Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
Pfaffenwaldring 5B
D-70569 Stuttgart

Bachelor thesis

Modeling the evaluative nature of German personal name compounds

Tana Deeg

Study program: B.Sc. Maschinelle Sprachverarbeitung

Examiner: Prof. Dr. Sabine Schulte im Walde
Supervisors: Annerose Eichel,
Prof. Dr. Sabine Schulte im Walde

Start date: 02.05.2023
End date: 02.11.2023

Statement of Authorship

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und dabei keine andere als die angegebene Literatur verwendet habe. Alle Zitate und sinngemäßen Entlehnungen sind als solche unter genauer Angabe der Quelle gekennzeichnet. Die eingereichte Arbeit ist weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen. Sie ist weder vollständig noch in Teilen bereits veröffentlicht. Die beigefügte elektronische Version stimmt mit dem Druckexemplar überein.[1]

Remseck, 18.10.2023 ____________________

(Tana Deeg)

[1]This thesis is the result of my own independent work, and any material from work of others which is used either verbatim or indirectly in the text is credited to the author including details about the exact source in the text. This work has not been part of any other previous examination, neither completely nor in parts. It has neither completely nor partially been published before. The submitted electronic version is identical to this print version.

Contents

1 Introduction

1.1 Motivation

German personal name compounds such as *Villen-Spahn* ('villa-Spahn'), *Gold-Rosi* ('gold-Rosi') and *Folter-Bush* ('torture-Bush') are a rather infrequent phenomenon in the German language. They have the structure of determinative compounds and serve as a nickname for a usually well-known person. According to Belosevic (2022), personal name compounds are mostly evaluative, i.e. they evaluate the person behind the name in a positive or negative way. Further research on an evaluation across different groups of compounds (politics, showbusiness, sports) is proposed. This work will investigate the evaluative nature of 413 German personal name compounds that mostly have the structure of noun as modifier and last name as head. The 131 corresponding full names will be considered as well, e.g. *Jens Spahn* would correspond to *Villen-Spahn*. The context data of compounds and names was collected from Twitter and the Leipzig Corpora Collection. The valence value of these context words, based on a valence database of Köper and Schulte im Walde (2016), will be used to investigate the evaluative nature of compounds in comparison to their names. Furthermore, the relation to and function of the modifier will be examined. The valence values will then be used to verify whether there are noticeable differences between the groups of compounds. Afterwards, a linear regression will be implemented to predict a 'delta' value: the difference between name valence and compound valence. Several predictor variables such as name valence, compound valence, modifier valence, age, gender, political party and nationality will be used.

The results reveal that compounds are both positively and negatively evaluative in comparison to their full name while highlighting the reason why they were created. Compound valence and modifier valence are only partially correlated due to modifiers being involved rather accidentally or interpreted ironically. Lastly, noticable differences between the groups can be observed with politicians being the most negative group regarding their valence values. Conducting the linear regression with different combinations of predictor variables shows that compound valence is a highly significant predictor. Also, other variables such as modifier valence, age or political party are able to compose models that predict the delta value very well.

1.2 Personal name compounds

Personal name compounds (in the following: compounds) such as *Villen-Spahn* ('Villa-Spahn'), *Lockdown-Lauterbach* ('Lockdown-Lauterbach'), *Plagiat-Giffey* ('Plagiarism-Giffey') or *Hummer-Wagenknecht* ('Lobster-Wagenknecht') are nominal compounds that have the structure and characteristics of determinative compounds. Determinative compounds consist of a modifier and a head. Personal name compounds have an appellative or onymic constituent as a first part. This acts as a modifier and defines the name of the person. The second part (head) is the name of a person, more precisely, a first name, last name or nickname (Belosevic, 2022). "The compound modifiers contribute some important properties of the name-bearer or of events in which the name-bearer is involved" (Belosevic and Arndt-Lappe, 2021). In other words, the meaning of the modifier is the reason or at least related to the reason why this compound was created. It can describe a party affiliation, the appearance or characteristics of the person, an event in which the person was involved, etc. Compounds serve as nicknames of a person (Belosevic, 2022). This thesis will mainly focus on nouns as modifier and last names as head, connected by a hyphen. Table 1 provides examples for a regular determinative compound (noun as head) compared to personal name compounds (first or last name as head): E.g. *Kernkraft-Merkel* refers to the fact that Angela Merkel published a government declaration in 2011 that proposed the nuclear phase-out.

Furthermore, Belosevic (2022) claims that personal name compounds bear an evaluative and knowledge-evoking function. This evaluative nature will be further examined in this thesis. Additionally, Belosevic and Arndt-Lappe (2021) proposed the approach to sort the compounds into different semantic frames using the German FrameNet[2] in order to identifiy extra-linguistic patterns. Semantic frames are conceptual knowledge units. They provide a schematic representation of different situa-

[2]https://gsw.phil.hhu.de

tions, events or entities, characterized by different participants and semantic roles[3]. The frames were identified by using "knowledge about a specific discoursive event", e.g. Villen-Spahn: Spahn bought an expensive Villa. Therefore, Villen-Spahn can be annotated with the frame COMMERCE_BUY with Spahn being the "BUYER" and Villen being "GOODS" (Belosevic and Arndt-Lappe, 2021). This categorization will be used for this thesis in Section 4 as well.

	compound	modifier	head	context
determinative compound	steamship	steam	ship	
personal name compound	Kernkraft-Merkel	Kernkraft ('nuclear power')	Merkel	political action
personal name compound	CDU-Laschet	CDU	Laschet	party affiliation
personal name compound	Leg-Angelina	Leg	Angelina	description of appearance

Table 1: Examples: determinative compounds and personal name compounds.

1.3 Related work

Kürschner (2020) conducted research on nickname formation in West Germanic. This also included German compounds, a combination of a legal name and a lexeme with the legal name being placed in the final position, e.g. *Partykarl* ('party Karl') or *Drogen Marc* ('drug Marc'). A great variety of word types in the first position, such as lexical items, were found. Nevertheless, these compounds occured with a very low frequency.

[3]Source: https://gsw.phil.hhu.de/documentation/glossary#glossary-12

A comprehensive study on the semantics of German personal name compounds was carried out by Belosevic and Arndt-Lappe (2021). They used 532 determinative compounds from Twitter with a personal name as second component, e.g. *Laber-Lindner* ('Babble-Lindner'). In order to investigate the meaning of the relation of proper name and common noun, a frame-semantic approach was used, precisely speaking a categorization of compounds into frames of German FrameNet. They argue that proper name components of compounds evoke different types of knowledge, e.g. discursive knowledge about the history or the actions of the name-bearer. This discursive knowledge on the event the compound is based on is then used to find the according frame. Subsequently, similar frames were grouped to a more general frame in order to generalize the patterns of semantic relations. This shows that the meaning of personal name compounds can be generalized and is determined by extra-linguistic and semantic knowledge. Furthermore, the authors claim that personal name compounds are mostly evaluative, sometimes mocking and exaggerating. Compounds highlight specific events that e.g. damaged the reputation of the name-bearer or characterised the political actions.

Another study that is focused on German personal name compounds was made by Belosevic (2022). Three hypotheses from previous work such as personal name compounds are infrequent, irregular and bear mainly an evaluative function were tested and could not or only partially be confirmed. To do so, 1194 personal name compounds from DWDS Corpus and Twitter were used. To investigate the irregularity, a classification of semantic patterns of noun compounds based on Ortner et al 1991 was performed. Even though "the patterns are too abstract and neglect the contextual factors and extra-linguistic knowledge about the name-bearer" (Belosevic, 2022), the assumption that personal name compounds are irregular could be discarded. This was achieved by enriching the context by taking into account the linguistic knowledge about the modifier, the linguistic context as well as knowledge about the name-bearer that is not included in the semantic pattern. It was also made clear that compounds can not only bear an evaluative function, which is based on looks or characteristics of name-bearer and can potentially be discriminatory, but compounds can also bear a knowledge-evoking function. This function arises from extralinguistic knowledge, e.g. about political affairs and the cotext. Knowledge about the activities of the name-bearer play a central role. Lastly, the hypothesis of compounds being infrequent was rejected after discovering a large frequency in Twitter. Social media may play a key role in the construction of evaluative meanings. Beyond that, domain specific differences (politics, sports, showbusiness) regarding the pragmatic function of personal name compounds were discovered which proposes further research on a classification of compounds into different, domain specific groups.

2 Data and preprocessing

2.1 Compounds and names

This section explains how the target lists of compounds and names were created and modified in order to suit this investigation of the evaluative nature on the basis of valence values best.

Belosevic (2022) investigates the semantics and pragmatics of 1194 personal name compounds. 80% of the data used was retrieved by specifically searching for the string *name or *-name in the DWDS corpus WebXL, based on a concrete name list. The rest of the data was collected via a manual search using the option *erweiterte Suche* ('extended search') on the micro-blogging platform Twitter. This list of compounds was used as a basis for the target list of this thesis, but in a slightly modified and reduced version: The list was filtered for compounds that fit into the scheme modifier = noun, head = last name to ensure that as many valence values as possible are available. Furthermore, spelling mistakes of modifier or head were corrected. These compounds make up approximately 95% of the final list for this thesis. The rest of the data was taken by filtering the fitting candidates[4] out of a target list created by André Blessing that was generated on the basis of a name list. The final target list includes **413 German personal name compounds** of celebrities or other well known people (politicians, athletes, climate activists etc.). According to this final compound list, a name list was created. It includes **131 full names**, with at least one compound in the compound list existing for each of these names. These compounds and names will be referred to as *targets* in the following. Table 10 (Appendix) provides a complete overview of all targets, sorted alphabetically by compounds.

To find the maximum possible number of sentences that contain a name compound, the compound list was modified at character level. Each compound was duplicated and then adapted in the following steps in all possible combinations:

[4] Pre-filtering was done by Sabine Schulte im Walde.

- **Umlauts:** Replace umlauts: ä → ae, ö → oe, ü → ue.
  Example: *Bätschi-Nahles* → *Baetschi-Nahles*

- **Eszett:** Replace ß → ss.
  Example: *Spaß-Guido* → *Spass-Guido* ('fun-Guido')

- **Interfix:** Add or delete the interfix accordingly.
  Example: *Hoffnungs-Obama* → *Hoffnung-Obama* ('hope-Obama')

- **Alternative spelling:** Included spelling variations of words.
  Example: *Gazprom-Schröder* → *Gasprom-Schröder*

- **Singular/Plural:** Added or deleted a letter to get the singular/plural form of the modifier.
  Example: *Tore-Klose* → *Tor-Klose* ('goal-Klose')

- **Wildcard search:** Added a wildcard (limited to 0 - 2 characters) between modifier and head to find compounds without a hyphen/with a space/with a hyphen and hashtag/etc. inbetween.

2.2 Corpora

The overall goal was to build two subcorpora to work with, one corpus where each sentence contains at least one compound (variation) of the compound list and one corpus where each sentence contains at least one name of the name list. To do so, 30,875,753 sentences of Wortschatz[5] and 56,208 sentences of Twitter[6] were used.

Tweets:[7] To download the tweets that include names and/or name compounds, the command line tool and Python library for archiving Twitter JSON data "twarc2" were used. "twarc2" allows to search for a string in the whole Twitter archive. All

[5] https://wortschatz.uni-leipzig.de/de

[6] https://twitter.com/

[7] All tweets were searched and downloaded by André Blessing.

matches can be downloaded in a json file that includes the tweets as well as metadata. Every compound and every name was searched in the Twitter archive using "twarc2". Concering the compounds, only perfect matches of modifier and head were found, but the character inbetween could be any symbol such as "-", " " or "#".

Wortschatz: Leipzig Corpora Collection (Goldhahn et al., 2012) is a collection of German news corpora. In this work, one million sentences per year from 1995 to 2021 were used.

The following modifications were made to the sentences: The data from Wortschatz had one sentence per line with a line number at the beginning. As the sentences were alphabetically ordered or not ordered at all, the maximum meaningful context of a target was only one sentence, considering the sentence before and after was not possible. Therefore the line numbers could be discarded. The maximum context for a target occurring in Twitter data was one tweet. Twitter represents "retweets" with an URL at the end which corresponds to the original tweet. These retweets caused duplicate sentences that only differed in this URL. To avoid these duplicates influencing the correct number of target sentences, all URLs that start with http(s): were deleted, as well as "\n". These two modifications were made via the sed command in a linux shell, the replace was left empty to delete the pattern (sed -e 's/http[s]\?:\/\/\S*//g' -e 's/\\n//g').

The names and the compounds (including their variations) were searched in the Twitter and Wortschatz sentences via a linux shell script using the grep command and ignoring case (grep -i). There was no separation made between Wortschatz and Twitter data. If a target was found, the corresponding line (one sentence in case of Wortschatz, one tweet in case of Twitter) was added to the correct subcorpus. All words occurring in this line or these lines (in cases of multiple occurrences of a target) will be refered to as "context words of the target" in the following. The final **corpus of names** has **242,622 sentences** and includes all **131 names** of the name list. The **corpus of compounds** has **24,858 sentences** and includes

321 compounds of the 413 original ones. Both corpora additionaly display the respective target at the beginning of each sentence, separated by a tab. See Table 2 for further corpus statistics (including the targets) and Figure 1 for most frequent compounds.

	total number of sentences	**sentences from Twitter**	**sentences from Wortschatz**	**words per line (mean)**	**characters per line (mean)**
corpus of compounds	24,858	24,688	170	22.26	180.29
corpus of names	242,622	9,145	233,477	21.57	162.44

Table 2: Statistics of corpus of names and corpus of compounds.

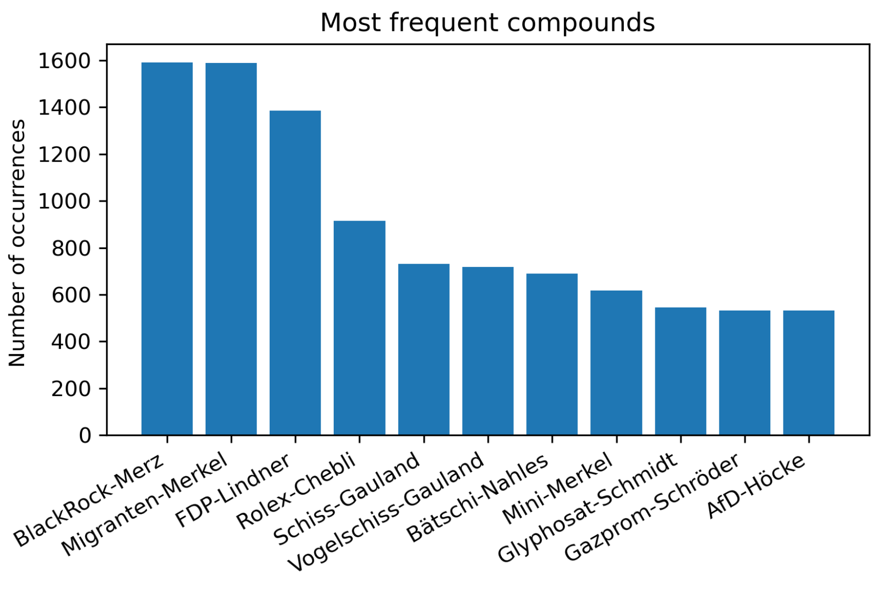


Figure 1: Most frequent compounds.

Important notes:
It is important to mention that the data is rather unbalanced. Wortschatz contains hardly any compounds, only 58 were found, most of them only once or twice. Consequently, the context words of the compounds come mainly from Twitter. Exactly speaking, 24,688 (99.32%) of the compound sentences are from Twitter, 170 (0.68%) are from Wortschatz. However, there is a huge number of names in Wortschatz data, so the context words of the names come mainly from Wortschatz. More precisely, 233,477 (96.23%) of the name sentences are from Wortschatz, 9,145 (3.77%) are from Twitter. Twitter and Wortschatz provide data from different genres, which possibly influences their evaluative nature in different ways. Unfortunately, the Twitter API was deprecated/disabled mid-research which made the collection of further data impossible. Deleting data to get a balanced distribution was no option either, considering the fact that personal name compounds are a rare phenomenon that comes with sparse data issues. Moreover, it was planned to search for all variations of targets in the Twitter archive, but facing this shutdown, tweets that include a variation could only be retrieved if a variation occurred in a tweet/sentence with an original target.

2.3 Valence database

Valence represents the pleasantness of a stimulus. Words with low valence are less pleasant, e.g. *böse* ('mean'), *Zahnschmerzen* ('toothache'), or *scheitern* ('to fail'). Words with high valence are more pleasant, e.g. *wunderbar* ('wonderful'), *Freude* ('joy'), or *begeistern* ('to excite'). Köper and Schulte im Walde (2016) created a database of 2,275,234 words with automatic ratings for abstractness/concreteness, arousal, anger, valency, disgust, fear, happiness, joy, and sadness. Valence ratings range from 0 to 10 with 0 and 10 denoting low and high valence, respectively. The words are all lowercased and not lemmatized. A feed forward neural network was trained to fit the human annotated gold rating for a given word. The model then predicted a rating score for every word. These valence values will be used to investigate and rate the evaluative nature of the targets and their context.

3 Experiments on Valence

To evaluate the pleasantness/evaluative nature of a name compound and its corresponding full name, the mean valence of the context of each target (name or compound) was computed. Additionally, the valence of the modifier of each compound was looked up in the valence database. Finally, the names and compounds were split into different subgroups, according to their profession. This allows comparisons between compound and name, between compound and modifier and an evaluation across different groups/professions.

3.1 Basic calculations

All context words of a target were tagged with part-of-speech (POS) labels using the probabilistic TreeTagger (German) by Helmut Schmid[8] (Schmid, 1999). It was developed in the TC project at the Institute for Computational Linguistics of the University of Stuttgart. This Treetagger outputs the corresponding POS-tag and the lemma of the input word. This lemma was used for all proceeding actions. In cases of a lemma being unknown to the TreeTagger, the full word form was used to avoid losing context data. The context words were then filtered to exclude words such as articles, prepositions, pronouns, modal verbs, punctuation, etc. of which the valence value has no useful or interpretable meaning.

Included POS-Tags:

- **NN**: simple noun
- **ADJA**: attributive adjective
- **ADJD**: predicative or adverbial adjective
- **VVFIN**: finite full verb

[8] https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

- **VVIMP**: imperative (full verb)
- **VVINF**: infinitive (full verb)
- **VVIZU**: infinitive with incorporated "zu" particle (full verb)
- **VVPP**: past participle (full verb)

In the following step, the valence score for each context word was looked up in the database created by Köper and Schulte im Walde (2016). If a valence value of a word was unavailable, the word was not considered. The valence of a target (t) is defined by the mean valence of all its context words (W_t). W_t represents all context words as a "bag of words" (duplicate entries allowed), therefore the valence of every context word is weighted by the number of its occurrences. The valence of a target word ($valence(t)$) was calculated as follows:

(1)
$$valence(t) = \frac{1}{|W_t|} * \sum_{w \in W_t} valence(w)$$

3.2 Name vs compound

Personal name compounds act as a nickname for a person. This nickname represents any kind of characteristic of the person or a special event/action in which the person was involved. Therefore, it can be assumed that personal name compounds are rather evaluative compared to their corresponding name. This section will look into the relation between name and compound, calculate the correlation and find reasons for big and small gaps between name valence and compound valence. In order to avoid outliers because of sparse data, only compounds that occured five times or more were considered in this section. Consequently, names were only included if they have at least one corresponding compound with five ocurrences or more.

3.2.1 Method

Compound valence and name valence were, as described in Section 3.1 (basic calculations), calculated based on the valence values of the context words. After excluding compounds that appeared 4 times or less and compounds without a valence value, there were a total of **216 name valence - compound valence pairs** left. The difference between name and compound was calculated by subtracting the name valence from the compound valence in order to compare compounds that are more positive than their corresponding name (positive difference) to compounds that are more negative than their corresponding name (negative difference). To investigate the correlation between name and compound, the python method pearsonr from scipy.stats was used. Details can be found in Table 10 (Appendix).

3.2.2 Results

Figures 2 and 3 respectively show the ten compounds with the highest and lowest valence, Figures 4 and 5 show the ten names with the highest and lowest valence. Additionally, the number of occurrences per target is provided as label of the x-axis. Figure 6 shows the valence of compounds (orange) in relation to their full name (blue), sorted by name valence (ascending). This provides an overview of compounds that are more positive or more negative than their name as well as an overview of the overall distribution of compound valence and name valence.

The lowest name valence is 4.44: *Karlheinz Schreiber*, the highest name valence is 5.21: *Bastian Schweinsteiger*. The arithmetic mean of all name valence values is 4.83. The name valence values therefore spread over the range of 4.44 to 5.21, whereas the compounds are distributed over the much bigger interval of 3.95 - 5.89. The lowest compound valence is 3.95: *Folter-Bush* ('torture-Bush'), the highest compound valence is 5.89: *Tore-Klose* ('goal-Klose'). The arithmetic mean of all compound valence values is 4.8. The compounds seem to be quite evenly distributed over and under the name valence line, but slightly more negative. More precisely, in 91 cases the compound valence is more positive than the name valence, in 125 cases it is more negative. The mean difference (compound - name) is -0.03 (standard devia-

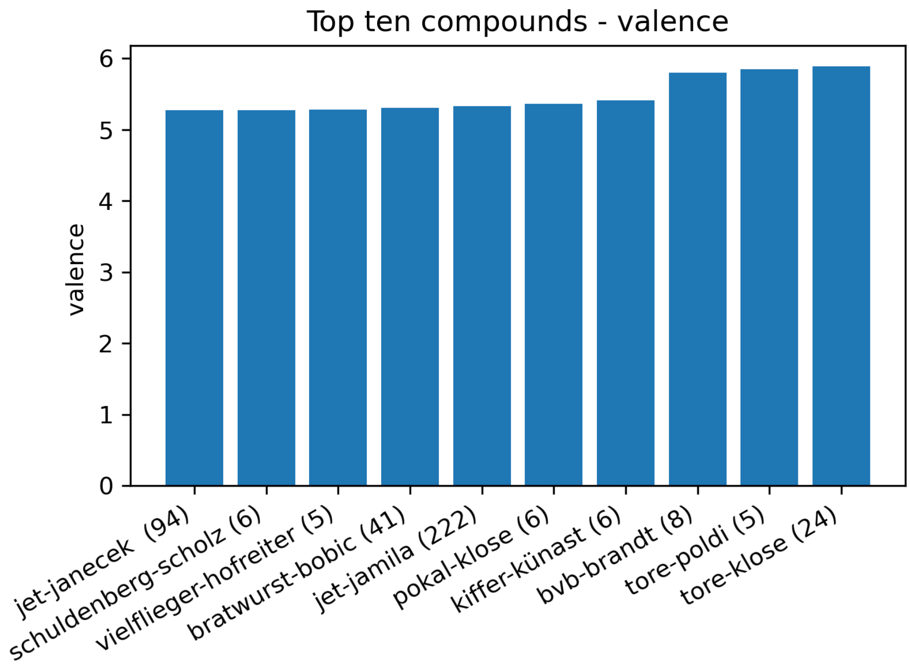


Figure 2: Ten compounds with highest valence value.

