Heinrich Heine
Universität
Düsseldorf

Philosophische Fakultät
Institut für Sprache und Information

Bachelor Thesis

zum Thema

Computerlinguistische Analyse empirischer neurolinguistischer Daten

von

Larissa Ferme
Feldstraße 41
40479, Düsseldorf
Matrikelnummer: 2780933

Erstprüferin:	Apl.-Prof. Dr. Wiebke Petersen
Zweitprüferin:	Dr. Laura Bechtold
Studiengang:	Computerlinguistik
Studiensemester:	Sommersemester 2022
Datum:	30.06.2022

Zusammenfassung

Das Ziel dieser Bachelorthesis ist es, eine computerlinguistische Analyse empirischer neurolinguistischer Daten der Wortlerner-Studie von Espey u. a. [Esp+21] durchzuführen. Dabei wird allem voran die Frage erforscht, wie viel Assoziationswissen in Sprachmodellen steckt. Auf Basis dessen wird zunächst eine Sentiment Analyse durchgeführt, um herauszufinden, ob das Modell vor einem emotionalisierten Hintergrund gelernt hat oder nicht.
Im zweiten Teil dieser Arbeit wird untersucht, ob komplexere Transfer-Modelle selbst Assoziationen erzeugen und ob sie die menschlich erstellten Assoziationen richtig zuordnen können.
Nach eingehenden Untersuchungen wurde festgestellt, dass anhand der Assoziationen nicht ermittelt werden kann, in welcher Form den Probanden das Wort vorlag. Nach Sichten der Ergebnisse kann die Vermutung aufgestellt werden, dass Menschen auch scheinbar neutral gelernte Worte mit emotionalen Situationen und somit mit emotionalen Wörtern verknüpfen. Im zweiten Teil der Bachelorthesis konnte beobachtet werden, dass die Assoziationen eher zufällig zugeordnet wurden und dass die vom Modell generierten Assoziationen den Assoziationen der Probanden ähnlich sind.

Abstract

The aim of this bachelor thesis is to conduct a computational linguistic analysis of empirical neurolinguistic data from the word learner study of Espey u. a. [Esp+21]. First and foremost, the question of how much associative knowledge is contained in language models is explored. Based on this, a sentiment analysis is first performed to find out whether the model learned against an emotionalized background or not.

. In the second part of this paper, we will investigate whether more complex transfer models generate associations themselves and whether they can correctly match the human-generated associations.
After thorough investigation, it was found that the associations cannot be used to determine in what form the subjects had the word. After viewing the results, the assumption can be made that people also associate seemingly neutral learned words with emotional situations and thus with emotional words. In the second part of the bachelor thesis, it was observed that the associations were assigned rather randomly and that the associations generated by the model were similar to the associations of the subjects.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

KDE K Desktop Environment

SQL Structured Query Language

JDK Java Development Kit

Kapitel 1

Einleitung

Seit Jahren entwickelt sich die Forschung der Künstlichen Intelligenz weiter. Ob maschinelle Übersetzer, wie Google Translator und DeepL, oder Sprachassistenten wie Siri und Alexa. In immer mehr Haushalten sind sie zu finden und sie haben eins gemeinsam: Sprachmodelle. Sowohl Sprachassistenten, als auch maschinelle Übersetzer, basieren auf den menschlich erstellten Textdaten, die es ihnen ermöglichen, die natürliche Sprache, die Menschen sprechen, zu verarbeiten.
Doch kann man diesen Sprachmodellen tatsächlich zu sprechen, ein Verständnis von Sprache und Emotionalität zu besitzen? Mit jedem neu erlernten Begriff speichert das semantische Gedächtnis in unserem Gehirn die Informationen, die während der Erfahrung mit dem Referenten des Wortes gesammelt wurden und bindet sie in Form eines Konzepts ein. Erlernen Menschen beispielsweise das Wort „Blume“ verknüpfen sie das neu erlernte Konzept eventuell mit Begriffen wie „schön“, „Frühling“, „Entspannung“ und „Farben“.
Die neurolinguistische Studie „*That means something to me: the effect of linguistic and emotional experience on the acquisition and processing of novel abstract concepts*“ von Espey et al. [Esp+21] lässt darauf hin deuten, dass emotional erlernte Begriffe eher mit anderen emotionalen Assoziationen zu einem Konzept verknüpft werden und neutrale Worte eher mit kognitiven Assoziationen verbunden werden.
Sprachmodelle besitzen diese Eigenschaft der Emotionalität nicht und können auch keine emotionalen Erfahrungen machen. Doch da sie auf Basis von Menschen erstellter Daten trainiert werden, lässt dies vermuten, dass auch sie eine gewisse Emotionalität erlernen können. Diese Tatsache lässt die Frage aufkommen, ob sich die Emotionalität, die beim Lernen neuer Wörter bei Menschen beobachtet werden kann, auch bei Sprachmodellen beobachten lässt?
Auf Basis der im oberen Abschnitt erwähnten Studie von [Esp+21] wird in

dieser Bachelorthesis untersucht, inwieweit ein sogenanntes Assoziationswissen in Sprachmodellen angenommen werden kann.
Es folgt zunächst ein Überblick über die wichtigsten Begriffe im Zusammenhang mit Sprachmodellen und dieser Bachelorthesis. Darauf aufbauend wird der Hintergrund dieser Arbeit mit Einblick in die neurolinguistische Studie näher beleuchtet. Im Anschluss daran beginnt der Methodikteil, in welchem zum einen dargestellt wird, wie das Assoziationswissen getestet wird und zum anderen die Ergebnisse präsentiert werden. Abschließend folgt ein Fazit und eine Aussicht auf mögliche zukünftige Arbeit.

Kapitel 2

Begriffserklärungen

Im Folgenden Kapitel werden sämtliche Begriffe, die für diese Bachelorthesis von Relevanz sind, näher beleuchtet. Ein gewisses Grundverständnis von Mathematik, Linguistik und maschinellem Lernen wird vorausgesetzt.

