# Adapting Entities across Languages and Cultures

**Denis Peskov**
Computer Science
University of Maryland
dpeskov@umd.edu

**Viktor Hangya**
Center for Information and Language Processing
LMU Munich
hangyav@cis.lmu.de

**Jordan Boyd-Graber**
CS, LSC, UMIACS, and iSchool
University of Maryland
jbg@umiacs.umd.edu

**Alexander Fraser**
Center for Information and Language Processing
LMU Munich
fraser@cis.lmu.de

## Abstract

How would you explain Bill Gates to a German? He is associated with founding a company in the United States, so perhaps the German founder Carl Benz could stand in for Gates in those contexts. This type of translation is called adaptation in the translation community (Vinay and Darbelnet, 1995). Until now, this task has not been done computationally. Automatic adaptation could be used in natural language processing for machine translation and indirectly for generating new question answering datasets and education. We propose two automatic methods and compare them to human results for this novel NLP task. First, a structured knowledge base adapts named entities using their shared properties. Second, vector arithmetic and orthogonal embedding mappings identify better candidates, but at the expense of interpretable features. We evaluate our methods through a new dataset[1] of human adaptations.

## 1 When Translation Misses the Mark

Imagine reading a translation from German, "I saw Merkel eating a Berliner from Dietsch on the ICE". This sentence is opaque without cultural context.

An extreme cultural *adaptation* for an American audience could render the sentence as "I saw Biden eating a Boston Cream from Dunkin' Donuts on the Acela", elucidating that Merkel is in a similar political post to Biden; that Dietsch (like Dunkin' Donuts) is a mid-range purveyor of baked goods; both Berliners and Boston Creams are filled, sweet pastries named after a city; and ICE and Acela are slightly ritzier high-speed trains. Human translators make this adaptation when it is appropriate to the translation (Gengshen, 2003).

[1] Available at https://go.umd.edu/adaptation

**Bill Gates**

| Top Adaptations: | | |
|---|---|---|
| **WikiData** | **3CosAdd** | **Human** |
| F. Zeppelin | M. Winterkorn | A. Bechtolsheim |
| Günther Jauch | Volkswagen_AG | Dietmar Hopp |
| N. Harnoncourt | DaimlerChrysler | Carl Benz |

Table 1: WikiData and unsupervised embeddings (**3CosAdd**) generate adaptations of an entity, such as Bill Gates. Human adaptations are gathered for evaluation. American and German entities are color coded.

Because adaptation is understudied, we leave the full translation task to future work. Instead, we focus on the task of cultural adaptation of entities: given an entity in a source, what is the corresponding entity in English? Most Americans would not recognize Christian Drosten, but the most efficient explanation to an American would be to say that he is the "German Anthony Fauci" (Loh, 2020). We provide top adaptations suggested by algorithms and humans for another American involved with the pandemic response, Bill Gates, in Table 1.

Can machines reliably find these analogs with minimal supervision? We generate these adaptations with structured knowledge bases (Section 3) and word embeddings (Section 4). We elicit human adaptations (Section 5) to evaluate whether our automatic adaptations are plausible (Section 5.3).

## 2 Wer ist Bill Gates?

We define cultural adaptation and motivate its application for tasks like creating culturally-centered training data for QA. Vinay and Darbelnet (1995) define adaptation as translation in which the relationship not the literal meaning between the receiver and the content needs to be recreated.

You could formulate our task as a traditional analogy Drosten::Germany as Fauci::United States (Turney, 2008; Gladkova et al., 2016), but despite this superficial resemblance (explored in Section 4), traditional approaches to analogy ignore the influence of culture and are typically *within* a language. Hence, analogies are tightly bound with culture; humans struggle with analogies outside their culture (Freedle, 2003).

We can use this task to identify named entities (Kasai et al., 2019; Arora et al., 2019; Jain et al., 2019) and for understanding other cultures (Katan and Taibi, 2004).

### 2.1 … and why Bill Gates?

This task requires a list of named entities adaptable to other cultures. Our entities come from two sources: a subset of the top 500 most visited German/English Wikipedia pages and the non-official characterization list (Veale, 2016, NOC), "a source of stereotypical knowledge regarding popular culture, famous people (real and fictional) and their trade-mark qualities, behaviours and settings". Wikipedia contains a plethora of singers and actors; we filter the top 500 pages to avoid a pop culture skew.[2] We additionally select all Germans and a subset of Americans from the Veale NOC list as it is human-curated, verified, and contains a broader historical period than popular Wikipedia pages. Like other semantic relationships (Boyd-Graber et al., 2006), this is not symmetric. Thus, we adapt entities in both directions; while Berlin is the German Washington, DC, there is less consensus on what is the American Berlin, as Berlin is both the capital, a tech hub, and a film hub. A full list of our entities is provided in Appendix D.