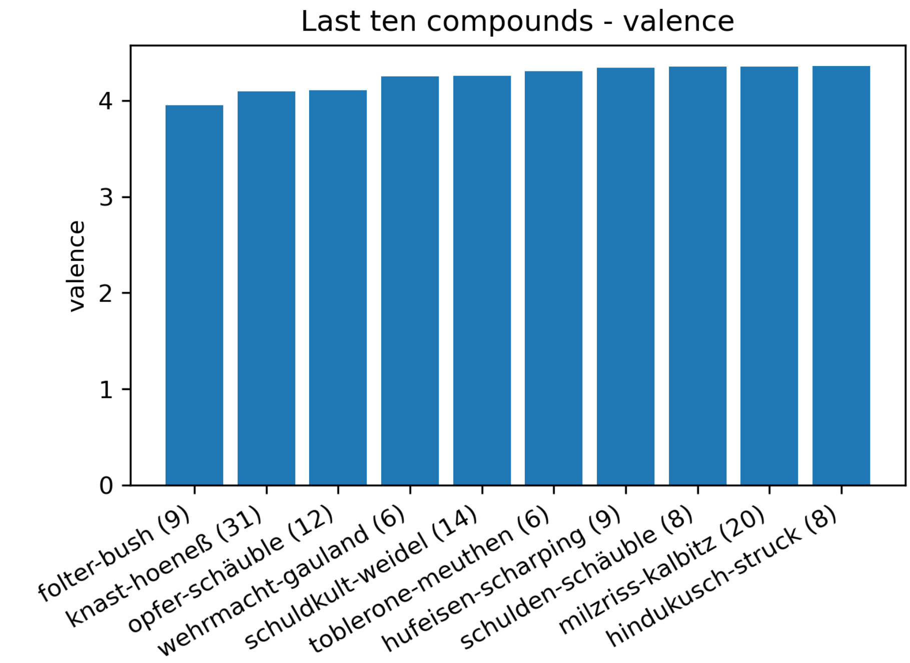


Figure 3: Ten compounds with lowest valence value.

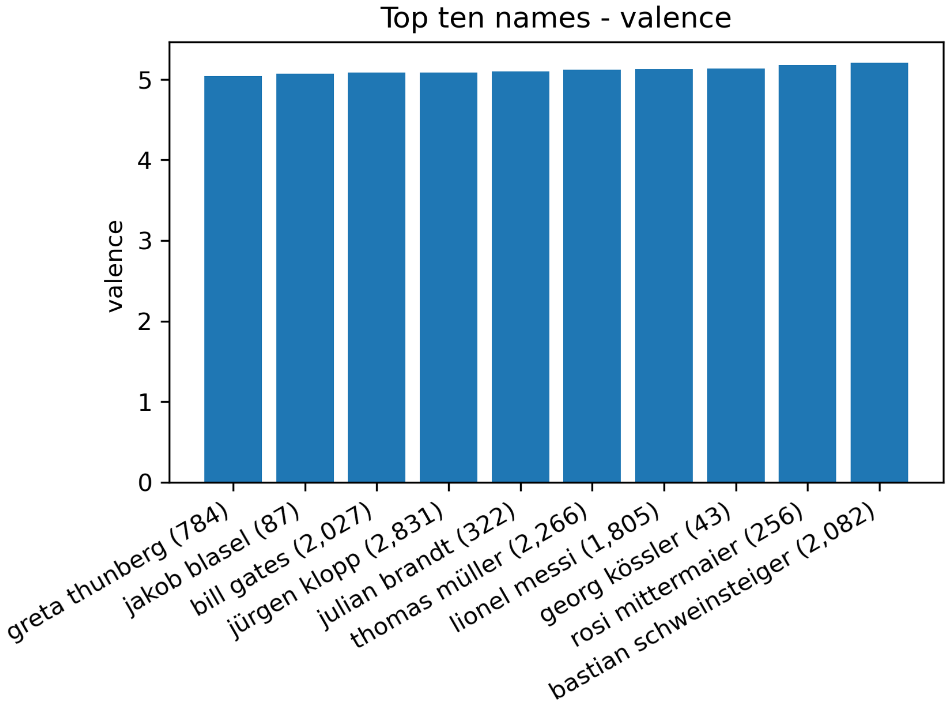


Figure 4: Ten names with highest valence value.

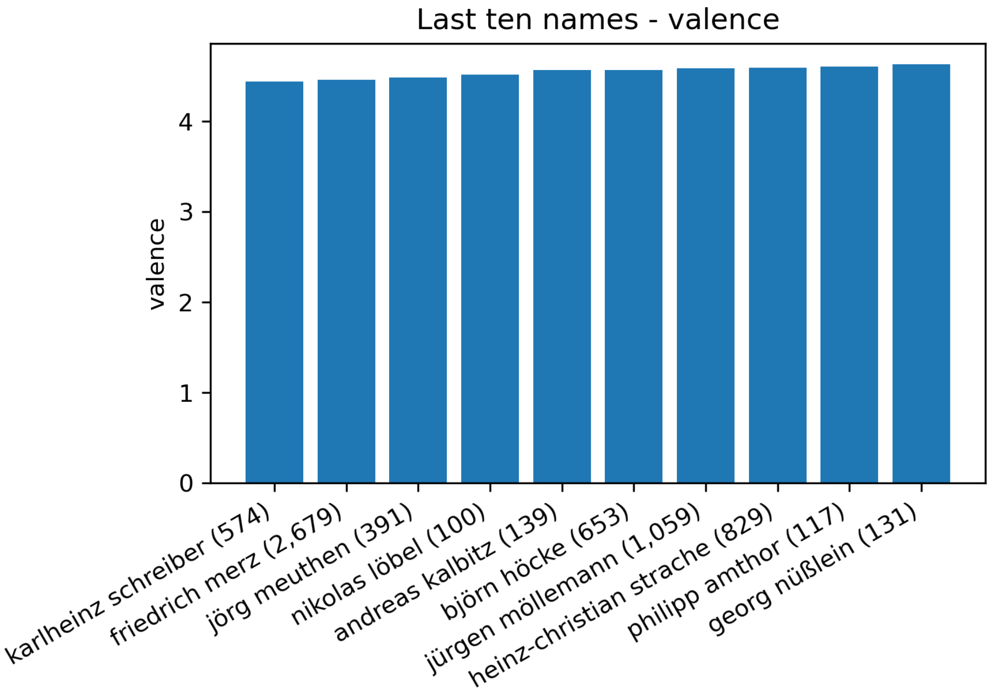


Figure 5: Ten names with lowest valence value.

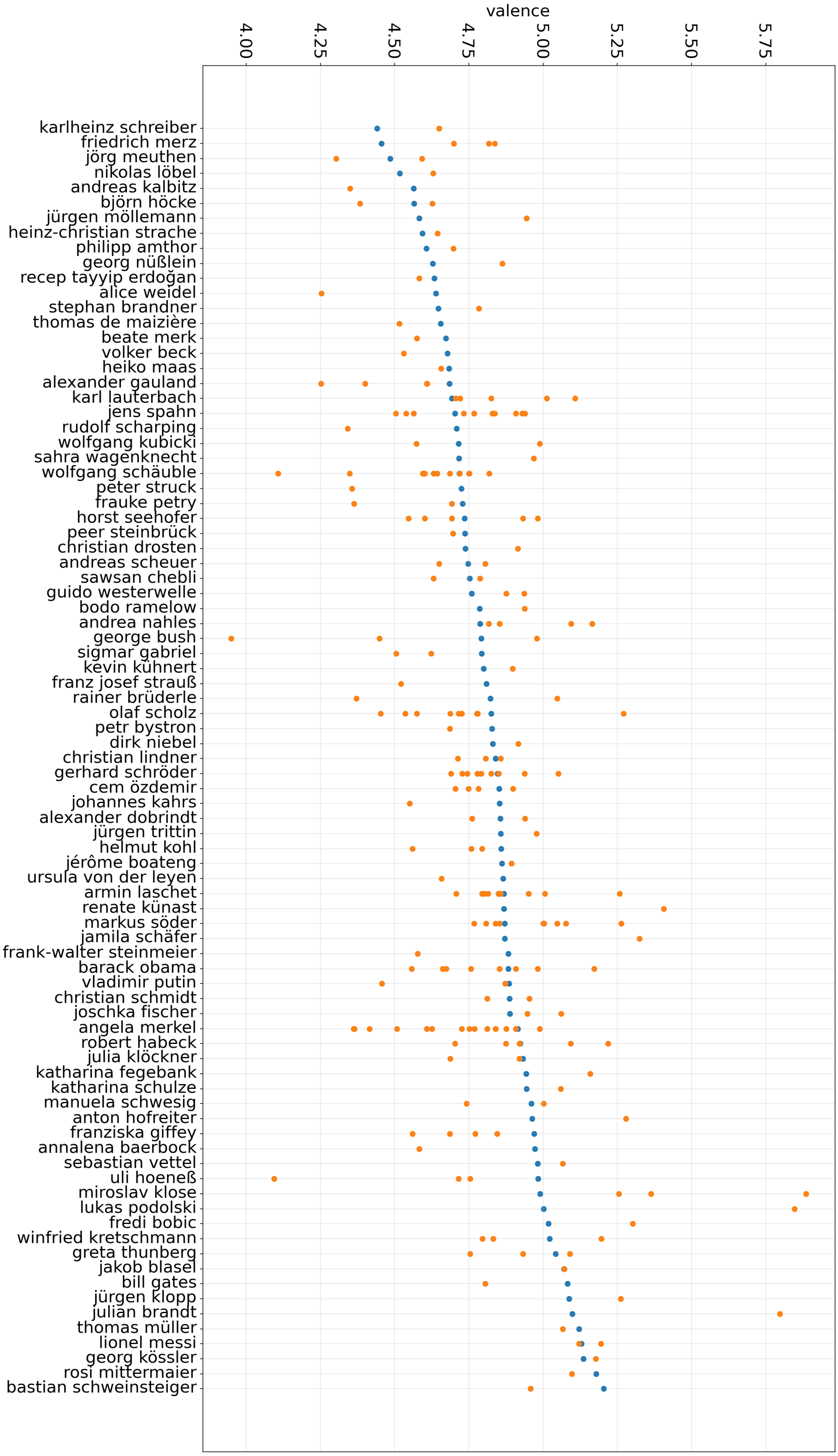


Figure 6: Comparison of name and compound valence.

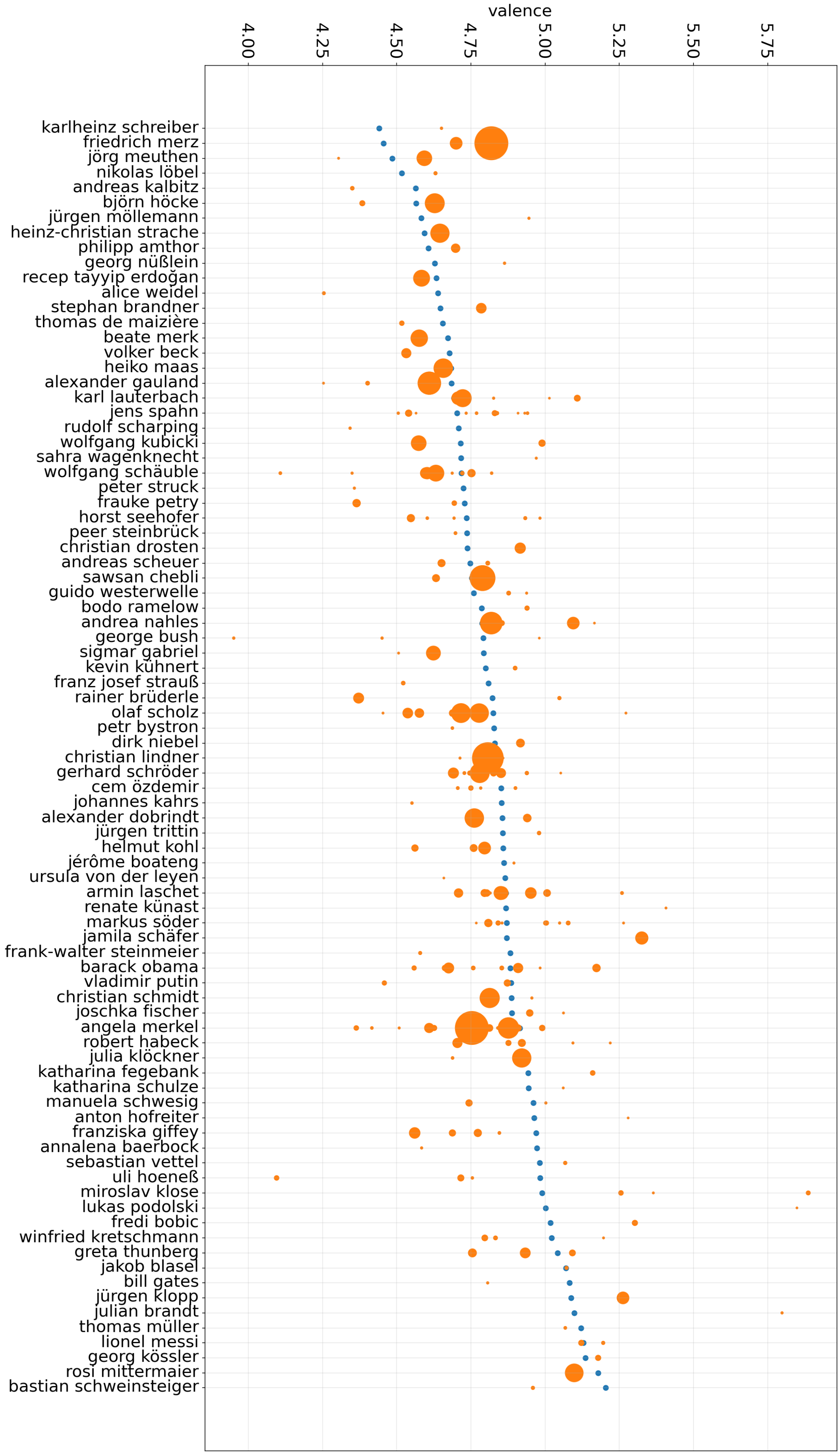


Figure 7: Comparison of name and compound valence including frequency information.

tion: 0.25). All in all, the compound values seem to increase slightly with increasing name valence.
Pearsons correlation coefficient was calculated to validate the assumption that name valence correlates with compound valence. The python method pearsonr from scipy.stats delivered a **correlation coefficient of 0.44** and a p-value of $2.17 * 10^{-11}$. Consequently, the moderately positive correlation of 0.44 is significant.
Figure 7 additionally displays the number of occurrences per compounds: the bigger the dot, the more occurrences there are. The absolute number was given as an argument to the dot size function of the plot. It becomes clear that most compounds do not appear too often, there are only few very big points. Additionally, it can be observed that most of the bigger dots are located near the name valence line. The outlier points, on the contrary, are mostly very small.
As a result, there are three important observations: Overall, the compound valence is slightly more negative than the name valence, the compounds are distributed over a bigger valence range than the names and the correlation is only moderately positive. To find out about the possible reasons for these three findings, it is crucial to look into name compound pairs in detail. The three pairs with the biggest differences between name valence and compound valence were picked out for each of the two cases "compound valence more positive than name valence" and "compound valence more negative than name valence". The context words considered in the following are part of the 15 most frequent ones per target.

Compound valence > name valence:

- *Miroslav Klose*: 4.99 - *Tore-Klose* ('goal-Klose'): 5.89
  - Difference: 0.9
  - 1,886 name occurrences, 24 compound occurrences
  - The context words of both compound and name are very positive, as Klose is a successful athlete. Considering the fact that *Tore-Klose* ('goal-Klose') refers to the specific positive act of scoring many goals, the compound is

more positively evaluative. Frequent context words of the compound such as *feiern* ('celebrate'), *herrlich* ('wonderful') and *gold* ('gold') support this assumption.

- *Lukas Podolski*: 5.00 - *Tore-Poldi* ('goal-Poldi'): 5.85
  - Difference: 0.85
  - 1,808 name occurrences, 5 compound occurrences
  - The context words of both compound and name are very positive, as Podolski is a successful athlete. Considering the fact that *Tore-Poldi* ('goal-Poldi') refers to the specific positive act of scoring many goals, the compound is more positive.
- *Julian Brandt*: 5.1 - *BVB-Brandt*: 5.8
  - Difference: 0.7
  - 322 name occurrences, 8 compound occurrences
  - The context words of both compound and name are very positive, as Brandt is a successful athlete of the football club BVB. The compound *BVB-Brandt* occurs in contexts that refer to his successful career in this club, thus the compound is slightly more positive.

Compound valence < name valence:

- *Uli Hoeneß*: 4.99 - *Knast-Hoeneß* ('prison-Hoeneß'): 4.1
  - Difference: -0.89
  - 4,007 name occurrences, 31 compound occurrences
  - Uli Hoeneß was a successfull football player. The compound refers to the fact that he spent nine months in prison due to tax evation, which explains the more negative value of the compound compared to the name itself.

- *George Bush*: 4.79 - *Folter-Bush* ('torture-Bush'): 3.95
  - Difference: -0.84
  - 1,580 name occurrences, 9 compound occurrences
  - The compound *Folter-Bush* ('torture-Bush') specifically refers to the fact that Bush supported the torture methods carried out by the CIA. This makes the compound more negative than the name. Context words of the compound such as *grausamkeit* ('cruelty'), *isolation* ('isolation'), *sanktion* ('sanction') support this.
- *Wolfgang Schäuble*: 4.72 - *Opfer-Schäuble* ('victim-Schäuble'): 4.11
  - Difference: 0.61
  - 6,692 name occurrences, 12 compound occurrences
  - The compound refers to a controversial statement of Schäuble about an attack that involved the word *Opfer* ('victim'). The compound valence is consequently slightly more negative than the name, whose most frequent context words like *bundesfinanzminister* ('federal finance minister'), *sagen* ('to say'), *deutsch* ('German') are in sum quite neutral.

3.2.3 Discussion

These specific examples show that the compounds actually highlight the reason for which they were created. This is reflected in the respective context words. The basis of the compounds are special, controversial or notable statements or actions, in short, things that draw attention. In order to achieve this, the action or statement either has to be very positive or very negative. This explains why the compound valence values are spread over a much bigger range than the name valence values - they are positively and negatively evaluative, compared to their name. There is also another factor here: The number of occurrences of compounds remain within the two-digit to three-digit range, only in some exceptional cases within the three to four-digit range. However, the number of occurrences of names almost entirely remains within the three to four-digit range. As a low number of occurrences leads to outliers, having less context words increases the probability of more extreme values, more context words introduce more neutral words that push the mean value towards 5, the middle of the scale. In addition to this, the difference in the origin of the context words must be considered. The distribution over Wortschatz and Twitter is not even close to equal and Twitter and Wortschatz provide data from different genres, which possibly influences their evaluative nature in different ways. As Twitter is more informal than the news corpus, it might produce more evaluative statements. Finally, most of the outlier points of the compounds don't have many occurrences. These outliers in turn influence the level of correlation, which is only weak positive. As more compounds are more negative than their corresponding name, more compounds seem to be negatively evaluative by representing a negative action or event.

3.3 Compound vs modifier

Determinative compounds are characterized by the fact that one constituent modifies the other. In case of personal name compounds, a noun modifies the head which is a first, last or nickname. Naturally, one would assume that the meaning of the modifier, apart from the name itself, will influence the way the whole compound is perceived. As the meaning of the modifier is the reason why the compound was created, its perception should be similar to the context in which the compound is going to be used. This section will therefore investigate whether there is a connection between the valence of a compound and the valence of its modifier and if yes, by which aspects this relation is characterized. In order to avoid outliers because of sparse data, only compounds that occured five times or more were considered in this section.

3.3.1 Method

The lemma of the modifier was used to make the following calculations. Lemmatization was done manually, as automatic lemmatization (TreeTagger/Spacy) yielded too many wrong results. The valence value was then looked up in the database. If a word appeared twice in the valence database, the first valence value was selected randomly. 374 values were found in the database. After excluding compounds that appeared four times or less and compounds without a valence value, there were a total of **203 compound valence modifier valence pairs** left. The difference between compound and modifier was calculated by subtracting the compound valence from the modifier valence. To investigate the correlation between compound and modifier, the python method pearsonr from scipy.stats was used. Details can be found in Table 11 (Appendix).

3.3.2 Results

Figures 8 and 9 show the valence of compounds (blue) in relation to their modifier (orange), sorted by compound valence (ascending). The following aspects can be observed: The lowest compound valence is 3.95: *Folter-Bush* ('torture-Bush'), the highest compound valence is 5.89: *Tore-Klose* ('goal-Klose'). The arithmetic mean of all compound valence values is 4.81. Thus, most of the compounds are located between 4 and 6. The modifiers, on the contrary, are spread over a much wider range: The lowest modifier valence is 0.89: *folter* ('torture') - *Folter-Bush*, the highest modifier valence is 7.9: *willkommen* ('welcome') - *Willkommens-Merkel*. The arithmetic mean of all modifier valence values is 4.22. There are many modifiers with a very low value: 44 modifiers have a value lower than 3. Especially the modifiers that belong to compounds with low valence values are very negative, 38 of those 44 modifiers belong to compounds in the lower half. Generally, the majority of modifiers is located under the compound valence line. To be precise, in 71 cases the modifier valence is more positive than the compound valence, in 132 it is more negative. The mean difference (modifier - compound) is -0.58 (standard deviation: 1.33). All in all, the modifier values seem to increase very slightly with increasing compound values.
Pearsons correlation coefficient was calculated to validate the assumption that compound valence correlates with modifier valence. The python method pearsonr from scipy.stats delivered a **correlation coefficient of 0.47** and a p-value of $2.19 * 10^{-12}$. Consequently, the moderately positive correlation of 0.47 is significant.