2.1 Sentimentanalyse

Die Sentimentanalyse repräsentiert ein Untergebiet des Text Minings und stellt präziser ein binäres Klassifikationsproblem dar. Dokumente, Rezensionen und Sätze können mithilfe dieser Analysetechnik in positive und negative Einheiten unterteilt werden. Vereinzelt existiert eine zusätzliche Unterteilung in neutrale Einheiten.
Diese Polaritätserkennung basiert auf zwei grundlegenden Vorgehensweisen:

- lexikonbasierte Ansätze und
- maschinelles Lernen.

Bei ersterem liegt ein Lexikon zugrunde, welches sprachspezifische Ausdrücke beinhaltet, die aufgrund von Wortbedeutungen bereits positiv oder negativ hinterlegt sind. Die Informationen zum Durchführen der Sentimentanalyse lassen sich sogenannten Sentimentlexika der jeweiligen Sprache entnehmen. Durch die Vorgehensweise mithilfe von Lexika ist eine Berücksichtigung von Kontext- oder Domänenabhängigkeiten wertender Ausdrücke, die kontextsensitiv ihre Polarität umkehren können (Beispiel: „gruselig “ ist ein negativ konnotiertes Wort, wenn dieser Begriff jedoch in Verbindung mit einer Horrofilm-Bewertung einhergeht, ist es durchaus positiv konnotiert) zwar kaum möglich, da dieser Ansatz jedoch ohne Trainingsdaten möglich ist, bietet sich hier ein großer Vorteil gegenüber dem Ansatz des maschinellen Lernens.

2.2 Transfer-Learning

Eine Methode des maschinellen Lernens, die immer mehr an Relevanz gewinnt, ist das Transfer-Learning. Dabei wird ein bereits vortrainiertes Modell optimiert und für Lösungen neuer Problemstellungen genutzt. In Analogie zum Lernprozess des Menschen, der nicht alles von Grund auf neu erlernt, sondern sein Wissen transferiert, funktioniert dieser Ansatz.
Ein Modell jedes Mal von Grund auf neu zu trainieren würde zum einen eine Menge von Ressourcen benötigen und teilweise wochenlang dauern und zum anderen Massen an CO2-Emissionen produzieren [Hao19].
Vorteile des Transfer-Learnings sind sowohl eine geringere Menge benötigter Ressourcen, als auch ein schnelleres Training.
Benötigt das Training eines neuen Modells teils Wochen, so kann dieses sehr aufwendige Training durch Anwendung von Transfer-Learning erheblich verkürzt werden, da ein gewisses Grundwissen bereits besteht und dieses lediglich erweitert wird.

2.3 Transformer-Modelle

Um die Möglichkeit zu bieten eine Art von Sequenzen in eine andere Art von Sequenzen zu transferieren, beispielsweise einen deutschen Satz in einen englischen Satz zu übersetzen, wurden lange Zeit rekurrente Modelle wie LSTMs und GRUs genutzt. Da diese neuronalen Netzwerke jedoch durch ihre Vorgehensweise zum einen ein langes Training benötigen und zum anderen Probleme mit längeren Sequenzen aufweisen, wurden neue Architekturen erforscht.
Mit der Veröffentlichung der Transformer-Modelle im Juni 2017 von Vaswani et al. in ihrer wissenschaftlichen Arbeit „Attention is all you need“ [Vas+17], boten sie der Machine Learning Gemeinschaft eine der rekurrenten Modellen ebenbürtige Architektur, die jegliche state-of-the-art Modelle übertraf.
Das ursprüngliche Forschungsziel war eine neue Art von *sequence-to-sequence*-Modellen einzuführen, die unter anderem für die zuvor genannte maschinelle Übersetzung hilfreich sein sollten. Durch die Struktur der Transformer-Modelle ist es jedoch möglich diese auch für andere Problemstellungen zu verwenden.
Transformer lassen sich in einen Encoder- (linke Hälfte der Architektur in Abbildung 2.1) und einen Decoder-Abschnitt (rechte Hälfte der Architektur in Abbildung 2.1) unterteilen, die beide jeweils für sich eigenständig funktionieren.
Transformer-Modelle werden für den Gebrauch von spezifischen Aufgaben wie der Text-Generierung nochmal einem supervised Fine-Tuning unterzogen,

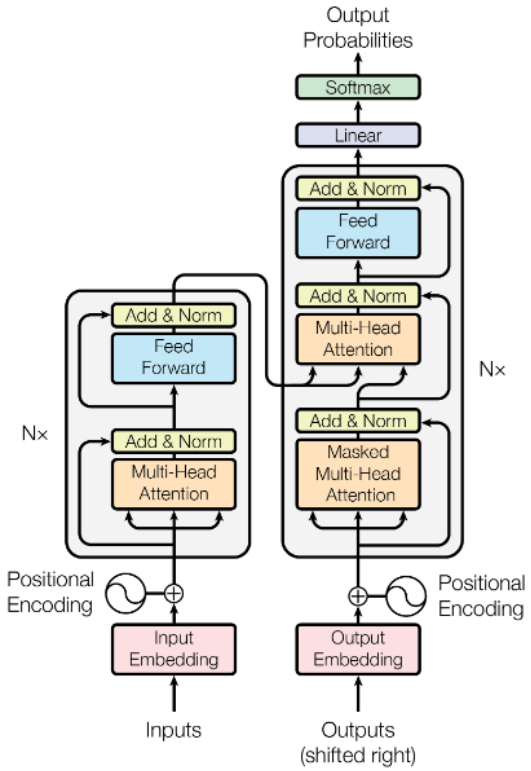


Abbildung 2.1: Darstellung der Transformer-Modell-Architektur aus „Attention Is All You Need“ [Vas+17]

da sie zuvor self-supervised ohne menschlich erstellte Labels trainiert wurden und noch keine ausreichend gute Performance bieten können.

2.3.1 Encoder-Modelle

Encoder-Modelle (oder auch *auto-encoding*-Modelle genannt) nutzen lediglich den Encoder-Teil eines Transformer-Modells. Bekannte Modelle dieser Art sind beispielsweise BERT und DistilBERT. Durch den Attention-Mechanismus kann diese Art von Modellen zu jedem Zeitpunkt des Prozesses auf jedes Wort des als Input verwendeten Satzes zugreifen. Folglich ist ein Wort-Vektor nicht nur eine numerische Repräsentation der Informationen des Wortes, sondern beinhaltet des Weiteren auch Kontextinformationen aus seiner Umgebung.

