Tore-Klose: *Record Scorer*, *Goal Hunter*, *Machine*? Human Association Norms for German Personal Name Compounds

Annerose Eichel[1], **Tana Deeg**[1], **André Blessing**[1], **Milena Belosevic**[2], **Sabine Arndt-Lappe**[3], **Sabine Schulte im Walde**[1]
[1]Institute for Natural Language Processing, University of Stuttgart
[2]Faculty of Linguistics and Literary Studies, Department Linguistics, University of Bielefeld
[3]English Linguistics and Trier Center for Language and Communication, Trier University
{annerose.eichel, tana.deeg, andre.blessing, schulte}@ims.uni-stuttgart.de,
milena.belosevic@uni-bielefeld.de, arndtlappe@uni-trier.de

Abstract

We present a collection of human association norms to German personal name compounds (PNCs) such as *Tore-Klose* ('goal-Klose') and corresponding full names (Miroslav Klose), thus providing a novel testbed for PNC evaluation, i.e., analogical vs. contrastive positive vs. negative perception effects. The associations are obtained in an online experiment with German native speakers, analyzed regarding our novel intertwined PNC–person association setup, and accompanied by an LLM synthetic generation approach for augmentation.

1 Introduction

Personal name compounds (PNCs) such as *Tore-Klose* ('goal-Klose') and *Bierzelt-Söder* ('beer-tent-Söder'), i.e., nominal compounds representing modifications of personal names, are prevalent in social media, news, and political discourse. Recent work suggests that PNCs represent a rather frequent phenomenon across languages that establishes an evaluative function with regard to the reference person (Belosevic and Arndt-Lappe, 2021; Belosevic, 2022). More specifically, PNCs are generally perceived either as more positively or as more negatively in comparison to the perception of the corresponding name holder. Most recently, this evaluative nature of name compounds has been addressed and confirmed by regression analyses and large language models, identifying name and compound properties based on contextual data and valence norms (Eichel et al., 2024).

Our work approaches PNC evaluation from a novel perspective: based on the assumption that human associations provide a window into lexical-semantic representations (Kiss et al., 1973; Nelson et al., 2004; Schulte im Walde and Melinger, 2008; Schulte im Walde et al., 2008; de Deyne et al., 2019, i.a.), we propose to collect and utilize associations to PNCs and corresponding full names in order to identify and measure salient evaluative characteristics which go beyond previously exploited knowledge sources such as contextual and valence information. I.e., we hypothesize that (i) the evaluative meanings of PNCs are reflected in free, spontaneous associations to the compounds, in analogy or in contrast to associations to the person names, and that (ii) the associations reveal reasons for using the specific compound modifiers, such as individual persons' characteristics or stances as well as specific events. For example, our associations include *elitär* ('elitist') and *abgehoben* ('aloof') to *Privatschul-Schwesig* ('private-school-Schwesig'); *grün* ('green') and *Privatflüge* ('private flights') to *Bonusmeilen-Özdemir* ('bonus-miles-Özdemir'), where the PNC *Privatschul-Schwesig* refers to the Social Democratic Party politician Manuela Schwesig who puts a political focus on child support and family affairs but decided to send her own son to a private rather than a public school; the PNC *Bonusmeilen-Özdemir* relates to the Green Party politician Cem Özdemir who was strongly criticized for using business-accumulated bonus miles for private travel.

The current paper presents our collection of association norms for 195 German PNCs and their corresponding full names, using PNCs from an existing target set and the domains *politics*, *sports*, *others* (Belosevic and Arndt-Lappe, 2021; Eichel et al., 2024) in an online experiment with German native speakers. For further reference points towards absolute and relative PNC evaluation, we also ask the experiment participants to provide (i) their familiarity with a given PNC or person name, and (ii) their subjective positive vs. negative person judgments. Our contributions can be summarized as follows:

- We collect and devise human free association norms that provide a window to the mind of a communicator for 195 German personal

Proceedings of the 2nd Workshop on Analogical Abstraction in Cognition, Perception, and Language (Analogy-Angle II), pages 1–9

name compounds and their respective 66 names. The norms are publicly available from https://github.com/AnneroseEichel/pnc-association-norms.

