit der brasilianischen Magnesita fusioniert und wird nun als RHI Magnesita N.V. geführt. Die RHI Magnesita N.V. hatte am 27.10.2017 ihren ersten Handelstag an der Londoner Börse (die presse, 2017). Aus diesem Grund und auch weil die Konzernabschlüsse nicht auf der Homepage verfügbar sind, wird die RHI AG bzw. die RHI Magnesita auch nicht in die Analyse aufgenommen. Ein Überblick der Unternehmen, welche für die Analyse herangezogen wurden, ist im Anhang ersichtlich.

4.3. Probleme bei der Analyse der Konzernabschlüsse

Beim Durchlauf der Schleife ist leider bei der Analyse der Konzernabschlüsse von 2014 eine Fehlermeldung aufgetreten, welche in der folgenden Abbildung ersichtlich ist. Diese folgende Fehlermeldung führt im Gegensatz zu einer Warnung dazu, dass der Code vom Programm nicht weiter durchgeführt werden kann.



Da der Code in diesem Fall alle Konzernabschlüsse von den Unternehmen analysiert, welche am Prime Market gelistet sind, war es nicht möglich, herauszufinden bei welcher Textdatei der Fehler immer wieder auftritt. Wie zu sehen ist, wurde in dem problembehafteten Textdokument ein Wort entdeckt, welches mehr als 200 Zeichen aufweist. Anscheinend ist bei der Formatierung in diesem Fall ein Fehler aufgetreten, welcher vom Code entdeckt worden ist. Dieser Fehler ist in meiner Analyse nur einmal aufgetreten, daher sehe ich keine Notwendigkeit die Problemlösung zu automatisieren.

Die Konzernabschlüsse wurden alphabetisch geordnet und mit laufenden Nummern versehen. Diese Nummerierung erleichtert die Gliederung und verschafft bei der Analyse einen genaueren Überblick.

Um zu erfahren, welche Textdatei zu dem Fehler führt, wurde der Löwenfangalgorithmus verwendet, welcher in der Mathematik unter anderem zur Nullstellensuche angewendet wird. Der Name stammt aus dem theoretischen Problembeispiel, welches zum Verständnis des Algorithmus dient. Einfach gesagt ist die Ausgangsituation jene, dass man einen Löwen in einer Wüste suchen muss. Um ihn zu finden, wird die Wüste in zwei Teile geteilt. Dann wird jene Hälfte gewählt, in

welcher sich der Löwe befindet und diese wird wiederum in zwei geteilt. Durch Iteration dieses Prozesses nähert man sich dem Löwen beliebig nahe. Dieses Prinzip wird bei dieser Problemstellung angewendet.

Zur Kontrolle wurde die erste Hälfte der Konzernabschlüsse, genauer gesagt, die Konzernabschlüsse eins bis 19 von 38 nochmals eingelesen, um zu erfahren, ob der Fehler unter den ersten 19 Konzernabschlüsse, vorliegt. Dies war dann auch der Fall und bestätigte die Vermutung. Anschließend wurden die ersten 19 Konzernabschlüsse wieder in zwei Teile geteilt und abermals wurde jener Teil, welcher den Fehler enthielt, weiter analysiert. Schlussendlich wurde die fehlerbehaftete Textdatei gefunden und das fehlerhafte Wort, welches nur aus t's bestanden hat, aus der Textdatei entfernt. Anschließend funktionierte der Durchlauf des Codes ohne Fehlermeldungen.

5. Literaturrecherche

In diesem Kapitel wird untersucht, ob es bereits wissenschaftliche Publikationen im Umfeld meiner Arbeit gibt. Es wird eine ausgewählte Anzahl an Studien behandelt und näher beschrieben, um einen besseren Einblick in die Thematik zu erhalten. Die Arbeit von Li (2008) war sehr hilfreich, da in dieser Veröffentlichung die Vorgehensweise sehr detailliert beschrieben wird und wie in dieser Arbeit der FOG-Index zur Analyse verwendet wird. Die anderen Texte verwenden teilweise etwas andere Methoden zur Analyse für Texte, sie sind aber dennoch sehr hilfreich, da ich auf die Erfahrungswerte von den Autoren zurückgreifen konnte und so meine Analyse optimieren konnte. Hervorheben möchte ich auch noch den Text von Loughran (2011), welcher sehr hilfreich war bei der Suche nach einer geeigneten Kennzahl für die wirtschaftliche Leistung eines Unternehmens. Aber auch die detaillierte Auseinandersetzung einer sprachlichen Analyse von Abschlüssen, war sehr interessant und hilfreich für meine Analyse.

5.1. Loughran, Tim “When is a liability not a liability?”

In Loughran (2011) wurde mit Wörterlisten der Tonfall von finanzwirtschaftlichen Texten überprüft. Die Autoren haben angemerkt, dass diese Methodik in anderen Disziplinen durchaus anerkannt ist, jedoch speziell für die Überprüfung von finanzwirtschaftlichen Zusammenhängen noch keinen Anklang gefunden hat. Im amerikanischen Raum müssen Unternehmen ein Formular an die Securities and Exchange Commission– kurz SEC –senden.

Das sogenannte Form 10-K ist eine jährliche Hinterlegung von Unternehmensinformationen, welche von börsenotierten Unternehmen auszufüllen und an die SEC zu senden ist. Die Dokumente sind frei verfügbar in Washington, D.C. hinterlegt. (Securities Exchange Act of 1934, 1934)

Die wichtigsten Informationen umfassen eine Erklärung der üblichen betrieblichen Tätigkeit, aber auch, wie das Unternehmen Umsatz sowie Gewinn generiert und in welchen Märkten das Unternehmen operiert. Weiters werden Risken und aktuelle Gerichtsverfahren des Unternehmens offengelegt. Zusätzlich enthalten ist der Konzernabschluss des jeweiligen Unternehmens, sodass Jahresüberschuss, Verbindlichkeiten und weitere Informationen eingesehen werden können. Abgesehen davon werden etwaige Leasingverträge, welche nicht in der Bilanz enthalten sind, extra im Formular aufgelistet. Abgeschlossen wird das Formular mit Unternehmensgrundsätzen (z.B.: *policies*), dem Kommentar vom Vorstand und jenem des Wirtschaftsprüfers.

Aus diesem Grund wollen die Autoren in Loughran (2011) die 10-K Formulare der Jahre 1994 bis 2008 analysieren, welche amerikanische Unternehmen verpflichtend an die SEC übermitteln müssen. Diese Überprüfung findet mithilfe einer Wörterliste, die von ihnen selbst entwickelt wurde, statt. Grund dafür ist, dass dreiviertel der Wörter in einem normalen Text als negativ behaftet bezeichnet werden, welche jedoch in einem finanzwirtschaftlichen Text als nicht negativ behaftet bezeichnet werden können. Sie wollen mit ihrer eigenen Wörterliste in Kombination mit fünf weiteren Wörterlisten besser den Tonfall und Inhalt von finanzwirtschaftlichen Texten analysieren. Die Wörterliste ist speziell für finanzwirtschaftliche Zusammenhänge erstellt worden. Ziel dieser Analyse ist es, die Wörterliste so zusammenzustellen, dass finanzwirtschaftliche Zahlen, wie Ertrag, Handelsvolumen, Volatilität des Ertrags, aber auch wesentliche Schwächen des Unternehmens, aus einem Text analysiert und teilweise prognostiziert werden können.

Im Gegensatz dazu stehen externe Wortlisten, die normalerweise von Autoren für lexikalische Analysen verwendet werden. Ein Beispiel für so eine externe Wortliste ist der Harvard's General Inquirer, welcher in Kategorien, wie etwa positiv, negativ, stark, schwach, aktiv, Freude oder etwa Schmerz eingeteilt wird. Für die Wortliste von Loughran (2011) für finanzwirtschaftliche Texte werden nur ein Teil der Kategorien übernommen, welche negative und positive Wortkategorien beschreiben.

Ein weiteres Problem ergibt sich dadurch, dass viele Wörterlisten urheberrechtlich geschützt sind. Der Harvard's General Inquirer ist jedoch nicht geschützt, dies war ausschlaggebend warum sich die Autoren in Loughran (2011) für die Harvard Liste entschieden haben. Infolgedessen sind die Wörter, welche in dieser Liste enthalten sind, frei zugänglich und demnach konnten Loughran (2011) selbst entscheiden, welche Wörter sie nun in ihre alternative Wörterliste für finanzwirtschaftliche Texte mitaufnehmen.