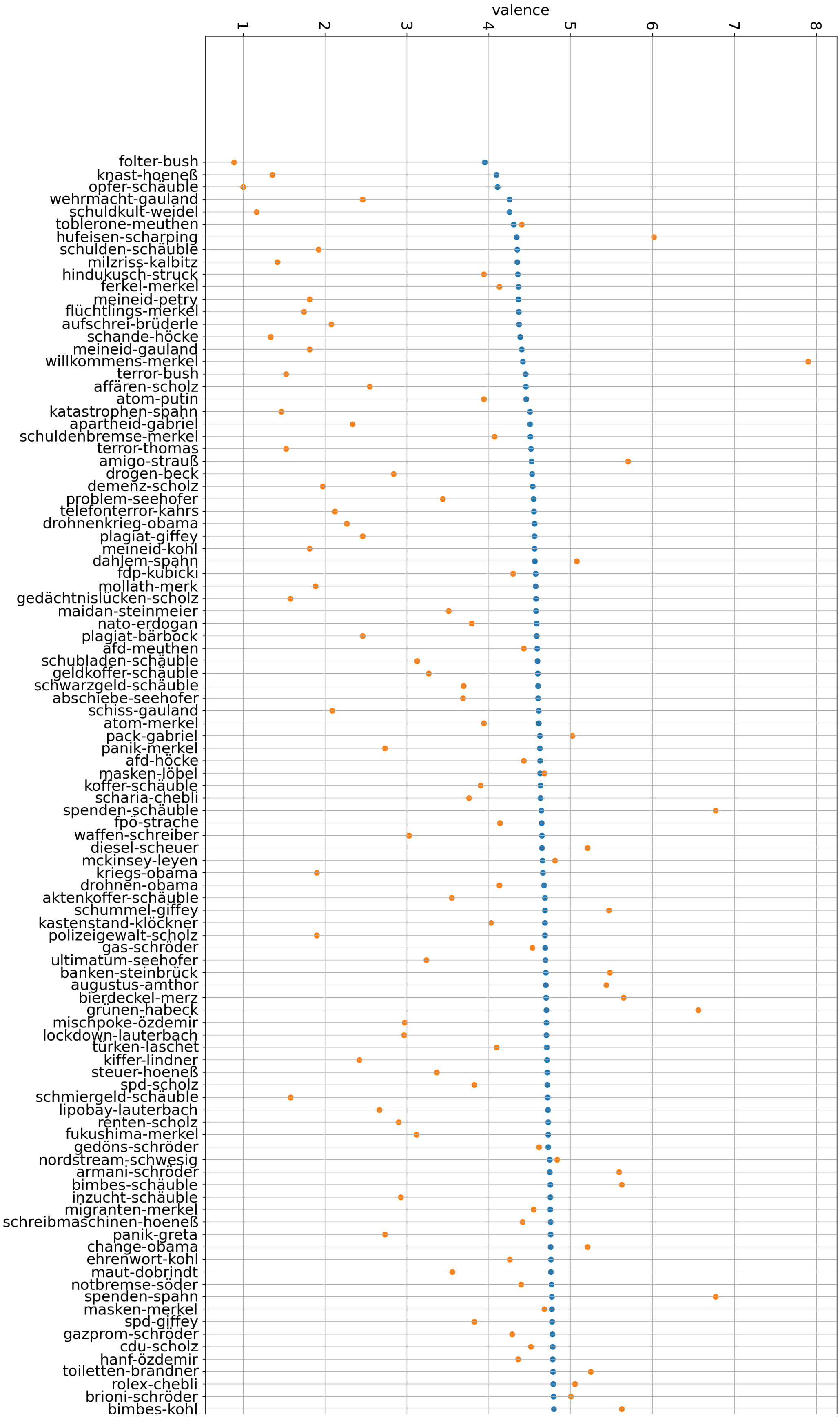


Figure 8: Comparison of compound and modifier valence, first half.

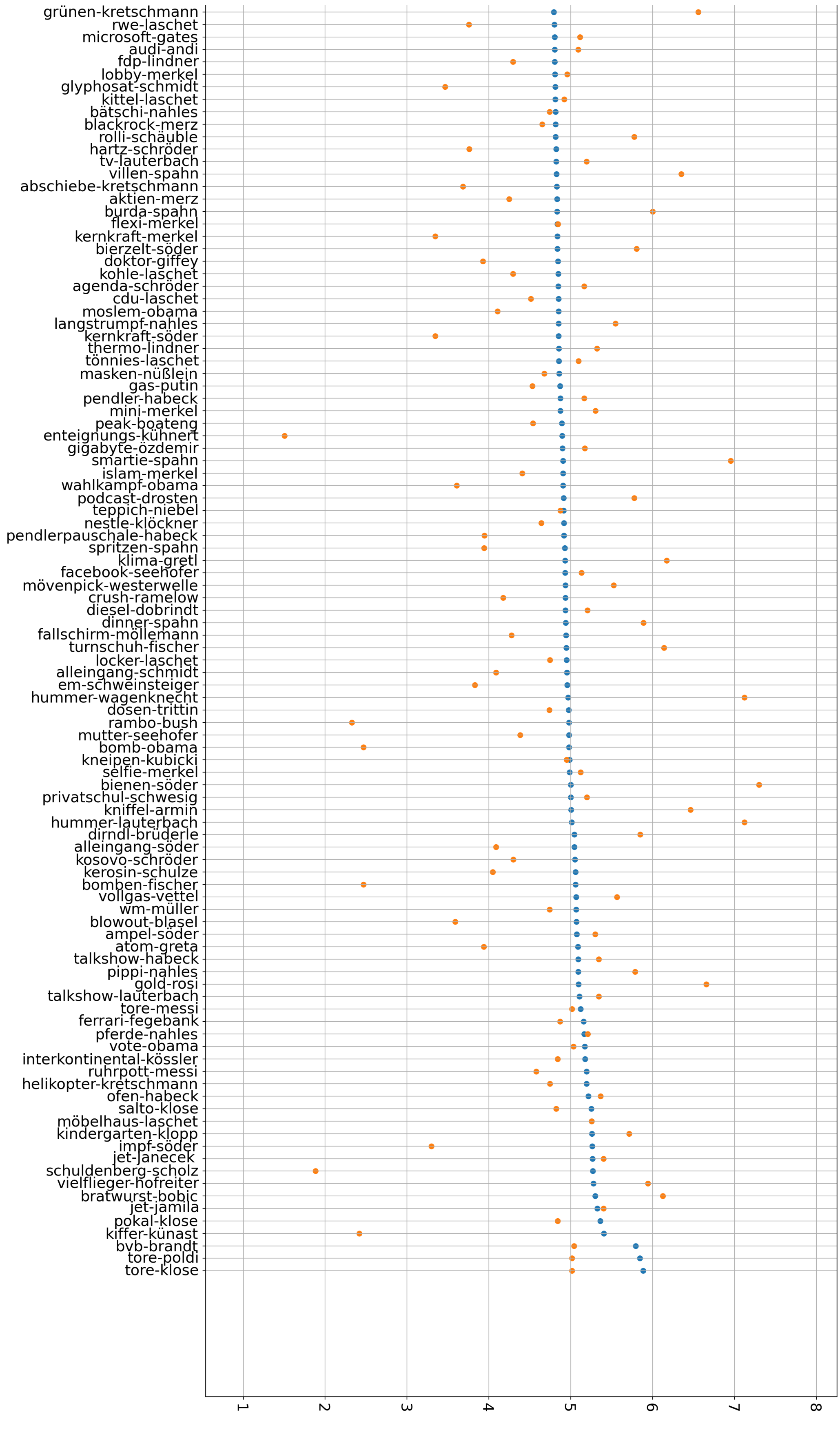


Figure 9: Comparison of compound and modifier valence, second half.

Again, there are three important observations: Overall, the modifier valence is slightly more negative than the compound valence, the modifiers are distributed over a bigger valence range than the compounds and the correlation is only moderately positive. To find out about the possible reasons for these three findings, it is crucial to look into compound modifier pairs in detail. The five pairs with the biggest differences between compound valence and modifier valence were picked out for each of the two cases "modifier valence more positive than compound valence" and "modifier valence more negative than compound valence". The context words considered in the following are part of the 15 most frequent ones per target.

Modifier valence > compound valence:

- *Willkommens-Merkel*: 4.42 - *willkommen* ('welcome'): 7.9
  - Difference: 3.48
  - 11 occurrences
  - Context words such as *abschiebung* ('deportation'), *verheerend* ('devastating'), *kritik* ('criticism'), *endgültig* ('final') indicate an ironic use of the word *willkommen*, as the context words convey the opposite meaning.
- *Bienen-Söder*: 5.00 - *biene* ('bee'): 7.3
  - Difference: 2.3
  - 27 occurrences
  - *Bienen-Söder* refers to the fact that Söder pleaded for a referendum to save bees and insects. As for any politician, some people support him and his actions, others do not. Context words such as *verteidigen* ('to defend'), *ausgepfiffen* ('booed') or *lachen* ('to laugh') support this mixture of meanings that results in a relatively neutral value.

- *Hummer-Wagenknecht*: 4.97 - *hummer* ('lobster'): 7.12
  - Difference: 2.15
  - 7 occurrences
  - *Hummer-Wagenknecht* refers to a specific event that (randomly) involved a lobster. Wagenknecht was once fotographed in a restaurant eating a lobster. This word therefore only has a symbolic meaning for a controversial event. Context words such as *partei* ('(political) party'), *blamabel* ('embarrassing'), *foto* ('photo') or *bestellen* ('to order') summarize this incident quite well.
- *Spenden-Schäuble*: 4.64 - *spende* ('donation'): 6.77
  - Difference: 2.13
  - 21 occurrences
  - Schäuble was involved in a donations affair in 2000, a clearly negative event. The word *Spende* ('donation') is intuitively quite positive, but in different contexts it can convey a more or less pleasant meaning. This compound occurred together with context words such as *politiker* ('politican'), *kriminell* ('criminal'), *finanzminister* ('finance minister') and *affäre* ('affair') which mirror the donations affair very well.
- *Hummer-Lauterbach*: 5.01 - *hummer* ('lobster'): 7.12
  - Difference: 2.11
  - 5 occurrences
  - Just like *Hummer-Wagenknecht*, *Hummer-Lauterbach* refers to a specific event that (randomly) involved a lobster. Lauterbach published a controversial tweet during a dinner that involved a lobster. This word therefore only has a symbolic meaning.

Modifier valence < compound valence:

- *Enteignungs-Kühnert*: 4.9 - *enteignung* ('expropriation'): 1.51
  - Difference: -3.39
  - 22 occurrences
  - The context words are, concerning their valence values, very mixed and often neutral: *partei* ('(political) party'), *sagen* ('to say'), *wahl* ('election'), *wohnung* ('apartment'). The valence value of the modifier, on the contrary, is very extreme.
- *Schuldenberg-Scholz*: 5.27 - *schuldenberg* ('debt mountain'): 1.89
  - Difference: -3.38
  - 6 occurrences
  - The context words are, concerning their valence values, again very mixed and often neutral. Some examples would be *milliarde* ('billion') *euro* ('euro'), *schuld* ('debt') or *erklären* ('to explain'). This results in a nearly neutral compound valence value and therefore a big contrast in comparison to the modifier appears.
- *Schmiergeld-Schäuble*: 4.72 - *schmiergeld* ('bribe'): 1.58
  - Difference: -3.14
  - 19 occurrences
  - The modifier is very extreme in its meaning, which is contrary to a relatively neutral compound valence value. The latter is composed of context words like *finanzminister* ('finance minister'), *deutsch* ('German'), *gehören* ('to belong') or *schmiergeld* ('bribe'). Nevertheless, the context words represent the event of being involved in a donations affair quite accurately.

- *Opfer-Schäuble*: 4.11 - *opfer* ('victim'): 1
  - Difference: -3.11
  - 12 occurrences
  - In 2006, Schäuble was criticised for a statement about an attack that involved the word *Opfer* ('victim'). Even though this event is rather negative, the extreme meaning and value of the modifier and nearly neutral compound valence make the difference considerably big. The low to neutral compound valence is formed out of context words such as *opfer* ('victim'), *pervers* ('perverse'), *weltkrieg* ('world war') or *bürgermeister* ('mayor').
- *Schuldkult-Weidel*: 4.26 - *schuldkult* ('culture of guilt'):1.17
  - Difference: -3.09
  - 14 occurrences
  - The modifier is very extreme in its meaning and its valence value. *Schuldkult* is a derogative term, mostly used by people of the far right political spectrum. It describes the culture of remembering the National Socialism. Consequently, the context words are predominantly negative, but several neutral and slightly positive words pull the mean value towards the middle: *schäbig* ('shabby'), *lügen* ('to lie'), *rassistisch* ('racist'), *helfen* ('to help') are some frequent examples.

Without excluding compounds with less than five occurrences, 40% of the ten compound-modifier pairs with the highest difference between modifier and compound had less than five occurrences.

3.3.3 Discussion

In summary, various aspects influence the moderate correlation, the bigger range of the modifiers and the lower values of modifiers compared to compounds: In a few cases, the modifier represents a special event that maybe accidentally involved the modifier, such as *Hummer-Wagenknecht* which refers to Wagenknecht being photographed while eating a lobster. Consequently, the valence of the modifier is not representative for the context in which the compound occurs. Also, there are modifiers that are not interpreted literally but ironically or interpreted differently depending on the context, e.g. *Willkommens-Merkel*. This explains the big gap between modifier valence and compound valence, which in turn results in the only moderately positive correlation. Another reason can be the very extreme valence values of the modifiers, which in many cases are very negative. This can happen due to two reasons: The word actually has a very extreme meaning or the valence value is not perfectly accurate/intuitive. In contrast to a very high or low modifier valence, the compound valence values seem to gather around 5, the middle of the 11 point scale. Although the context words were filtered for nouns, verbs and adjectives, there is still a big amount of relatively neutral words that pull the mean valence towards the middle of the scale. These could be reasons for the wider range of the modifiers. Lastly, sparse data is, of course, a problem. Compounds with only a few occurrences are hardly representative - very few context words more likely have a very high or very low arithmetic mean. Even when excluding compounds with less than five occurrences, compounds with few occurrences almost exclusively have context words that occur only once or twice.

3.4 Comparison of different groups

The compound list and corresponding name list contain people from very different professions and groups. According to Belosevic (2022), a cursory examination showed that targets from sports or showbusiness tend to have a positive evaluation. To validate if there are noticeable differences between the groups, this section will investigate the differences in the evaluative nature according to valence values of both compounds and names. As with excluding compounds with less than five occurrences one category will drop out completely, the filtered as well as the non-filtered (all data) results will be considered and compared.

3.4.1 Method

In order to compare between different groups, the compound list and the name list were both split into one of four subgroups. The groups were worked out manually from the lists. Each target was sorted into exactly one subgroup:

- **Politics:** This subgroup includes names/compounds of mainly German politicians such as Angela Merkel, Wolfgang Schäuble or Olaf Scholz, but also international politicians like Barack Obama, Boris Johnson or François Hollande.
- **Sports:** Athletes of different sports such as Lionel Messi (football), Sebastian Vettel (formula 1) or Rosi Mittermaier (skiing) can be found in this subgroup.
- **Showbusiness:** This group consists of celebrities in the music, fashion or acting industry like Angelina Jolie, Karl Lagerfeld or Lady Gaga.
- **Others:** Climate activists (e.g. Greta Thunberg), virologists (e.g. Christian Drosten) and lobbyists (e.g. Karlheinz Schreiber) are part of this subgroup.

Table 3 provides an overview of the numbers of compounds and names per group. Since this section considers compounds with more than five occurrences and their corresponding names apart from the whole data, the filtered numbers are shown as well.

3.4.2 Results

Figures 10 and 11 show an overall comparison of target valence values across all groups and a complete group including all targets for both compounds and names.

	politics	sports	showbusiness	others	total
compounds	354	43	5	11	**413**
compounds filtered	193	17	0	6	**216**
names	95	25	5	6	**131**
names filtered	68	12	0	4	**84**

Table 3: Number of targets per group.

The targets are not filtered and all targets, independent of their numbers of occurrences, are displayed. The size of the box reaches from first to third quartile and therefore contains 50 % of the data. The length of the whiskers is 1.5 * inter-quartile range (default value). The median is shown as an orange line, the arithmetic mean as a green triangle. Figure 10 additionally shows the outlier points that do not lie within the box plus 1.5 * inter-quartile range. The settings of Figures 12 and 13 are analogous to 10 and 11, the only difference is that compounds with less than five occurrences and their corresponding names were excluded. In Figure 10, the outlier points are striking: Concerning compounds politicians in both directions (2.91 to 6.48) and concerning compounds athletes in the negative direction, starting at 2.19. Figure 11 shows that the (arithmetic) mean differs a lot between the groups, but is fairly similar from compound to name within the groups. Politicians (compounds and names) have the lowest mean value, both under 5. The group *others* is slightly more positive. Athletes, both compounds and names, have a mean value of more than 5 and showbusiness are the most positive concerning the mean value. Details can be found in Table 4. This order of groups from negative to positive is reflected in the position of the boxes (50% of data) as well. In all groups, the range (length of box + whiskers) of the compounds is bigger than the range of the names. In all groups, the compounds reach more into the positive than the names, except for the category *others*, in which they also reach more into the negative. The overall category is very similar to the politicians. After excluding compounds with less than five occurrences and their corresponding names, there are still some outliers left, but in a much smaller range than before, as can be seen in Figure 12. Outliers of politician compounds spread over the range of 3.95 to ca. 5.41, outliers of athlete compounds from 4.1 to 5.89. Furthermore, the category showbusiness is empty, as there are no compounds with enough occurrences in this subgroup. Figure 13 shows that the (arithmetic) mean differs a lot between the groups, but is fairly similar from compound to name within the groups. Again, politicians are (concerning the mean value) the most negative and athletes the most positive. Also, the range (length of box + whiskers) of the compounds is bigger than the range of the names in all groups. Overall, this range is slightly smaller than in Figure 11, concerning the compounds. As the names all occur very often and cutting compounds affects random names, their range does not change remarkably.

Figures 14 (politicians), 15 (athletes) and 16 (*others*) compare name valence (blue) and compound valence (orange) per group, sorted by name valence (ascending). The dot size of the compounds refers to the number of the occurrences. Figure 14 shows that many compounds of politicians are very similar to their name. Most of the outlier points are relatively small. The majority of compounds is more negative than their names. In the case of the athletes, the compounds are nearly evenly distributed since they are more positive or negative than their names, but slightly more positive, as Figure 15 shows. The same finding applies to Figure 16, *others*.

	politics	**sports**	**showbusiness**	**others**	**total**
compounds	4.80	5.07	5.12	5.04	4.84
compounds filtered	4.77	5.16	-	4.90	4.80
names	4.80	5.05	5.22	4.85	4.87
names filtered	4.79	5.06	-	4.82	4.83

Table 4: Arithmetic mean per group.

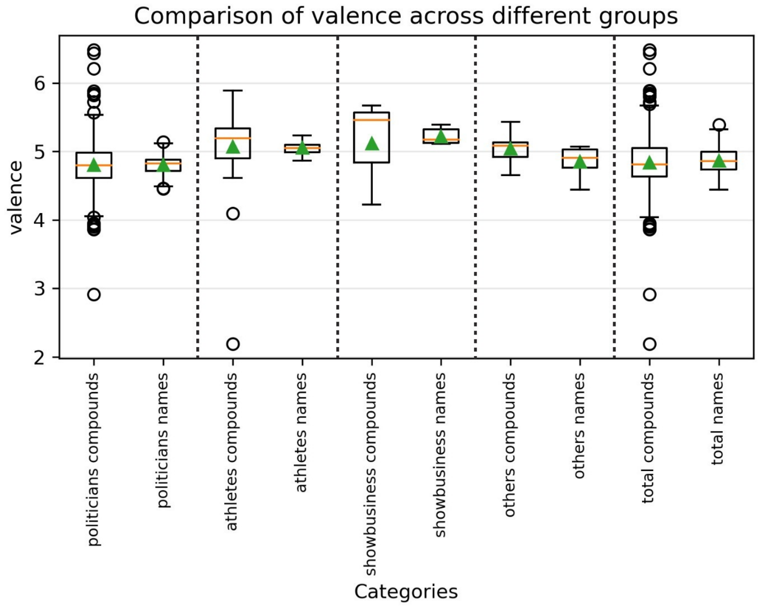


Figure 10: Comparison of all groups including all targets and outlier points.

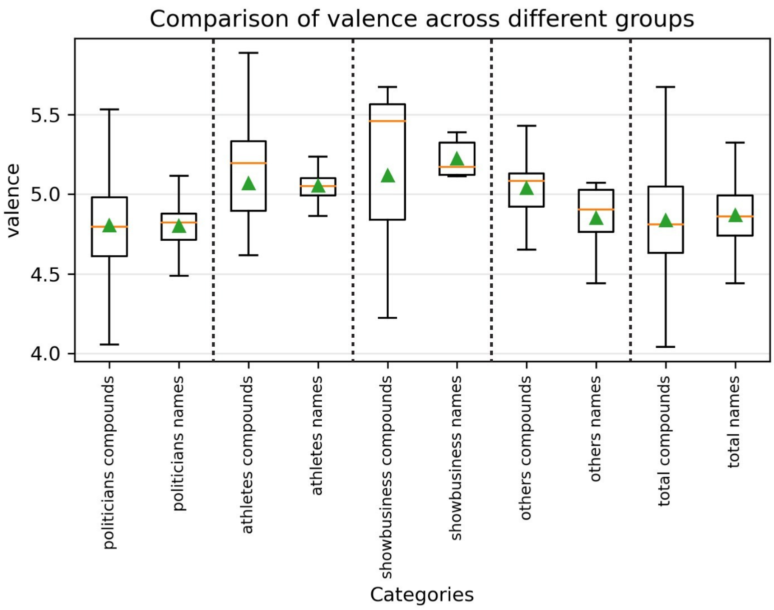


Figure 11: Comparison of all groups including all targets without outlier points.

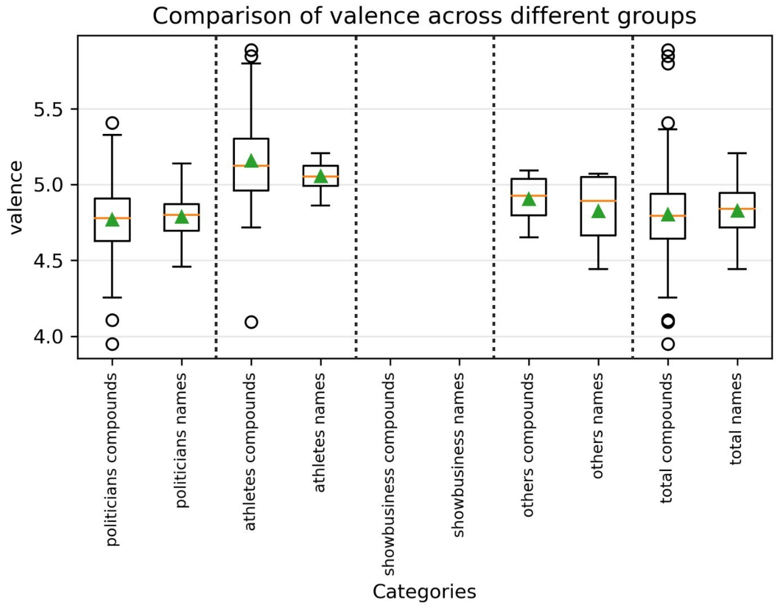


Figure 12: Comparison of all groups with filtered targets including outlier points.