## 3 Adaptation from a Knowledge Base

We first adapt entities with a knowledge base. We use WikiData (Vrandečić and Krötzsch, 2014), a structured, human-annotated representation of Wikipedia entities that is actively developed. This resource is well-suited to the task as features are standardized both within and across languages.

Many knowledge bases explicitly encode the nationality of individuals, places, and creative works. Entities in the knowledge base are a discrete sparse vector, where most dimensions are unknown or not applicable (e.g., a building does not have a spouse).

[2]We discuss the applicability of using Wikipedia (i.e., what proportion of the English Wikipedia is visited from the United States) in Appendix B.

For example, Angela Merkel is a human (instance of), German (country of citizenship), politician (occupation), Rotarian (member of), Lutheran (religion), 1.65 meters tall (height), and has a PhD (academic degree). How would we find the "most similar" American adaptation to Angela Merkel? Intuitively, we should find someone whose nationality is American.

Some issues immediately present themselves; contemporary entities will have more non-zero entries than older entities. Some characteristics are more important than others: matching unique attributes like "worked as journalist" is more important than matching "is human".

Each entity in WikiData has "properties", which we can think about as the dimension of a sparse vector and "values" that those properties can take on. For example, Merkel has the properties "occupation" and "academic degree". *Values* for those properties are that her "occupation" is "politician" and her "academic degree" is a "doctorate". To match entities across cultures, we focus on matching properties rather than values; many of the values are more relevant inside a culture. For example, we cannot find American politicians who belong to the Christian Democratic Union, but we can find politicians who have an academic degree and a dissertation title.

As a toy example, if Beethoven, Merkel, and Bach all have only two *properties*: Beethoven has an "occupation" and "genre", Merkel has an "Erdős number" and "political party", and Bach has a "occupation" and "genre", then Beethoven and Bach has a distance of zero and are the closest entities while Merkel has a distance of two since {"Erdős number", "political party"} is two away from {"occupation", "genre"}.

First, we bifurcate WikiData into two sets: an American set $\mathcal{A}$ for items which contain the *value* "United States of America" and a German set $\mathcal{D}$ for those with German values.[3] This is a liberal approximation, but it successfully excludes roughly seven out of the eight million items in WikiData. Then we explore the *properties* from WikiData. We create entity vectors with dimensions corresponding to frequently-occurring properties.

[3]While the geopolitical definition of American is straightforward, the German nation state is more nuanced (Schulze, 1991). Following Green (2003), we adopt members of the Zollverein or the German Confederation as "German" *as well as their predecessor and successor states*. This approach is a more inclusive (Großdeutschland) definition of "German" culture.

The *properties* are discrete and categorical; Merkel either has an "occupation" or she does not. Each entity then has a sparse vector. We calculate the similarity of the vectors with Faiss's $L_2$ distance (Johnson et al., 2021) and for each vector in $\mathcal{A}$ find the closest vector in $\mathcal{D}$ and *vice versa*.

So who is the American Angela Merkel? One possible answer is Woodrow Wilson, a member of a "political party", who had a "doctoral advisor" and a "religion", and ended up with "awards". This answer may be unsatisfying as it was Barack Obama who sat across from Merkel for nearly a decade. To capture these more nuanced similarities, we turn to large text corpora in Section 4.

## 4 An Alternate Embedding Approach

While the classic NLP vector example (Mikolov et al., 2013c) isn't as magical as initially claimed (Rogers et al., 2017), it provides useful intuition. We can use the intuitions of the cliché:

$$\overrightarrow{\text{King}} - \overrightarrow{\text{Man}} + \overrightarrow{\text{Woman}} = \overrightarrow{\text{Queen}} \tag{1}$$

to adapt between languages.

This, however, requires relevant embeddings. First, we use the entire Wikipedia in English and German, preprocessed using Moses (Koehn et al., 2007). We follow Mikolov et al. (2013b) and use named entity recognition (Honnibal et al., 2020) to tokenize entities such as Barack_Obama.

We use word2vec (Mikolov et al., 2013b), rather than FastText (Bojanowski et al., 2017), as we do not want orthography to influence the similarity of entities. Angela Merkel in English and in German have quite different neighbors, and we intend to keep it that way by preserving the distinction between languages.