Der bidirektionale Charakter ist typisch für Encoder-Modelle. Dank des Attention-Layers sind nicht nur Informationen links oder rechts des aktuellen Wortes verfügbar, sondern in beide Richtungen (siehe Abbildung 2.2). Aufgrund dieser Eigenschaft eignen sich Encoder-Modelle besonders für die

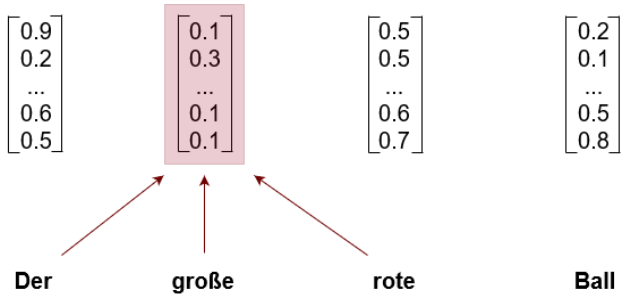


Abbildung 2.2: Darstellung des bidirektionalen Attention-Mechanismus der Encoder-Modelle

Aufgaben der Sentence Classification, Named Entity Recognition und der Extractive Question Answering.

2.3.2 Decoder-Modelle

Decoder-Modelle (oder auch *auto-regressive*-Modelle genannt) nutzen den Decoder-Teil eines Transformer-Modells. Ein bekanntes Decoder-Modell ist *Generative Pretrained Transformers* (GPT). Der gravierendste Unterschied zu Encoder-Modellen besteht im Attention-Mechanismus. Während Encoder-Modelle durch ihren Attention-Layer eine bidirektionale Besonderheit besitzen, handelt es sich bei Decoder-Modellen um einen Masked-Attention-Mechanismus. Diese sogenannte Maske, die Teile des Inputs links oder rechts des aktuellen Wortes verdeckt, sorgt für eine uni-direktionale Verarbeitung der Sequenzen und macht den Decoder somit besonders attraktiv für Text-Generierungs-Aufgaben (siehe Abbildung 2.3).

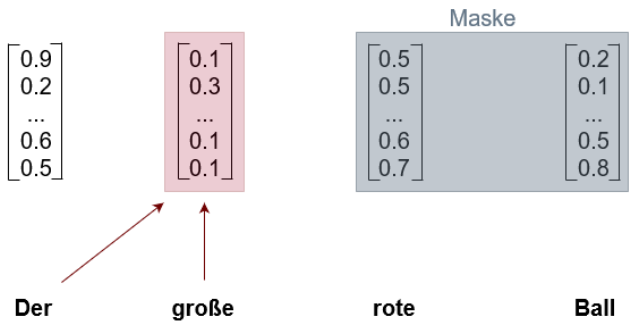


Abbildung 2.3: Darstellung des uni-direktionalen Attention-Mechanismus der Decoder-Modelle

2.3.3 Sequence-to-sequence-Modelle

Sequence-to-sequence-Modelle nutzen im Gegensatz zu den bisher erwähnten Modellen beide Elemente der Transformer-Architektur (Abbildung 2.1). Aufgrund dessen sind sie auch teils unter dem Namen Encoder-Decoder-Modelle bekannt.
Der erste Teil des Prozesses gleicht dem eines Encoder-Modells, nachdem jedoch Feature-Vektoren für die einzelnen Token der Input-Sequenz generiert worden sind, wird der Decoder-Teil anders als bisher verwendet. Dem Decoder wird nicht nur eine Initialsequenz, sondern zusätzlich der Output des Encoders als Input übergeben. Nachdem das erste initiale Wort vom Decoder generiert wurde, ist der Encoder-Teil nicht weiter von Nutzen, da dieser Output in einer auto-regressiven Weise wieder als Input verwendet werden kann. Wird dem Encoder beispielsweise ein englischer Satz übergeben, berechnet dieser die numerischen Repräsentationen der einzelnen Token und übergibt diese dem Decoder. Dieser erhält als Input zusätzlich eine Initialsequenz und genereiert daraufhin das erste übersetzte Wort. Auf Basis der numerischen Repräsentationen des Encoders und des ersten übersetzten Wortes, ist er in der Lage, das darauffolgende Wort vorherzusagen.
Durch die Eigenschaft beide Elemente der Architektur zusammen zu nutzen ist diese Art von Modellen besonders für die Text-Übersetzung, Text Summarization und Generative Question Answering geeignet. Bekannte Modelle sind BART und mBART.

Kapitel 3

Hintergrund

Der Hintergrund dieser Bachelorthesis liegt in der Studie „That means something to me: the effect of linguistic and emotional experience on the acquisition and processing of novel abstract concepts“ von Espey et al. zu den „Auswirkungen sprachlicher und emotionaler Erfahrungen auf den Erwerb und der Verarbeitung neuer abstrakter Konzepte“ [Esp+21].
Ziel der Studie war es den Erwerb neuer abstrakter Konzepte und deren Darstellung in sprachlichen und emotionalen Erfahrungen zu untersuchen, da eine fortlaufende Debatte darüber existiert, ob Erfahrungen in amodale neuronale Repräsentationen [1] übersetzt werden oder ob die semantische Verarbeitung in Erfahrungen verankert bleibt.
Dass Wörter wie „laufen “ oder „Pflanze “ während des Erlernens mit emotionalen Erfahrungen verknüpft werden, ist durchaus zu beobachten, da ersteres eine Handlung und letzteres eine Entität in der erfahrbaren Welt darstellt. Die Forschenden stellten sich jedoch die Frage, wie es sich mit abstrakteren Konzepten wie „Schadenfreude “ oder „Glück “ verhält.
Um dieser Frage auf den Grund zu gehen, erstellten die Forschenden 30 deutschsprachige abstrakte Konzepte in emotionaler und neutraler Form mit zugehörigen Definitionen und je fünf Situationsbeschreibungen.
In fünf Trainingseinheiten lernten 66 Studienteilnehmer 30 abstrakte Konzepte (pro Proband je 15 in emotionaler und 15 in neutraler Form) und beschäftigten sich entweder mit der Erzeugung mentaler Bilder oder der lexikalisch-semantischen Verarbeitung. In einem Pre-Test haben die Probanden die Definition eines jeweiligen neutralen oder emotionalen Konzepts gezeigt bekommen und sollten im Anschluss daran auf einer Skala von eins bis fünf einschätzen wie vertraut ihnen dieses Konzept ist. Auf einer anderen

[1] Amodale Repräsentationen bezeichnen in der Psychologie die „Vorstellung, dass das Format der begrifflichen Darstellungen abstrakt ist und keine Wahrnehmungseigenschaften mehr hat“ [Arn21].