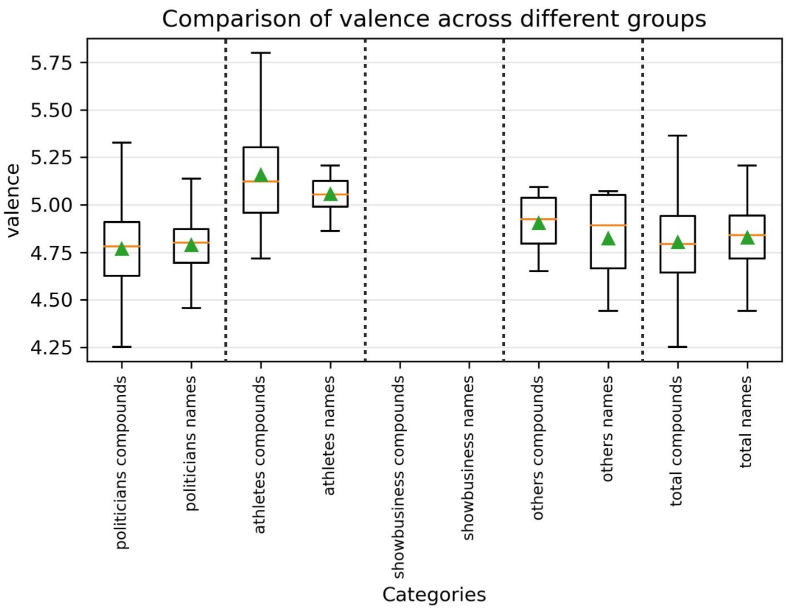


Figure 13: Comparison of all groups with filtered targets excluding outlier points.

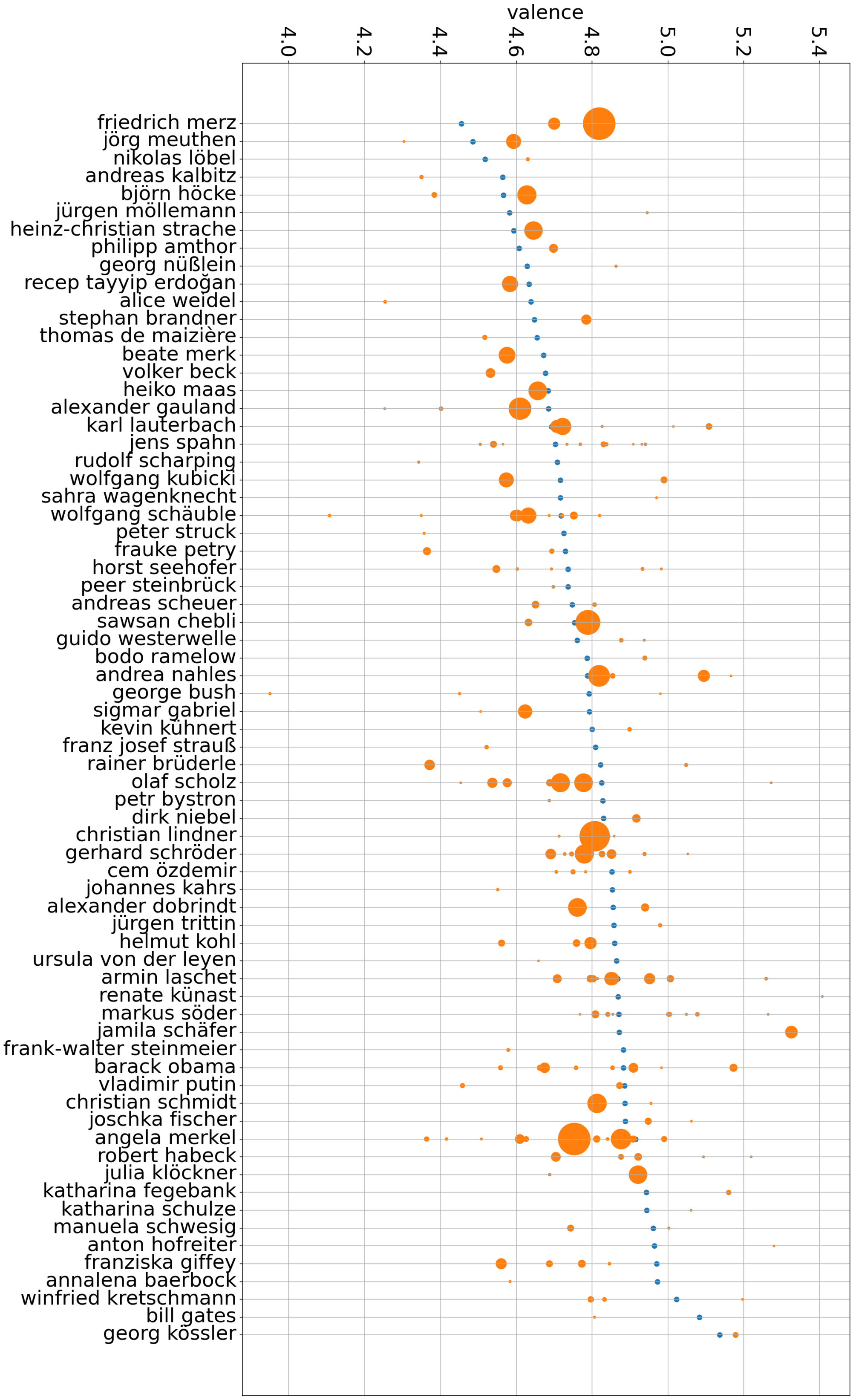


Figure 14: Politicians: Comparison of name and compound valence including frequency information.

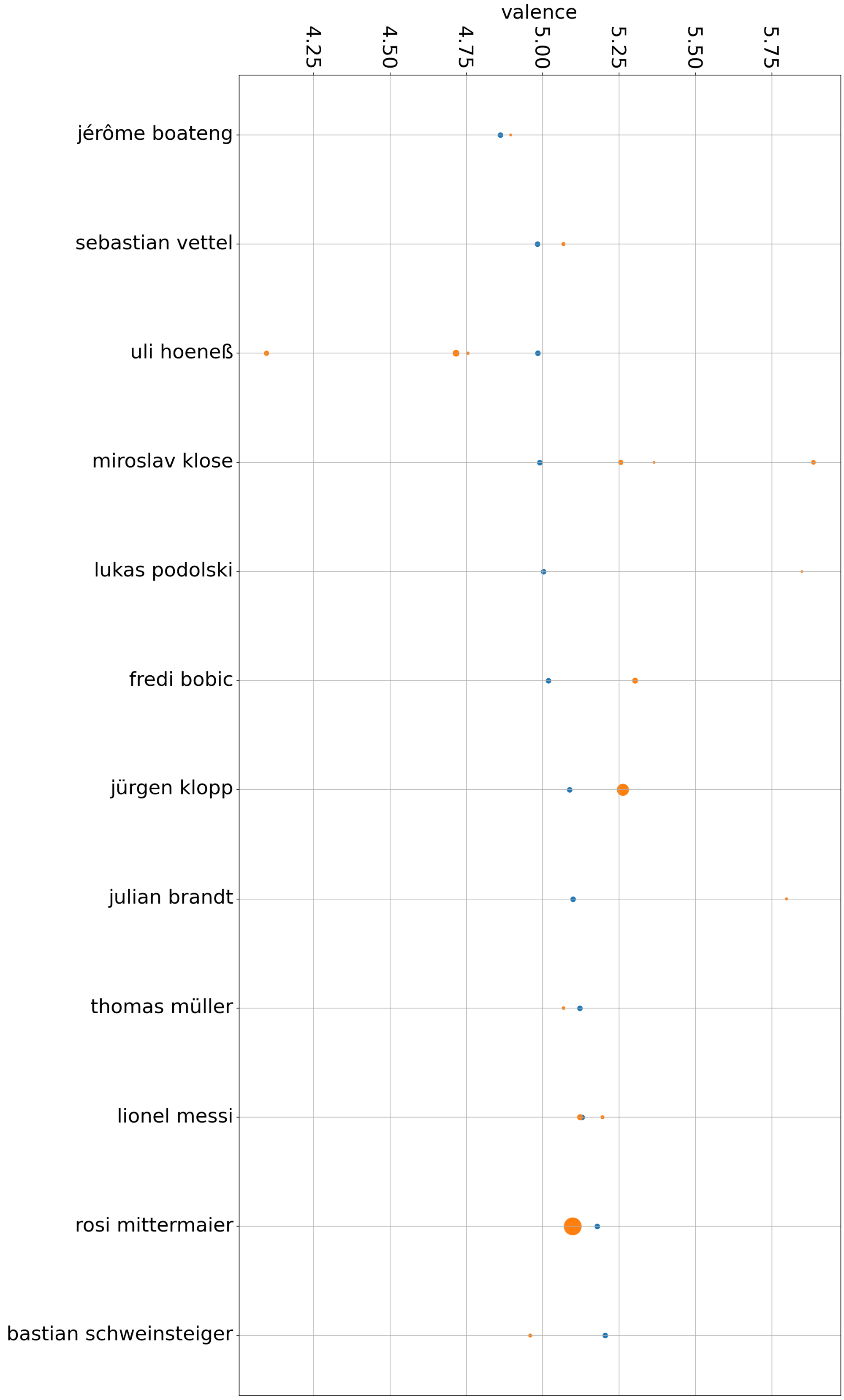


Figure 15: Athletes: Comparison of name and compound valence including frequency information.

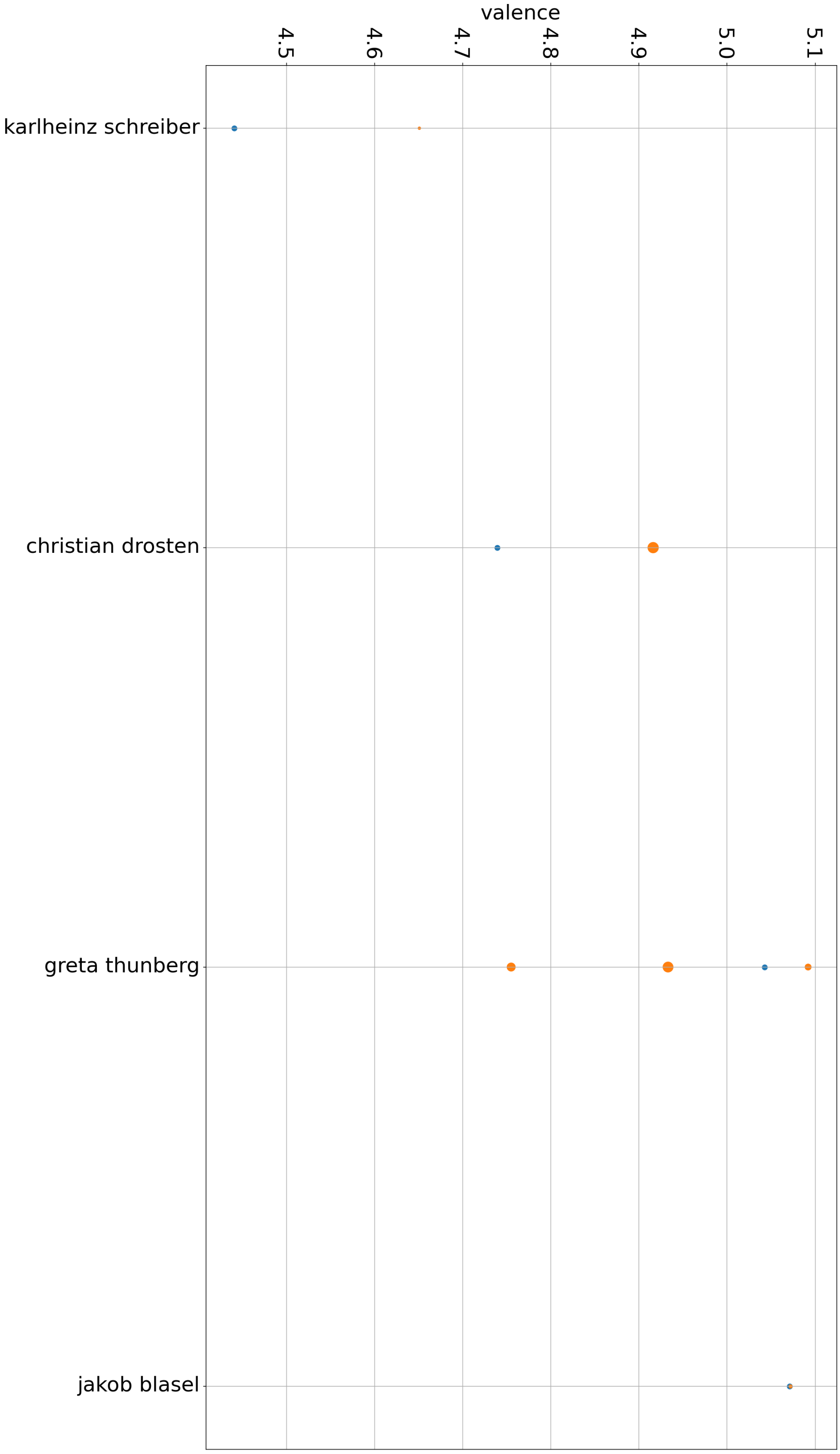


Figure 16: Others: Comparison of name and compound valence including frequency information.

3.4.3 Discussion

The most important finding is that the valence values definitely differ between the groups. Politicians seem to be perceived as the most negative, for both compounds and names. The group *others* is also more negative than 5, the middle of the valence scale. Showbusiness and sports, on the contrary, are on average more positive than 5 with showbusiness having the most positive values. This confirms the results of Belosevic (2022). Furthermore, it shows that the creation of compounds about politicians is based on more negative events than the creation of all other compounds. Also, politicians in general are perceived more negatively than athletes or people from showbusiness. Excluding the compounds and corresponding names with less than five occurrences mainly changed the number and range of the outliers. Both were decreased drastically, which shows the importance of excluding sparse data candidates. The box size and whisker length were slightly decreased, further supporting this decision. Furthermore, the majority of politicians compounds is more negative than their corresponding name. This shows that these compounds represent more negative events or actions than positive ones - they are predominantly negative evaluative. On the other hand, compounds involving athletes or *others*, such as virologists or climate activists, are often based on positive events or actions, even though there are still a small number of negative ones. Having positive and negative compounds of virologists and climate activists, compared to their name, shows the topicality and controversy of the COVID-19 crisis and the climate crisis.

3.5 Conclusion

This section investigated the evaluative nature of personal name compounds, based on the valence of their context, in comparison to the corresponding name and modifier, as well as between different subgroups. This revealed the following key findings: Personal name compounds can be both positively and negatively evaluative in comparison to their name. Nevertheless, more compounds have a lower valence value than their corresponding name. There is a moderately positive correlation between name valence and compound valence.

The context words of the compounds respectively show that they highlight reason for which they were created. This is mostly very positive or negative which results in the compound valence values spreading over a bigger range than the name valence values. However, some irregularities like outliers appear because of sparse data. Thus, personal name compounds take up special political or sporting events and draw attention to them while being evaluative at the same time. As different genres, e.g. Wortschatz vs Twitter, could possibly influence the evaluative nature in different ways, it could be interesting to further investigate the influence of the genre.

There is also a moderately positive correlation between the compounds and their respective modifier. The valence values of the modifiers are partially very extreme or not literally interpretable, but there is still a connection between compound and modifier.

Splitting the targets into four subgroups revealed that there are noticable differences between the groups. Compounds and names of politicians are perceived as the most negative. Also, the majority of politicians compounds have a lower value than their corresponding name. Athletes, on the contrary, are more positively evaluative in comparison to politicians. There are slightly more compounds with a higher value in comparison to the name than with a lower value. The group of *others* is similarly distributed, but overall somewhat more negative than the athletes. This represents current crises and their controversy. The compounds and names sorted into the group showbusiness are the most positive, but there is not enough data to support this finding, as all compounds occurred four times or less.

Regarding future work, it would be interesting to investigate the influence of the semantic frame on the evaluative nature of compounds, based on valence values of the context.

4 Linear Regression

The preceding sections have shown that different factors such as the modifier valence or the belonging to a specific group (politics, sports, showbusiness or *others*) influenced the valence of a personal name compound. Being a politician for example seemed to have a negative influence on the compound valence. Furthermore, a connection between a compound and its corresponding full name was visible, e.g. through a moderately positive correlation of name valence and compound valence. The difference between name and compound valence was of importance quite often, especially when investigating the very extreme values of compounds in comparison to their corresponding full names. In order to examine which factors, apart from name valence, compound valence, modifier valence and group, have an influence on this name-compound gap, a linear regression with the value of this gap (in the following: "delta") as response variable will be conducted. As additional categories there will be age, gender, nationality, political party, origin and a classification of the compound into different frames of the German FrameNet. A number of different models will be trained using lasso regression and stepwise regression and then compared using ANOVA to investigate the influence of the different factors as well as the best combination of predictor variables. Models containing both compound valence and name valence will be left out as this logically produced a (nearly) perfect fit. Lastly, several interaction terms will be tested to improve the models even further.

4.1 Data

The targets used for the linear regression are name-compound pairs that are based on the target lists (names and compounds) that were created for this thesis, see Section 2.1 (Compounds and names). This resulting list of name-compound pairs was then filtered for all targets with a valence value for both compounds and names, see Section 3.1 (basic calculations), and for all compounds with a valence value of the modifier, see Section 3.2.1 (Compound vs modifier - method). After filtering, 289 targets were left. For each target, a *delta value* was calculated. It represents the difference between name and compound valence (calculated as compound valence - name valence). A negative *delta value* therefore represents a pair with a compound that is more negative than the corresponding name and a positive *delta value* represents a pair with a more positive compound in comparison to the name. Ten factors that might have an influence on the size and direction of the name compound gap (*delta value*) were worked out manually for each target. The following list provides an overview of all predictor variables that were considered. All nominal variables were factorized, the type and frequency information of their levels will be provided as well. Except for *Origin*, all possible levels per factor that were encountered through a Google search were included, even if the number of targets per level was very low.

- **Name valence** (Numerical)
  Valence score calculated according to basic calculations.
- **Compound valence** (Numerical)
  Valence score calculated according to basic calculations.
- **Modifier valence** (Numerical)
  Valence score calculated according to method of compound vs modifier.
- **Age** (Numerical)
  Current age of the person in full years. If person is deceased: age at time of death.
- **Gender** (Factor with 2 levels)
  - **Female**: 65 (22%)
  - **Male**: 224 (78%)
- **Profession** (Factor with 4 levels)
  Groups according to manual classification, siehe 3.4.1
  - **Others**: 10 (3%)
  - **Showbusiness**: 3 (1%)
  - **Politics**: 250 (87%)
  - **Sports**: 26 (9%)

- **Political Party** (Factor with 16 levels)
  Current or former political party the person is/was a member of. "Independent" if the person is a politician, but not a member of a party. "No party" if the person is neither a politician nor a member of a political party.
  - **AfD** (Germany): 14 (5%)
  - **AKP** (Turkey): 1 (<1%)
  - **CDU** (Germany): 72 (25%)
  - **Conservatives** (UK): 2 (<1%)
  - **CSU** (Germany): 29 (10%)
  - **Democrats** (USA): 10 (3%)
  - **FDP** (Germany): 15 (5%)
  - **The Greens** (Germany): 36 (12%)
  - **Independent**: 3 (1%)
  - **The Left** (Germany): 3 (1%)
  - **No party**: 38 (13%)
  - **Republicans** (USA): 6 (2%)
  - **SPD** (Germany): 53 (18%)
  - **Team HC Strache** (Austria): 1 (<1%)
  - **United Russia** (Russia): 4 (1%)
  - **Centre Party** (Germany): 2 (<1%)
- **Nationality** (Factor with 9 levels)
  - **Argentina**: 2 (<1%)
  - **Austria**: 1 (<1%)
  - **France**: 1 (<1%)
  - **Germany**: 254 (88%)
  - **Russia**: 4 (1%)
  - **Sweden**: 3 (1%)
  - **Turkey**: 1 (<1%)
  - **UK**: 3 (1%)
  - **USA**: 20 (7%)
- **Origin** (Factor with 3 levels)
  This category considers the origin of a person, more precisely the place of birth. In order to restrict the variety of places, the places of birth were sorted into East or West Germany ('new federal states' or 'old federal states') or Outside of Germany.

- **East Germany**: 21 (7%)
- **West Germany**: 222 (77%)
- **Outside Germany**: 46 (16%)

- **German FrameNet** (Factor with 20 levels)
  All compounds were sorted manually[9] into frames of the German FrameNet[10] using contextual knowledge about the compound i.e. knowledge about the reason why the compound was created. *Nicht eventiv* ('not eventive') was assigned, if the compound doesn't represent an event that can be sorted into a frame. *Unbekannt* ('unknown') was assigned if the event the compound is based on is unknown. Table 5 provides an overview of all frames, their meaning and their frequency information as well as an example for each frame.

Frame	Example	Frequency information
absichtliche_Täuschung ('deliberate deception')	Plagiat-Giffey ('Plagiarism-Giffey')	3 (1%)
Ähnlichkeit ('similarity')	Ruhrpott-Messi ('Ruhr area-Messi')	3 (1%)
Aktivität ('activity')	Tore-Messi ('Goal-Messi')	37 (13%)
Besitz ('possession')	Microsoft-Gates ('Microsoft-Gates')	2 (<1%)
besuchen ('to visit')	Talkshow-Lindner ('Talk show-Lindner')	7 (2%)
erkranken ('to fall ill')	Demenz-Scholz ('Dementia-Scholz')	1 (<1%)
erzählen ('to tell')	Schande-Höcke ('Shame-Höcke')	42 (15%)
geben ('to give')	Kiffer-Lindner ('Stoner-Lindner')	1 (<1%)

[9] The annotation was carried out by Milena Belosevic.
[10] https://gsw.phil.hhu.de

Handel_kaufen ('Commerce_buy')	Villen-Spahn ('Villa-Spahn')	3 (1%)
Mitgliedschaft ('membership')	FPÖ-Strache ('FPÖ-Strache')	17 (6%)
Nahrungsaufnahme ('food intake')	Burrito-Bieber ('Burrito-Bieber')	6 (2%)
nicht eventive ('not eventive')	Ferkel-Merkel ('Piglet-Merkel')	1 (<1%)
reisen ('to travel')	Vielflieger-Hofreiter ('Frequent flyer-Hofreiter')	12 (4%)
Schaden_verursachen ('to cause damage')	Katastrophen-Spahn ('Catastrophe-Spahn')	2 (<1%)
Teilnahme ('participation')	Fallschirm-Möllemann ('Parachute-Möllemann')	75 (26%)
Übergang_zu_Zustand ('transition to state')	Corona-Gnabry ('Corona-Gnabry')	1 (<1%)
unbekannt ('unknown')	Facebook-Seehofer ('Facebook-Seehofer')	2 (<1%)
unterstützen ('to support')	Seenotrettungs-Seehofer ('Sea rescue-Seehofer')	65 (22%)
verwenden ('to use')	Kerosin-Krause ('Kerosene-Krause')	4 (1%)
Zusammenarbeit ('cooperation')	Mckinsey-Leyen ('McKinsey-Leyen')	5 (2%)

Table 5: Overview of all frames including one example and frequency information.