However, the standard word2vec model assumes a single monolingual embedding space. We use unsupervised Vecmap (Artetxe et al., 2018), a leading tool for creating cross-lingual word embeddings, to build bilingual word embeddings. We propose two approaches for adaptation.

**3CosAdd** We follow the word analogy approach of 3CosAdd[4] (Levy and Goldberg, 2014; Köper et al., 2016). American→German adaptation takes the source entity's ($v$) embedding in the English vector space and looks for its adaptation ($u^*$) based on embeddings in the German space. This is like the word analogy task, i.e., what entity has the role in the German culture as $v$ does in American culture. As an example, Merkel has a similar role in the German culture as Biden. Formally, the adaptation of the English entity $v$ into German is

$$\vec{a} \equiv \text{avg}\left(\overrightarrow{E^{en}}_{\text{United\_States}}, \overrightarrow{E^{de}}_{\text{USA}}\right) \tag{2}$$

$$\vec{d} \equiv \text{avg}\left(\overrightarrow{E^{en}}_{\text{Germany}}, \overrightarrow{E^{de}}_{\text{Deutschland}}\right) \tag{3}$$

$$u^* = \underset{u \in V^{de}}{\arg\max}\, \text{sim}\left(\overrightarrow{E^{de}_u}, \overrightarrow{E^{en}_v} - \vec{a} + \vec{d}\right), \tag{4}$$

where $\overrightarrow{E^l_w}$ is the embedding of word $w$ in language $l$, $V^{de}$ is the German vocabulary and sim is the cosine similarity. The American anchor word $\vec{a}$ and German anchor $\vec{d}$ represent the American and German cultures.[5] We average the English and German embeddings of the individual word types for robust anchor vectors. In standard analogies, as in Equation 1, the $\vec{a}$ and $\vec{d}$ vectors are different for each test pair; here they are the same for each example, as we always are pivoting between the two cultures.

**Learned adaptation** To eliminate the need for manual anchor selection for both cultures, our second approach learns the adaptation as a linear transformation of source embeddings to the target culture given a few adaptation examples. Specifically, we use the human adaptations sourced for the Wikipedia entities as training for the Veale NOC ones. We follow the work of Mikolov et al. (2013a) and learn a transformation matrix $\mathbf{W}_{en\to de}$ for American→German by minimizing the $L_2$ distance of $\mathbf{W}_{en\to de}\overrightarrow{E^{en}_{v_i}}$ and $\overrightarrow{E^{de}_{u_i}}$ over gold adaptation $v_i, u_i{}_{i=1}^{n}$ entity pairs. The adaptation of a source entity $v$ is $u^* = \mathbf{W}_{en\to de}\overrightarrow{E^{en}_v}$. Likewise, we learn the reverse mapping $\mathbf{W}_{de\to en}$ for German→American adaptation. This requires supervised training data—but not much (Conneau et al., 2018)—which we collect in Section 5.

## 5 Comparing Automation to Human Judgment

The automated methods can generate entities at scale, but humans have to evaluate their relevance.

### 5.1 Adaptation by Locals

Since quality control is difficult for generation (Peskov et al., 2019), we need users who

[4] We experiment with 3CosMul as well but found 3CosAdd generally more robust.

[5] USA is used to refer to the United States in German. Der Spiegel, the largest newspaper, calls their US section USA.

will answer the task accurately. We recruit five American citizens educated at American universities and five German citizens educated at German ones. These human annotations serve as a gold standard against which we can compare our automated approaches. To improve the user experience, we create an interface that provides a brief summary of each source entity from Wikipedia and asks the users to select a target adaptation that autocompletes Wikipedia page titles (all entities; targets are not limited to the lists in Section 2) in a text box *a la* answer selection in Wallace et al. (2019). The annotation task requires two hours for our users to complete. Obviously, German annotators are more familiar with German culture than the Americans, and vice-versa. Annotators translate into their native language. Since we are focusing on popular entities, they are often known despite the cultural divide, but the introductory paragraph from Wikipedia reminds users if not.

### 5.2 Are the Adaptations Plausible?

To validate and compare all our adaptation strategies' precision, five German translators[6] who understand American culture assess the adaptations. The top five adaptations from WikiData, 3CosAdd, learned adaptation, and humans—as well as five randomly selected options from the human pool—are evaluated for plausibility on a five-level Likert scale.[7] Fleiss' Kappa (0.382) and Krippendorf's Alpha (0.381) assess interannotator Agreement; this "fair" agreement suggests that vetting an adaptation is challenging and sometimes subjective, even for translators.