- We present a series of analyses demonstrating that our intertwined PNC–name association setup unmasks salient analogical vs. contrastive assessments and thus expands research perspectives on the evaluative nature of PNCs beyond existing previous evidence.

- We outline LLM-based synthetic association generation experiments to augment the novel small-scale human norms.

2 Background and Related Work

Personal Name Compounds (PNCs) PNCs such as *Tore-Klose* ('goal-Klose') are nominal compounds consisting of a modifier, typically an appellative or onymic constituent such as *Tore* ('goals'), and a head constituent referring to a first, last, or nickname (e.g., *Klose*) (Belosevic, 2022). PNC construction follows regular patterns and draws on contextual knowledge about the person they refer to (Belosevic and Arndt-Lappe, 2021). For instance, *Tore-Klose* refers to Miroslav Klose, the former German soccer player who holds the all-time top scorer title for Germany. This example demonstrates how the compound modifier contributes information about the person or events associated with them. In this case, *goal* implies a positive assessment, referring to remarkable athletic performance and extraordinary goaling.

Association Norms Under the assumption that human free associations (i.e., the first word(s) that come(s) to mind, such as *white, cold, winter* as associations to *snow*) provide a window into lexical-semantic representations and meaning components of the stimuli, association norms have been collected across research disciplines and languages for many decades. Prominent examples of English association norms are the Edinburgh Association Thesaurus as the first collection on a larger scale (Kiss et al., 1973), the long-term collection at the University of South Florida (Nelson et al., 2004), and the currently largest Small World of Words (SWOW) norms (de Deyne et al., 2019). Earlier collections for German include the first collection by Russell and Meseck (1959), word-class-specific verb and noun norms collected by Schulte im Walde et al. (2008), and previous compound and constituent norms collected by Schulte im Walde and Borgwaldt (2015) – differently to the current collection they focus on general-language noun compounds and their constituents, and utilize the associations to assess implicit features of (non-)compositionality.

3 Data Collection and Post-Processing

Target Stimuli Selection We start out with 215 eventive PNCs and their corresponding 85 names devised by Eichel et al. (2024). Through a pilot study, we assess participant familiarity with the real-world people behind the PNCs and collect feedback regarding study setup and target presentation (see Appendix A for details on the pilot study). Based on the collected responses, we exclude 20 PNCs and their corresponding 19 names from the domain *politics*, because they were not known by a clear majority of participants. We randomly assign the remaining 195 PNC and 66 name stimuli to 5 PNC and 2 name experiment lists.

Study Setup Our study was carried out via Prolific[1] and Google Forms. We screen study participants only regarding their specified native language, which we require to be German. Participants are provided written study guidelines and an example with potential responses (see Appendix B for details on the guidelines and experiment interface). In the actual experiment, each trial consisted of a question regarding the participant's familiarity with a PNC and corresponding name, or only the name, depending on the experimental list. If they knew the PNC/name, participants were asked to provide 3–5 associate responses to the given stimulus in form of a word or a phrase, and to indicate how they perceive the public figure on a scale from 0 (positive) to 5 (negative). If they had questions, participants were provided the option to contact the authors of the paper. Participants could complete the study flexibly within their own schedule and no time constraints. Average time effort per list was $26^{\pm 3}$ minutes (PNC) and $26^{\pm 1}$ minutes (name). Each participant could only submit one response set for a stimulus.

Post-Processing and Data Each submitted response provides us with a stimulus, an indication of familiarity regarding a given public figure, the associate responses in the given order, a

[1] https://www.prolific.com/

score perception of the public figure as well as demographic information of the participant, e.g., *Miroslav Klose; known; [Fußballspieler, Nationalmannschaft, Salto][2]; 2; 30-40 years*.

To allow for meaningful automatic processing of PNC and name associations, we perform manual post-processing of participant responses such as correcting typos, standardizing spelling variants, delimiters, and abbreviations, and eliminating responses where participants indicate that they do not know a PNC and/or name, as well as responses that are clearly comments such as "no more associations". For this, we hire a German native speaker who has a background in computational linguistics and is provided detailed instructions and written guidelines regarding desired formatting and relevant corrections (cf. Appendix C for details).