4.2 Method

This section provides an overview of all verifications of conditions needed for a linear regression as well as all steps of combining predictor variables to train and evaluate different linear models.

Firstly, the normal distribution of the dependent variable (*delta value*) will be checked. Then, the linear relationship between dependent and all numerical independent variables will be analyzed.
As the *delta value* is calculated on the basis of *name valence* and *compound valence*, a model including both led to a (nearly) perfect fit. Thus, there will always be three scenarios: the model either includes *compound valence* or *name valence* or neither of these two variables.
As a first step, ten different linear models using all 289 targets will be fitted. Each model uses exactly one independent variable to see which variable will lead to significant results predicting the *delta value*. Concerning the factor variables, the reference category is left at the default value which is always the lowest (first in alphabet) value. The significance of the results will be further examined by performing a Tukey post-hoc test on an analysis of variance (aov) fitted model object for each model with a significant factorial predictor variable in order to find significant differences in pairs of means of the different levels.
As a next step, several independent variables will be combined on an intuitive and theoretical basis: The first two models are based on personal information including *age*, *gender* and *age*, *gender*, *nationality*, *origin*. A third approach will aim at capturing extra-linguistic and semantic knowledge of the compound by fitting a model with *modifier valence* and *FrameNet* as predictors. *Compound valence* will also be added in a fourth model to include all information that is available for compounds. Then, a model capturing the *profession* and *political party* will be fitted and analyzed. In the final step, one model for each of the three cases (exclude *name valence*, exclude *compound valence*, exclude both) will be fitted using all remaining independent variables.
As this intuitive, manual combination of independent variables led to rather moder-

ate results and only including all variables worked better, two automatic approaches to detect a good combination of variables for each of the three cases will be carried out next. Firstly, lasso (least absolute shrinkage and selection operator) regression will be carried out. Lasso regression performs variable regularization and selection (shrinkage and removement of variables) in order to find an accurate model. Additionally, a forward, backward and both-direction stepwise regression will be performed to serve as a starting point, i.e. the number of independent variables will be incrementally increased, reduced or both until the best fitting model is reached. Stepwise regression is based on the Akaike information criterion (AIC) which compares the fit of regression models by penalizing a large number of variables and rewarding models that explain the most variation in the data. Increasing or decreasing the number of variables is done as long as the AIC value decreases significantly doing so. As stepwise regression is criticized for possibly excluding variables that actually have effects on the dependent variable and including other variables that are coincidentally significant, these three models will be only used as a baseline and then improved manually. Also, these models will be compared to the models resulting from lasso regression and the models including all predictors using ANOVA (anova() function in R for nested models). Finally, the best three models for each case will be presented.
In order to improve these three models for each case even further, several interactions will be included as a last step and the best results will be identified by comparing with the R anova() function.

4.3 Results and Discussion

4.3.1 Prerequisites

Figure 17 shows that the *delta value* is suitable as a dependent variable as it is approximately normally distributed. Figure 18 shows an overview of all numerical predictor variables plotted against the *delta value* including a smooth line. *Name valence*, *modifier valence* and *age* seem to be at least roughly linear. *Delta value* tends to decrease with both decreasing and increasing *name valence* and seems to

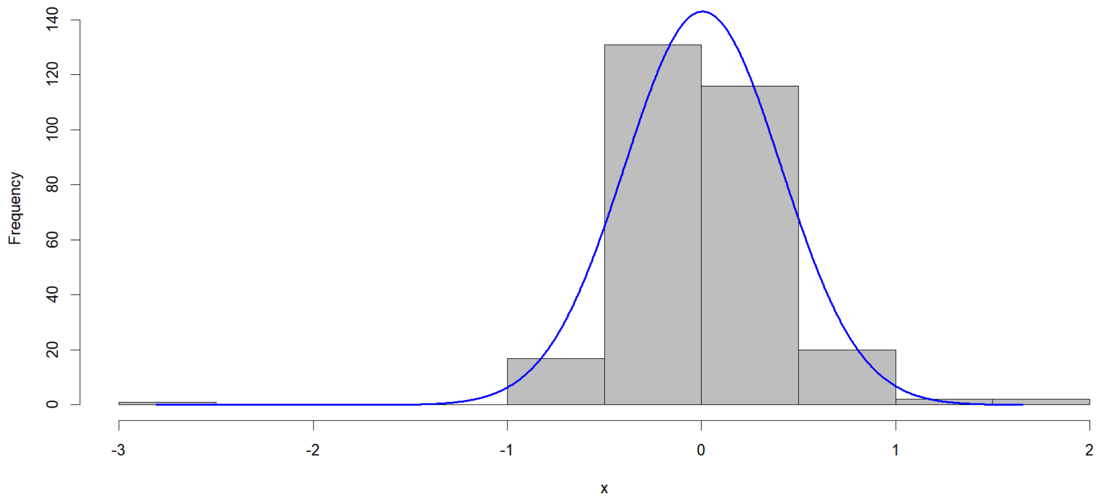
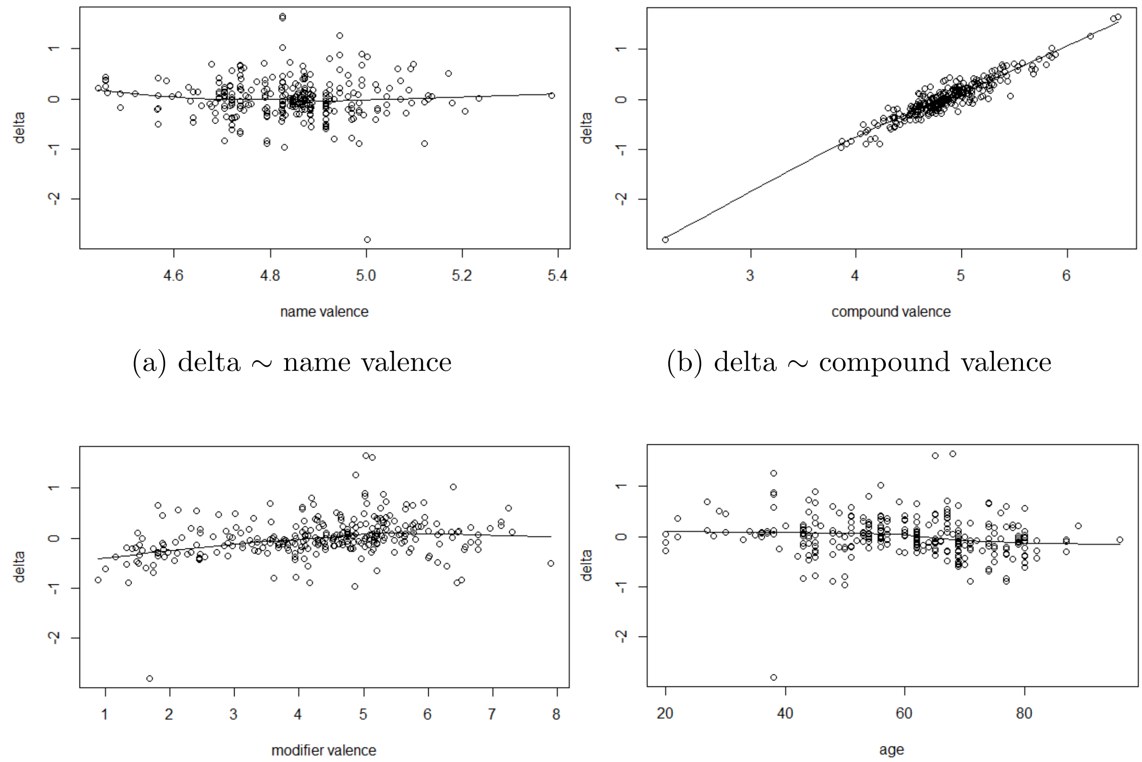


Figure 17: Normal distribution of the dependent variable (*delta value*).

decrease with both decreasing and increasing *modifier valence*. A decreasing *delta value* could be recognized with a decreasing *age*. Only *compound valence* has a well fitting, positive linear relationship with the *delta value*.

4.3.2 Single predictor variable

Fitting ten linear models with one predictor variable each revealed that five out of these ten predictors have a significant effect on the *delta value*. Exactly speaking, the models using *compound valence*, *modifier valence*, *age*, *political party* or *FrameNet categorization* yield a p-value <0.05. An overview of all results is presented in Tables 6 and 7. Table 6 provides the results of all numerical predictors including the t-value, p-value and Multiple R-Squared. The regression line can be estimated as $y = Intercept + Slope * x$. Concerning the factor variables in Table 7, t-value, p-value and Multiple R-Squared will be presented and only significant levels including their respective p-value are displayed. The prediction for the reference level is represented as the intercept of the model. The shown values can be read per level as $y = Intercept + Slope * 1$ if the target person is part of the current level. E.g.:



(a) delta $\sim$ name valence (b) delta $\sim$ compound valence

(c) delta $\sim$ modifier valence (d) delta $\sim$ age

Figure 18: Linear relationship of the *delta value* with all numerical variables.

$y = -0.22 + 0.31 * 1$ if the target is member of the party *FDP*.

The results reveal that the *compound valence* is a highly significant predictor for the *delta value* with a positive linear relationship, also explaining around 88% of the variance of the *delta value*. The *modifier valence* is a significant predictor with a positive linear relationship as well, but with only around 10% of the variance being explained. *Age*, on the contrary, has a negative linear relationship with the *delta value*. It is also significant, but with only around 2% of variance being explained. *Gender*, *profession* and *nationality* did not yield any significant results. Furthermore, being an *AfD* party member will influence the *delta value* in a negative direction, being a member of any other (significant) party will influence the *delta value* in a more positive direction compared to the reference category *AfD*. Especially *Conservatives*, *The Greens* and *The Left* have a higher positive influence. Being born

in *West Germany* has a low positive influence on the *delta value* compared to the reference category *Outside Germany*.
The Tukey post-hoc test was performed on top of an (aov) fitted model object for each model with a significant factorial predictor variable (*party* and *FrameNet*). This revealed that no pair of levels has a significant difference in mean values.

Predictor	Intercept	Slope	t-value	p-value	R-Squared
Name valence	0.61	-0.12	-0.75	0.45	0.00
Compound valence	-4.35	0.90	46.87	$< 2.2 * 10^{-16}$	0.88
Modifier valence	-0.37	0.09	5.66	$< 3.7 * 10^{-8}$	0.10
Age	0.24	-0.00	-2.47	0.01	0.02

Table 6: Regression results of numerical predictors.

Discussion:
The analysis of name valence vs. compound valence in Section 3.2 has shown that the compound valence values spread over a much bigger range than the name valence values due to more controversial/highlighting events these compounds are based on. As name valence values more or less gather around 5, the middle of the 11 point scale, compounds with a lower valence value tend to have a lower *delta value* and, accordingly, compounds with a higher valence value have a higher *delta value*. This explains the very well-fitted linear relationship between *delta value* and *compound valence*. Nevertheless, there can of course be cases of compounds with a relatively high valence value that are still more negative compared to their name, although these cases appear to be exceptions, considering the nice linear fit of *delta value* and *compound valence*.
On the other hand, *name valence* was not identified as a significant predictor for the *delta value*. This seems to happen due to the exact same reason - name valence values gather around the middle of the scale with their corresponding compound valence values distributing over both more negative and more positive sides.

Predictor	Intercept	Slope	t-value	p-value	R-Squared
Gender				0.50	0.00
Profession				0.67	0.01
Political Party				0.01	0.10
- AfD (reference)	-0.22		-2.13	0.03	
- Conservatives		0.63	2.12	0.04	
- CSU		0.25	1.99	0.05	
- FDP		0.31	2.15	0.03	
- The Greens		0.42	3.40	0.00	
- No party		0.25	2.05	0.04	
- The Left		0.59	2.38	0.02	
- SPD		0.25	2.08	0.04	
Nationality				0.13	0.04
Origin	-0.11			0.08	0.02
- West Germany		0.14	2.22	0.03	
FrameNet				0.04	0.11

Table 7: Regression results of factorial predictors.

Furthermore, *modifier valence* appears to have a positive linear relationship - a higher modifier valence increases the *delta value*. Even though this predictor has a very low R-Squared value and only has a low positive linear relationship, it corresponds perfectly to the findings of compound vs. modifier in Section 3.3. The comparison of compound and modifier revealed that the modifier highlights the specific reason why the compound was created and this consequently is mirrored in the valence value of the compound. Thus, a very negative or controversial event that leads to the creation of a personal name compound tends to have a low-valued modifier and overall a more negative compound valence. This goes with a lower *delta value* that represents a more negative compound in comparison to the full name. This works analogously for compounds based on a very positive event.

The last numerical predictor variable is *age* with a very low R-Squared value of 2%, but a still significant p-value of 0.01 and a negative slope value. Consequently, a person increasing in age seems to have a compound that is more negative in comparison to the full name.
Considering the factor variables, neither the *gender*, nor the *profession*, nor the *nationality*, nor the *origin* yielded significant results when being used as a single predictor variable, although the level *West Germany* is significant and seems to have a more positive influence on the *delta value* compared to the reference category *Outside Germany*.
The *political party* a person is member of has a significant influence on the *delta value*, although only 1% of the variance of the dependent variable can be explained. Eight different levels (*AfD*, *Conservatives*, *CSU*, *FDP*, *The Greens*, *No party*, *The Left* and *SPD*) were significant as well. Compared to the reference category *AfD*, all parties have a positive influence on the *delta value*. Considering the fact that the *AfD* is a political party on the far right of the political spectrum and often causes controversial news, this specific relation to the *delta value* is not surprising.
The last significant predictor is *FrameNet* with a relatively low R-Squared value. No level is significant which can be caused by the high number of levels, low number of targets per level and only weak significant global effect (p-value: 0.04). These three factors could also explain why neither of the pairwise comparisons of levels in the Tukey post-hoc test yielded a significant difference.

4.3.3 Multiple predictor variables

After investigating which of the predictor variables actually influence the *delta value* when being used as a single predictor, several predictors were now combined in order to find even better models to predict the *delta value*.

The first two models using personal information *age*, *gender* and *age*, *gender*, *nationality*, *origin*) were both significant (p-value: 0.02 and 0.03) with a Standard Error of 0.4 for both but could hardly explain the variance of the delta value (Multiple R-squared: 0.03 and 0.08, Adjusted R-squared: 0.02 and 0.04), see Table 8.

The next two models were based on information about the compound using *modifier valence*, *FrameNet* and *compound valence*, *modifier valence*, *FrameNet*. Both models are highly significant (p-value: $4.76*10^{-5}$ and $< 2.2*10^{-16}$), but including *compound valence* lead to a huge improvement in R-Squared values (Multiple R-squared: 0.18 and 0.9, Adjusted R-squared: 0.11 and 0.89) and a drop in the Standard Error (0.38 and 0.13), see Table 8.
The profession-based approach including *profession* and *political party* as predictors also resulted in a significant model (p-value: 0.03) with a Standard Error of 0.39 but yielded rather low R-Squared values (Multiple R-squared: 0.11, Adjusted R-squared: 0.05), see Table 8.
Lastly, all three models (one per scenario) that included all remaining independent variables were significant. The first model that used all predictors except for *name valence* yielded a p-value of $< 2.2*10^{-16}$, a Standard Error of 0.09 and Multiple and Adjusted R-Squared values of 0.96. The second model that used all independent variables except for *compound valence* had a p-value of 0.002, a Standard Error of 0.38, a Multiple R-squared of 0.27 and an Adjusted R-squared of 0.12. The last model excluded both *name valence* and *compound valence* but used all other remaining predictors. This resulted in a significant model (p-value: 0.003) with a Standard Error of 0.38, a Multiple R-squared of 0.26 and an Adjusted R-squared of 0.11, see Table 8.

Exclude name valence: As *compound valence* has a very well-fitting linear relationship with the *delta value* and also worked best as a single predictor variable, the best models by far are the ones including compound valence.
The best model according to lasso regression uses *compound valence*, *modifier valence*, *age*, *gender*, *profession*, *nationality*, *political party* and *FrameNet* as predictors. *Origin* was shrunk to zero and is therefore left out. It is highly significant (p-value: $< 2.2 * 10^{-16}$), with a Standard Error of only 0.09 and R-Squared values of 0.96, see Table 8. Also, when compared to the model that additionally includes *origin* using ANOVA, this more complex model does not work significantly better.
Forward, backward and both-direction stepwise regression all proposed the same

model which uses *compound valence, political party, gender, profession* and *nationality*. It is highly significant (p-value: $< 2.2 * 10^{-16}$) with a Standard Error of 0.09, a Multiple R-squared of 0.96 and an Adjusted R-squared of 0.95, see Table 8. According to ANOVA, the lasso model is not significantly better than this less complex model resulting from stepwise regression. Furthermore, the model with all predictors is not significantly better than this model from stepwise regression (with *compound valence, political party, gender, profession* and *nationality* as predictors). Surprisingly, neither *modifier valence* nor *age* are included here, although Section 4.3.2 revealed that both are good predictors, at least when used as a single predictor. This seems to happen because of the exact, already mentioned reasons why stepwise regression can be criticized. Thus, this case needs to be further examined.

Neither adding *age*, nor *modifier*, nor both to the model from stepwise regression resulted in a significantly better model. However, an addition of compound based information (*modifier valence, FrameNet*), which worked quite well at the beginning of this section, also led to significantly better results. Precisely speaking, this improved model using *compound valence, profession, nationality, political party, gender, FrameNet* and *modifier* is highly significant (p-value: $< 2.2 * 10^{-16}$) with a Standard Error of 0.09 and R-Squared values of 0.96. Details can be found in Table 9, column 1 (Hlavac, 2022). For every predictor, its estimate, significance and standard error will be reported. Concerning the individual predictors, *compound valence* has a positive effect on the *delta value* and *modifier valence* has a very low negative effect. The latter is rather surprising as the results from Section 4.3.1 (Prerequisites) and 4.3.2 (Single predictor variable) revealed a rather positive linear relationship between *delta value* and *modifier valence*. *Gender male* appears to have a positive influence on the *delta value* in comparison to the baseline category *female* (with all other factors being held constant). Compared to the baseline category *profession*: *others, politics* and *showbusiness* have a positive effect on the *delta value* while *sports* has a negative effect (with all other factors being held constant). Furthermore, regarding the factor *political party*, many parties such as *The Greens* or *FDP* now reveal a negative influence on the *delta value* compared to the baseline category *AfD*, again if all other factors remain constant.

Exclude compound valence: Lasso regression proposed to not exclude any of the predictors. This model is consequently equivalent to the model including all independent variables except for *compound valence*.
Forward, backward and both-direction stepwise regression again resulted in the same best model. It includes *name valence, modifier valence, age* and *origin* with a p-value of $4.93 * 10^{-8}$, a Standard Error of 0.38, a Multiple R-squared of 0.14 and an Adjusted R-squared of 0.12, see Table 8. The model including all variables, which was also proposed by lasso regression, was not significantly better according to an ANOVA analysis. Thus, this model resulting from stepwise regression was manually improved further. Adding different predictors did not seem to have any significant effects, according to ANOVA. However, excluding *origin* worked very well, as the model using *name valence, modifier valence, age* and *origin* is not significantly better than the model that excludes *origin*. Details of this model can be found in Table 9, column 2. Analogously to the results from Section 4.3.1 (Prerequisites) and Section 4.3.2 (Single predictor variable), *name valence* and *age* have a negative influence on the *delta value* while *modifier valence* has a positive linear relationship with the *delta value*.

Exclude name valence and compound valence: Analogously to the case of excluding *compound valence*, lasso regression proposed to not exclude any of the predictors. This model is consequently equivalent to the model including all independent variables except for *name valence* and *compound valence*.
Forward, backward and both-direction stepwise regression resulted in a highly significant model (p-value: $9.98 * 10^{-8}$) with *modifier valence, age, gender* and *origin* as predictors and a Standard Error of 0.38, Multiple R-squared of 0.13, and Adjusted R-squared of 0.12, see Table 8. The model including all predictors, which was also proposed by lasso regression, does not predict the *delta value* significantly better, according to an ANOVA analysis. Consequently, the model resulting from stepwise regression was improved manually. Similarly to the case of excluding *compound valence*, adding different combinations of variables to the model did not seem to lead to a significant improvement. Thus, several variables were removed in order

to find an even better model. The current model still predicts better than models using only *modifier valence, age* or *modifier, age, gender*, according to an ANOVA analysis. However, it is not significantly better than using only *age, modifier* and *origin*, leaving out the *gender* variable. This reduced model is also significant with a p-value of $7.1 * 10^{-8}$, a Standard Error of 0.38, a Multiple R-squared of 0.13 and an Adjusted R-squared of 0.12. Detailed results can be found in Table 9, column 3. As in Section 4.3.1 (Prerequisites), *modifier valence* has a positive influence on the *delta value* and *age* has a slightly negative impact. In addition to that, both *East* and *West Germany* have a positive influence on the *delta value*, compared to the reference category *Outside Germany*.

Model	**p-value**	**Multiple R-Squared**	**Adjusted R-Squared**	**Standard Error**
Personal information (1)	0.02	0.03	0.02	0.40
Personal information (2)	0.03	0.08	0.04	0.40
Compound information (1)	$4.76 * 10^{-5}$	0.18	0.11	0.38
Compound information (2)	$< 2.2 * 10^{-16}$	0.90	0.89	0.13
Profession-based	0.03	0.11	0.05	0.39
Exclude name valence				
- All predictors	$< 2.2 * 10^{-16}$	0.96	0.96	0.09
- Lasso regression	$< 2.2 * 10^{-16}$	0.96	0.96	0.09
- Stepwise regression	$< 2.2 * 10^{-16}$	0.96	0.95	0.09
Exclude compound valence				
- All predictors	0.00	0.27	0.12	0.38
- Stepwise regression	$4.93 * 10^{-8}$	0.14	0.12	0.38
Exclude both				
- All predictors	0.00	0.26	0.11	0.38
- Stepwise regression	$9.98 * 10^{-8}$	0.13	0.12	0.38

Table 8: Regression results of different models using multiple predictor variables.