### 5.3 Why Adaptation is Difficult

Embedding adaptations are better than Wikidata's, and human adaptations are better still (Figure 1). Thus, we use human adaptations as the gold standard for evaluating recall. Only the learned embedding method uses training data, so we use human adaptations from Wikipedia to train the projection matrix and evaluate (for all methods) using human adaptations the NOC list. Given that the task is subjective, we take our results with a grain of salt given cultural variation (e.g., some people view Angela Merkel's conservatism as a defining characteristic, while others focus on her science pedigree).

[6] Recruited through Upwork for $40 each.

[7] Our custom Qualtrics survey is provided in Appendix C. The order of adaptations is randomized and assessed on a Likert scale with anchors from Jurgens et al. (2014).

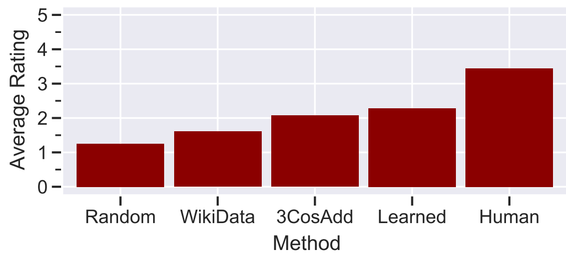


Figure 1: We validate adaptation strategies with expert translators on a five-point Likert scale. The human-generated adaptations are rated best—between "related" (3) and "similar" (4). These human adaptations become the reference for evaluation in Table 2.

| Data | Metric | WikiData | 3CosAdd | Learned |
|---|---|---|---|---|
| *American→ German* | | | | |
| Wikipedia | Rec@5 | 7.5% | **14.2%** | - |
| | Rec@100 | 34.4% | **52.8%** | - |
| | MRR | 0.05 | **0.10** | - |
| Veale NOC | Rec@5 | 3.0% | 22.9% | **28.6%** |
| | Rec@100 | 42.4% | **51.4%** | 45.7% |
| | MRR | 0.03 | 0.17 | **0.24** |
| *German→ American* | | | | |
| Wikipedia | Rec@5 | 3.1% | **17.2%** | - |
| | Rec@100 | 15.4% | **40.5%** | - |
| | MRR | 0.01 | **0.12** | - |
| Veale NOC | Rec@5 | 0.0% | **25.0%** | **25.0%** |
| | Rec@100 | 25.0% | **70.0%** | 55.0% |
| | MRR | 0.02 | 0.12 | **0.15** |

Table 2: If we consider human adaptations as correct, where do they land in the ranking of automatic adaptation candidates? In this recall-oriented approach, learned mappings (which use a small number of training pairs), rate highest.

We use the mean reciprocal rank (Voorhees, 1999, MRR) to measure how high the gold adaptations are ranked by our other adaptation strategies. Since MRR decreases geometrically and our gold standard is not exhaustive, the Recall@5, and @100 metrics are more intuitive. We calculate Recall@$n$ by measuring what fraction of the correct adaptations of a source entity is retrieved in the top $n$ predictions.[8] Table 2 validates that the human annotations are near the top of the automatic adaptations; the precision-oriented evaluation (Figure 1) validates whether the top of the list is reasonable. All human annotations and a sample of the automatic adaptations are provided in Appendix D.

[8] This is often referred to as P@$n$ in bilingual lexicon induction literature (Conneau et al., 2018).

### 5.4 Qualitative Analysis

There is no single answer to what makes a good adaptation. Let us return to the question of who Bill Gates is, which underlines how there is often no one right answer to this question but several context-specific possibilities. The human adaptations show the range of plausible adaptations, each appropriate for a particular facet of the position Bill Gates has in US society. As previously mentioned, Carl Benz represents a larger than life founder who created an entire industry with his company. However, Carl Benz made cars, not computers.

Even within technology, different adaptations highlight different aspects of Bill Gates. Like the implementer of the BASIC programming language, Konrad Zuse contributed to computers that were more than single-purpose machines. Just as as Bill Gates's Microsoft is seen as a stodgy tech giant, Dietmar Hopp founded SAS, a giant German tech company that is more often discussed in board rooms than in living rooms. And because the epicenter of modern tech is America's West Coast, Andreas von Bechtolsheim represents a German founder of Sun Microsystems and early Google investor that made his way to Silicon Valley.