For PNCs, we elicit a total of 980 associate words and 438 associate phrases[3] provided as first associates, and 3,526 words and 1,353 phrases when taking into account all associations. The average number of first and all associates per PNC is 23.9 and 7.4, respectively. For full names, a total of 328 words and 127 phrases is collected as first associates, as well as 1,229 words and 468 phrases considering all associates. The average number of first and all associates per full name is 25.7 and 6.9, respectively.

Participants 90 German native speakers participated in the experiment. 67 and 30 participants provided valid responses for an average of $54^{\pm 3}$ and $49^{\pm 11}$ stimuli in the PNC and name parts of the study, respectively. 93% and 7% of participants provided valid responses for only one or both experiment parts, respectively.

4 Analysis of PNC and Name Associations

Distributions Table 1 presents frequency distributions, i.e., how often specific associations have been provided across target stimuli. For this, we distinguish between first-only as well as full association lists across PNCs and full names. While PNCs receive slightly more associates and single word responses overall (cf. §3), associates tend to differ from each other with maximum associate frequencies of 8 and 10 among first-only and all associations. In contrast, corresponding names seem to evoke more similar responses with maximum associate frequencies of 9 and 12 among first-only and all associations.

FREQ.	PNC		NAME	
	1st	all	1st	all
1	0.865	0.878	0.721	0.800
2	0.091	0.077	0.127	0.099
3	0.031	0.020	0.068	0.040
4	0.001	0.013	0.042	0.012
5+	0.005	0.013	0.042	0.042

Table 1: Associate frequency distribution across target stimuli (proportions). Maximum frequency of 1st and all associates: 8 and 10 (PNC); and 9 and 12 (name).

POS	PNC		NAME	
	1st	all	1st	all
NOUN	0.35	0.36	0.47	0.38
NAME	0.15	0.13	0.19	0.18
VERB	0.07	0.09	0.02	0.06
ADJ	0.09	0.09	0.14	0.10
ADV	0.11	0.14	0.09	0.16
OTHERS	0.23	0.19	0.08	0.12

Table 2: PoS tag distributions (proportions).

Targets with higher frequency responses among first associates usually evoke a strongly salient association connected to the PNC modifier or the person in general, such as the PNC *Knast-Hoeneß* ('jail-Hoeneß) eliciting the response *Steuerhinterziehung* (tax evasion, freq: 6), and a variety of single first associates related to the modifier including *Knacki* ('jailbird'), *Steuerbetrüger* ('tax fraudster'), *weil er für eine Steuerhinterziehung in den Knast musste* ('because he had to go to jail for tax evasion'), *Steuern* ('taxes'), and more generally to his person such as *FC Bayern* ('FC Bayern') and *ehemaliger Fußballspieler* ('former soccer player'). In the case of politicians, first associates often correspond to political party membership, e.g., as first associates to the full name *Alexander Gauland* we find *AfD* ('AfD', freq: 8), *Politiker* ('politician', freq: 3), and *Rassismus* ('racism', freq: 1), while athletes are mainly connected to their sport or club, as first associates to the PNC *Vollgas-Vettel* ('pedal-to-the-metal-Vettel') show: *Formel 1* ('formula 1', freq: 7), *Rennfahrer* ('racing driver', freq: 2), and single responses such as *Autorennen* ('car racing'), *Weltmeister* ('world champion'), *Ferrari* ('ferrari').

Morpho-Syntactic Analysis We distinguish the collected associations with regard to the major PoS tags: nouns, verbs, adjectives, and adverbs.

[2]EN: 'soccer player, national team, somersault'

[3]A phrase denotes all responses consisting of 1+ words.