Die Klassifizierung der Worte in einem analysierten Text sollte Verflechtungen bzw. verschiedene Abwandlungen desselben Wortes erkennen. Analysiert man das englische Wort für Unfall, *accident*, dann sollte dies genauso wie *accidentally,* aber auch *accidents*, als negativ erkannt werden. Die Autoren verwendeten die sogenannte „Harvard's General Inquirer H4N"- Liste, welche sie soweit erweitert haben, sodass jede Abwandlung eines Wortes auf den Wortstamm zurückzuführen ist. Dies führt dazu, dass die Anzahl der ursprünglichen Liste von 2500 Wörter auf 4187 Wörter der neuen Liste angewachsen ist. Ziel der Analyse ist es, die 4187 Wörter nun mit ihrer sogenannten Fin-Neg Liste zu vergleichen.

Weiters inkludieren sie fünf weitere Kategorien in ihre neue Wörterliste, welche positive, Unsicherheit ausdrückende, prozessfreudige, starke und schwache Modalwörter umfassen. Die klassische Methodik um mit empirischen Methoden die einflussreichsten Wörter herauszufiltern, wurde in diesem Fall abgelehnt, denn würden die Manager jene einflussreichen negativen Wörter kennen, würden sie schließlich diese Wörter vermeiden, sodass das Computerprogramm keine negativen Schlüsse aus der Nachricht ziehen kann. Dieser Fehler des Modells wird endogener Bias genannt, da er direkt durch das Modell selbst entsteht. Eine andere Möglichkeit wäre, eine so umfangreiche Wörterliste zu erstellen, die es nahezu unmöglich, beziehungsweise sehr schwierig für Manager macht, diese zu vermeiden. (Loughran, 2011)

In ihrer Arbeit haben Loughran (2011) sich schließlich dazu entschieden, eine Art von Wörterbuch anzulegen, welches in alle oben beschriebenen Kategorien eingeteilt

wird. Die Wörter und die empirischen Daten bezüglich dem Auftreten einzelner Wörter, welche sie für ihre alternative Wörterliste benötigt haben, sind aus den 10-K Formularen aus den Jahren 1994-2008. Wörter, die in mindestens 5% der Dokumente enthalten waren, wurden in Betracht gezogen und bezeichneten sie als essentiell im finanzwirtschaftlichen Sprachgebrauch, weil sie in diesen Texten am meisten frequentiert werden. Diese Wörter inkludierten auch etwaige Wortverflechtungen bzw. Abwandlungen.

Verneinungen werden mit einem der englischen Wörter: *no, not, none, neither, never, nobody* erkannt. Eine Verneinung wird vom Computer erfasst, wenn eines dieser Wörter kurz – bis zu drei Wörter – vor einem positiven Wort auftritt. (Loughran, 2011) nehmen an, dass Negationen von negativen Wörtern auftreten, wie am Beispiel *„not terrible earnings",* ersichtlich ist. Aus diesem Grund wird dieser Fall nicht weiter behandelt.

In dem Wörterbuch von (Loughran, 2011) wurde auch eine eigene Wörterliste für Finanzbegriffe erstellt, da im finanzwirtschaftlichen Bereich Vokabular angewendet wird, welches im normalen Sprachgebrauch einen positiven oder negativen Tonfall miteinschließt, jedoch aber im finanzwirtschaftlichen und buchhaltungstechnischen Umfeld neutral angewandt wird. Als Beispiel für diesen Sachverhalt wurde Anstieg (engl. *increase*) beziehungsweise Reduzierung (engl. *decrease*) verwendet. Nimmt man an, dass Kosten ansteigen, dann hat das einen deutlich negativen Beigeschmack. Im Gegensatz dazu steht der Anstieg von Gewinnen, welcher meistens eine positive Nachricht übermittelt. Aus diesem Grund ist es, wie bereits oben erwähnt, entscheidend in welchem Zusammenhang Wörter auftreten. Folglich werden die Wörter im Umfeld gescreent und aus dem Konsortium von Wörtern, welche in der unmittelbaren Nähe auftreten, der richtige Inhalt erfasst. Aus der Wortkombination errechnet der Computer dann den Tonfall der Nachricht und, ob die Nachricht positiv oder negativ zu werten ist.

Für die richtige Deutung der Sätze einer finanzwirtschaftlichen Analyse von Texten ist entscheidend, dass eine Wörterliste enthalten ist, welche Begriffe aus dem

eigenen Umfeld richtig interpretieren kann. Daher ist es auch nicht möglich, eine herkömmliche Wörterliste für die Analyse von finanzwirtschaftlichen Texten zu verwenden, da durch die Missdeutung von finanztechnischen Begriffen ein nicht zu vernachlässigbarer Bias entsteht.

Es ist notwendig, eine Anpassung der Wörterlisten an die Texte herzustellen, sodass die Fehldeutung erfolgreich vermieden und die Power der Methodik erhöht werden kann. Autoren wissen, dass mithilfe von finanzwirtschaftlichen Texten, welche das Unternehmen betreffen, eine Bewertung des entsprechenden Unternehmens durchgeführt werden kann und aus diesem Grund passen die Autoren ihren Sprachgebrauch auch dementsprechend an. Wörter, welche einen negativen Tonfall repräsentieren, werden vermieden. (Loughran, 2011)

Die Auswahl der Wörter, welche von Loughran (2011) in die alternative Wörterliste für finanzwirtschaftliche Texte aufgenommen wurde, wurde nach Auftreten in der Stichprobe von 50.115 10-K Formularen ausgewählt, welche etwa 2,5 Milliarden Wörter enthält.

Die Autoren haben ihre Ergebnisse aus der sehr umfangreichen Analyse nach Industrien aufgeteilt und beobachteten wie sich die Frequenzen in der jeweiligen Industrie verteilten. Manche Wörter sind sehr spezifisch für eine bestimmte Industrie und oft ist es so, dass solch ein neutrales Wort, welches im fachspezifischen Hintergrund einen neutralen Tonfall bedeutet, im normalen Sprachgebrauch negativ behaftet ist. Ein Beispiel dafür sind buchhaltungstechnische Begriffe, wie die englischen Wörter *tax, cost, loss, capital* und *expenses.* Im normalen Kontext vermitteln diese Wörter einen eindeutig negativen bzw. positiven Unterton, in einem Konzernabschluss ist es jedoch ein neutrales Vokabular um die operativen Aktivitäten in einem Unternehmen zu beschreiben. Die oben erwähnten Vokabeln können hier negativ aber auch positiv behaftet sein.

Grundsätzlich ist zu sagen, dass in einer literarischen Analyse eine höhere Frequenz von negativ behafteten Wörtern ein Indiz für eine negative Nachricht ist. Aus diesem

Grund werden Wörter, welche bei einer Analyse für normale Texte einen negativen Tonfall repräsentieren würden, wie *loss, losses, impairment, against* und *adverse* aus der Textanalyse entfernt, sodass keine Missinterpretation des Textes durch die, für finanzwirtschaftliche Texte neutralen, Wörter entsteht.

Nach der Wortanalyse der 10-K Formulare haben die Autoren festgestellt, dass aus der negativen Wortliste nur etwa 25 Wörter übriggeblieben sind, was zu bedeuten hat, dass fast 50% der Wörter dieser Wortliste in normalen Texten negativ behaftet sind, aber eben in einem finanzwirtschaftlichen Kontext einen neutralen Tonfall innehaben. Dieser Effekt verstärkt sich noch bei der Untersuchung der 25 Wörter, welche am meisten verwendet werden. In diesem Fall werden 73,8% der Worte, welche eigentlich einen negativen Sachverhalt beschreiben, aber in finanzwirtschaftlichen Berichten nicht negativ gedeutet werden sollten, falsch interpretiert.

Weiters konnten die Autoren feststellen, dass die Wörterverteilung dem Zip'schen Gesetz unterliegt beziehungsweise folgt. Dieser Sachverhalt wurde als erstes von Baayen (2001) festgestellt. Das Gesetz besagt, dass in der Wortverteilung grundsätzlich wenige Wörter existieren, die sehr oft verwendet werden und eine sehr große Anzahl an Wörtern, welche nur selten auftreten. Die Schlussfolgerung daraus ist, dass die Gewichtung der Wörter entscheidend ist. Wenn beispielsweise ein Wort zehnmal sooft auftritt wie ein anderes, dann enthält es höchstwahrscheinlich nicht zehnmal so viel Information. Mithilfe einer Gewichtung nach der Frequentierung der Wörter, kann dieser Effekt verkleinert werden.

Auffallend ist außerdem, dass speziell jene finanzwirtschaftlichen Wörter, welche im finanzwirtschaftlichen Kontext eine andere Bedeutung haben als in einem normalen Text, weitaus häufiger in 10-K Formularen auftreten. Weiters ist anzumerken, dass es für jede Industrie Fachbegriffe gibt, die im fachspezifischen Kontext eine andere Bedeutung haben. Aus diesem Grund empfehlen die Autoren diese Wörter für die jeweiligen Unternehmen herauszufiltern, um den Effekt zu minimieren.

Somit ist die oben behandelte Adaptierung der Wörterlisten eine sehr wichtige Methodik zur richtigen Analyse der vorliegenden Texte.