Skala sollten sie ebenfalls von eins bis fünf beurteilen wie häufig dieses Konzept in den letzten vier Wochen in ihrem Leben aufgetreten ist.
Im Anschluss an den Pre-Test folgten die fünf Trainingseinheiten, welche in einem Abstand von 24 Stunden erfolgten. In der ersten und fünften Trainingseinheit sahen die Studienteilnehmer ein emotionales oder neutrales Pseudowort (= *Vul*) zusammen mit seiner Definition (= *Die Beunruhigung und Anspannung, wenn man sich den Kopf über etwas zerbricht.*) und sollten mithilfe einer 6-Punkt-Likert-Skala bewerten wie emotional positiv oder emotional negativ dieses Konzept für sie behaftet ist. Danach sahen sie noch einmal das Pseudowort, jedoch dieses Mal zusammen mit einer der fünf Situationsbeschreibungen (= *Du grübelst, ob du deine letzte Klausur bestanden hast und dein Studium in Regelstudienzeit beenden kannst. Du bist zum Zerreißen gespannt und nervös.*), die beispielhafte Situationen konstruieren, in denen dieses Konzept auftritt.
In der zweiten und vierten Trainingseinheit unterzogen sich die Probanden einem *Forced Choice Semantic Judgement*. Bei dieser Aufgabe wird eins der Pseudowörter zusammen mit der richtigen dazugehörigen Definition und einer falsch zugeordneten Definition eingeblendet. Hierbei muss entschieden werden, welche nun die für das Pseudowort vorgesehene richtige Definition darstellt. Im Anschluss daran wurden wieder die Pseudowörter einzeln mit einer der fünf Situationsbeschreibungen eingeblendet.
Nach dem Training unterzogen sich die Probanden den Post-Tests. In dem ersten Test sollten sie entscheiden, ob das eingeblendete Wort ein Pseudowort ist, welches sie gelernt haben, oder ob es ein fremdes Pseudowort ist. Im Anschluss daran folgte eine *Feature Production* Aufgabe. Diese Aufgabe motivierte die Probanden und Probandinnen bestimmte Wörter, sogenannte Merkmale (im Folgenden *Features* genannt), aufzulisten, die sie mit den erlernten Konzepten in Verbindung bringen. In den Instruktionen wurden Features so definiert, dass sie intern (körperliche Empfindungen) oder extern wahrnehmbar (Sinneseindrücke, wie Geräusche, Geschmäcker, etc.), funktional assoziierend (wie, wo, wann) oder nicht-physisch (Gefühle, Gedanken, Meinungen) sein können. Die Probanden hatten pro Konzept 90 Sekunden Zeit diese Features zu produzieren und durften innerhalb dieses Zeitraums beliebig viele Features aufschreiben. Allerdings blieb ihnen auch die Möglichkeit vorzeitig zum nächsten Wort fort zu schreiten.

Espey et al. stellen anhand ihrer Forschungsergebnisse fest, dass eine gewisse Bedeutung sprachlicher Erfahrung für den Erwerb und die Darstellung neuer abstrakter Konzepte existiert. Des Weiteren konnten sie in ihren Resultaten beobachten, dass emotionale Konzepte eher mit emotionalen Features in Verbindung gebracht werden und für neutrale Konzepte eher kognitive

Merkmale produziert werden.

Auf Basis dieser Studienergebnisse rückten zwei computerlinguistische Fragestellungen in den Vordergrund. Ein Sprachmodell wie ein Sentiment Analyse Modell, aber auch komplexere Transfer-Modelle wie BERT oder BART sind nicht in der Lage dazu Emotionen zu erleben und emotionale Erfahrungen abzuspeichern. Trotzdessen sind sie mit emotionalisierten Daten trainiert und kreiert worden. Daher stellt sich die Frage, wie viel Merkmalswissen in einem Sprachmodell steckt, das auf menschlich erstellten Daten trainiert wurde? Ist ein primitives Sprachmodell dazu in der Lage die Emotionalität in den von den Studienteilnehmern produzierten Features zu erkennen? Inwiefern ist es einem komplexeren Sprachmodell möglich eine Verbindung zwischen den produzierten Merkmalen und den abstrakten Konzepten zu ziehen?
Diese Fragen werden in den folgenden Kapiteln aufgegriffen und unter Zuhilfenahme diverser Sprachmodelle werden einge Tests durchgeführt.

Kapitel 4

Durchführung

Kapitel X beschreibt die Implementierung der Sentimentanalyse, sowie der Transfer-Modelle. Es wird mit der Umsetzung der Sentimentanalyse und der anschließenden Präsentation der Ergebnisse begonnen. Nachfolgend werden die Transfer-Modelle näher beleuchtet und die zweite Phase der computerlinguistischen Auswertung der Studie von Espey et al. eingeleitet. Auch hier folgt eine intensive Begutachtung der Ergebnisse.