	Dependent variable:		
	delta		
	(1)	(2)	(3)
Constant	−4.94***(0.14)	1.73 **(0.82)	−0.28 **(0.13)
Compound valence	0.98***(0.01)		
Name valence		−0.39 **(0.16)	
Modifier valene	0.00(0.00)	0.09 ***(0.02)	0.08 ***(0.02)
Age		0.00 **(0.002)	0.00 **(0.00)
Male	0.09***(0.01)		
Politics	0.17**(0.07)		
Showbusiness	0.21(15)		
Sports	−0.33***(0.04)		
AKP			
CDU	−0.13***(0.03)		
Conservatives	−0.07(0.13)		
CSU	−0.16***(0.03)		
Democrats	0.46***(0.15)		
United Russia			
FDP	−0.16***(0.03)		
The Greens	−0.23***(0.03)		
No party	0.11(0.08)		
The Left	−0.13**(0.06)		
Independent	0.11*(0.06)		
Republicans	0.53***(0.15)		
SPD	−0.17***(0.03)		
Team HC Strache			
Centre Party	0.15**(0.07)		
Germany	0.09(0.07)		

France	0.09(0.11)		
UK	0.10(0.11)		
Austria	0.14(0.12)		
Russia	$-0.18^{**}(0.09)$		
Sweden	$-0.15(0.10)$		
Turkey	0.06(0.12)		
USA	$-0.61^{***}(0.16)$		
East Germany			0.09(10)
West Germany			$0.15\ ^{**}(0.06)$
Deliberate deception	0.00(0.08)		
Similarity	0.04(0.09)		
Activity	0.09(0.06)		
Possession	$0.21^{*}(0.11)$		
To visit	0.10(0.07)		
To fall ill	0.04(0.10)		
To tell	0.08(0.06)		
To give	0.01(0.11)		
Commerce_buy	0.08(0.08)		
Membership	0.06(0.06)		
Food intake	$0.16^{**}(0.07)$		
Not eventive	$-0.01(0.11)$		
To Travel	0.01(0.07)		
To cause damage	0.07(0.09)		
Participation	0.06(0.06)		
Transition to state	$-0.13(0.11)$		
To support	0.05(0.06)		
To use	0.10(0.08)		
Cooperation	0.09(0.08)		
Observations	289	289	289
R^2	0.96	0.13	0.13

Adjusted R^2	0.96	0.12	0.12
Residual Std. Error	$0.09(df = 243)$	$0.38\ (df = 285)$	$0.38\ (df = 284)$
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 9: Linear Regression results for all three models (1: exclude *name valence*, 2: exclude *compound valence*, 3: exclude both).

Discussion: A number of different combinations of variables have revealed that several models are able to predict the *delta value* quite well. However, including *compound valence* as predictor is certainly a great advantage, as it has the best fitting linear relationship with the *delta value*. This was proven by the models that were composed on an intuitive basis: The compound-based approach (including *compound valence, modifier vaelnce* and *FrameNet*) worked far better than the models using personal information (*age, gender, nationality* and *origin*) or the profession-based approach (using *profession* and *political party*).
When using *compound valence* to predict the *delta value*, it is advisable to add *political party, gender, profession, nationality, modifier* and *FrameNet* in addition to the *compound valence*. It is the best working model compared to the other two scenarios, because *compound valence* is a very good predictor. As it has a positive linear relationship with the *delta value*, the estimate shows a positive effect on the *delta value*. The fact that the estimate of the *modifier valence* is slightly negative could be caused by having added multiple predictors to the model.
When predicting the *delta value* including only *name valence*, it is advisable to use *modifier valence, name valence* and *age* as predictor variables. Analogously to the models using a single predictor, *modifier valence* has a positive influence on the *delta value* and *age* has negative influence, thus, young people with low modifier values will most likely have a relatively high *delta value*.
Lastly, in order to predict the gap between *compound valence* and *name valence* without knowing either of the values, a model using *modifier, origin* and *age* as independent variables works best. Analogously to the findings of Section 4.3.2 (Single predictor variable), *modifier valence* has a positive influence on *delta value, age* has

a negative influence on *delta value*. *East* and *West Germany* having a more positive influence on the *delta value* compared to the reference category *Outside Germany* shows that compounds of people born outside of Germany tend to be more negative in comparison to their corresponding name when compared to people from Germany. As most targets in the list used for this thesis are people from Germany and the compounds in this thesis are German as well, non-German targets probably only appear in cases of very controversial events that draw attention internationally which could explain the more negative impact on the *delta value*.

4.3.4 Interactions

After identifying three different, relatively good models to predict the *delta value* for each scenario, this section will investigate which interactions could possibly improve these models even further. The best improvements per category will be presented in the following.

Exclude name valence: The great number of predictors in the model allowed to test an even bigger number of interactions. Many of them yielded significant results, including *compound valence* and *profession, origin* and *profession, profession* and *modifier valence* or *compound valence* and *gender*, to name a few. The Standard Error could be mostly reduced to 0.8 and R-Squared values raised to 0.96 or 0.97.

Exclude compound valence: The interaction *age*modifier valence* makes the model significantly better. Multiple R-squared is now at 0.14, Adjusted R-squared at 0.13.
Also, the model using *modifier valence*name valence* shows an improvement: Multiple R-squared rises to 0.14, Adjusted R-squared to 0.13.

Exclude name valence and compound valence: When predicting the delta value without *name valence* and *compound valence*, including the interaction *age*modifier valence* leads to a significant better model with a Multiple R-squared value of 0.15

and an Adjusted R-squared of 0.13.

Discussion: Several interactions were able to improve the models a bit further, e.g. *compound valence* and *profession, age* and *modifier valence* as well as *modifier valence* and *name valence*. All of these combinations seem to have a clear connection: In our data, *age* and *modifier* seem to have a negative linear relationship: The *delta value* increases when *age* is increased and *modifier valence* is decreased. Also, *modifier valence* and *name valence* both tend to increase the *delta value* when being increased. To sum up, including interactions into the models predicting the *delta value* could increase the performances of all three models.

4.4 Conclusion

This section investigated whether it is possible to predict a *delta value* - the gap between name valence and compound valence - with a linear regression model and futhermore examined which independent variables and combinations work best to do so. Combining *name valence* and *compound valence* led to a (nearly) perfect model, so only the three cases involving either *name valence* or *compound valence* or none of them were considered.

The first step of using each independent variable individually to predict the *delta value* showed that *compound valence, modifier valence, age, political party* and *FrameNet* work as a significant predictor while *name valence, gender, profession, nationality* and *origin* do not.

An intuitive combination of independent variables led to moderately good results. Nevertheless, all considered models (based on personal information / based on compound information / profession-based) were significant, which shows that many combinations of predictors are already able to predict the *delta value* quite well.

Combining different variables on the basis of a lasso regression model and stepwise regression model (further improved manually) for each case provided three different, highly significant models. Considering the well fitting, linear relationship of

compound valence and *delta value*, models including *compound valence* as predictor always achieve the best results. When predicting the *delta value* using *compound valence*, including *profession, nationality, political party, gender, FrameNet* and *modifier valence* as additional variables has proved to be a suitable decision. Adding different interactions such as *compound valence* and *profession* or *origin* and *profession* could improve this model even further. For a prediction of the *delta value* without *compound valence*, a model including *name valence, modifier valence* and *age* worked very well, even improving by adding *age* and *modifier valence* or *modifier valence* and *name valence* interactions. Lastly, to predict the *delta value* without knowing the *name valence* nor the *compound valence*, a model using *modifier valence, age* and *origin* works best. This could also be improved by adding an interaction of *age* and *modifier valence*.
In summary, several aspects of personal information such as *age, gender, political party, nationality* etc. can be used to successfully predict the difference between the *name valence* and the *compound valence*. Knowing the *name valence* or *compound valence* additionally, the respective other value can be predicted through the *delta value*. Regarding future work, it would be interesting to investigate the relationship of the predictors even further and check which more complex interactions could improve the prediction of the *delta value*.

5 Summary

This thesis investigated the evaluative nature of German personal name compounds on the basis of valence values that were calculated for 413 compounds and their corresponding 131 full names, extracted from Wortschatz and Twitter.

Personal name compounds highlight the reason for which they were created and therefore draw attention to special political or sports events. They are both positively and negatively evaluative in comparison to their corresponding full name. A moderately positive correlation between compound valence values and their respective modifiers showed a significant connection, even though some modifiers are very extreme or not literally interpretable, e.g. *Hummer-Wagenknecht* ('lobster-Wagenknecht') which refers to an event that randomly involved a lobster. In addition to that, extending the proposal of Belosevic (2022) to investigate the evaluative function across politicians, athletes and people from showbusiness revealed that politicians (both compounds and names) are perceived as the most negative. Also, their compounds are on average more negative than the names while the compounds from athletes and the group *others* are more positive, compared to their respective names. Showbusiness did not have enough data to provide valuable results.

In a last step, a linear regression in order to predict the *delta value* (difference of name and compound valence) was carried out. A number of different independent variables (compound valence, name valence, modifier valence, age, gender, profession, nationality, origin and FrameNet) were used. *Compound valence* was identified as the best predictor due to its very well fitting positive linear relationship with the *delta value*. A number of different combinations of independent variables were able to predict the *delta value* very well, excluding the case of using both *compound valence* and *name valence* as this led to a perfect fit. Several interactions could improve these models even further.

Bibliography

Milena Belosevic. Veggie-renate und Merci-Jens. *Zeitschrift für germanistische Linguistik*, 50(2):289–319, 2022. doi: 10.1515/zgl-2022-2056. URL https://doi.org/10.1515/zgl-2022-2056.

Milena Belosevic and Sabine Arndt-Lappe. Merci-jens and Lösch-Leyen The Semantics of Personal Name Compounds in German. In *Third International Symposium of Morphology (ISMo 2021)*, page 28, 2021.

Dirk Goldhahn, Thomas Eckart, and Uwe Quasthoff. Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 2012.

Marek Hlavac. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Social Policy Institute, Bratislava, Slovakia, 2022. URL https://CRAN.R-project.org/package=stargazer. R package version 5.2.3.

Maximilian Köper and Sabine Schulte im Walde. Automatically Generated Affective Norms of Abstractness, Arousal, Imageability and Valence for 350 000 German Lemmas. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 2595–2598, Portoroz, Slovenia, 2016.

Sebastian Kürschner. Nickname formation in West Germanic: German Jessi and Thomson meet Dutch Jess and Tommie and English J-Bo and Tommo. *German and Dutch in Contrast*, page 15, 2020.

H. Schmid. *Improvements in Part-of-Speech Tagging with an Application to German*, pages 13–25. Springer Netherlands, Dordrecht, 1999. ISBN 978-94-017-2390-9. doi: 10.1007/978-94-017-2390-9_2. URL https://doi.org/10.1007/978-94-017-2390-9_2.

A Appendix

Table 10: All name-compound pairs including number of compound occurrences (count), compound valence (c_val), name valence (n_val) and difference (diff), rounded to two decimal places.

count	compound	name	c_val	n_val	diff
23	abschiebe-kretschmann	winfried kretschmann	4.83	5.02	-0.19
11	abschiebe-seehofer	horst seehofer	4.60	4.74	-0.13
0	abschottungs-merkel	angela merkel		4.92	
1	abschreib-franziska	franziska giffey	4.18	4.97	-0.79
0	abseits-kroos	toni kroos		5.11	
531	afd-höcke	björn höcke	4.63	4.57	0.06
318	afd-meuthen	jörg meuthen	4.59	4.49	0.11
6	affären-scholz	olaf scholz	4.45	4.83	-0.37
122	agenda-schröder	gerhard schröder	4.85	4.85	
0	airline-analena	annalena baerbock		4.97	
8	aktenkoffer-schäuble	wolfgang schäuble	4.69	4.72	-0.03
29	aktien-merz	friedrich merz	4.84	4.46	0.38
2	akw-merkel	angela merkel	4.33	4.92	-0.59
8	alleingang-schmidt	christian schmidt	4.96	4.89	0.07
8	alleingang-söder	markus söder	5.05	4.87	0.18
21	amigo-strauß	franz josef strauß	4.52	4.81	-0.29
1	amok-seehofer	horst seehofer	5.39	4.74	0.65
23	ampel-söder	markus söder	5.08	4.87	0.21
1	anatolien-özoguz	aydan özoğuz	5.17	4.83	0.34
3	anden-özdemir	cem özdemir	5.16	4.85	0.30
6	apartheid-gabriel	sigmar gabriel	4.51	4.79	-0.29
1	arizona-jakob	jakob blasel	5.43	5.07	0.36
27	armani-schröder	gerhard schröder	4.75	4.85	-0.10
0	armutsbericht-rössler	philipp rössler		5.03	

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
2	astra-spahn	jens spahn	4.75	4.70	0.05
81	asyltourismus-söder	markus söder	4.81	4.87	-0.06
52	atom-greta	greta thunberg	5.09	5.04	0.05
126	atom-merkel	angela merkel	4.61	4.92	-0.31
28	atom-putin	vladimir putin	4.46	4.89	-0.43
23	audi-andi	andreas scheuer	4.81	4.75	0.06
149	aufschrei-brüderle	rainer brüderle	4.37	4.82	-0.45
106	augustus-amthor	philipp amthor	4.70	4.61	0.09
0	australien-rahmstorf	stefan rahmstorf		4.83	
0	bagdad-angela	angela merkel		4.92	
0	bahnhofs-palmer	boris palmer		4.79	
13	banken-steinbrück	peer steinbrück	4.70	4.74	-0.04
690	bätschi-nahles	andrea nahles	4.82	4.79	0.03
0	besatzer-hollande	françois hollande		4.84	
2	bhutan-beckham	david beckham	5.66	5.07	0.60
27	bienen-söder	markus söder	5.00	4.87	0.13
205	bierdeckel-merz	friedrich merz	4.70	4.46	0.24
28	bierzelt-söder	markus söder	4.84	4.87	-0.03
2	big-brother-westerwelle	guido westerwelle	4.52	4.76	-0.24
208	bimbes-kohl	helmut kohl	4.80	4.86	-0.06
8	bimbes-schäuble	wolfgang schäuble	4.75	4.72	0.03
2	bingo-drosten	christian drosten	5.31	4.74	0.57
1591	blackrock-merz	friedrich merz	4.82	4.46	0.36
0	blasphemie-bobic	fredi bobic		5.02	
0	blockade-rössler	philipp rössler		5.03	
15	blowout-blasel	jakob blasel	5.07	5.07	
6	bomben-fischer	joschka fischer	5.06	4.89	0.17
6	bomb-obama	barack obama	4.98	4.88	0.10

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
0	bonusmeilen-habeck	robert habeck		4.92	
30	bonusmeilen-özdemir	cem özdemir	4.75	4.85	-0.10
0	bonusmeilen-schulze	katharina schulze		4.94	
3	bosporus-putin	vladimir putin	4.71	4.89	-0.17
41	bratwurst-bobic	fredi bobic	5.30	5.02	0.28
0	brechstangen-poldi	lukas podolski		5.00	
1	brexit-dave	david cameron	4.80	4.69	0.11
2	brezel-bush	george bush	3.91	4.79	-0.89
0	bringschuld-boateng	jérôme boateng		4.86	
40	brioni-schröder	gerhard schröder	4.79	4.85	-0.06
0	brustraus-brüderle	rainer brüderle		4.82	
0	bundestrainer-schröder	gerhard schröder		4.85	
4	bunga-bunga-brüderle	rainer brüderle	4.64	4.82	-0.18
17	burda-spahn	jens spahn	4.84	4.70	0.13
1	burrito-bieber	justin bieber	5.67	5.17	0.50
8	bvb-brandt	julian brandt	5.80	5.10	0.70
237	cdu-laschet	armin laschet	4.85	4.87	-0.01
7	cdu-scholz	olaf scholz	4.78	4.83	-0.05
22	change-obama	barack obama	4.76	4.88	-0.13
1	chaos-johnson	boris johnson	5.38	4.68	0.70
1	charlie-hollande	françois hollande		4.84	
0	chefsachen-schröder	gerhard schröder		4.85	
0	cocain-westerwelle	guido westerwelle		4.76	
1	container-westerwelle	guido westerwelle	4.66	4.76	-0.10
1	copy-and-paste-giffey	franziska giffey	5.10	4.97	0.13
4	corona-gnabry	serge gnabry	5.25	5.23	0.02
28	crush-ramelow	bodo ramelow	4.94	4.79	0.15
0	crystall-beck	volker beck		4.68	

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
5	dahlem-spahn	jens spahn	4.57	4.70	-0.14
141	demenz-scholz	olaf scholz	4.54	4.83	-0.29
0	denim-britney	britney spears		5.11	
87	diesel-dobrindt	alexander dobrindt	4.94	4.86	0.08
76	diesel-scheuer	andreas scheuer	4.65	4.75	-0.10
0	digital-lindner	christian lindner		4.84	
11	dinner-spahn	jens spahn	4.94	4.70	0.24
17	dirndl-brüderle	rainer brüderle	5.05	4.82	0.22
12	doktor-giffey	franziska giffey	4.85	4.97	-0.12
2	doppelpass-koch	roland koch	5.24	4.80	0.44
19	dosen-trittin	jürgen trittin	4.98	4.86	0.12
0	dosen-weidel	alice weidel		4.64	
125	drogen-beck	volker beck	4.53	4.68	-0.15
0	drogen-roth	claudia roth		4.83	
28	drohnenkrieg-obama	barack obama	4.56	4.88	-0.33
149	drohnen-obama	barack obama	4.68	4.88	-0.21
0	dussmann-strieder	peter strieder		5.07	
73	ehrenwort-kohl	helmut kohl	4.76	4.86	-0.10
0	einlass-merkel	angela merkel		4.92	
3	eisbällchen-trittin	jürgen trittin	5.12	4.86	0.26
3	elternrat-esken	saskia esken	4.90	4.85	0.05
16	em-schweinsteiger	bastian schweinsteiger	4.96	5.21	-0.25
22	enteignungs-kühnert	kevin kühnert	4.90	4.80	0.10
15	facebook-seehofer	horst seehofer	4.93	4.74	0.20
8	fallschirm-möllemann	jürgen möllemann	4.95	4.58	0.36
320	fdp-kubicki	wolfgang kubicki	4.57	4.72	-0.14
1386	fdp-lindner	christian lindner	4.81	4.84	-0.03
0	feinesache-steinmeier	frank-walter steinmeier		4.88	

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
0	feinkost-federer	roger federer		5.05	
31	ferkel-merkel	angela merkel	4.36	4.92	-0.55
32	ferrari-fegebank	katharina fegebank	5.16	4.94	0.22
28	ffp2-söder	markus söder	5.00	4.87	0.13
0	finpronil-schmidt	christian schmidt		4.89	
1	fischfillet-steinmeier	frank-walter steinmeier	4.96	4.88	0.08
3	fleisch-gaga	lady gaga	5.46	5.39	0.07
9	flexi-merkel	angela merkel	4.84	4.92	-0.07
19	flüchtlings-merkel	angela merkel	4.37	4.92	-0.55
1	flüchtlings-palmer	boris palmer	4.58	4.79	-0.21
2	flüchtlings-seehofer	horst seehofer	4.04	4.74	-0.69
0	flug-habeck	robert habeck		4.92	
1	flugmeilen-schulze	katharina schulze	5.07	4.94	0.12
9	folter-bush	george bush	3.95	4.79	-0.84
493	fpö-strache	heinz-christian strache	4.65	4.59	0.05
1	freiheits-ramelow	bodo ramelow	5.49	4.79	0.70
0	frühjahrs-vidal	arturo vidal		4.97	
76	fukushima-merkel	angela merkel	4.73	4.92	-0.19
0	funkloch-altmeier	peter altmeier		4.87	
507	g20-scholz	olaf scholz	4.78	4.83	-0.05
60	gas-putin	vladimir putin	4.87	4.89	-0.01
152	gas-schröder	gerhard schröder	4.69	4.85	-0.16
532	gazprom-schröder	gerhard schröder	4.78	4.85	-0.07
110	gedächtnislücken-scholz	olaf scholz	4.58	4.83	-0.25
13	gedöns-schröder	gerhard schröder	4.73	4.85	-0.12
173	geldkoffer-schäuble	wolfgang schäuble	4.60	4.72	-0.12
4	gender-schwesig	manuela schwesig	5.43	4.96	0.46
0	generalklausel-merkel	angela merkel		4.92	

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
13	gigabyte-özdemir	cem özdemir	4.90	4.85	0.05
0	glücks-schweinsteiger	bastian schweinsteiger		5.21	
545	glyphosat-schmidt	christian schmidt	4.81	4.89	-0.07
0	goldmann-merz	friedrich merz		4.46	
3	gold-ribery	franck ribéry	5.26	5.05	0.21
463	gold-rosi	rosi mittermaier	5.10	5.18	-0.08
1	großspenden-weidel	alice weidel	4.06	4.64	-0.58
126	grünen-habeck	robert habeck	4.70	4.92	-0.22
50	grünen-kretschmann	winfried kretschmann	4.80	5.02	-0.23
0	guardiola-boateng	jérôme boateng		4.86	
3	guatanamo-steinmeier	frank-walter steinmeier	5.37	4.88	0.49
0	guidomobil-westerwelle	guido westerwelle		4.76	
3	gülle-klöckner	julia klöckner	4.14	4.93	-0.80
15	gummistiefel-schröder	gerhard schröder	4.78	4.85	-0.07
0	halal-meuthen	jörg meuthen		4.49	
9	hanf-özdemir	cem özdemir	4.78	4.85	-0.07
1	hartz4-fischer	joschka fischer	2.91	4.89	-1.98
18	hartz4-schröder	gerhard schröder	4.94	4.85	0.09
54	hartz-schröder	gerhard schröder	4.83	4.85	-0.02
2	helikopter-habeck	robert habeck	5.00	4.92	0.07
6	helikopter-kretschmann	winfried kretschmann	5.20	5.02	0.17
0	hemden-kubicki	wolfgang kubicki		4.72	
8	hindukusch-struck	peter struck	4.36	4.73	-0.37
2	hoffnungs-obama	barack obama	4.70	4.88	-0.18
2	horizont-seehofer	horst seehofer	5.42	4.74	0.68
0	hosenschiss-steinbrück	peer steinbrück		4.74	
2	hotelbar-brüderle	rainer brüderle		4.82	
1	hubschrauber-habeck	robert habeck	5.21	4.92	0.28