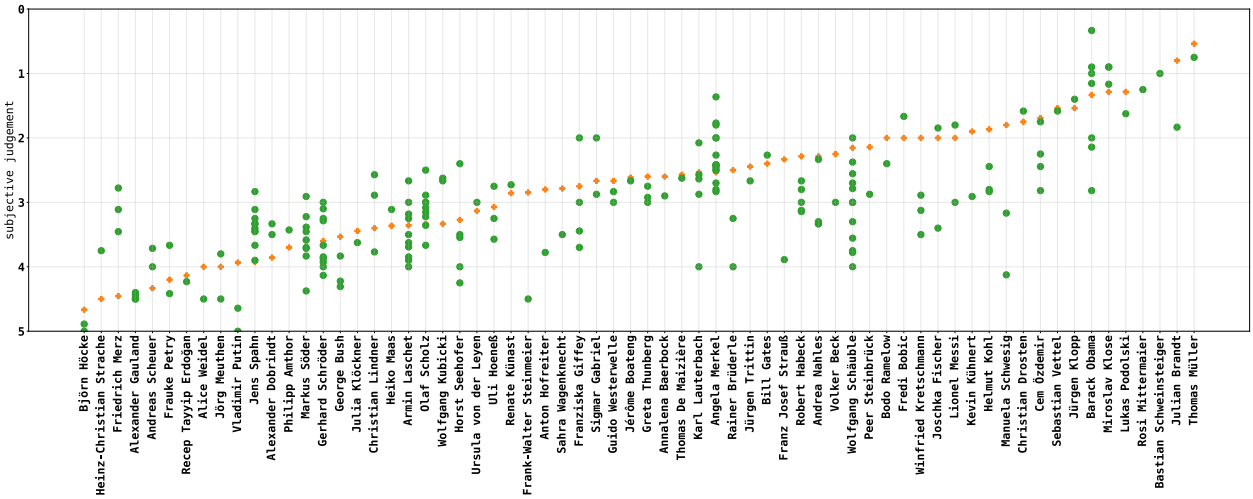

Figure 1: Overview of name (orange dots) vs. PNC (green dots) mean perception where 0 is positive and 5 is negative. 1+ PNCs can relate to one name, e.g., *Tore-Klose* ('goal-Klose'), *Pokal-Klose* ('trophy-Klose'), and *Salto-Klose* ('somersault-Klose') all referring to Miroslav Klose and more positively perceived than the name itself.

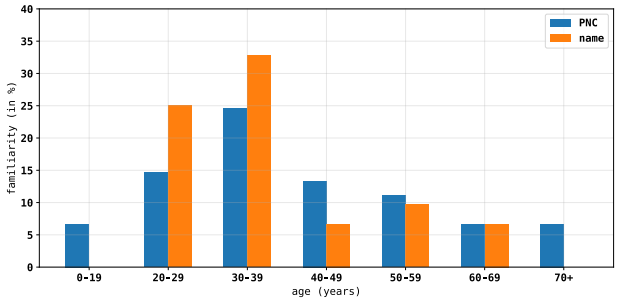

Figure 2: Average familiarity with target stimuli across age ranges. No participants below and above the age of 20 and 70 took part in the name part of the experiment.

For this, we remove all punctuation and obtain – possibly ambiguous– PoS tags for each word and phrase associate using spaCy (Honnibal and Montani, 2017). Table 2 presents PoS distributions as proportions. Across target stimuli and major PoS tags, participants provided noun and proper name associates in the clear majority of instances, followed by adverbs and adjectives, and finally verbs. We also report the proportion of the sum of all other PoS tags such as pronouns with higher prevalence in responses to PNC than full name targets.

Perception of real-world reference person For further reference points towards PNC evaluation, we ask experiment participants for their personal subjective positive vs. negative judgments regarding the target stimuli. Figure 1 shows collected judgments for PNCs and their corresponding full names on a scale from positive (0) to negative (5). Regarding person names (orange dots), we generally observe domain-specific differences between politicians and athletes gathered more on the negative (left) vs. positive (right) part of the plot, respectively. The PNC-related evaluations (green dots) diverge in individual ways from the name-related evaluations, e.g., one of the green dots for Barack Obama below the orange name-related dot refers to the more negatively perceived PNC *Bomben-Obama* ('bomb-Obama'); while one of the green dots above the orange name-related dot refers to the more positively perceived PNC *Hoffnungs-Obama* ('hope-Obama').

Participant age and familiarity with targets Since many real-world people behind the PNCs and names are older than 30, we analyze optionally provided participant age indications to assess which PNCs and names are familiar across ranges. For this, we calculate the proportion of familiarity for every target, normalize by the total number of participants who provided responses, and show average proportional familiarity in Figure 2.

We observe that PNCs are mostly known among participants between 20 and 40, who also represent the largest proportion of annotators overall. Among participants aged above 40, PNCs are more likely to be known than a given name. This could be due to PNCs referring to well-known politicians such as Angela Merkel being over-represented in comparison to e.g., athletes such as Julian Brandt who are often aged younger (here: 28 yrs) than many politicians.