4.1 Datenaufbereitung

In diesem Abschnitt wird ein kurzer Überblick über die Datenaufbereitung gegeben. Die zugrundeliegenden Daten der Studie umfassten insgesamt 1980 Feature-Ketten, die die Studienteilnehmer produziert hatten. Da einige Probanden in der Feature Prdocution Task zu manchen Wörtern keine Merkmale aufgeschrieben haben, wurden diese aus dem Datensatz für diese Bachelorthesis entfernt. Nach der Aussortierung dieser Einträge umfasst der Datensatz eine Anzahl von 1953 Feature-Ketten.
Des Weiteren wurden Ketten, bei denen das letzte Wort abgeschnitten, also nicht ganz ausgeschrieben war, um dieses letzte Wort reduziert. Ein Fehler, der ebenfalls behoben wurde, ist der, dass in der Datenerstellung seitens der Studie ein Fehler bei dem Wort 'nach' entstanden sein muss. Denn statt 'nach' fanden sich in den Ketten nur die Buchstaben 'ch'. Beispiele dafür sind 'chdenklich' und 'ch Worten ringen'. Um diese Ausdrücke trotzdem mit in den Datensatz zu übernehmen, wurde das 'ch' durch ein 'nach' ausgetauscht.
Im Anschluss an die Bereinigung wurden die deutschsprachigen Daten ins Englische übersetzt. Dies geschah im Zuge dessen, dass englische Sprachmodelle besser trainiert und präzisere Ergebnisse erzielen als ihre deutschen Pendants.

4.2 Sentimentanalyse

4.2.1 Vorgehensweise

Im Rahmen dieser Bachelorthesis stellen sich für den Teil der Sentimentanalyse folgende zwei Fragen in den Vordergrund:

1. Kann anhand der Features, die die Probanden aufgeschrieben haben, darauf geschlossen werden, ob die Probanden das Wort als emotionales oder neutrales Wort gelernt haben?

2. Kann anhand der Situationsbeschreibungen darauf geschlossen werden, ob das Wort zugehörig zu der jeweiligen Beschreibung emotional oder neutral dargestellt wird?

In Anbetracht der beiden Fragestellungen, wurden die zugrundeliegenden Daten mittels einer Sentimentanalyse näher betrachtet. Die hier und im weiteren Verlauf dieser Bachelorthesis verwendeten Daten der neurolinguistischen Studie umfassen die Features, Situationsbeschreibungen und Valens-Bewertungen der Probanden aus der fünften Trainingseinheit. Zur Durchführung der Sentimentanalyse wurde der SentimentIntensityAnalyzer von VADER gewählt. VADER bedeutet Valence Aware Dictionary and sEntiment Reasoner und wurde im Jahr 2014 veröffentlicht Hutto und Gilbert [HG14]. Dieser SentimentIntensityAnalyzer basiert auf dem lexikonbasierten Ansatz. Doch statt die Wörter des Lexikons lediglich in positiv, *negativ* und *neutral* zu clustern, wurde die Intensität der einzelnen Wörter von Menschen bewertet. Auf einer Skala von -4 bis 4 unter Berücksichtigung der 0, die die Neutralität darstellt, sollten die Rater bewerten wie emotional positiv oder emotional negativ das jeweilige Wort für sie behaftet ist. Auf Basis dessen erstellten die Forschenden ein Sentiment Intensity (Valence-based) Lexikon, das zur Durchfühurng der Sentimentanalyse genutzt wird.
Dadurch, dass das Lexikon aus weitestgehend Social Media bezogenen Daten besteht, eignet sich der Analyzer von VADER vor allem für kurze Sequenzen mit Umgangssprache und Abkürzungen. Da im Rahmen dieser Bachelorarbeit produzierte Features analysiert werden, stellt sich VADER daher als sehr geeignet heraus.
Dem SentiemntIntesityAnalyzer wurde jede Feature-Kette einzeln zur Bewertung übergeben. Das gleiche geschah für die Situationsbeschreibungen.
Nach der Bewertung der Daten durch das Sprachmodell lagen für jede Feature-Kette und für jede Situationsbeschreibung vier von dem Modell berechnete Werte mit den deskriptiven Labels *positiv*, *negativ*, *neutral* und *compound* vor. Der compound-Wert ist eine Metrik, die die Summe aller lexikalischen

Ratings, die zwischen -1 und 1 normalisiert worden sind, berechnet. Die ersten drei Werte geben lediglich an, wie hoch der Anteil an Negativität, Neutralität und Positivität in dem jeweiligen Satz ist.
Der compound-Wert wird durch die Addition der Valenz-Werte jedes Wortes im Lexikon berechnet und anschließend mithilfe folgender Formel zwischen -1 und 1 normalisiert:

$$x = \frac{x}{(\sqrt{x^2 + \alpha})} \tag{4.1}$$

wobei x die Summe der Valenz-Werte der Wortbestandteile und die Normalisierungskonstante ist. Die liegen zwischen -1 und 1, wobei positive Werte die Positivität und negative Werte die Negativität ausdrücken. Damit eine aussagekräftige Auswertung durchgeführt werden kann, wurden diese compound-Werte für den weiteren Verlauf der Sentimentanalyse verwendet. Um die Werte der einzelnen Daten entsprechend auswerten zu können, wurden sogenannte Thresholds eingesetzt. Die Thresholds dienen dazu die Werte komparativ zu interpretieren.
Da im Rahmen der Studie primär zwischen Neutralität und Emotionalität unterschieden wurde, wird dies ebenfalls im Hinblick der computerlinguistischen Auswertung berücksichtigt und infolgedessen die Negativität, sowie die Positivität einer Feature-Kette oder Situationsbeschreibung unter der Emotionalität zusammen gefasst.
Die jeweils emotionalsten Bewertungen werden durch die beiden äußersten Zahlen der compound-Skala (-1 und 1) ausgedrückt. Die Werte zur Mitte hin werden als neutraler geltend betrachtet. Daher wurden die Thresholds folglich so gesetzt, dass Werte zwischen -0.5 und 0.5 liegend als neutral betrachtet werden und Werte zwischen -1 und -0.5 und 0.5 und 1 als emotional konnotiert gelten.
Da sich diese recht simple Vorgehensweise zum Setzen der Thresholds jedoch wenig an den Daten orientierte, sondern einem universal anwendbarem Schema folgte, wurden die Thresholds im Nachhinein mit zwei weiteren Verfahren angepasst und die Ergebnisse verglichen, um eine Art Underfitting ausschließen zu können. Mittels Bestimmung der Mediane des Medians der compound-Werte aller emotionalen und neutralen Situationsbeschreibungen erfolgte das zweite Setzen der Thresholds. Es wurde der Median der zuvor erwähnten Situations-compound-Werte berechnet. Daraufhin wiederholte sich der Vorgang für die zwei entstandenen Hälften. Die Mediane der beiden Hälften bildeten die neuen Thresholds für Methode 2. Mithilfe dieser Vorgehensweise konnte ein Underfitting ausgeschlossen werden und die daraus resultierenden Werte lauteten -0.24555 und 0.5844.