Im nächsten Schritt der Studie war nun zu überprüfen, ob die alternativen Wörterlisten zur Analyse von 10-K Formularen geeignet sind. Es ist anzunehmen, dass in Fällen in welchen die 10-K Formulare eines Unternehmens eher negativ formuliert werden und eher einen negativen Tonfall besitzen, das Unternehmen einen negativen *excess return* haben wird. Der *excess return* ist die Differenz zwischen Rendite des Unternehmens und dem Mittelwert der Rendite aller Marktteilnehmer. Genauer gesagt, ist der *excess return* der "common stock buy-and-hold return" von welchem der "CRSP value-weighted market index buy-and-hold return" subtrahiert wird (Loughran, 2011). Ist der *excess return* eines Unternehmens negativ, dann performt das Unternehmen schlechter als die übrigen Marktteilnehmer.

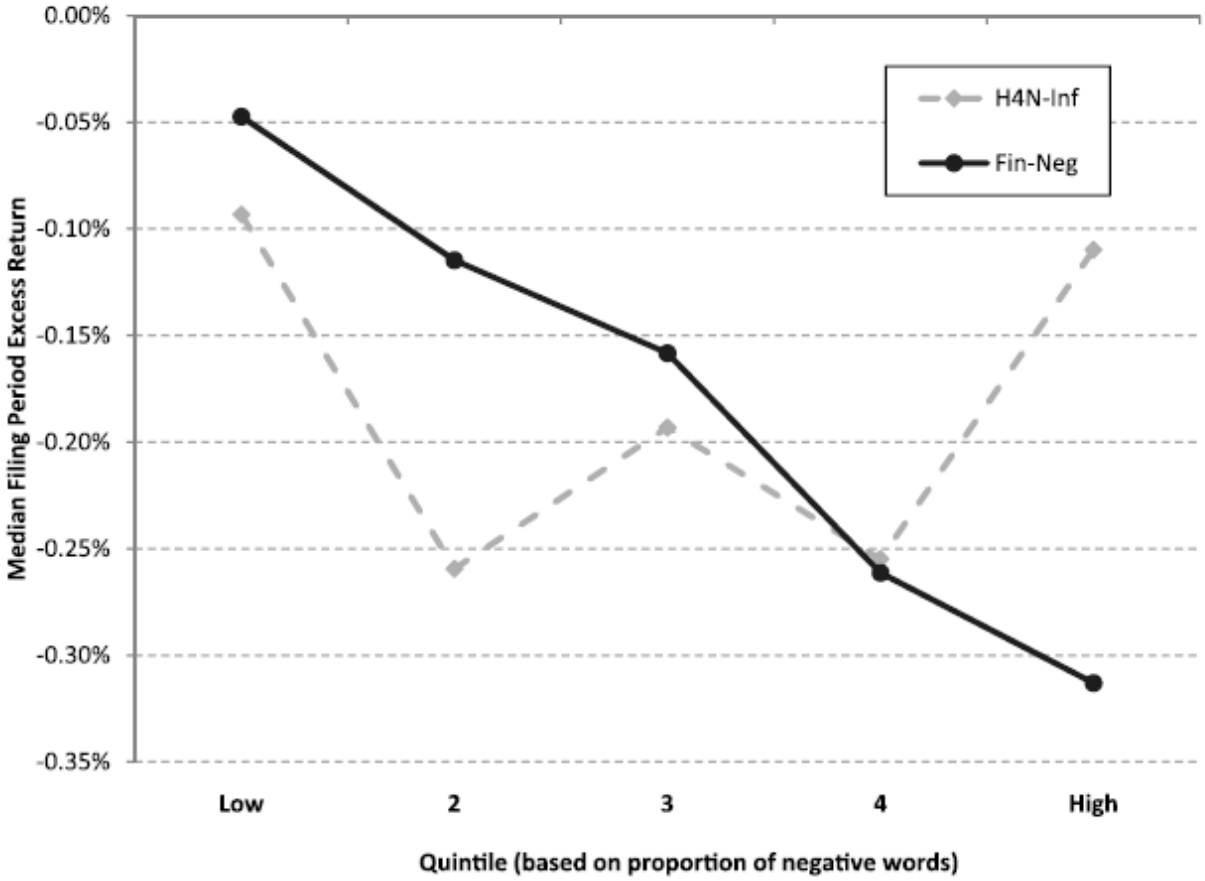


Abbildung 3 Quelle: Loughran, Tim, and Bill McDonald. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." The Journal of Finance 66.1, 2011: 51

Abbildung 3 zeigt den Median der *excess returns* der inkludierten 10-K Formulare nach Quintilen für die Wörterlisten H4N-Inf, welche die Wörterliste mit Verflechtungen darstellt, und Fin-Neg, welche die Wörterliste speziell negativer Wörter im finanzwirtschaftlichen Kontext umfasst. Zur Erklärung: das erste Quintil umfasst Werte des ersten Fünftels, also zwischen 0-20%, das zweite Quintil beschreibt das zweite Fünftel, also Werte zwischen 20-40% und so weiter. Genauer gesagt, wurde jede Wörterliste des Stichprobenumfangs, welche 50.115 10-K Formulare umfasst, in fünf Abschnitte geteilt. Die Unterteilung wurde anhand des Anteils an negativen Worten im gesamten Text bestimmt, das heißt im gesamten 10-K Formular. Die Periode, für welche der *excess return* berechnet wurde, begann am

Tag der Veröffentlichung des 10-K Formulars und umfasste die drei darauffolgenden Handelstage.

Der Median der *returns* für die H4N-Inf Wörterliste weist keinen Zusammenhang mit dem Anteil negativer Wörter auf. Unternehmen mit einem hohen Anteil an negativen Wörtern der Harvard Wörterliste, haben nur minimal höhere *returns* im Beobachtungszeitraum verglichen mit Unternehmen, welche nur einen kleinen Anteil an negativen Wörtern in ihrer Veröffentlichung aufweisen.

Im Gegensatz steht die sogenannte Fin-Neg Wörterliste, welche genau das gefragte Muster erfüllt und somit die richtigen Informationen der Formulare erkennt. Erstens ist für diese Wörterliste zu beobachten, dass Unternehmen, die einen niedrigeren Anteil an pessimisstischer Ausdrucksweise verwenden, auch einen leicht negativen *excess return* in der beobachteten Periode aufweisen. Zweitens ist auffallend, dass Unternehmen mit einem höheren Anteil an negativen Wörtern einen etwas niedrigeren Median *excess return* aufweisen, als jene mit einem niedrigen Anteil. Kurz gesagt, weist die Funktion der *returns* für die Fin-Neg Wörterliste einen monotonen Verlauf auf. Dieses Ergebnis war von (Loughran, 2011) erwartet worden und bedeutet, dass ein negativer Zusammenhang zwischen dem Anteil an negativer Formulierung in der Fin-Neg Wörterliste und den *excess returns* der Beobachtungsperiode, welche sich vom Tag der Veröffentlichung bis drei Handelstage danach erstreckt, besteht.

Als nächstes wollten die Autoren den Zusammenhang zwischen der Rendite, welche in den 10-K Formularen ausgewiesen wurde und einem negativen Tonfall mit einer multivariaten Regression überprüfen. Abermals wurde als Indiz für die Rendite der sogenannte *excess return* der beobachteten Periode gewählt, welcher gleichzeitig auch die abhängige Variable in der Regression darstellt. Die Kontrollvariablen wurden mit einigen finanzwirtschaftlichen Kennzahlen, wie *size, book-to-market und share turnover* definiert. Abermals konnten sie feststellen, dass kein signifikanter Zusammenhang zwischen der H4N-Inf Wörterliste, aber ein negativer Zusammenhang mit der Fin-Neg Wörterliste besteht. Es kann also gesagt werden,

dass ein höherer Anteil an negativen Wörtern, welche mit der Fin-Neg Wörterliste identifiziert werden, stark mit einem niedrigeren *excess return* in derselben Periode korreliert ist.

Obwohl die Wörterlisten mit deren Auftreten in den 10-K Formularen gewichtet worden sind, ist anzumerken, dass die *power* der Regression niedrig ist. Dieses Ergebnis wird durch ein niedriges R^2mit etwa 2,6% unterstrichen. Es wird also nur ein kleiner Teil der tatsächlichen Variation der Beobachtungsperiode mit den unabhängigen Variablen beschrieben.

In einem weiteren Test überprüfen die Autoren, ob der Management Discussion & Analysis (kurz: MD&A) Abschnitt allein ein besserer den Tonfall eines 10-K Dokuments wiederspiegelt. Der MD&A Abschnitt enthält zusätzliche Informationen zu finanziellen Kennzahlen, wie Liquidität, Kapital oder Betriebsergebnis. Weiters wird ein Ausblick des Management Boards enthalten sein, welcher die zukünftige Entwicklung, sowie Risiken und Chancen beschreibt.