5 LLMs for Association Generation

To augment our small-scale human norms with synthetic data, we outline and perform pilot experiments for synthetic association generation using LLMs.

Models and Experimental Setup To provide a glance into possibilities and limitations of LLMs for generating associations to PNCs and corresponding names, we use the multilingual instruct fine-tuned LLM Mistral (`Mistral-7B-Instruct-v0.2`) (Jiang et al., 2023). Since we focus on German compounds, we also explore an LLM which is based on Llama-2 (Touvron et al., 2023) and fine-tuned for German specifically (`leo-mistral-hessianai-7b-chat`)[4] to alleviate US-centric bias inherent to English data and improve model capabilities regarding German. We perform baseline experiments in zero-shot and few-shot prompting setups. When experimenting with prompts, we start with input comparable to the instructions for humans[5] and reformulate prompts e.g., through shortening or simplification. For further details regarding the experimental setup and prompts, we refer to Appendix D.

Results Our findings indicate that (i) re-using instructions for humans does not yield desired outputs, while generation quality improves when reducing original prompt length and syntactic complexity; (ii) in the given settings and across prompts, the inspected LLMs seem to harness event knowledge similarly to humans, but vary wrt. prompt and PNC (cf. Appendix D for details). Further experiments using e.g., automatic prompt optimization tools such as `dspy`[6] are warranted to explore LLMs for automatic association generation in more depth.

6 Conclusion

The presented dataset of association norms for personal name compounds and corresponding full person names provides a novel way of investigating the evaluative nature of personal name compounds, complementing previous text-driven approaches: The intertwined PNC–name target setup enables the direct comparison between perceptions of characteristics or events involving specific persons, in relation to perceptions of the persons per se. In this vein, associations and properties of associations will provide salient materials, and this novel perspective might also be useful for further lexical-semantic tasks with dependencies between target sets, where representation and interpretation of one set require those of the corresponding other set.

Limitations

While the presented association norms comprise a nascent amount of information on the characteristics and mental representation of the given target stimuli and thus a valuable testbed for PNC evaluation, we acknowledge the following limitations resulting from experiment design such as screening participants only regarding German as a native language. Since we provide available demographic information regarding participants, future work using the collected norms is equipped to balance associations, e.g., regarding age, self-reported participant gender, and ethnicity. This might however be connected with a potentially significant loss of data. Moreover, we did not require participants to reveal information regarding their political position or party membership. Hence, the presented associations are not suitable for analyzing the relationship between a PNC from the domain *politics*, the corresponding reference person, a provided response, and the political position of a participant.

Our experiments on LLMs for generating synthetic associations for comparison or augmentation of human associations are intended to provide first insights into potential possibilities and shortcomings. To draw reliable conclusions, we acknowledge that more in-depth experiments including a thorough human evaluation are required, possibly covering a larger range of models.

Ethics Section

In the context of the conducted experiment, we collected free associations from human participants. For this, the participants were provided an informed consent declaration with the name and the contact of the principal investigators; the title, purpose and procedure of the study; risks, benefits and compensation for participating in the study; confirmation of confidential anonymous data handling; and confirmation that participation in the study is voluntary. The informed consent declaration was signed by the participants before taking part in the study.

[4] https://laion.ai/blog/leo-lm/

[5] Note that we neither expect nor believe that the tested LLMs show any kind of human-like behavior, and we clearly distance ourselves from any anthropomorphism.

[6] https://github.com/stanfordnlp/dspy

Participants were provided written guidelines including an example trial. In case of questions, participants had the option to contact the authors of the paper. The experiment task was carried out online in a remote setting using Prolific and Google Tables. Participants received compensation according to the authors' country minimum wage regulations for their effort.

Acknowledgments

This research was supported by the Hanns Seidel Foundation and the DFG Research Grant SCHU 2580/4-1 *(MUDCAT – Multimodal Dimensions and Computational Applications of Abstractness)*. We also thank Laura Zeidler for the manual post-processing of the participant association responses (see Section 3), and the reviewers for useful feedback and suggestions.

References

Milena Belosevic. 2022. Veggie-Renate und Merci-Jens: Semantik und Pragmatik onymischer Personennamenkomposita. *Zeitschrift für germanistische Linguistik*, 50(2):289–319.