Other times, there is more consensus: a majority of raters declare Angela Merkel is the German Hilary Clinton, and Joseph Smith is the American Martin Luther. There are even some unanimous adaptations: Bavaria is the German California. Adaptations of fictional characters seem particularly difficult, although this may represent the supremacy of American popular culture; Superman and Homer Simpson are so well known in Germany that there are no clear adaptations; Till Eulenspiegel, Maverick, Bibi Blocksberg are not superheroes from a dying world and Heidi is not a dumb, bald everyman.

## 6 A New Computational Task

We formally introduce entity **adaptation** as a new computational task. Word2vec embeddings and WikiData can be used to figuratively—not just literally—translate entities into a different culture. Humans are better at generating candidates for this task than our computational methods (Figure 1). These methods are well-motivated, but have room for improvement. Knowledge bases improve over time and increased coverage of entities—as well as improved information about each entity—would improve the method. Alternate word embedding approaches—perhaps those that discard orthography—may provide better candidates. Even humans occasionally disagree with other humans on this task, so evaluation for this task is nontrivial.

Our new dataset of machine-generated adaptations, human adaptations, and human evaluation of these adaptations can serve as an evaluation for future automatic methods.

People need NLP systems that reflect their language **and** culture, but datasets are lacking: adaptation can help. There has been an explosion of English-language QA datasets, but other languages continue to lag behind. Several approaches try to transfer English's bounty to other languages (Lewis et al., 2020; Artetxe et al., 2019), but most of the entities asked about in major QA datasets are American (Gor et al., 2021). Adapting entire questions will require not just adapting entities and non-entities in tandem but will also require integration with machine translation (Kim et al., 2019; Hangya and Fraser, 2019). Our automatic methods did not create precise adaptations, but the alternative "incorrect" adaptations may be useful for low-precision tasks, such as generating numerous simple open-ended questions or gauging the popularity of a entity.

Given the existence of robust datasets in high resource languages can we **adapt**, rather than literally translate, them to other cultures and languages?

## 7 Acknowledgments

Peskov was supported by a DAAD Research Fellowship and by the wonderful faculty and students of Ludwig Maximilians Universität München. Fraser is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 640550) and by German Research Foundation (DFG; grant FR 2829/4-1). Boyd-Graber is supported by NSF Grant IIS-1822494. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the view of the sponsors.

We thank Sander Schulhoff for his web development expertise and Pedro Rodriguez for his help with processing WikiData. Thanks to Phillip Dufter, Benno Krojer, Stephan Huber, Janika Linke, Andrew Guo, Adam Visokay, Christian Rice, Connor Knight, and others for help with adaptations.

## Ethics

We worked with human participants to collect our data. They are all adults who participated of their own volition and no payment was made. No personal data was collected or used for the dataset. For evaluation of the adaptations, we hired translators through Upwork. They were paid $40 for a task that took roughly between one and two hours.

The broad motivation of this work is to spread cultural understanding. Humans must be kept in-the-loop for making claims about cultural relevance. Having multiple diverse opinions is necessary for supporting any cultural claim. Like with language, nationality is often correlated with culture, but is not synonymous. Large countries contain multitudes, while some nationalities (e.g., Kurds) lack a *de jure* nation but span many nations. We elide this detail and focus on information often available in knowledge bases.

These lists contain figures that are controversial. From a research perspective, research datasets should reflect the real world and prior work, thus we include prominent entities as identified by Veale NOC and Wikipedia. Any list may contain biases in the collection processes, and this should not be thought of as an exclusive and definitive list, but as a start that can be refined and ultimately expanded to other cultures.

## References

Ravneet Arora, Chen-Tse Tsai, Ketevan Tsereteli, Prabhanjan Kambadur, and Yi Yang. 2019. A semi-Markov structured support vector machine model for high-precision named entity recognition. In *Proceedings of the Association for Computational Linguistics*.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the Association for Computational Linguistics*.

Mikel Artetxe, Sebastian Ruder, and Dani Yogatama. 2019. On the cross-lingual transferability of monolingual representations. *CoRR*, abs/1910.11856.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. In *Proceedings of the Association for Computational Linguistics*.

Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson, and Robert Schapire. 2006. Adding dense, weighted, connections to WordNet. In *Proc. Global WordNet Conference 2006*. Global WordNet Association.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word translation without parallel data. In *Proceedings of the International Conference on Learning Representations*.

Roy Freedle. 2003. Correcting the SAT's ethnic and social-class bias: A method for reestimating sat scores. *Harvard Educational Review*, 73(1):1–43.