Nun löst diese Fragestellung jedoch nicht das Problem der zu geringen power des Tests, ganz im Gegenteil, es verschlimmert es nur noch weiter, da die Stichprobengröße weiter verkleinert wird, weshalb das Ergebnis als nicht aussagekräftig gewertet werden kann.

Zusammenfassend ist zu sagen, dass diese Methodik nicht die herkömmliche Analyse eines Konzernabschlusses ersetzen, jedoch kann sie ein wichtiges Hilfsmittel sein, die Konzernabschlüsse und die Aktienkurse der Unternehmen besser verstehen zu können. Besonders bei der Verwendung von Wörterlisten im Zuge von literarischen Analysen ist auf die Art des Textes zu achten, weil eben wie schon vorher erwähnt im finanzwirtschaftlichen Kontext manche Begriffe eine andere Bedeutung haben und somit der Inhalt und der Sinn des Textes falsch interpretiert werden könnten. Dieser Fakt spielte auch bei der Auswahl der Methodik, die in dieser Arbeit getroffen wurde, eine große Rolle. Insbesondere da die Analyse möglichst simpel und einfach nachvollziehbar gestaltet werden, sowie eine standardisierte Testmethode

verwendet werden sollte, wurden Methoden, welche Wörterlisten miteinbeziehen, ausgeschlossen.

Diese Publikation wurde für diese Analyse gewählt, weil die Vorgehensweise sehr detailliert und der Stichprobenumfang an getesteten Unternehmen sehr umfangreicht ist. Obwohl für diese Analyse im Zuge der Arbeit keine spezifizierte Wörterliste in Betracht gezogen wurde, stimmt grundsätzlich die Vorgangsweise mit der Loughran (2011) Studie überein, denn die Autoren analysierten Veröffentlichungen von Unternehmen mit einer computerbasierten Methodik.

Zudem wurde somit bestätigt, dass ein Zusammenhang mit der Wortwahl in einem Konzernabschluss und dem Ergebnis des Unternehmens besteht. Auch wenn das Ergebnis mit *excess return* definiert wurde und die literarische Analyse etwas anders durchgeführt wurde, stehen die beiden Analysen in direktem Zusammenhang.

5.2. Li, Feng. "Annual report readability, current earnings, and earnings persistence."

Das nächste Paper, welches hier näher behandelt wird, ist von Li (2008) und es untersucht abermals die 10-K Formulare, welche von amerikanischen Unternehmen an die Securities and Exchange Commission gesendet werden müssen und in Folge veröffentlicht werden. Im Gegensatz zum vorherigen Paper untersucht Li (2008) die Lesbarkeit der 10-K Formulare und überprüft, ob ein Zusammenhang mit dem Jahresergebnis der Unternehmen besteht und der wirtschaftliche Erfolg langfristig ist.

Der Autor nahm an, wenn die wirtschaftliche Leistung eines Unternehmens schlecht ist, die Manager mehr Gründe haben, Informationen zu verstecken bzw. zu verschleiern. Folglich gilt auch der Umkehrschluss, dass Unternehmen, welche eine gute wirtschaftliche Leistung in der Beobachtungsperiode erreicht haben, ihre Konzernabschlüsse besser verständlich formulieren. Zusätzlich wurde vom Autor noch überprüft, ob der Erfolg eines Unternehmens langfristig ist und welche Auswirkung dies auf die Formulierung hat.

Die Lesbarkeit wurde mithilfe von zwei Variablen getestet. Als erstes verwendete der Autor den FOG-Index, welcher auch in dieser Analyse verwendet wird, um die Komplexität der Sätze eines Textes zu berechnen. Die zweite Variable misst die Länge der Konzernabschlüsse. Der Grund für die Berücksichtigung dieser Variable ist, dass längere Texte abschreckend für den Leser sind und mehr Zeit aufgewendet werden muss um die Informationen aus dem Text zu extrahieren. Nachdem ein längerer Konzernabschluss mehr Zeit für die Analyse benötigt, sind die Kosten für eine Unternehmensanalyse für potentielle Investoren, höher.

Hinzukommend besagt die sogenannte „incomplete revelation hypothesis" nach Bloomfield (2002), wenn es teurer ist, Informationen zu verarbeiten, dann wird durch den Marktpreis dieser Informationsgehalt wahrscheinlich weniger wiedergespiegelt. Wenn diese Informationen vergleichsweise schlechte Nachrichten

enthalten, dann beeinflusst die schlechte Nachricht die Marktpreise weniger, wenn sie so im Text verpackt wird, dass es mit höheren Ressourcenaufwand verbunden ist, diesen Informationsgehalt aus dem Text zu extrahieren. Diese Hypothese bewegt Manager dazu Informationen zu verstecken, welche den Wert des Unternehmens nicht steigern, sodass sie weniger Einfluss auf den Marktpreis haben. Manager können entweder durch besonders lange Texte, oder durch Texte die eine hohe Komplexität aufweisen, die schlechten Nachrichten verstecken.

Die „managerial obfuscation hypothesis“ besagt, dass ein negativer Zusammenhang zwischen wirtschaftlichen Erfolg eines Unternehmens und dem Level an Komplexität des Konzernabschlusses besteht. An dieser Hypothese gibt es nach (Li, 2008) jedoch zahlreiche Zweifel. Erstens könnte es sein, dass der Effekt statistisch nicht signifikant ist. Zweitens enthalten Konzernabschlüsse nicht ausschließlich Informationen aus der aktuellen Periode, sondern auch aus Vorperioden. Der Vorteil für Manager, welcher durch die Verfassung von komplexen Konzernabschlüssen entsteht, um schlechte Nachrichten der aktuellen Periode zu verstecken, könnte daher eher gering ausfallen.

Letztendlich ist auch noch zu sagen, dass es strategisch vorteilhaft sein kann, dass ein Unternehmen ein gutes Jahresergebnis veröffentlicht, obwohl das Geschäftsjahr schlecht ausgefallen ist. In diesem Fall kann es sein, dass Manager die Konzernabschlüsse absichtlich komplexer formulieren, sodass ein schlechtes operatives Ergebnis, weil etwa der Umsatz rückläufig war, hinter dem guten Jahresergebnis, welches etwa nur durch Einmaleffekte, wie zum Beispiel den Verkauf von Beteiligungen entsteht, versteckt werden kann. Der Zusammenhang zwischen der Komplexität des Konzernabschlusses und dem Jahresergebnis ist problematisch, da die Sachverhalte, welche im Konzernabschluss enthalten sind sowohl aus der aktuellen Periode stammen, sowie auch aus Vorperioden. In manchen Teilen eines Konzernabschlusses sind Prognosen enthalten, welche folglich auch nicht dem Beobachtungszeitraum entsprechen. Aus diesem Grund verwendet der Autor die sogenannte *earnings persistence*, welche besser die tatsächliche wirtschaftliche Leistung des Unternehmens in einem längerfristigen Zeithorizont wiederspiegelt.

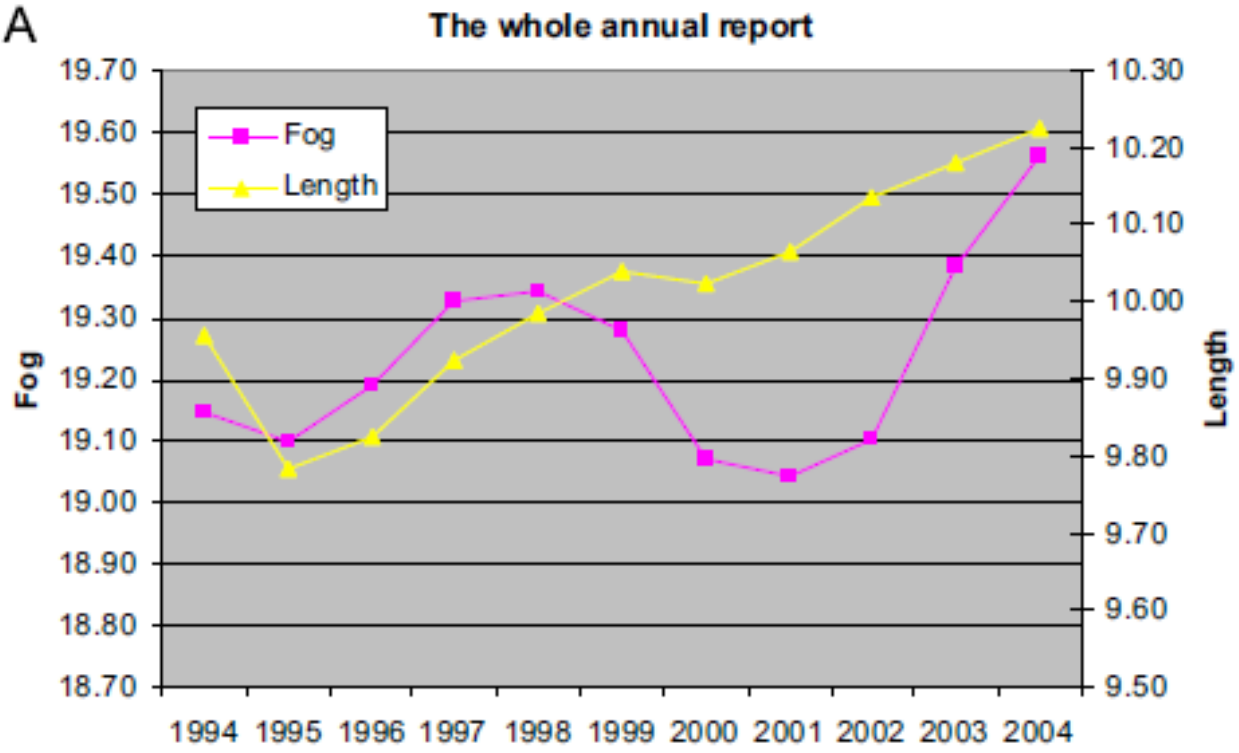
Ein interessanter Punkt in diesem Zusammenhang ist die zukünftige wirtschaftliche Leistung eines Unternehmens und die daraus resultierende strategische Beeinflussung auf die Formulierung der Konzernabschlüsse. Für den Manager ist es jedoch besser, ein gutes Jahresergebnis zu veröffentlichen, da das Unternehmen attraktiver für Investoren erscheint und auch etwaige Bonuszahlungen von den Managern positiv beeinflusst werden.