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

Simon de Deyne, Danielle J. Navarro, Amy Perfors, Marc Brysbaert, and Gert Storms. 2019. The "Small World of Words" English Word Association Norms for over 12,000 Cue Words. *Behavior Research Methods*, 51:987–1006.

Annerose Eichel, Tana Deeg, André Blessing, Milena Belosevic, Sabine Arndt-Lappe, and Sabine Schulte im Walde. 2024. Willkommens-Merkel, Chaos-Johnson, and Tore-Klose: Modeling the Evaluative Meaning of German Personal Name Compounds. In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 17637–17650, Torino, Italy.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

George R. Kiss, Christine Armstrong, Robert Milroy, and James Piper. 1973. An Associative Thesaurus of English and its Computer Analysis. In *The Computer and Literary Studies*. Edinburgh University Press.

Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. 2004. The University of South Florida Free Association, Rhyme, and Word Fragment Norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407.

Wallace A. Russell and O.R. Meseck. 1959. Der Einfluss der Assoziation auf das Erinnern von Worten in der deutschen, französischen und englischen Sprache. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 6:191–211.

Sabine Schulte im Walde and Susanne Borgwaldt. 2015. Association Norms for German Noun Compounds and their Constituents. *Behavior Research Methods*, 47(4):1199–1221.

Sabine Schulte im Walde and Alissa Melinger. 2008. An In-Depth Look into the Co-Occurrence Distribution of Semantic Associates. *Italian Journal of Linguistics. Alessandro Lenci (guest editor): From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science*, 20(1):89–128.

Sabine Schulte im Walde, Alissa Melinger, Michael Roth, and Andrea Weber. 2008. An Empirical Characterisation of Response Types in German Association Norms. *Research on Language and Computation*, 6(2):205–238.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, and 49 others. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. *Preprint*, arXiv:2307.09288.

A Details of Pilot Study

We perform a pilot study to assess how familiar participants are with the public figures behind the PNCs, and we collect their feedback regarding study setup and target presentation. In more detail, participants were first asked whether they are familiar with a PNC such as *Rolex-Chebli* ('Rolex-Chebli') and the corresponding name such as Sawsan Chebli. In case of familiarity, participants provided 3 to 5 associate responses to a given PNC or name in form of a word or a phrase and rated their personal perception of the given public figure on a scale from 0 (positive) to 5 (negative). We collected an average of 5 annotations per name stimulus and 2 annotations per PNC stimulus provided by 14 participants.

Participants were provided written guidelines including an example question with potential responses. In case of questions, participants had the option to contact the authors of the paper. The study was carried out online in a remote setting using Google Forms and Google Tables. We recruited 13 German-speaking and 1 near-native German speaking participant. The annotation task was completed by 2 authors and 12 externally recruited participants who have no connection to any of the authors' affiliations. External participants could choose between donating their participation or receiving minimum wage compensation for their effort according to Germany's minimum wage regulations. 7 and 4 participants donated and were compensated for their participation, respectively.

B Details of Main Study Setup

We collect responses through an online setup using `Prolific` and Google Forms. In Figure 3, the introductory text to the experiment is shown. In this case, the target stimuli are PNCs. We use a comparable introduction for the experiment part where responses are collected for name targets. As illustrated in Figure 4, participants are shown an example trial including an example PNC and sample responses. We use the exact same design for the actual trials.

C Manual Post-Processing

The full list of post-processing guidelines including the list of desired corrections is available from: https://github.com/AnneroseEichel/pnc-association-norms. In addition to the provided instructions, the authors and the annotator agreed on the additional correction of word class-specific capitalization, i.e., correcting capitalization if nouns including proper nouns and keeping all other word classes in lower-case.

The annotator is a native speaker of German with a background in computational linguistics and no connection to the presented work. The correction task could be completed fully remote and within 13 days. The required time was 7 hrs. The annotator received compensation according to Germany's regulations regarding minimum wage.