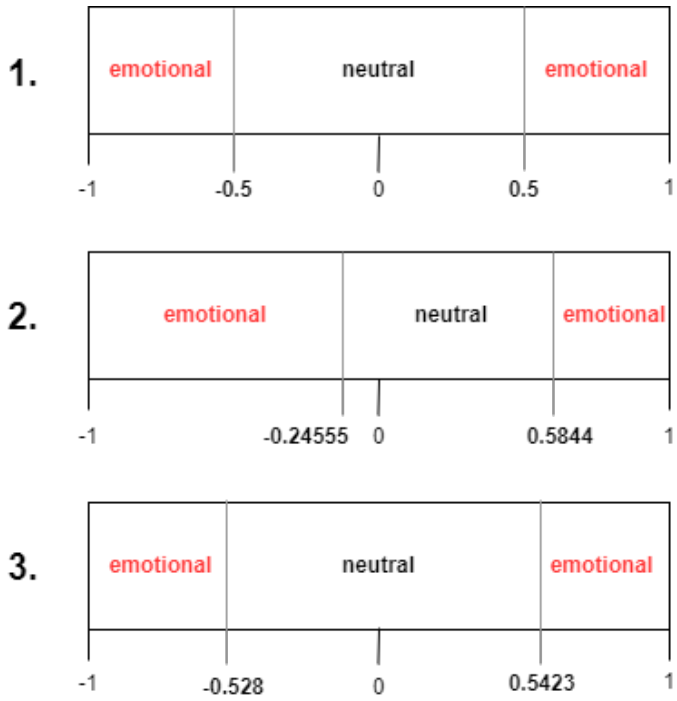


Abbildung 4.1: Darstellung der drei verschiedenen Varianten der Thresholds

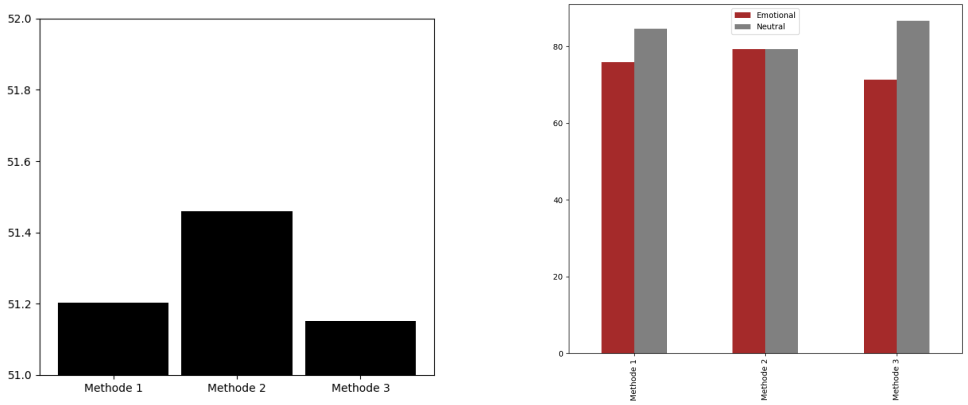
Da eine weitere Frage offen blieb, wie das Verhältnis von negativen zu positiven Beschreibungen innerhalb der emotionalen Situationen ist und ob eine zu ungleiche Verteilung eventuellen Einfluss auf die Ergebnisse haben könnte, wurden die prozentualen Anteile von negativen und positiven Situationen innerhalb der emotionalen Beschreibungen berechnet und auf Basis dessen wiederholt die Mediane ermittelt. Das Verhältnis wurde ermittelt, indem alle compound-Werte < 0 als negativ gelten und alle Werte > 0 als positiv. Der untere Threshold entspricht dem Median aller negativ konnotierten Situationen. Gleichermaßen fand der Median der positiven Situationsbeschreibungen als oberer Threshold Verwendung. Nach Durchführung dieser Berechnungen ergaben sich folgende Werte: -0.528 und 0.5423.
Da die Studie Daten zur Valenz der produzierten Merkmale erhoben hatte, indem die Studienteilnehmer in der fünften Trainingseinheit auf einer 7-Punkt-Likert-Skala bewerten sollten wie positiv oder negativ das erlernte Konzept für sie ist. Auf Basis dieser Werte und der Sentiment-Scores, die das Modell berechnet hat, wurde die Korrelation mittels Pearson berechnet, um den linearen Zusammenhang der Werte zu bestimmen. Sowohl für jedes Wort einzeln, indem beide Werte aller Studienteilnehmer eines Wortes zur Berechnung der Korrelation verwendet wurden, als auch für alle Wörter zusammen.

4.2.2 Ergebnisse

Pseudowort	Einstufung der Studie	Assoziationen der Studienteilnehmer	Berechnungen des NLTK-Modells
Mölauzegt	neutral	Informationen suchen + Entscheidung + schlau + Frustration + abwarten	{'neg': 0.316, 'neu': 0.408, 'pos': 0.276, 'compound': -0.1027}
Zimerhubst	emotional	Mühe + Freude + Befriedigung + Sinn	{'neg': 0.0, 'neu': 0.23, 'pos': 0.77, 'compound': 0.7717}
Rugliebast	emotional	dreist + Wut + Unsicherheit + Einmischen + aufdringlich	{'neg': 0.705, 'neu': 0.082, 'pos': 0.213, 'compound': - 0.7184}

Abbildung 4.2: Auszug aus der Datei *features_results.tsv* mit Pseudowörtern, Assoziationen und zugehörigen Modell-Output