Andererseits sind Unternehmen, die gute zukünftige Aussichten haben, bereit, mehr Informationen zu veröffentlichen, um diese positiven und ertragreichen Prognosen zu untermauern, sodass sie sich bewusst von Unternehmen mit einer schlechteren *performance* abheben können. Dies führt jedoch dazu, dass die Konzernabschlüsse umfangreicher werden und die Analyse von den Konzernabschlüssen für Investoren teurer wird.

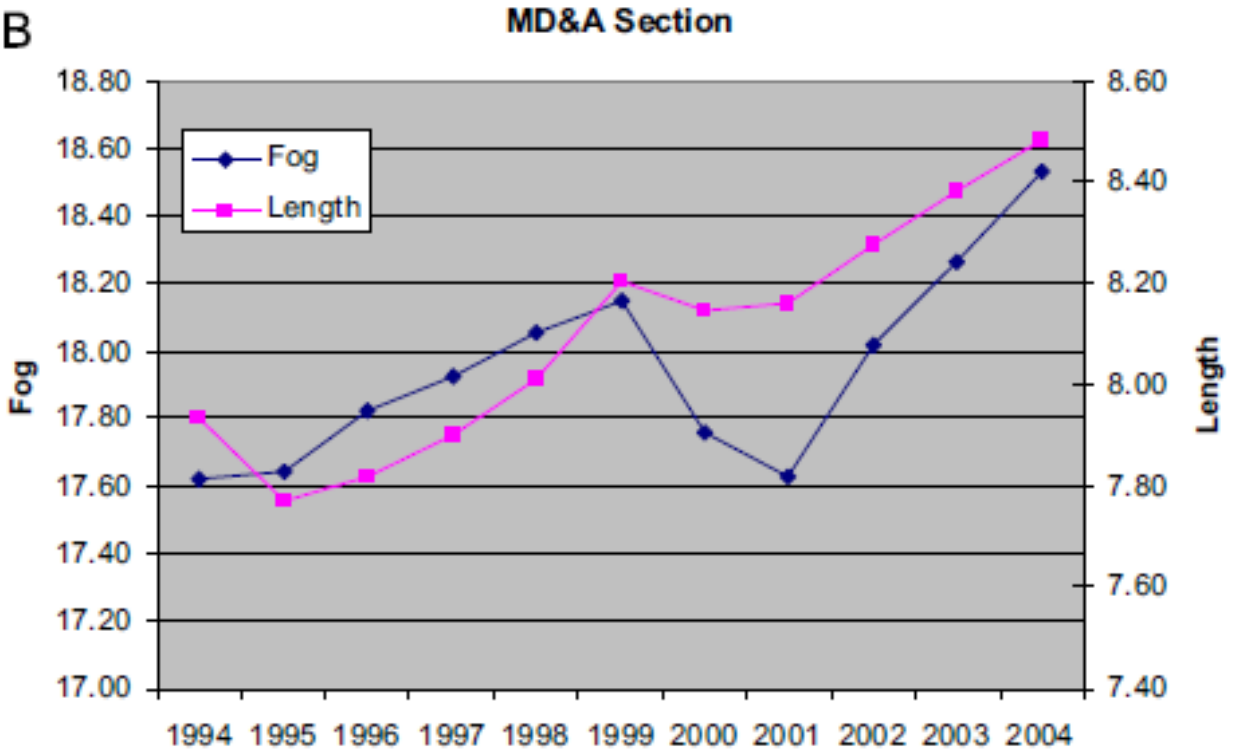
Die meisten vorherigen Studien behandeln die sogenannte *disclosure quality* als Vorhersagefaktor für die wirtschaftliche Leistung eines Unternehmens. Das Problem mit *disclosure quality* ist jedoch, dass dieser Wert für jedes Unternehmen spezifisch bestimmt werden muss und nicht einfach abrufbar ist. Daher ist es nahezu unmöglich einen größeren Stichprobenumfang zu erreichen. Aus diesem Grund wird für die Studie von Li (2008) dieser Faktor nicht inkludiert, stattdessen verwendet der Autor *die earnings persistence*, welche die langfristige wirtschaftliche Leistungsfähigkeit des Unternehmens abbilden soll. Folglich kann der Autor auch eine größere Stichprobengröße bereitstellen. Abermals werden die 10-K Formulare aus den Geschäftsjahren 1993 bis 2004 analysiert. Zusammenfassend umfasst der Stichproblemumfang der Studie somit 55.719 10-K Formulare.

Interessant an der Studie von Li (2008) ist auch, dass die 10-K Formulare zerlegt werden und die Notes sowie der MD&A Abschnitt einzeln analysiert werden. Anschließend werden die Teile mit dem ganzen Konzernabschluss verglichen.

A


The whole annual report
Fog
Length
19.70
19.60
19.50
19.40
19.30
19.20
19.10
19.00
18.90
18.80
18.70
10.30
10.20
10.10
10.00
9.90
9.80
9.70
9.60
9.50
1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004

B


MD&A Section
Fog
Length
18.80
18.60
18.40
18.20
18.00
17.80
17.60
17.40
17.20
17.00
8.60
8.40
8.20
8.00
7.80
7.60
7.40
1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004

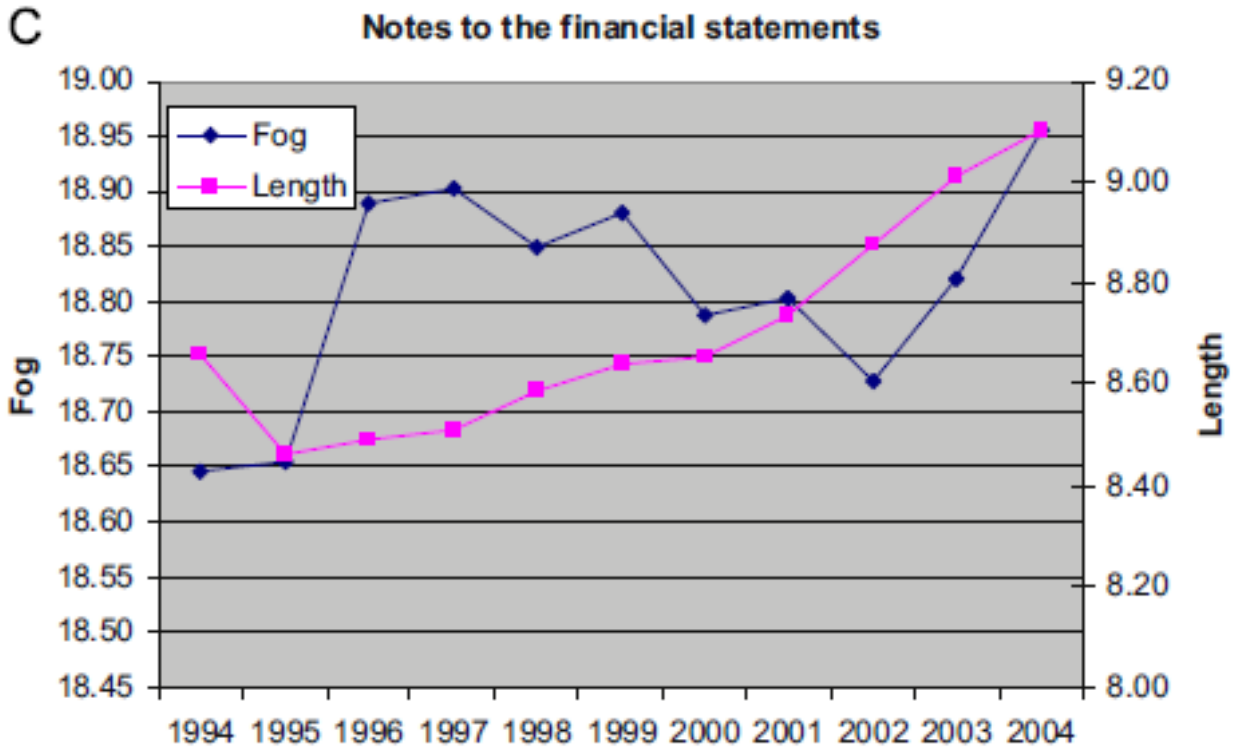


Abbildung 4 Die Diagramme zeigen den Median FOG-Index und die Länge der Konzernabschlüsse nach Kalenderjahren. In A werden die Ergebnisse der vollständigen Konzernabschlüsse abgebildet, in B nur der MD&A Abschnitt und in C die Notes der Konzernabschlüsse.