Hu Gengshen. 2003. Translation as adaptation and selection. *Perspectives: Studies in Translatology*, 11(4):283–291.

Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuoka. 2016. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In *Proceedings of the NAACL Student Research Workshop*.

Maharshi Gor, Kellie Webster, and Jordan Boyd-Graber. 2021. Toward deconfounding the influence of subject's demographic characteristics in question answering. In *Proceedings of Empirical Methods in Natural Language Processing*.

Abigail Green. 2003. Representing Germany? the zollverein at the world exhibitions, 1851–1862. *The Journal of Modern History*, 75(4):836–863.

Viktor Hangya and Alexander Fraser. 2019. Unsupervised parallel sentence extraction with parallel segment detection helps machine translation. In *Proceedings of the Association for Computational Linguistics*.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Alankar Jain, Bhargavi Paranjape, and Zachary C. Lipton. 2019. Entity projection via machine translation for cross-lingual NER. In *Proceedings of the Association for Computational Linguistics*.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2021. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547.

David Jurgens, Mohammad Taher Pilehvar, and Roberto Navigli. 2014. SemEval-2014 Task 3: Cross-level semantic similarity. In *Proceedings of the Workshop on Semantic Evaluation*. Association for Computational Linguistics.

Jungo Kasai, Kun Qian, Sairam Gurajada, Yunyao Li, and Lucian Popa. 2019. Low-resource deep entity resolution with transfer and active learning. In *Proceedings of the Association for Computational Linguistics*.

David Katan and Mustapha Taibi. 2004. *Translating cultures: An introduction for translators, interpreters and mediators*. Routledge.

Yunsu Kim, Yingbo Gao, and Hermann Ney. 2019. Effective cross-lingual transfer of neural machine translation models without shared vocabularies. In *Proceedings of the Association for Computational Linguistics*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the Association for Computational Linguistics*.

Maximilian Köper, Sabine Schulte im Walde, Max Kisselew, and Sebastian Padó. 2016. Improving zero-shot-learning for german particle verbs by using training-space restrictions and local scaling. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*.

Omer Levy and Yoav Goldberg. 2014. Linguistic regularities in sparse and explicit word representations. In *Conference on Computational Natural Language Learning*.

Patrick Lewis, Barlas Oğuz, Ruty Rinott, Sebastian Riedel, and Holger Schwenk. 2020. MLQA: Evaluating cross-lingual extractive question answering. In *Proceedings of the Association for Computational Linguistics*.

Tim Loh. 2020. Germany has its own Dr. Fauci—and actually follows his advice. *Bloomberg*.

Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. 2013a. Exploiting Similarities among Languages for Machine Translation. *CoRR*, abs/1309.4.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Proceedings of Advances in Neural Information Processing Systems*.

Tomáš Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013c. Linguistic regularities in continuous space word representations. In *Conference of the North American Chapter of the Association for Computational Linguistics*.

Denis Peskov, Nancy Clarke, Jason Krone, Brigi Fodor, Yi Zhang, Adel Youssef, and Mona Diab. 2019. Multi-domain goal-oriented dialogues (MultiDoGO): Strategies toward curating and annotating large scale dialogue data. In *Proceedings of Empirical Methods in Natural Language Processing*.

Anna Rogers, Aleksandr Drozd, and Bofang Li. 2017. The (too many) problems of analogical reasoning with word vectors. In *Proceedings of the Joint Conference on Lexical and Computational Semantics*.

Hagen Schulze. 1991. *The Course of German Nationalism: From Frederick the Great to Bismarck 1763–1867*. Cambridge University Press.

Peter D Turney. 2008. A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of International Conference on Computational Linguistics*.

Tony Veale. 2016. Round up the usual suspects: Knowledge-based metaphor generation. In *Proceedings of the Fourth Workshop on Metaphor in NLP*.

Jean-Paul Vinay and Jean Darbelnet. 1995. *Comparative stylistics of French and English: A methodology for translation*, volume 11. John Benjamins Publishing.

Ellen M Voorhees. 1999. The TREC-8 question answering track report. In *Proceedings of the Text REtrieval Conference*, volume 99.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Eric Wallace, Pedro Rodriguez, Shi Feng, Ikuya Yamada, and Jordan Boyd-Graber. 2019. Trick me if you can: Human-in-the-loop generation of adversarial question answering examples. *Transactions of the Association of Computational Linguistics*, 10.

## A Appendix

Our appendix contains our entire human-collected dataset, as well as a sample of our WikiData and embedding approaches for adaptation.