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
9	hufeisen-scharping	rudolf scharping	4.34	4.71	-0.37
5	hummer-lauterbach	karl lauterbach	5.01	4.69	0.32
7	hummer-wagenknecht	sahra wagenknecht	4.97	4.72	0.25
8	immo-spahn	jens spahn	4.73	4.70	0.03
6	impf-söder	markus söder	5.26	4.87	0.39
41	interkontinental-kössler	georg kössler	5.18	5.14	0.04
81	inzucht-schäuble	wolfgang schäuble	4.75	4.72	0.03
62	islam-merkel	angela merkel	4.91	4.92	-0.01
0	israel-petry	frauke petry		4.73	
222	jet-jamila	jamila schäfer	5.33	4.87	0.45
0	jet-janecek	dieter janecek		5.12	
0	jetleg-janecek	dieter janecek		5.12	
0	jogging-joschka	joschka fischer		4.89	
0	kanonenboot-trittin	jürgen trittin		4.86	
11	kastenstand-klöckner	julia klöckner	4.69	4.93	-0.25
9	katastrophen-spahn	jens spahn	4.51	4.70	-0.20
11	kernkraft-merkel	angela merkel	4.84	4.92	-0.07
7	kernkraft-söder	markus söder	4.85	4.87	-0.02
2	kerosina-schulze	katharina schulze	5.53	4.94	0.59
2	kerosin-krause	günther krause	5.37	4.75	0.62
6	kerosin-schulze	katharina schulze	5.06	4.94	0.12
6	kiffer-künast	renate künast	5.41	4.87	0.54
7	kiffer-lindner	christian lindner	4.71	4.84	-0.13
4	kilometerpauschale-habeck	robert habeck	5.73	4.92	0.80
1	kinderaugen-gauland	alexander gauland	4.85	4.69	0.17
204	kindergarten-klopp	jürgen klopp	5.26	5.09	0.17
2	kita-poggenburg	andre poggenburg	4.59	4.46	0.13
8	kittel-laschet	armin laschet	4.81	4.87	-0.05

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
3	klarnamen-schäuble	wolfgang schäuble	5.28	4.72	0.56
1	klartext-scholz	olaf scholz	6.44	4.83	1.61
0	kli-kla-klasnic	ivan klasnic		4.92	
0	klimaerdbeben-künast	renate künast		4.87	
145	klima-gretl	greta thunberg	4.93	5.04	-0.11
31	knast-hoeneß	uli hoeneß	4.10	4.98	-0.89
61	kneipen-kubicki	wolfgang kubicki	4.99	4.72	0.27
69	kniffel-armin	armin laschet	5.01	4.87	0.14
373	koffer-schäuble	wolfgang schäuble	4.63	4.72	-0.09
252	kohle-laschet	armin laschet	4.85	4.87	-0.02
1	kohlelobby-laschet	armin laschet	5.88	4.87	1.01
1	kohle-merz	friedrich merz	4.88	4.46	0.43
0	kohle-rössler	philipp rössler		5.03	
0	koks-bush	george bush		4.79	
5	kosovo-schröder	gerhard schröder	5.05	4.85	0.20
1	kreuzzug-bush	george bush	4.54	4.79	-0.25
46	kriegs-obama	barack obama	4.66	4.88	-0.22
1	kriegsspiel-klingbeil	lars klingbeil	4.68	5.00	-0.31
0	krypto-erdogan	recep tayyip erdoğan		4.63	
0	kugeleis-tritin	jürgen trittin		4.86	
1	kultur-özoguz	aydan özoğuz	5.85	4.83	1.02
1	kz-weidel	alice weidel	4.17	4.64	-0.47
36	langstrumpf-nahles	andrea nahles	4.85	4.79	0.07
2	last-minute-neuer	manuel neuer	5.19	5.08	0.11
0	lazarus-lindner	christian lindner		4.84	
1	leg-angelina	angelina jolie	4.22	5.12	-0.90
0	lillenhammer-schäfer	jamila schäfer		4.87	
428	lipobay-lauterbach	karl lauterbach	4.72	4.69	0.03

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
66	lobby-merkel	angela merkel	4.81	4.92	-0.10
231	lockdown-lauterbach	karl lauterbach	4.71	4.69	0.01
168	locker-laschet	armin laschet	4.95	4.87	0.08
4	lockerungs-söder	markus söder	4.82	4.87	-0.05
1	lock-spahn	jens spahn	3.87	4.70	-0.84
3	lösch-leyen	ursula von der leyen	4.74	4.87	-0.12
15	maidan-steinmeier	frank-walter steinmeier	4.58	4.88	-0.30
60	maskendeal-spahn	jens spahn	4.54	4.70	-0.16
16	masken-löbel	nikolas löbel	4.63	4.52	0.11
26	masken-merkel	angela merkel	4.77	4.92	-0.15
9	masken-nüßlein	georg nüßlein	4.86	4.63	0.23
505	maut-dobrindt	alexander dobrindt	4.76	4.86	-0.10
5	mckinsey-leyen	ursula von der leyen	4.66	4.87	-0.21
1	meilen-roth	claudia roth	6.48	4.83	1.65
22	meineid-gauland	alexander gauland	4.40	4.69	-0.28
61	meineid-kohl	helmut kohl	4.56	4.86	-0.30
83	meineid-petry	frauke petry	4.36	4.73	-0.37
8	microsoft-gates	bill gates	4.81	5.08	-0.28
1589	migranten-merkel	angela merkel	4.75	4.92	-0.16
1	milchkannen-karlicek	anja karliczek	5.01	5.09	-0.09
0	miles-and-more-özdemir	cem özdemir		4.85	
20	milzriss-kalbitz	andreas kalbitz	4.35	4.57	-0.21
617	mini-merkel	angela merkel	4.88	4.92	-0.04
13	mischpoke-özdemir	cem özdemir	4.71	4.85	-0.15
0	mittelstands-westerwelle	guido westerwelle		4.76	
13	möbelhaus-laschet	armin laschet	5.26	4.87	0.39
1	model-lindner	christian lindner	5.34	4.84	0.50
401	mollath-merk	beate merk	4.58	4.67	-0.10

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
23	moslem-obama	barack obama	4.85	4.88	-0.03
7	mövenpick-westerwelle	guido westerwelle	4.94	4.76	0.18
8	mutter-seehofer	horst seehofer	4.98	4.74	0.25
2	nachrüstungs-schmidt	helmut schmidt	4.80	4.87	-0.06
375	nato-erdogan	recep tayyip erdoğan	4.58	4.63	-0.05
496	nestle-klöckner	julia klöckner	4.92	4.93	-0.01
508	netzdg-maas	heiko maas	4.66	4.68	-0.03
60	nordstream-schwesig	manuela schwesig	4.74	4.96	-0.22
5	notbremse-söder	markus söder	4.77	4.87	-0.10
1	notenwendels-laschet	armin laschet	5.11	4.87	0.24
0	nullen-scholz	olaf scholz		4.83	
6	ofen-habeck	robert habeck	5.22	4.92	0.29
1	ohrfeigen-poldi	lukas podolski	2.19	5.00	-2.81
1	öko-seehofer	horst seehofer		4.74	
12	opfer-schäuble	wolfgang schäuble	4.11	4.72	-0.61
285	pack-gabriel	sigmar gabriel	4.62	4.79	-0.17
0	palästinensertuch-trittin	jürgen trittin		4.86	
95	panik-greta	greta thunberg	4.76	5.04	-0.29
40	panik-merkel	angela merkel	4.63	4.92	-0.29
7	peak-boateng	jérôme boateng	4.89	4.86	0.03
2	peitschen-steinbrück	peer steinbrück	5.20	4.74	0.47
40	pendler-habeck	robert habeck	4.88	4.92	-0.05
75	pendlerpauschale-habeck	robert habeck	4.92	4.92	
0	petro-bush	george bush		4.79	
5	pferde-nahles	andrea nahles	5.17	4.79	0.38
1	pillen-spahn	jens spahn	4.72	4.70	0.01
0	pipelineendstation-merkel	angela merkel		4.92	
203	pippi-nahles	andrea nahles	5.09	4.79	0.31

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
8	plagiat-bärbock	annalena baerbock	4.58	4.97	-0.39
165	plagiat-giffey	franziska giffey	4.56	4.97	-0.41
1	plastikeis-schulze	katharina schulze	4.81	4.94	-0.14
1	plastik-schulze	katharina schulze	5.82	4.94	0.88
0	plastiktüte-hofreiter	anton hofreiter		4.96	
0	pleiten-mourinho	josé mourinho		4.96	
157	podcast-drosten	christian drosten	4.92	4.74	0.18
6	pokal-klose	miroslav klose	5.36	4.99	0.37
68	polizeigewalt-scholz	olaf scholz	4.69	4.83	-0.14
8	privatschul-schwesig	manuela schwesig	5.00	4.96	0.04
79	problem-seehofer	horst seehofer	4.55	4.74	-0.19
1	promotionsbetrugs-giffey	franziska giffey	5.16	4.97	0.19
0	provokations-petry	frauke petry		4.73	
2	puh-putin	vladimir putin	4.32	4.89	-0.57
5	rambo-bush	george bush	4.98	4.79	0.19
0	raser-franz	franz untersteller		4.97	
1	reise-neubauer	luisa neubauer	5.10	4.97	0.12
2	renten-bush	george bush	4.44	4.79	-0.35
10	renten-scholz	olaf scholz	4.73	4.83	-0.10
1	rettungs-schäuble	wolfgang schäuble	4.32	4.72	-0.40
0	risiko-khedira	sami khedira		5.02	
1	roadtrip-schulze	katharina schulze	6.21	4.94	1.27
914	rolex-chebli	sawsan chebli	4.79	4.75	0.03
0	rolex-rummennige	karl-heinz rummenigge		5.04	
10	rolli-schäuble	wolfgang schäuble	4.82	4.72	0.10
1	rückrunden-klose	miroslav klose	5.67	4.99	0.68
16	ruhrpott-messi	lionel messi	5.20	5.13	0.07
64	rwe-laschet	armin laschet	4.80	4.87	-0.07

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
2	sachsenhausen-weidel	alice weidel	4.79	4.64	0.15
31	salto-klose	miroslav klose	5.26	4.99	0.26
39	schande-höcke	björn höcke	4.38	4.57	-0.18
2	schande-schäuble	wolfgang schäuble	4.20	4.72	-0.52
74	scharia-chebli	sawsan chebli	4.63	4.75	-0.12
32	schießbefehl-petry	frauke petry	4.69	4.73	-0.04
11	schießübung-bystron	petr bystron	4.69	4.83	-0.14
730	schiss-gauland	alexander gauland	4.61	4.69	-0.08
0	schlieffen-gauland	alexander gauland		4.69	
19	schmiergeld-schäuble	wolfgang schäuble	4.72	4.72	
0	schnee-fischer	joschka fischer		4.89	
10	schreibmaschinen-hoeneß	uli hoeneß	4.75	4.98	-0.23
2	schreibtisch-niebel	dirk niebel	4.51	4.83	-0.32
12	schubladen-schäuble	wolfgang schäuble	4.60	4.72	-0.12
6	schuldenberg-scholz	olaf scholz	5.27	4.83	0.45
7	schuldenbremse-merkel	angela merkel	4.51	4.92	-0.41
8	schulden-schäuble	wolfgang schäuble	4.35	4.72	-0.37
0	schuld-gauck	joachim gauck		4.94	
14	schuldkult-weidel	alice weidel	4.25	4.64	-0.39
59	schummel-giffey	franziska giffey	4.69	4.97	-0.28
0	schussbefehl-von-storch	beatrix von storch		4.82	
0	schwarzekassen-schäuble	wolfgang schäuble		4.72	
193	schwarzgeld-schäuble	wolfgang schäuble	4.60	4.72	-0.12
0	schweiz-schäuble	wolfgang schäuble		4.72	
0	security-schäuble	wolfgang schäuble		4.72	
1	seenotrettungs-seehofer	horst seehofer	4.09	4.74	-0.65
2	segel-söder	markus söder	5.28	4.87	0.41
0	seitenaus-bobic	fredi bobic		5.02	

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
45	selfie-merkel	angela merkel	4.99	4.92	0.07
0	siegelring-fischer	joschka fischer		4.89	
5	smartie-spahn	jens spahn	4.91	4.70	0.20
1	sommertraum-beckenbauer	franz beckenbauer	5.69	5.09	0.60
23	spaß-guido	guido westerwelle	4.88	4.76	0.12
79	spd-giffey	franziska giffey	4.77	4.97	-0.20
526	spd-scholz	olaf scholz	4.72	4.83	-0.11
1	spendengeld-laschet	armin laschet	5.53	4.87	0.66
1	spenden-hoeneß	uli hoeneß	4.76	4.98	-0.22
21	spenden-schäuble	wolfgang schäuble	4.64	4.72	-0.08
11	spenden-spahn	jens spahn	4.77	4.70	0.06
1	spielzeugauto-haderthauer	christine haderthauer	5.12	4.77	0.35
6	spritzen-spahn	jens spahn	4.93	4.70	0.23
1	staatsglauben-schwesig	manuela schwesig	3.93	4.96	-1.03
2	steak-drosten	christian drosten	5.14	4.74	0.40
56	steuer-hoeneß	uli hoeneß	4.72	4.98	-0.27
2	südafrika-bystron	petr bystron	3.86	4.83	-0.97
7	talkshow-habeck	robert habeck	5.09	4.92	0.17
53	talkshow-lauterbach	karl lauterbach	5.11	4.69	0.41
2	talkshow-lindner	christian lindner	5.57	4.84	0.73
3	tampon-scholz	olaf scholz	5.14	4.83	0.32
4	tätervolk-hohmann	martin hohmann	4.06	4.57	-0.51
10	telefonterror-kahrs	johannes kahrs	4.55	4.85	-0.30
1	telefon-wulff	christian wulff	5.06	4.83	0.23
0	telegram-steinmeier	frank-walter steinmeier		4.88	
93	teppich-niebel	dirk niebel	4.92	4.83	0.09
9	terror-bush	george bush	4.45	4.79	-0.34
0	terrorparanoia-schäuble	wolfgang schäuble		4.72	

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
30	terror-thomas	thomas de maizière	4.52	4.66	-0.14
5	thermo-lindner	christian lindner	4.86	4.84	0.02
6	toblerone-meuthen	jörg meuthen	4.30	4.49	-0.18
139	toiletten-brandner	stephan brandner	4.78	4.65	0.14
9	tönnies-laschet	armin laschet	4.86	4.87	-0.01
24	tore-klose	miroslav klose	5.89	4.99	0.90
42	tore-messi	lionel messi	5.12	5.13	-0.01
5	tore-poldi	lukas podolski	5.85	5.00	0.85
0	toscana-lafontaine	oskar lafontaine		4.72	
0	toscana-trittin	jürgen trittin		4.86	
2	transfer-bobic	fredi bobic	4.62	5.02	-0.40
3	tribünen-hoeneß	uli hoeneß	4.90	4.98	-0.09
0	turban-trittin	jürgen trittin		4.86	
105	türken-laschet	armin laschet	4.71	4.87	-0.16
2	türken-roth	claudia roth	4.87	4.83	0.04
63	turnschuh-fischer	joschka fischer	4.95	4.89	0.06
8	tv-lauterbach	karl lauterbach	4.83	4.69	0.13
0	überlebens-sammer	matthias sammer		5.07	
0	überschuldungs-schäuble	wolfgang schäuble		4.72	
0	überwachungsterrorist-schäuble	wolfgang schäuble		4.72	
0	überwachungswahn-schäuble	wolfgang schäuble		4.72	
9	ultimatum-seehofer	horst seehofer	4.69	4.74	-0.04
3	umwelt-söder	markus söder	4.66	4.87	-0.21
74	vanlaack-laschet	armin laschet	4.80	4.87	-0.07
4	veggie-künast	renate künast	4.38	4.87	-0.49
0	verfassungs-schäuble	wolfgang schäuble		4.72	
0	verschärfungs-maas	heiko maas		4.68	
0	verschwörungs-mourinho	josé mourinho		4.96	

Table 10 continued from previous page

count	compound	name	c_val	n_val	diff
5	vielflieger-hofreiter	anton hofreiter	5.28	4.96	0.31
1	vielflieger-lauterbach	karl lauterbach	4.94	4.69	0.24
0	vielflieger-özdemir	cem özdemir		4.85	
40	villen-spahn	jens spahn	4.83	4.70	0.13
2	viren-söder	markus söder	4.88	4.87	0.01
0	viren-spahn	jens spahn		4.70	
719	vogelschiss-gauland	alexander gauland	4.61	4.69	-0.07
17	vollgas-vettel	sebastian vettel	5.07	4.98	0.08
0	vorsaison-boateng	jérôme boateng		4.86	
84	vote-obama	barack obama	5.17	4.88	0.29
0	wachstums-merkel	angela merkel		4.92	
9	waffen-schreiber	karlheinz schreiber	4.65	4.44	0.21
128	wahlkampf-obama	barack obama	4.91	4.88	0.03
0	waterboarding-bush	george bush		4.79	
6	wehrmacht-gauland	alexander gauland	4.25	4.69	-0.43
2	wende-höcke	björn höcke	4.98	4.57	0.41
0	wende-westerwelle	guido westerwelle		4.76	
1	west-rühe	volker rühe	4.52	4.75	-0.22
0	wiedergeburt-bush	george bush		4.79	
11	willkommens-merkel	angela merkel	4.42	4.92	-0.50
2	wm-merkel	angela merkel	5.04	4.92	0.13
12	wm-müller	thomas müller	5.07	5.12	-0.05
0	zeichentrick-lagerfeld	karl lagerfeld		5.32	
4	zigarren-clinton	bill clinton	4.61	4.82	-0.21
2	zoten-brüderle	rainer brüderle	4.99	4.82	0.17

Table 11: All compound-modifier (lemma) pairs including compound valence (c_val), modifier (lemma) valence (m_val) and difference (diff), rounded to two decimal places.

compound	modifier	c_val	m_val	diff
abschiebe-kretschmann	abschieben	4.83	3.69	-1.15
abschiebe-seehofer	abschieben	4.60	3.69	-0.92
abschottungs-merkel	abschottung		3.10	
abschreib-franziska	abschreiben	4.18	3.96	-0.22
abseits-kroos	abseits		4.46	
afd-höcke	afd	4.63	4.43	-0.20
afd-meuthen	afd	4.59	4.43	-0.17
affären-scholz	affäre	4.45	2.55	-1.90
agenda-schröder	agenda	4.85	5.17	0.32
airline-analena	airline		5.48	
aktenkoffer-schäuble	aktenkoffer	4.69	3.55	-1.14
aktien-merz	aktie	4.84	4.25	-0.59
akw-merkel	akw	4.33	1.63	-2.69
alleingang-schmidt	alleingang	4.96	4.09	-0.87
alleingang-söder	alleingang	5.05	4.09	-0.96
amigo-strauß	amigo	4.52	5.70	1.18
amok-seehofer	amok	5.39	1.82	-3.57
ampel-söder	ampel	5.08	5.30	0.22
anatolien-özoguz	anatolien	5.17	4.07	-1.10
anden-özdemir	anden	5.16	5.30	0.15
apartheid-gabriel	apartheid	4.51	2.34	-2.17
arizona-jakob	arizona	5.43	5.30	-0.13
armani-schröder	armani	4.75	5.59	0.85
armutsbericht-rössler	armutsbericht		4.46	
astra-spahn	astra	4.75	5.28	0.52
asyltourismus-söder	asyltourismus	4.81		

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
atom-greta	atom	5.09	3.94	-1.15
atom-merkel	atom	4.61	3.94	-0.67
atom-putin	atom	4.46	3.94	-0.52
audi-andi	audi	4.81	5.10	0.29
aufschrei-brüderle	aufschrei	4.37	2.08	-2.29
augustus-amthor	augustus	4.70	5.44	0.74
australien-rahmstorf	australien		5.17	
bagdad-angela	bagdad		3.90	
bahnhofs-palmer	bahnhof		4.30	
banken-steinbrück	bank	4.70	5.48	0.78
bätschi-nahles	bätschi	4.82	4.74	-0.07
besatzer-hollande	besatzer		1.51	
bhutan-beckham	bhutan	5.66	5.26	-0.40
bienen-söder	biene	5.00	7.30	2.30
bierdeckel-merz	bierdeckel	4.70	5.65	0.95
bierzelt-söder	bierzelt	4.84	5.81	0.97
big-brother-westerwelle	big-brother	4.52		
bimbes-kohl	bimbes	4.80	5.63	0.83
bimbes-schäuble	bimbes	4.75	5.63	0.87
bingo-drosten	bingo	5.31	5.64	0.33
blackrock-merz	blackrock	4.82	4.65	-0.16
blasphemie-bobic	blasphemie		1.20	
blockade-rössler	blockade		2.19	
blowout-blasel	blowout	5.07	3.59	-1.48
bomben-fischer	bombe	5.06	2.47	-2.59
bomb-obama	bombe	4.98	2.47	-2.51
bonusmeilen-habeck	bonusmeile			
bonusmeilen-özdemir	bonusmeile	4.75		

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
bonusmeilen-schulze	bonusmeile			
bosporus-putin	bosporus	4.71	4.04	-0.67
bratwurst-bobic	bratwurst	5.30	6.13	0.83
brechstangen-poldi	brechstange		3.50	
brexit-dave	brexit	4.80	3.09	-1.71
brezel-bush	brezel	3.91	6.44	2.54
bringschuld-boateng	bringschuld		3.44	
brioni-schröder	brioni	4.79	5.00	0.21
brustraus-brüderle	brustraus			
bundestrainer-schröder	bundestrainer		4.55	
bunga-bunga-brüderle	bunga-bunga	4.64		
burda-spahn	burda	4.84	6.00	1.16
burrito-bieber	burrito	5.67	4.58	-1.09
bvb-brandt	bvb	5.80	5.04	-0.76
cdu-laschet	cdu	4.85	4.52	-0.34
cdu-scholz	cdu	4.78	4.52	-0.26
change-obama	change	4.76	5.21	0.45
chaos-johnson	chaos	5.38	3.55	-1.83
charlie-hollande	charlie		4.86	
chefsachen-schröder	chefsache		4.46	
cocain-westerwelle	cocaine		4.01	
container-westerwelle	container	4.66	4.31	-0.35
copy-and-paste-giffey	copy-and-paste	5.10		
corona-gnabry	corona	5.25	6.37	1.12
crush-ramelow	crush	4.94	4.18	-0.76
crystall-beck	crystal		5.86	
dahlem-spahn	dahlem	4.57	5.08	0.51
demenz-scholz	demenz	4.54	1.97	-2.57