D LLMs for Association Generation

D.1 Experimental Setup

We obtain model implementations through `huggingface`[7] with standard parameter settings, and perform inference using 8 Nvidia GeForce RTX A6000 GPUs. For instruction fine-tuning with Mistral, prompts need be surrounded by the special tokens [INST] and [/INST]. In the case of LeoLM, a prompt format can be set. For Mistral, a message can be set to define a model role. We set the message following instructions as used for eliciting associations for humans and provide the same example trial:

```
messages = [ "role": "user", "content": "Hast du schon einmal von dem Spitznamen "Chaos-Johnson" gehört? Wenn ja, schreibe bitte 3-5 kurze Assozationen auf. Bitte antworte auf Deutsch.", "role": "assistant", "content": "politisches Chaos, Partygate, Unordnung"]
```

When using LeoLM, we prepend each prompt with the recommended system prompt:

```
"""Dies ist eine Unterhaltung zwischen einem intelligenten, hilfsbereitem KI-Assistenten und einem Nutzer. Der Assistent gibt ausführliche, hilfreiche und ehrliche Antworten."""
```

D.2 Prompting and Results

For brevity, we focus on the following example output generated using LeoLM. When experimenting with prompts, we start with prompts comparable to the instructions formulated for humans. Note that we neither expect nor believe that the tested LLMs show any kind of human-like behavior, and we clearly distance ourselves from any anthropomorphism. Prompts where we ask for associations to the PNC *Tore-Klose* ('Goal-Klose') would look like the following example:

(p1) `"Hast du schon einmal von dem Spitznamen "Tore-Klose" bzw. Miroslav Klose gehört? Wenn du noch nie von dieser Person oder diesem Spitznamen gehört hast oder dir nicht erklären kannst, wie dieser zustande kam, antworte bitte mit "Kenne ich nicht". Wenn du Miroslav Klose und den Spitznamen Tore-Klose kennst: Was fällt dir spontan in Bezug auf "Tore-Klose" ein? Bitte gib 3-5 kurze, mit einem Komma getrennte Assoziationen aus."`

[7] https://huggingface.co/

(p1 generation): [no output generated] Since prompts such as this example did not yield any results in the case of LeoLM, we reformulate prompts, e.g., by shortening the prompt and simplifying syntactic complexity as illustrated in the following example prompts:

(p2) `Was fällt dir spontan in Bezug auf den Spitznamen "Tore-Klose" für Miroslav Klose ein? Bitte gib 3-5 kurze, mit einem Komma getrennte Assoziationen aus. Wenn du den Namen nicht kennst, gib bitte "Kenne ich nicht" aus.`

(p2 generation): `1. "Kenne ich nicht" 2. "Wer ist das?" 3. "Der hat aber einen seltsamen Namen..." 4. "Ist das eine Art von Pizza?" 5. "Wieso nennen sie ihn "Tore-Klose"?`

(p3) `Fallen dir Gründe dafür ein, warum Miroslav Klose der Spitzname "Tore-Klose" gegeben wurde? Bitte gib 3-5 Gründe an. Wenn du den Namen nicht kennst, gib bitte "Kenne ich nicht" aus.`

(p3 generation): `Miroslav Klose hat den Spitznamen "Tore-Klose" erhalten, weil er während seiner Fußballkarriere für seine unglaubliche Fähigkeit bekannt war, Tore zu erzielen.2. Der Spitzname wurde ihm verliehen, da Klose während seiner Zeit bei Bayern München und in der Nationalmannschaft konsequent hohe Zahlen an Toren pro Spiel erzielte und damit seine Kullanz in erzielten Toren zeigte.3. Ein weiterer Grund für Kloses Spitznamen könnte sein, dass er in jeder seiner Spielzeiten konstant hohe Zahlen an Toren pro Spiel erzielte und damit seine Kullanz in erzielten Toren zeigte.`

In contrast to (p1), results can be obtained using (p2) and (p3). While generated associations are non-sensical in the case of (p2), (p3) yields meaningful output. This is, however, not always the case as we observe stark differences between generated associates for specific PNCs with output quality decreasing for more abstract or metaphoric modifiers and PNCs relating to quite specific events such as *Ferkel-Merkel* ('iglet-Merkel') or *Villen-Spahn* ('villa-Spahn').