Figure 2 shows our collection tool. Table 3 shows German→American Veale NOC items. Table 4 shows American→German Veale NOC items. Table 5 shows German→American Veale NOC items. Table 6 shows American→German Veale NOC items.

Table 7 shows our WikiData predictions and Table 8 shows our embedding-based predictions. We pose several background questions about Wikipedia and WikiData as well:

## B Wikipedia Analysis

**Are the Wikipedia pages in German and English visited from the associated country?** Yes; the Wikipedias for the respective languages are most used by visitors located in those countries: 63% of German wikipedia was visited from Germany and 32% of English Wikipedia was visited from the United States in the past year.[9]

**Are the top Wikipedia topics notably different across languages?** Yes; less than a quarter of top 500 searches for 2019 are identical across English and German.

**Does WikiData cover areas outside of the United States?** *Wikipedia* coverage does not mean that *WikiData* annotations are conducted equally across German and American entities. Analyzing WikiData[10] reveals a discrepancy in coverage of Germans and Americans.

Out of 8,126,559 titles, 1,030,762 include a reference to the United States in any capacity. However, only 184,692 contain a reference to (broader) Germany. This imbalance is significant but has enough German items for our methodology. As WikiData is a maintained resource, there is room for future additional coverage and standardization of fields.

Countries use different names throughout history. While the United States of America is straightforward, Germany includes several variations, such as: German Empire, the Kingdom of Bavaria, the Kingdom of Prussia, etc. The WikiData feature-based approach can be used for other countries as well (. . . or anything that is consistently coded). For example, there are 65,957 Russian, 152,701 French, and 48,026 Chinese items in WikiData.[11]

**Are the top Wikipedia topics necessarily belonging to the culture?** No; the top 10 most visited German Wikipedia includes a cultural potpurri: Germany, Greta Thurnberg, Asperger Syndrome, Game of Thrones, and Freddie Mercury. While there are *uniquely* German entities in the longer list—ZDF, Capital Bra, The Cratez, Niki Lauda—we **cannot** conclude that all top entities in a language belong culturally to a given country. Therefore, we need a stricter methodology.

**Where does one find entities?** We rely on a human-sourced dataset: Veale's Non-Official Characterization list (Veale, 2016). This list contains 1031 people, real and fictional, such as Daniel Day-Lewis, Anton Chekhov, and Bridget Jones. These people are annotated with properties, one of which is conveniently their address. There are 25 people with a German location and 575 with an American one. Removing fictional characters written by non-nationals causes the German leaves the list with 20 entities. An American author filters the list of Americans down to 35 iconic ones with achievements that span politics, music, activism, athletics, and pop culture.

Wikipedia provides another avenue for gauging popular topics in a language. We manually filter the top 500 German/English Wikipedia topics to remove non-German/non-American entities; Game of Thrones and Unix-Shell are popular in the German Wikipedia, but they are not culturally idiosyncratic. For the 2019 German Wikipedia we are left with roughly 200 items, which we further reduce down to 120 after putting a cap on pop culture entities. For the American counterpart, over 300 items are culturally American. We add a three-year filter to remove pop items to make it comparable to the German one.