Quelle: Li, Feng. "Annual report readability, current earnings, and earnings persistence." Journal of Accounting and economics 45.2, 2008: 221-247.

In den Diagrammen ist zu erkennen, dass der gesamte Konzernabschluss eine höhere Komplexität aufweist, als nur der MD&A-Abschnitt. Der MD&A-Abschnitt ist daher für einen Leser einfacher zu verstehen, als der gesamte Konzernabschluss. Die Notes sind etwas leichter zu lesen als der gesamte Konzernabschluss. Interessanterweise ist ein Abfall der Komplexität ab dem Jahr 1999 deutlich zu erkennen. Dies steht wahrscheinlich stark im Zusammenhang mit den sogenannten „plain english" Richtlinien, welche 1998 verabschiedet wurden und die Unternehmen zu einer einfacheren Formulierung in den Konzernabschlüssen bewegen soll. Die Richtlinien zeigen vor allem in den ersten zwei Jahren eine deutliche Reduzierung der Komplexität der Sprache in den Konzernabschlüssen. Leider ist dieser Trend nicht von Langfristigkeit geprägt, da in den folgenden Jahren, ab 2000, die Komplexität

sehr stark anstieg und das Niveau von vor der Veröffentlichung der Richtlinien überstieg. Die Länge der Konzernabschlüsse wurde von den Richtlinien nicht beeinflusst und stieg über alle Jahre des Beobachtungszeitraums stetig an.

Weiters ist auffallend, dass eine signifikante positive Korrelation (Pearson Korrelation Koeffizient von 0,377) zwischen FOG-Index und Länge der Konzernabschlüsse besteht. Es kann daher festgestellt werden, dass mit der Länge des Konzernabschlusses die Komplexität der Sprache in einem Konzernabschluss steigt. Dieser Effekt gilt nicht nur für den gesamten Konzernabschluss, sondern auch für den MD&A Abschnitt sowie die Notes.

Grundsätzlich ist noch anzumerken, dass größere Unternehmen tendenziell längere Konzernabschlüsse aufweisen. In diesem Fall wurde die Größe eines Unternehmens mit der Marktkapitalisierung am Ende des fiskalen Jahres bemessen. Der Autor hat herausgefunden, dass ein negativer Zusammenhang zwischen der wirtschaftlichen Leistung und der Komplexität der Konzernabschlüsse besteht. Diese Beobachtung ist die Grundlage für meine These und ich möchte diesen Sachverhalt am Prime Market der Wiener Börse mit einer etwas anderen Methodik überprüfen.

5.3. Wayne Guay, Delphine Samuels, Daniel Taylor "Guiding through the Fog: Financial statement complexity and voluntary disclosures"

Guay (2016) behandelt den Zusammenhang zwischen der freiwilligen Selbstauskunft von unternehmensinternen Informationen und der Komplexität von Konzernabschlüssen.

Wie bereits in den vorherigen Studien, werden die 10-K Formulare jedes Unternehmens zur Analyse herangezogen. Die Variable, welche die freiwillige Selbstauskunft beschreibt, ist die Anzahl an management forecasts des Unternehmens nach der 10-K Formular Veröffentlichung.

Die Autoren haben drei verschiedene Tests durchgeführt um die Zusammenhänge zu erforschen. Als erstes wurde eine Querschnittstudie für jedes Unternehmen erstellt. Dabei wurde herausgefunden, dass ein positiver Zusammenhang zwischen der Komplexität der 10-K Formulare und der freiwilligen Selbstauskunft in der Zeitperiode ein bis 12 Monate nach der Veröffentlichung der 10-K Formulare besteht. Innerhalb der Unternehmen steigt die freiwillige Selbstauskunft somit mit der Komplexität der Konzernabschlüsse.

Zweitens konnten Guay (2016) feststellen, dass der Zusammenhang zwischen der freiwilligen Selbstauskunft und der Komplexität der Konzernabschlüsse sinkt, wenn die Liquidität des Unternehmens sinkt.

In ihrer Analyse haben die Autoren eine Hauptkomponentenanalyse mit einem selbsterstellten Lesbarkeitsindex und sechs anderen Lesbarkeitsindizes (LIX, RIX, Flesch Reading Ease, Fog, ARI, SMOG) durchgeführt. Der selbsterstelle Lesbarkeitsindex ist aus den sechs anderen entstanden. Dieses Paper bestätigt die Ergebnisse der vorherigen Studien und, wie auch in meiner empirischen Analyse, werden die Lesbarkeitsindizes Fog, Flesch Reading Ease sowie SMOG-Index

angewendet. Diese Lesbarkeitsindizes sind ein Indiz zur Analyse von finanzwirtschaftlichen Texten.

5.4. Tetlock, Paul C. "Giving content to investor sentiment: The role of media in the stock market."

Tetlock P. C. (2007) hat den Inhalt der Kolumne "Abreast of the Market" des Wall Street Journals verwendet um einen Zusammenhang mit Aktienrendite und Handelsvolumen aus dem Inhalt herauszulesen und gleichzeitig die Werte für den folgenden Tag zu prognostizieren. Tetlock P. C. (2007) fand heraus, wenn ein Text mit einem starken pessimistischen Wert publiziert wurde, so folgte ein Tag an dem allgemein weniger Rendite erzielt wird. Auch in dieser Publikation wird der Harvard's General Inquirer als Basis für die Analyse benutzt. In diesem Fall werden 77 Kategorien direkt übernommen, also wieder eine leicht angepasste Wortliste.

Bei einer weitergehenden Analyse, welche sich speziell mit negativ und positiv behafteten Worten befasst, stellt Tetlock P. C. (2007) jedoch fest, dass ein starker Zusammenhang von Pessimismus sowohl mit negativen jedoch auch mit positiven Worten besteht. Zu diesem Sachverhalt sagt Tetlock P. C. (2007), dass negative Wörter ein Indiz für qualitative Informationen sind. Negativ behaftete Wörter dienen dazu eine positive Nachricht etwas abzuschwächen. Daraus folgte die Studie von Tetlock P. C. (2008), welcher sich ausschließlich mit negativ behafteten Wörtern der Harvard Wörterliste befasst hatte.

Wörterlisten gehen zurück bis in die 1950er Jahre, als Computer gerade anfingen für eine größere Anzahl an Menschen verfügbar zu werden. (Stone, 1966) Es existieren eigene Methoden für verschiedene wissenschaftliche Disziplinen, wie zum Beispiel Psychologie, Anthropologie, Linguistik, Politikwissenschaft, Journalismus sowie Computerwissenschaft. Mit der Weiterentwicklung des Computers wurden auch die Möglichkeiten zur literarischen Analyse vergrößert.