Die mithilfe der im vorherigen Kapitel erwähnten Methoden erzielten Ergebnisse werden in dieser Sektion berichtet. Abbildung 4.2 illustriert einen beispielhaften Auszug aus der Datei, mit den berechneten Sentiment-Werten, wobei nur der compound-Wert beachtet wurde. In dieser Tabelle sind drei Beispiel-Pseudowörter aus dem Datensatz mit ihren zugehörigen Einsortierungen in das neutrale oder emotionale Spektrum abgebildet. Des Weiteren repräsentiert die dritte Spalte die Assoziationen, welche die Studienteilnehmer mit dem erlernten Konzept verbinden. Die letzte Spalte illustriert den Output des Modells mit dem compound-Wert.
Vergleichsweise verhält es sich mit der Output-Datei der Situationsbeschreibungen. Die Spalte „Assoziationen der Studienteilnehmer “ wird dort lediglich gegen eine Spalte mit den Situationsbeschreibungen ausgetauscht.
Für eine komparative Auswertung der Ergebnisse wurde primär die Accuracy aller Methoden berechnet (siehe Abbildung 2). Wie in Abbildung 2 zu erkennen ist, wurde mithilfe der Mediane des Medians Methode (Methode 2) die höchste Accuracy von circa 51.46% erzielt.
Die anderen beiden Methoden brachten jedoch ebenfalls mit 51.20% und 51.15% Accuracies hervor, die sich sehr nah an der höchsten Accuracy bewegen.
Bei genauerer Begutachtung der Ergebnisse für die Situationsbeschreibungen, ist zum einen eine allgemein höhere Accuracy im Vergleich der Assoziationsaccuracies zu beobachten. Die prozentualen Ergebnisse liegen im oberen Drittel, während die Accuracies der Assoziationen im mittleren Bereich liegen.
Zum anderen ist eine etwas höhere Accuracy der neutralen Situationen gegenüber der emotionalen Beschreibungen zu erkennen.



(a) Accuracies der Assoziationsdaten

(b) Accuracies der Situationsbeschreibungen

Abbildung 4.3: Darstellung der drei verschiedenen Varianten der Thresholds

In Abbildung 4.4 ist die Korrelation auf den Daten nach Pearson zu sehen. Auf der y-Achse sind die Pearson R-Werte zwischen -1 und 1 dargestellt. Die x-Achse repräsentiert die Anzahl der Werte. Insgesamt sind 60 Werte in diesem Graphen dargestellt, da für jedes Pseudowort in neutraler und emotionaler Form jeweils ein Korrelationswert berechnet wurde. Die orange Linie bildet die Durchschnittskorrelation aller Wörter ab. Aus anderen Quellen ist hervorgegangen, dass die Thresholds zur Bewertung der Korrelationswerte so gewählt wurden, dass Werte zwischen -0.33 und 0.33 als *weak*, Werte zwischen -0.33 und -0.66 und zwischen 0.33 und 0.66 als *moderate* und alles andere als *strong* bewertet wird. Auf diese Weise wurden die Thresholds für diese Bachelorthesis gewählt. In der Abbildung sind diese Bereiche durch die grauen und hellblauen Markierungen im Hintergrund zu betrachten.
Anhand der orangenen Linie ist bereits zu erkennen, dass der durchschnittliche Korrelationswert nahezu 0 ist. Eine Korrelationswert von nahezu 0 drückt aus, dass die Daten nicht korrelieren. Der hier abgebildete Wert liegt bei -0.04536415140543944.
Die blaue Linie verzeichnet neben den Werten, die im schwachen Bereich liegen jedoch auch vier Ausschläge, die im moderaten Bereich anzuordnen sind. Ein Beispiel dafür ist die neutrale Version des Pseudowortes *Faube*.

4.2.3 Diskussion

In diesem Teil der Bachelorthesis wurde untersucht inwiefern anhand der Situationsbeschreibungen und Features, die die Probanden aufgeschrieben

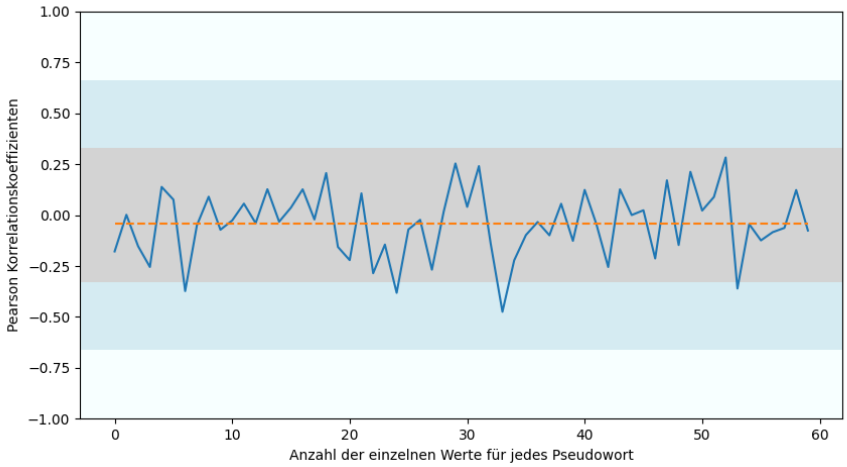


Abbildung 4.4: Darstellung der Pearson Korrelation auf den Daten

haben, darauf geschlossen werden kann, ob das dazugehörige Pseudowort in emotionaler oder neutraler Form vor lag. Die Erwartungen waren, dass sowohl die Situationsbeschreibungen als auch die Features eher emotional konnotiert sind, die Vermutung naheliegt, dass eine komplette Neutralität nahezu unmöglich ist. Jede Person hat emotionale Erfahrungen in ihrem Leben gemacht und wenn sie ein neues Wort lernt, verknüpft sie diese emotionalisierten Erfahrungen mit dem neuen Wort. Die Vermutung hat sich auch weitestgehend bestätigt. Fast 70% der Feature-Ketten wurden als emotional markiert. Wenn man sich die Beispiele in Abbildung 4.5 anschaut, ist eine Tendenz der Emotionalisierung in den Daten zu erkennen. Die Feature-Ketten würden objektiv betrachtet als eher emotional eingestuft werden, da Wörter wie *anxiety, nervousness, negative, positive, being annoyed* etc. darin vorkommen.