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
denim-britney	denim		4.59	
diesel-dobrindt	diesel	4.94	5.21	0.27
diesel-scheuer	diesel	4.65	5.21	0.56
digital-lindner	digital		5.58	
dinner-spahn	dinner	4.94	5.89	0.95
dirndl-brüderle	dirndl	5.05	5.85	0.80
doktor-giffey	doktor	4.85	3.93	-0.91
doppelpass-koch	doppelpass	5.24	5.17	-0.07
dosen-trittin	dose	4.98	4.74	-0.24
dosen-weidel	dose		4.74	
drogen-beck	droge	4.53	2.84	-1.69
drogen-roth	droge		2.84	
drohnenkrieg-obama	drohnenkrieg	4.56	2.27	-2.29
drohnen-obama	drohne	4.68	4.13	-0.54
dussmann-strieder	dussmann		5.22	
ehrenwort-kohl	ehrenwort	4.76	4.26	-0.50
einlass-merkel	einlass		4.97	
eisbällchen-trittin	eisball	5.12	5.76	0.64
elternrat-esken	elternrat	4.90	4.82	-0.09
em-schweinsteiger	em	4.96	3.83	-1.13
enteignungs-kühnert	enteignung	4.90	1.51	-3.39
facebook-seehofer	facebook	4.93	5.14	0.20
fallschirm-möllemann	fallschirm	4.95	4.28	-0.67
fdp-kubicki	fdp	4.57	4.30	-0.28
fdp-lindner	fdp	4.81	4.30	-0.51
feinesache-steinmeier	feinesache			
feinkost-federer	feinkost		5.42	
ferkel-merkel	ferkel	4.36	4.13	-0.23

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
ferrari-fegebank	ferrari	5.16	4.87	-0.29
ffp2-söder	ffp2	5.00		
finpronil-schmidt	finpronil			
fischfillet-steinmeier	fischfilet	4.96	5.65	0.69
fleisch-gaga	fleisch	5.46	4.64	-0.82
flexi-merkel	flexi	4.84	4.85	0.01
flüchtlings-merkel	flüchtling	4.37	1.74	-2.62
flüchtlings-palmer	flüchtling	4.58	1.74	-2.84
flüchtlings-seehofer	flüchtling	4.04	1.74	-2.30
flug-habeck	flug		5.41	
flugmeilen-schulze	flugmeile	5.07		
folter-bush	folter	3.95	0.89	-3.06
fpö-strache	fpö	4.65	4.14	-0.51
freiheits-ramelow	freiheit	5.49	5.93	0.43
frühjahrs-vidal	frühjahr		5.52	
fukushima-merkel	fukushima	4.73	3.12	-1.61
funkloch-altmeier	funkloch		4.68	
g20-scholz	g20	4.78		
gas-putin	gas	4.87	4.54	-0.34
gas-schröder	gas	4.69	4.54	-0.16
gazprom-schröder	gazprom	4.78	4.29	-0.49
gedächtnislücken-scholz	gedächtnislücke	4.58	1.58	-3.00
gedöns-schröder	gedöns	4.73	4.62	-0.11
geldkoffer-schäuble	geldkoffer	4.60	3.27	-1.33
gender-schwesig	gender	5.43	4.64	-0.79
generalklausel-merkel	generalklausel		3.34	
gigabyte-özdemir	gigabyte	4.90	5.17	0.27
glücks-schweinsteiger	glück		7.86	

Table 11 continued from previous page

compound	**modifier**	**c_val**	**m_val**	**diff**
glyphosat-schmidt	glyphosat	4.81	3.47	-1.35
goldmann-merz	goldmann		6.59	
gold-ribery	gold	5.26	6.66	1.40
gold-rosi	gold	5.10	6.66	1.56
großspenden-weidel	großspende	4.06		
grünen-habeck	grüne	4.70	6.56	1.86
grünen-kretschmann	grüne	4.80	6.56	1.77
guardiola-boateng	guardiola		4.82	
guatanamo-steinmeier	guantanamo	5.37	3.02	-2.36
guidomobil-westerwelle	guidomobil		5.89	
gülle-klöckner	gülle	4.14	3.61	-0.53
gummistiefel-schröder	gummstiefel	4.78		
halal-meuthen	halal		4.84	
hanf-özdemir	hanf	4.78	4.36	-0.42
hartz4-fischer	hartz4	2.91		
hartz4-schröder	hartz4	4.94		
hartz-schröder	hartz	4.83	3.76	-1.06
helikopter-habeck	helikopter	5.00	4.75	-0.25
helikopter-kretschmann	helikopter	5.20	4.75	-0.45
hemden-kubicki	hemd		4.67	
hindukusch-struck	hindukusch	4.36	3.94	-0.42
hoffnungs-obama	hoffnung	4.70	6.57	1.86
horizont-seehofer	horizont	5.42	5.66	0.24
hosenschiss-steinbrück	hosenschiss			
hotelbar-brüderle	hotelbar		6.19	
hubschrauber-habeck	hubschrauber	5.21	4.36	-0.84
hufeisen-scharping	hufeisen	4.34	6.02	1.67
hummer-lauterbach	hummer	5.01	7.12	2.11

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
hummer-wagenknecht	hummer	4.97	7.12	2.15
immo-spahn	immobilie	4.73		
impf-söder	impfen	5.26	3.30	-1.96
interkontinental-kössler	interkontinental	5.18	4.85	-0.33
inzucht-schäuble	inzucht	4.75	2.93	-1.83
islam-merkel	islam	4.91	4.41	-0.50
israel-petry	israel		4.51	
jet-jamila	jet	5.33	5.41	0.08
jet-janecek	jet	5.27	5.41	0.14
jetleg-janecek	jetlag		2.87	
jogging-joschka	jogging		5.53	
kanonenboot-trittin	kanonenboot		4.32	
kastenstand-klöckner	kastenstand	4.69	4.03	-0.66
katastrophen-spahn	katastrophe	4.51	1.47	-3.04
kernkraft-merkel	kernkraft	4.84	3.35	-1.49
kernkraft-söder	kernkraft	4.85	3.35	-1.51
kerosina-schulze	kerosin	5.53	4.05	-1.48
kerosin-krause	kerosin	5.37	4.05	-1.32
kerosin-schulze	kerosin	5.06	4.05	-1.01
kiffer-künast	kiffer	5.41	2.42	-2.99
kiffer-lindner	kiffer	4.71	2.42	-2.29
kilometerpauschale-habeck	kilometerpauschale	5.73	4.19	-1.54
kinderaugen-gauland	kinderauge	4.85		
kindergarten-klopp	kindergarten	5.26	5.72	0.45
kita-poggenburg	kita	4.59	6.45	1.86
kittel-laschet	kittel	4.81	4.92	0.11
klarnamen-schäuble	klarname	5.28	2.13	-3.15
klartext-scholz	klartext	6.44	5.14	-1.30

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
kli-kla-klasnic	kli-kla			
klimaerdbeben-künast	klimaerdbeben			
klima-gretl	klima	4.93	6.18	1.24
knast-hoeneß	knast	4.10	1.36	-2.74
kneipen-kubicki	kneipe	4.99	4.95	-0.04
kniffel-armin	kniffel	5.01	6.47	1.46
koffer-schäuble	koffer	4.63	3.90	-0.73
kohle-laschet	kohle	4.85	4.30	-0.55
kohlelobby-laschet	kohlelobby	5.88		
kohle-merz	kohle	4.88	4.30	-0.59
kohle-rössler	kohle		4.30	
koks-bush	koks		3.26	
kosovo-schröder	kosovo	5.05	4.30	-0.75
kreuzzug-bush	kreuzzug	4.54	3.31	-1.23
kriegs-obama	krieg	4.66	1.90	-2.76
kriegsspiel-klingbeil	kriegsspiel	4.68	2.47	-2.21
krypto-erdogan	krypto		4.00	
kugeleis-tritin	kugeleis		4.18	
kultur-özoguz	kultur	5.85	6.39	0.54
kz-weidel	kz	4.17	1.49	-2.68
langstrumpf-nahles	langstrumpf	4.85	5.55	0.70
last-minute-neuer	last-minute	5.19		
lazarus-lindner	lazarus		6.03	
leg-angelina	leg	4.22	4.16	-0.06
lillenhammer-schäfer	lillehammer		5.09	
lipobay-lauterbach	lipobay	4.72	2.67	-2.06
lobby-merkel	lobby	4.81	4.96	0.15
lockdown-lauterbach	lockdown	4.71	2.97	-1.74

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
locker-laschet	lockern	4.95	4.75	-0.20
lockerungs-söder	lockerung	4.82	3.64	-1.18
lock-spahn	locken	3.87	6.52	2.65
lösch-leyen	löschen	4.74	4.75	0.01
maidan-steinmeier	maidan	4.58	3.51	-1.07
maskendeal-spahn	maskendeal	4.54		
masken-löbel	maske	4.63	4.68	0.05
masken-merkel	maske	4.77	4.68	-0.09
masken-nüßlein	maske	4.86	4.68	-0.18
maut-dobrindt	maut	4.76	3.56	-1.21
mckinsey-leyen	mckinsey	4.66	4.81	0.15
meilen-roth	meile	6.48	5.03	-1.45
meineid-gauland	meineid	4.40	1.81	-2.59
meineid-kohl	meineid	4.56	1.81	-2.75
meineid-petry	meineid	4.36	1.81	-2.55
microsoft-gates	microsoft	4.81	5.12	0.31
migranten-merkel	migranten	4.75	4.55	-0.21
milchkannen-karlicek	milchkanne	5.01	4.67	-0.34
miles-and-more-özdemir	miles-and-more			
milzriss-kalbitz	milzriss	4.35	1.42	-2.93
mini-merkel	mini	4.88	5.31	0.43
mischpoke-özdemir	mischpoke	4.71	2.98	-1.73
mittelstands-westerwelle	mittelstand		5.42	
möbelhaus-laschet	möbelhaus	5.26	5.26	0.00
model-lindner	model	5.34	5.01	-0.33
mollath-merk	mollath	4.58	1.89	-2.69
moslem-obama	moslem	4.85	4.11	-0.75
mövenpick-westerwelle	mövenpick	4.94	5.53	0.59

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
mutter-seehofer	mutter	4.98	4.39	-0.60
nachrüstungs-schmidt	nachrüstung	4.80	4.07	-0.74
nato-erdogan	nato	4.58	3.79	-0.79
nestle-klöckner	nestle	4.92	4.64	-0.28
netzdg-maas	netzdg	4.66		
nordstream-schwesig	nordstream	4.74	4.84	0.09
notbremse-söder	notbremse	4.77	4.40	-0.37
notenwendels-laschet	notenwendel	5.11		
nullen-scholz	null		4.22	
ofen-habeck	ofen	5.22	5.37	0.15
ohrfeigen-poldi	ohrfeige	2.19	1.69	-0.50
öko-seehofer	öko		6.13	
opfer-schäuble	opfer	4.11	1.00	-3.11
pack-gabriel	pack	4.62	5.02	0.40
palästinensertuch-trittin	palästinensertuch		3.91	
panik-greta	panik	4.76	2.73	-2.02
panik-merkel	panik	4.63	2.73	-1.89
peak-boateng	peak	4.89	4.54	-0.35
peitschen-steinbrück	peitsche	5.20	2.76	-2.45
pendler-habeck	pendler	4.88	5.17	0.29
pendlerpauschale-habeck	pendlerpauschale	4.92	3.95	-0.97
petro-bush	petro		4.64	
pferde-nahles	pferd	5.17	5.21	0.05
pillen-spahn	pille	4.72	4.24	-0.47
pipelineendstation-merkel	pipelineendstation			
pippi-nahles	pippi	5.09	5.79	0.70
plagiat-bärbock	plagiat	4.58	2.46	-2.12
plagiat-giffey	plagiat	4.56	2.46	-2.10

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
plastikeis-schulze	plastikeis	4.81		
plastik-schulze	plastik	5.82	5.31	-0.51
plastiktüte-hofreiter	plastiktüte		2.71	
pleiten-mourinho	pleite		1.92	
podcast-drosten	podcast	4.92	5.78	0.86
pokal-klose	pokal	5.36	4.84	-0.52
polizeigewalt-scholz	polizeigewalt	4.69	1.90	-2.79
privatschul-schwesig	privatschule	5.00	5.20	0.20
problem-seehofer	problem	4.55	3.44	-1.11
promotionsbetrugs-giffey	promotionsbetrugs	5.16		
provokations-petry	provokation		3.19	
puh-putin	puh	4.32	6.29	1.97
rambo-bush	rambo	4.98	2.33	-2.65
raser-franz	raser		3.66	
reise-neubauer	reise	5.10	6.20	1.10
renten-bush	rente	4.44	2.90	-1.54
renten-scholz	rente	4.73	2.90	-1.83
rettungs-schäuble	rettung	4.32	6.01	1.68
risiko-khedira	risiko		2.19	
roadtrip-schulze	roadtrip	6.21	4.87	-1.34
rolex-chebli	rolex	4.79	5.05	0.26
rolex-rummennige	rolex		5.05	
rolli-schäuble	rolli	4.82	5.78	0.96
rückrunden-klose	rückrunde	5.67	4.22	-1.45
ruhrpott-messi	ruhrpott	5.20	4.58	-0.62
rwe-laschet	rwe	4.80	3.76	-1.04
sachsenhausen-weidel	sachsenhausen	4.79	2.87	-1.91
salto-klose	salto	5.26	4.83	-0.43

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
schande-höcke	schande	4.38	1.34	-3.05
schande-schäuble	schande	4.20	1.34	-2.87
scharia-chebli	scharia	4.63	3.76	-0.87
schießbefehl-petry	schießbefehl	4.69		
schießübung-bystron	schießübung	4.69		
schiss-gauland	schiss	4.61	2.09	-2.52
schlieffen-gauland	schlieffen		4.38	
schmiergeld-schäuble	schmiergeld	4.72	1.58	-3.14
schnee-fischer	schnee		5.04	
schreibmaschinen-hoeneß	schreibmaschine	4.75	4.41	-0.34
schreibtisch-niebel	schreibtisch	4.51	4.27	-0.24
schubladen-schäuble	schublade	4.60	3.13	-1.47
schuldenberg-scholz	schuldenberg	5.27	1.89	-3.39
schuldenbremse-merkel	schuldenbremse	4.51	4.07	-0.43
schulden-schäuble	schulden	4.35	1.92	-2.43
schuld-gauck	schuld		0.90	
schuldkult-weidel	schuldkult	4.25	1.17	-3.09
schummel-giffey	schummeln	4.69	5.47	0.78
schussbefehl-von-storch	schussbefehl		2.53	
schwarzekassen-schäuble	schwarzekassen			
schwarzgeld-schäuble	schwarzgeld	4.60	3.69	-0.91
schweiz-schäuble	schweiz		5.40	
security-schäuble	security		4.52	
seenotrettungs-seehofer	seenotrettung	4.09	5.21	1.12
segel-söder	segel	5.28	5.95	0.66
seitenaus-bobic	seitenaus		4.57	
selfie-merkel	selfie	4.99	5.12	0.13
siegelring-fischer	siegelring		4.79	

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
smartie-spahn	smartie	4.91	6.96	2.05
sommertraum-beckenbauer	sommertraum	5.69	7.24	1.55
spaß-guido	spaß	4.88		
spd-giffey	spd	4.77	3.82	-0.95
spd-scholz	spd	4.72	3.82	-0.89
spendengeld-laschet	spendengeld	5.53	5.76	0.23
spenden-hoeneß	spende	4.76	6.77	2.01
spenden-schäuble	spende	4.64	6.77	2.13
spenden-spahn	spende	4.77	6.77	2.00
spielzeugauto-haderthauer	spielzeugauto	5.12	5.48	0.36
spritzen-spahn	spritze	4.93	3.95	-0.99
staatsglauben-schwesig	staatsglauben	3.93		
steak-drosten	steak	5.14	5.21	0.07
steuer-hoeneß	steuer	4.72	3.37	-1.35
südafrika-bystron	südafrika	3.86	4.87	1.01
talkshow-habeck	talkshow	5.09	5.35	0.25
talkshow-lauterbach	talkshow	5.11	5.35	0.24
talkshow-lindner	talkshow	5.57	5.35	-0.22
tampon-scholz	tampon	5.14	3.57	-1.57
tätervolk-hohmann	tätervolk	4.06	1.51	-2.55
telefonterror-kahrs	telefonterror	4.55	2.12	-2.43
telefon-wulff	telefon	5.06	5.22	0.16
telegram-steinmeier	telegram		4.50	
teppich-niebel	teppich	4.92	4.88	-0.04
terror-bush	terror	4.45	1.53	-2.92
terrorparanoia-schäuble	terrorparanoia			
terror-thomas	terror	4.52	1.53	-2.99
thermo-lindner	thermo	4.86	5.32	0.46

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
toblerone-meuthen	toblerone	4.30	4.40	0.10
toiletten-brandner	toilette	4.78	5.25	0.46
tönnies-laschet	tönnies	4.86	5.10	0.24
tore-klose	tor	5.89	5.02	-0.87
tore-messi	tor	5.12	5.02	-0.10
tore-poldi	tor	5.85	5.02	-0.83
toscana-lafontaine	toscana		6.87	
toscana-trittin	toscana		6.87	
transfer-bobic	transfer	4.62	4.96	0.34
tribünen-hoeneß	tribüne	4.90	5.16	0.26
turban-trittin	turban		4.97	
türken-laschet	türke	4.71	4.10	-0.61
türken-roth	türke	4.87	4.10	-0.77
turnschuh-fischer	turnschuh	4.95	6.14	1.19
tv-lauterbach	tv	4.83	5.20	0.37
überlebens-sammer	überleben		4.26	
überschuldungs-schäuble	überschuldung		1.44	
überwachungsterrorist-schäuble	überwachungsterrorist			
überwachungswahn-schäuble	überwachungswahn		1.53	
ultimatum-seehofer	ultimatum	4.69	3.24	-1.45
umwelt-söder	umwelt	4.66	5.73	1.07
vanlaack-laschet	vanlaack	4.80		
veggie-künast	veggie	4.38	6.15	1.77
verfassungs-schäuble	verfassung		4.62	
verschärfungs-maas	verschärfung		3.02	
verschwörungs-mourinho	verschwörung		2.31	
vielflieger-hofreiter	vielflieger	5.28	5.95	0.67
vielflieger-lauterbach	vileflieger	4.94		

Table 11 continued from previous page

compound	modifier	c_val	m_val	diff
vielflieger-özdemir	vielflieger		5.95	
villen-spahn	villa	4.83	6.36	1.52
viren-söder	virus	4.88	2.19	-2.69
viren-spahn	virus		2.19	
vogelschiss-gauland	vogelschiss	4.61		
vollgas-vettel	vollgas	5.07	5.57	0.50
vorsaison-boateng	vorsaison		4.50	
vote-obama	vote	5.17	5.04	-0.14
wachstums-merkel	wachstum		5.16	
waffen-schreiber	waffe	4.65	3.03	-1.62
wahlkampf-obama	wahlkampf	4.91	3.61	-1.30
waterboarding-bush	waterboarding		2.64	
wehrmacht-gauland	wehrmacht	4.25	2.46	-1.79
wende-höcke	wende	4.98	4.65	-0.33
wende-westerwelle	wende		4.65	
west-rühe	west	4.52	5.15	0.63
wiedergeburt-bush	wiedergeburt		5.96	
willkommens-merkel	willkommen	4.42	7.90	3.49
wm-merkel	wm	5.04	4.75	-0.30
wm-müller	wm	5.07	4.75	-0.32
zeichentrick-lagerfeld	zeichentrick		6.03	
zigarren-clinton	zigarre	4.61	4.79	0.18
zoten-brüderle	zote	4.99	3.22	-1.77

B Summary (German)

Deutsche Personennamenkomposita wie *Villen-Spahn*, *Gold-Rosi* und *Folter-Bush* sind ein eher seltenes Phänomen der deutschen Sprache. Sie weisen die Struktur von Determinativkomposita auf und dienen als Spitznamen für bekannte Persönlichkeiten. Personennamenkomposita haben eine evaluative Funktion, d. h. die Person, auf die sich das Kompositum bezieht, wird positiv oder negativ gewertet (Belosevic, 2022). Diese Arbeit untersucht die evaluative Funktion von 413 deutschen Personennamenkomposita, die hauptsächlich Substantive als Modifikator sowie Nachnamen als Kopf aufweisen. Die 131 dazugehörigen Vor- und Nachnamen werden ebenfalls betrachtet. *Jens Spahn* wäre hierbei der entsprechende Name zu *Villen-Spahn*. Die Kontextsätze der Komposita und Namen werden aus *Twitter* und *Deutscher Wortschatz* herangezogen. Um die evaluative Funktion der Komposita im Vergleich zu den dazugehörigen Namen zu untersuchen, wird der Valenzwert der Kontexte, basierend auf einer Valenzdatenbank von Köper und Schulte im Walde (2016), berechnet. Auch die Relation zwischen Kompositum und Modifikator sowie die Funktion der Modifikatoren werden betrachtet. Außerdem wird mithilfe der Valenzwerte untersucht, ob es wahrnehmbare Unterschiede der Komposita- und Namensvalenz zwischen verschiedenen Gruppen (Politik, Sport, Showbusiness, Andere) gibt. Abschließend wird eine lineare Regression durchgeführt, die einen "Delta-Wert" vorhersagt. Dieser repräsentiert die Differenz von Kompositumvalenz und Namensvalenz. Hierfür werden verschiedene unabhängige Variablen wie Namensvalenz, Kompositumvalenz, Modifikatorvalenz, Alter, Geschlecht, Partei und Nationalität herangezogen.

Die Ergebnisse zeigen, dass Personennamenkomposita im Vergleich zu den dazugehörigen Namen sowohl positiv als auch negativ evaluativ sind und darüber hinaus den Grund ihrer Entstehung verdeutlichen. Komposita und ihre jeweiligen Modifikatoren weisen nur teilweise eine Korrelation auf, da einige Modifikatoren mit Ironie behaftet sind oder lediglich eine symbolische Bedeutung besitzen. Außerdem sind deutliche Unterschiede zwischen den Gruppen zu beobachten. Komposita und Namen von Politikern werden am negativsten wahrgenommen. Die Ergebnisse der

linearen Regression zeigen, dass die Kompositumvalenz eine hochsignifikante unabhängige Variable ist. Andere Variablen wie Modifikatorvalenz, Alter oder Partei können den "Delta-Wert" ebenfalls sehr gut vorhersagen.