6. Resultate

6.1. Darstellung der Resultate

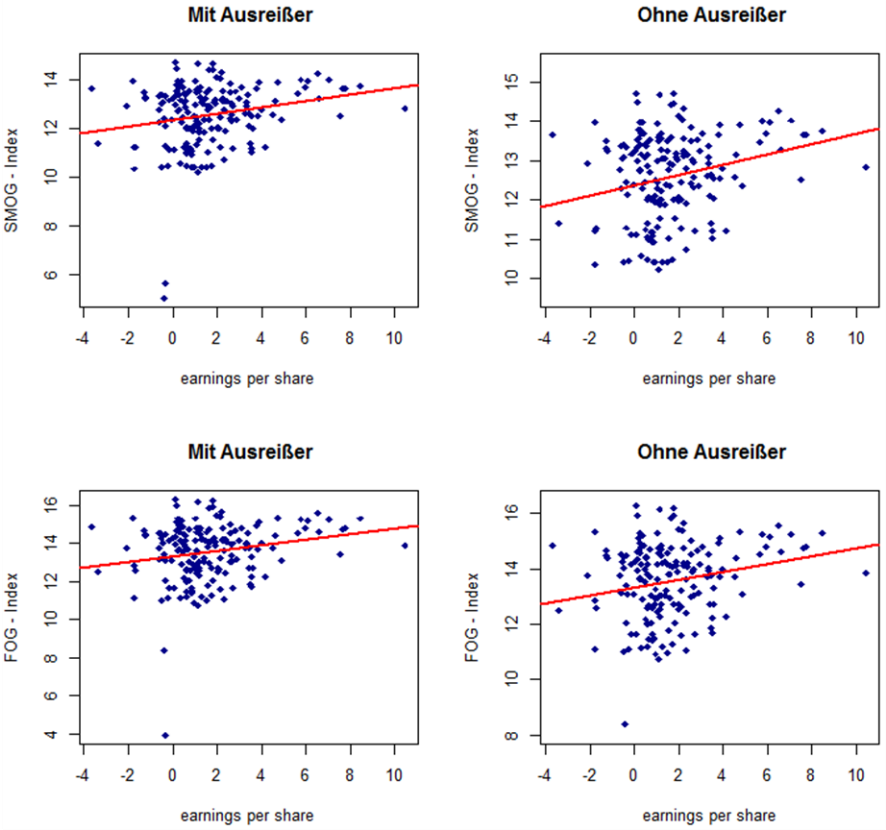


Abbildung 5 Vergleich der Ergebnisse von SMOG- und FOG Index mit und ohne Ausreißer

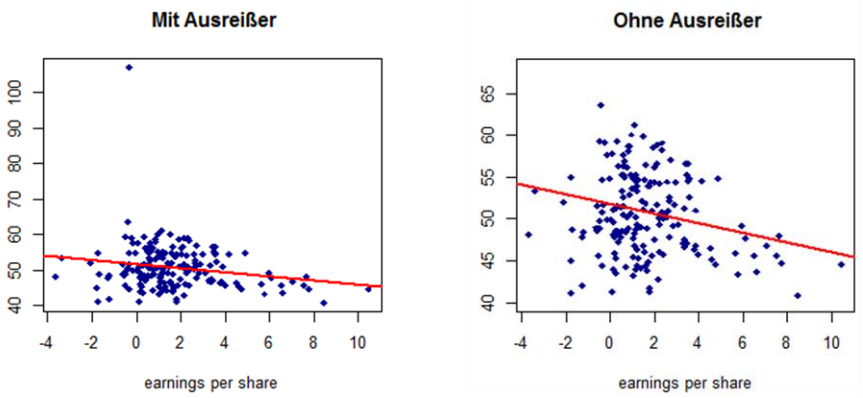


Abbildung 6 Vergleich der Ergebnisse der Flesch Reading Ease Formula mit und ohne Ausreißer

Wie auf dem ersten Blick auf den Scatterplot ersichtlich ist, gibt es bei jedem Lesbarkeitsindex einige Ausreißer, welche die Grafiken merklich beeinflussen. Für die statische Analyse sind diese Ausreißer zu vernachlässigen. Bei genauerer Durchsicht der Daten der einzelnen Unternehmen ist nämlich ersichtlich, dass für alle Ausreißer derselbe Abschluss verantwortlich ist. Es handelt sich um den Konzernabschluss von der Wolford AG für das Jahr 2013. Laut Presseaussendung von der Wolford AG, lief das Geschäftsjahr jedoch stabil auf Vorjahresniveau. *Quelle:*https://www.ots.at/presseaussendung/OTS_20140718_OTS0089/wolford-ag-umsatz-und-ergebnisse-des-geschaeftsjahres-201314-im-rahmen-der-erwartungen-bild

Aus diesem Grund kann keine wirtschaftliche Erklärung für den Ausreißer gefunden werden. Es ist möglich, dass es in diesem Abschluss eine spezielle Formatierung gewählt wurde, welche die technische Analyse beeinflusst haben könnte. Es wurde aber kein offensichtlicher Fehler in der Analyse entdeckt. Für die weitere Analyse ist der einzelne Ausreißer nicht von großer Bedeutung und kann daher vernachlässigt werden.

Aus optischen Gründen wurden die Grenzen der Grafik verkleinert, sodass die Ausreißer nicht mehr abgebildet werden. Dies führt dazu, dass die Mehrzahl der Ergebnisse und deren Verteilung besser ersichtlich ist, wie im Vergleich der Ergebnisse in *Abbildung* 5 Vergleich der Ergebnisse von SMOG- und FOG Index mit und ohne Ausreißer und Abbildung 6 Vergleich der Ergebnisse der Flesch Reading Ease Formula mit und ohne Ausreißer zu sehen ist. Jeder Punkt in der Grafik steht für einen Lesbarkeitsindex eines Unternehmens aus der Periode von 2012 bis 2016. In der Abbildung werden die Ergebnisse mit und ohne Ausreißer gegenübergestellt.

Bei Durchsicht der Ergebnisse ist anzunehmen, dass für die SMOG und FOG ein positiver Zusammenhang mit den *Earnings per Share* der Unternehmen besteht. Das bedeutet, dass mit der Komplexität der Sprache die EPS eines Unternehmens steigen.

Für die Flesch-Ease Formula ist ein negativer Zusammenhang anzunehmen, da bei der Flesch-Ease Formula ein höherer Wert eine leichtere Lesbarkeit suggestiert, bestätigen diese drei Resultate bestätigen einander und liefern eine konsistente Aussage. Dieser Sachverhalt lässt sich aus der Steigung der Trendline welche in den Grafiken rot dargestellt ist, ableiten. Diese Hypothese wird im folgenden Schritt mithilfe eines linearen Regressionsmodells überprüft. Genauer gesagt wird angenommen, dass die *Earnings per Share* jedes Unternehmens mithilfe von den Lesbarkeitsindizes vorherzusagen sind. Die Lesbarkeitsindizes sollen einen Prädiktor für die *Earnings per Share* für jedes Unternehmen darstellen. Die Hypothese, welche im folgenden Schritt überprüft werden soll, besagt, dass der Lesbarkeitsindex die *Earnings per Share* des Unternehmens für die angegebene Periode gut prognostizieren kann und eine hohe Korrelation zwischen den EPS eines Unternehmens und der Ergebnisse der Lesbarkeitsindizes besteht.

Das lineare Modell, welches für die Analyse verwendet wurde, weist folgende Formel auf:

$$Y = X\,\beta + u\,,$$

wobei *Y* die endogene Variable, *Earnings per Share* und *X* die exogene Variable, der Lesbarkeitsindex, darstellt. Der Fehler, auch Residualwert genannt, wird mit *u* beschrieben und umfasst jene Werte, welche nicht durch *X* erklärt werden.

Auf der folgenden Seite sind mehre Outputs aus dem Statistik Programm R ersichtlich, welche die soeben beschriebene Hypothese überprüfen. Der erste Output überprüft, ob ein signifikanter Zusammenhang zwischen Earnings per Share und dem FOG-Index besteht. Die Teststatistik weist einen p-Wert von 0,0081 auf. Das bedeutet, dass der Zusammenhang der beiden Variablen auf einem 99,9% Signifikanzniveau signifikant ist. Auch beim SMOG-Index ist der Zusammenhang mit demselben Signifikanzniveau statistisch signifikant. Die Wahrscheinlichkeit für ein extremes Ergebnis ist 0,001%, der FOG- sowie der SMOG-Index dienen daher sehr gut als Prädiktor für die Earnings per Share, kurz EPS, für prime market Unternehmen

der Wiener Börse. Im dritten Output ist zu erkennen, dass ein statistisch signifikanter negativer Zusammenhang zwischen EPS und den Ergebnissen des Flesch Reading Ease besteht, jedoch nur auf einem 99% Signifikanzniveau, was jedoch immer noch ein sehr gutes Ergebnis darstellt.