Da die Feature-Ketten vom Modell alle eher emotional bewertet worden sind, und auch eine gewisse Emotionalität in den für neutrale Konzepte produzierten Features beobachtet werden konnte, lag die Vermutung nahe, dass die Valenz-Ratings der Probanden ebenfalls eher emotional ausfällt. Nach intensiver Begutachtung der Korrelationswerte, konnte jedoch festgestellt werden, dass die beiden Daten kaum miteinander korrelieren. Ein Beispiel aus den Daten lässt jedoch einen Einblick darin geben, warum die Daten nicht so stark korrelieren wie anfänglich gedacht.Das Pseudowort *Wupforau* bedeutet in seiner emotionalen Form „*The special feeling that you will fall in love with the person you just met.*“. Der Studienteilnehmer hat zu diesem Konzept

Pseudowort	Definition	Emotionalität	Features	Sentiment-Wert
Skibt	the state one gets when sharing other people's feelings.	neu	anxiety + nervousness + sadness + loneliness + tired + bed + lying + stress	-0.9493
Vul	the realization that more information is needed to solve a problem.	neu	evaluation + self-image + criticism + positive + negative + harsh	-0.7096
Bingsemöl	when one repeats an activity in order to be able to perform it sufficiently well.	neu	anger + being annoyed + other people + unpleasant + annoying person + knowing everything better + distinguishing oneself + advice + unwanted	-0.8779

Abbildung 4.5: Beispiele für neutrale Konzepte, die emotionalisiert sind

folgende Features aufgeschrieben: „*beautiful + pleasant feeling + exciting + morning + sometimes embarrassing + subconscious + processing things + detail + feeling*“. Die Features machen einen eher positiv konnotierten Eindruck. Das Sprachmodell hat diesen Features einen compound-Wert von 0.8689 zugeschrieben. Der Proband hat in dem Valenz-Rating, jedoch eine negative Bewertung von-2 gegeben.
Ein anderes Beispiel, ebenfalls für Wupforau in emotionaler Form lautet wie folgt. Die Features eines Studienteilnehmers sind „*family + friendship + trust + being able to tell everything*“. Das Sprachmodell hat für diese Kette einen Wert von 0.7351 berechnet. Der Proband hat das Konzept jedoch mit einer 1 bewertet, die im unteren Drittel der Positivität liegt. Während das Modell Ausdrücke wie *trust*, *being able to tell everything*, etc. als etwas Positives bewertet hat, hat der Proband das Konzept beinahe als neutral eingestuft.
Dieser Unterschied in den Bewertungen könnte an einer zeitlichen Konstante liegen. Während die Valenz-Ratings während des fünften Trainings vollzogen wurden, sind einige Tage bis zur Feature Production Task vergangen. Es könnte sein, dass die Probanden sich umentschieden haben.

4.3 Verwendung von Transformer-Modellen

Nachdem nun untersucht wurde, ob das Lernen von Sprachmodellen vor einem emotionalisierten Hintergrund passiert und ob anhand der Assoziationen und Situationsbeschreibungen darauf geschlossen werden kann, ob die Probanden das Wort als emotionales oder neutrales Wort gelernt haben, wird nun näher untersucht, wie viel Assoziationswissen in einem komplexeren Sprachmodell steckt. Steckt etwas von dem, was ein Mensch assoziiert, auch in Sprachmodellen?
Unter Verwendung von Transfer-Modellen wurden diese Fragestellungen hauptsächlich mithilfe zweier Aufgabenstellungen näher beleuchtet:

1. Zero Shot Classification
2. Text Generation

Bei der Zero Shot Classification werden Texte, die bisher nicht manuell gelabelt wurden, von dem Sprachmodell mit einem Label versehen. Als Input werden beispielsweise der Satz „Angela Merkel reist für eine Konferenz nach Paris.“ und die Labels „Politik“, „Schule“, „Freundschaft“, „Ausland“ dem Modell übergeben, welches daraufhin unter den Labels berechnet, welches der Labels am besten zu dem Satz passt.
Im Rahmen dieser Bachelorthesis wird ein ähnliches Problem behandelt. Das Sprachmodell soll auf Basis der vorgegebenen Definitionen zuordnen, welche Assoziationen die Probanden zu den Definitionen hatten. Dazu werden die Definitionen als Text und die Assoziationen als Labels verwendet. Um zu bestätigen, dass komplexere Sprachmodelle eine Art Assoziationswissen besitzen, sollten die von den Probanden generierten Assoziationen einen möglichst hohen Score erreichen.

4.3.1 Vorgehensweise

In einem ersten Schritt wurden also die Assoziationen der Studienteilnehmer, sowie die Definitionen extrahiert. Da englischsprachige Modelle zum aktuellen Zeitpunkt besser performen als deutsch- oder anders sprachige Modelle, wurden die Daten mithilfe der GoogleTranslator-API ins Englische übersetzt und englisch-sprachigen Modellen übergeben. Bevor dies jedoch passiert, werden die Daten bereinigt. Die Assoziationsketten liegen in einer Form vor, in der jede einzelne Assoziation mit einem plus-Zeichen getrennt ist. Dieses Zeichen wird durch ein Komma ersetzt und alle Wörter werden klein geschrieben.

4.3.2 Ergebnisse

4.3.3 Diskussion

Kapitel 5

Fazit

Literatur

[HG14] C.J. Hutto und Eric Gilbert. *Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. 2014.

[Vas+17] Ashish Vaswani u. a. *Attention Is All You Need*. 2017.

[Hao19] Karen Hao. *Training a single AI model can emit as much carbon as five cars in their lifetimes*. Abgerufen am 27. März 2022. 2019.

[Arn21] Frank Arnould. *amodal representation*. Abgerufen am 01. März 2022. 2021.

[Esp+21] Linda Espey u. a. *That means something to me: the effect of linguistic and emotional experience on the acquisition and processing of novel abstract concepts*. 2021.

Anhang A

Textbausteine

A.1 Erster Baustein

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt habe. Ich habe alle Stellen, die ich aus den Quellen wörtlich oder inhaltlich entnommen habe, als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Larissa Ferme, Düsseldorf, den 30.06.2022