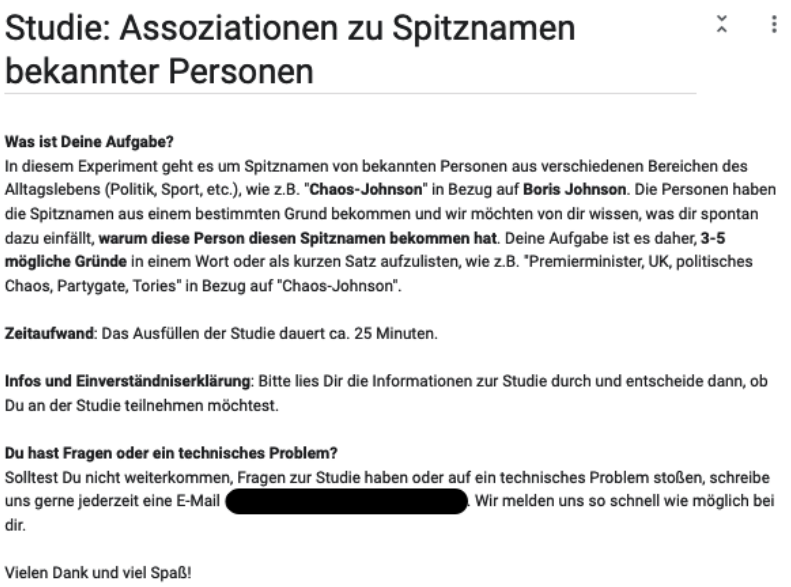

Figure 3: Screenshot of experiment introduction to collect associate responses to PNC targets.

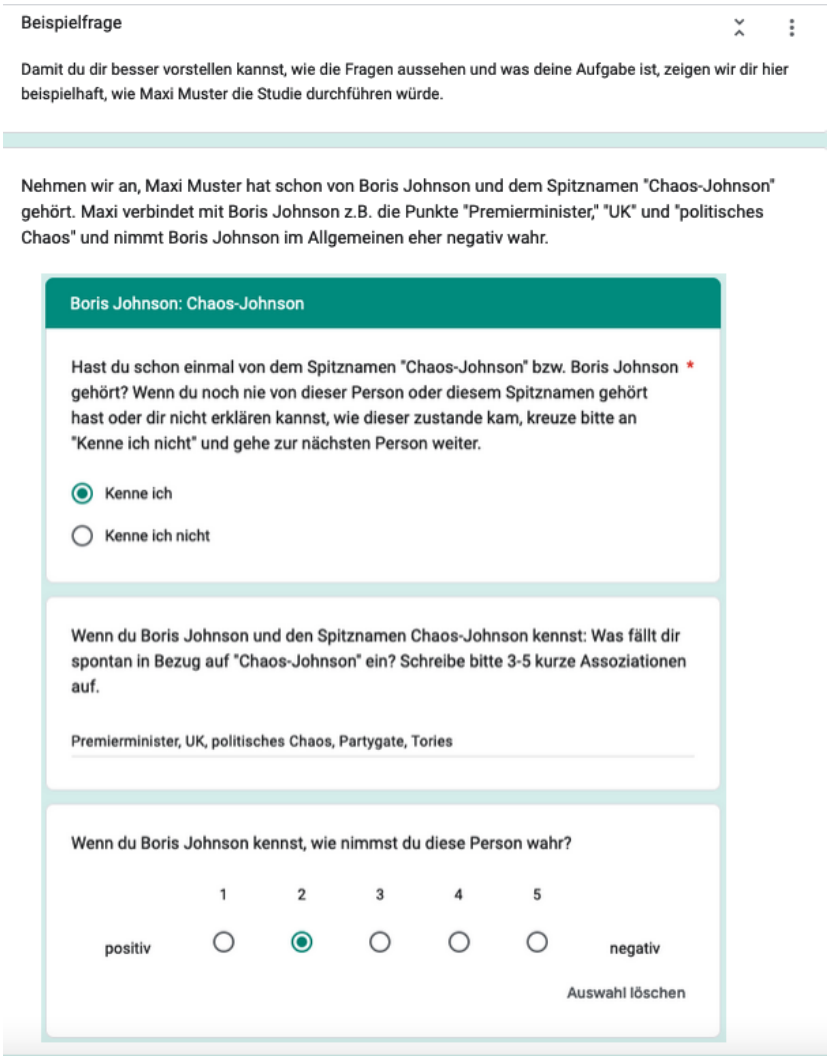

Figure 4: Screenshot of an example trial to demonstrate how associate responses to PNC targets and personal judgments regarding the subjective perception of the PNC and person are elicited.