Zusammenfassend ist nach der Auswertung der Resultate zu sagen, dass ein starker Zusammenhang zwischen der Ergebnisse der Lesbarkeitsindizes und den EPS der Unternehmen des prime market der Wiener Börse festzustellen ist.

```
Call:
lm(formula = EPS ~ FOG)

Residuals:
    Min      1Q  Median      3Q     Max
-5.6306 -1.2496 -0.2824  0.8673  8.7444

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.0548     1.3864  -1.482   0.1401
FOG           0.2722     0.1016   2.679   0.0081 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.07 on 176 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.03917,   Adjusted R-squared:  0.03371
F-statistic: 7.174 on 1 and 176 DF,  p-value: 0.008096

> lmESMOG

Call:
lm(formula = EPS ~ SMOG)

Residuals:
    Min      1Q  Median      3Q     Max
-5.6337 -1.2894 -0.2637  0.8493  8.7350

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.5382     1.4721  -1.724  0.08642 .
SMOG          0.3318     0.1164   2.851  0.00488 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.064 on 176 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.04414,   Adjusted R-squared:  0.03871
F-statistic: 8.128 on 1 and 176 DF,  p-value: 0.00488


Call:
lm(formula = EPS ~ Flesch)

Residuals:
   Min     1Q Median     3Q    Max
-5.446 -1.286 -0.355  1.009  8.430

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.79972    1.24853   3.844 0.000169 ***
Flesch      -0.06225    0.02437  -2.554 0.011487 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.074 on 176 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.03575,   Adjusted R-squared:  0.03027
F-statistic: 6.525 on 1 and 176 DF,  p-value: 0.01149
```

Abbildung 7 R-Output - lineare Regressionsmodelle für alle Lesbarkeitsindizes

6.2. Interpretation der Ergebnisse:

Die Ergebnisse waren überraschend. Statt eines negativen Zusammenhangs wurde herausgefunden, dass ein positiver Zusammenhang für die EPS eines Unternehmens und der Komplexität der Konzernabschlüsse besteht. Dies bestätigt zwar nicht die Hypothese, aber es lässt sich dennoch ein Mehrwert aus der Analyse ziehen, weil die Indizes durchaus als legitimes Mittel zur Prognose der EPS verwendet werden können.

Die überraschenden Ergebnisse deuten darauf hin, dass die Unternehmen eine komplexere Sprache anwenden, wenn die wirtschaftliche Leistung des Unternehmens steigt. Ein möglicher Grund dafür könnte sein, dass die Gründe für das erfolgreiche Wirtschaften sehr komplex sind und die Beschreibung dieser eine komplexere Sprache verlangen. Oder wenn ein Industrieunternehmen etwa ein neues Produkt auf den Markt bringt, welches maßgeblich für den Erfolg des Unternehmens verantwortlich ist, dann kann angenommen werden, dass das Unternehmen einen Vorsprung gegenüber seinen Konkurrenten besitzt. In weiterer Folge ist anzunehmen, dass dieser Vorsprung ein Unternehmensgeheimnis ist und aus diesem Grund wird höchstwahrscheinlich nur eine wage Formulierung der Vorgänge stattfinden.

Weiters ist anzumerken, dass die Aussage aller drei Lesbarkeitsindizes ist konsistent ist. Das heißt, dass alle drei Lesbarkeitsindizes zum selben Ergebnis gekommen sind. Es wurde festgestellt, dass die Korrelation zwischen den einzelnen Lesbarkeitsindizes hoch signifikant ist, was dazu führt, dass die Lesbarkeitsindizes zum selben konsistenten Ergebnis kommen und die Lesbarkeitsindizes sich gegenseitig bestätigen.

Folglich kann gesagt werden, dass die Komplexität der Sprache ansteigt, welche mittels FOG- und SMOG-Index, sowie Flesch-Ease-Formula dargestellt wird, wenn

die EPS eines Unternehmens steigen. Die Lesbarkeitsindizes können somit als Prädiktoren für die EPS eines Unternehmens dienen.

Literaturverzeichnis

AG, W. (Juli 2014). *APA-OTS*. Von https://www.ots.at/presseaussendung/OTS_20140718_OTS0089/wolford-ag-umsatz-und-ergebnisse-des-geschaeftsjahres-201314-im-rahmen-der-erwartungen-bild abgerufen

Arbeiterkammer, W. (15. 2 2018). *Arbeiterkammer Wien*. Abgerufen am 15. 2 2018 von https://wien.arbeiterkammer.at/service/betriebsrat/ifamaufsichtsrat/ifamneu/Konzernabschluss/Offenlegung_von_Konzernabschluessen.html

Baayen, R. H. (2001). Word frequency distributions. Vol. 18. *Springer Science & Business Media*.

Bloomfield, R. J. (2002). The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons 16.3*, 233-243.

Carver, R. P. (1974). *Improving Reading Comprehension: Measuring Readability. Final Report.*

Caylor, J. S. (1973). *Methodologies for Determining Reading Requirements of Military Occupational Specialties.*

die presse. (28. 10 2017). *http://diepresse.com*. Von http://diepresse.com/home/wirtschaft/boerse/5310541/Ruhiger-Start-fuer-RHI-Magnesita abgerufen

Duffy, T. M. (1978). *Reading Skill Levels in the Navy.* Navy Personnel Research and Development Center, San Diego, CA.

Duffy, T. M. (1982). Testing a readable writing approach to text revision. *Journal of educational psychology, 74(5)*, 733.

Duffy, T. M. (1985). Readability formulas: What's the use? *Designing usable texts, 104*, 113-143.

Entin, E. B. (1978). Some inter-relationships of readability, cloze and multiple choice scores on a reading comprehension test. *journal of Reading Behavior, 10(4)*, 417-436.

Flesch, R. (1948). "A new readability yardstick. *Journal of applied psychology 32.3*, 221.

Francis, W. N. (1979). *Brown corpus manual.* Brown University.

Francis, W. N. (1982). Frequency analysis of English usage. *Houghton Mifflin*.

Gomber, P. a. (2015). High frequency trading." Encyclopedia of Information Science and Technology, Third Edition. *IGI Global*.

Grefenstette, G. a. (1994). What is a word, what is a sentence?: problems of Tokenisation. *Rank Xero Research Centre, Grenoble Laboratory*.

Guay, W. S. (2016). Guay, Wayne, Delphine Samuels, and Daniel Taylor. "Guiding through the fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics 62.2*, 234-269.

Gunning, R. (1952). *The technique of clear writing.* McGraw Hill Higher Education.

Gunning, R. (1969). The fog index after twenty years. *Journal of Business Communication 6.2*, 3-13.

International Standard Classification of Education Fields of Education and Training. (2015). *UNESCO Institute for Statistics*.

Kern, R. P. (1980). *Usefulness of Readability Formulas for Achieving Army Readability Objectives: Research and State-of-the-Art-Applied to the Army's Problem (No. ARI-TR-437).* ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA.

Kincaid, J. P. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (No. RBR-8-75).* Naval Technical Training Command Millington TN Research Branch.

Kincaid, J. P. (1975). Kincaid, J. Peter, et al. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Naval Technical Training Command Millington TN Research Branch*.

Klare, G. (1979). Readability Standards for ARMY-Wide Publications(Evaluation Report 79-1). *US Army Administrative Center, Fort Benjamin Harrison*.

Klare, G. R. (1976). A Second Look at the Validityl of Readability Formulas a. *Journal of reading behavior 8.2*, 129-152.

Labonte, S. (2014). OTS-APA.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics 45.2*, 221-247.

Loughran, T. a. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance 66.1*, 35-65.

Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. *Journal of reading 12.8*, 639-646.

McClure, G. M. (1987). Readability formulas: Useful or useless? *IEEE Transactions on Professional Communication*, 12-15.

Mockovak, W. P. (1974). *Literacy Skills and Requirements in Air Force Career Ladders (No. AFHRL-TR-74-90).* AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX.

pdf2doc. (15. 03 2018). Von https://pdf2doc.com/de/ abgerufen

Pressman, R. (1979). Legislative and regulatory progress on the readability of insurance policies. *Document Design Center, American Institutes for Research*.

Sadasivam, G. S. (2016). Corporate governance fraud detection from annual reports using big data analytics. *International Journal of Big Data Intelligence 3.1*, 51-60.

Securities Exchange Act of 1934. (1934). *Securities and Exchange Commission*.

Smallpdf. (15. 2 2018). *Smallpdf GmbH*. Abgerufen am 15. 2 2018 von https://smallpdf.com/de/

Sprache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal 53.7*, 410-413.

Stone, P. J. (1966). The general inquirer: A computer approach to content analysis. *MIT Press*.

Tetlock, P. C. (2007). "Giving content to investor sentiment: The role of media in the stock market.". *The Journal of Finance 62.3*, 1139-1168.

Tetlock, P. C.-T. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance 63.3*, 1437-1467.

Anhang:

Liste der untersuchten Unternehmen:

Agrana AG

AMAG

Andritz AG

AT & S AG

BUWOG AG

CA Immo AG

Conwert AG

Do&Co AG

Erste Group AG

EVN AG

FACC AG

Flughafen Wien AG

Immofinanz AG

KapschTC AG

Lenzing AG

Mayr Melnhof AG

OMV AG

Österreichische Post AG

Palfinger AG

Polytec AG

PORR AG

RBI AG

Rosenbauer AG

SBO AG

Semperit AG

S-Immo AG

STRABAG SE

Telekom Austria AG

UBM AG

Uniqua AG

Verbund AG

Vienna Insurance Group AG

Voestalpine AG

Warimpex Finanz- und Beteiligungs AG

Wienerberger AG

Wolford AG

Zumtobel AG