[9] https://stats.wikimedia.org/

[10] we use a full 1.2 Terabyte dump as of 10.26.20

[11] the modern day name countries only

# C Interfaces

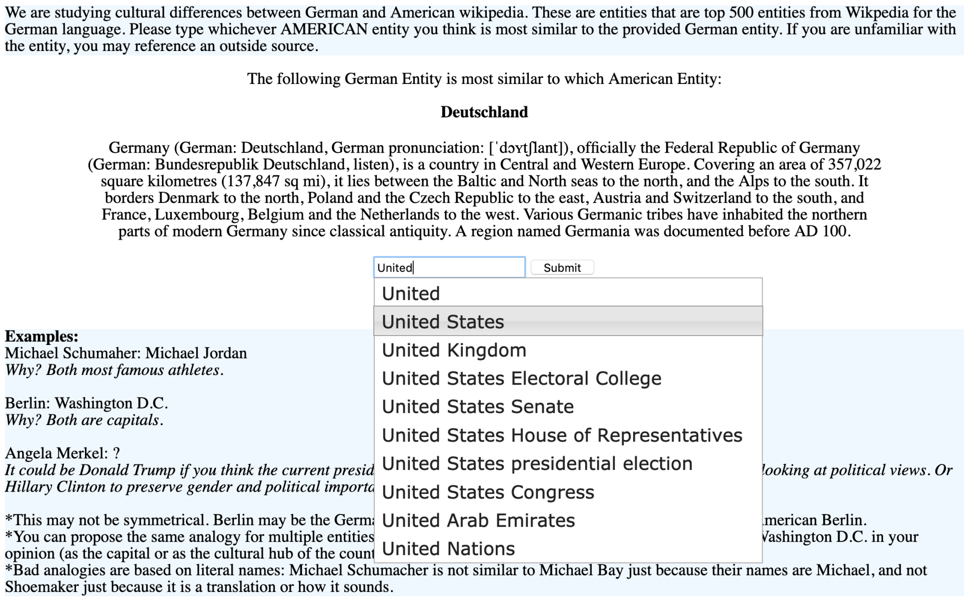


Figure 2: Our interface provides users with information about the entity and asks them to select an option from possible Wikipedia pages

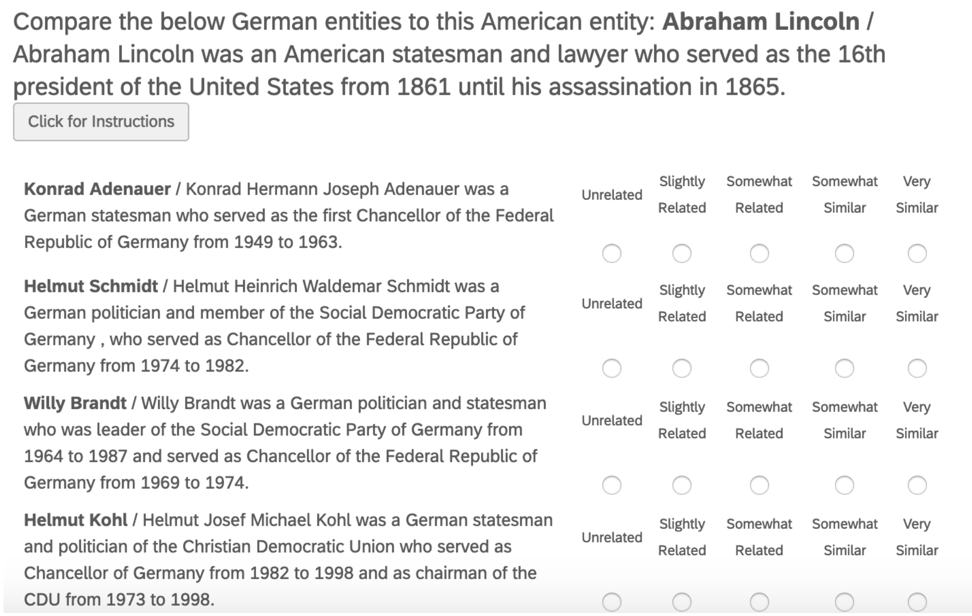


Figure 3: Our Qualtrics survey

## D Data

| Entity | Human Adaptation: NOC German→American |
|---|---|
| Adolf Eichmann | Andrew Jackson, Andrew Jackson, Franklin D. Roosevelt, Nathan Bedford Forrest, Steve Bannon |
| Angela Merkel | Barack Obama, Donald Trump, Hillary Clinton, Hillary Clinton, Hillary Clinton, Hillary Clinton, Joe Biden |
| Baron Munchausen | Captain America, Daniel Bolger, Joseph Smith, Paul Bunyan, Robert Jordan , Yankee Doodle |
| Carl von Clausewitz | Alfred Thayer Mahan, Dwight D. Eisenhower, Henry Knox, Robert E. Lee, Ulysses S. Grant |
| Friedrich Nietzsche | Ayn Rand, Henry David Thoreau, Henry Thoreau, Jordan Peterson, William James |
| Henry Kissinger | Henry Kissinger, Henry Kissinger, John Kerry, Madeleine Albright, Richard Nixon |
| Immanuel Kant | Benjamin Franklin, John Dewey, John Locke, John Rawls, Robert Nozick |
| Johann Sebastian Bach | Aaron Copland, Elvis Presley, Elvis Presley, Irving Berlin, Johnny Cash, Scott Joplin |
| Johann Wolfgang von Goethe | Edgar Allan Poe, Ernest Hemingway, Walt Whitman |
| Johannes Gutenberg | Benjamin Franklin, Bill Gates, Eli Whitney, Thomas Edison |
| Joseph Goebbels | David Duke, Franklin D. Roosevelt, George Rockwell, Rupert Murdoch, david duke |
| Karl Lagerfeld | Anna Wintour, Anna Wintour, Marc Jacobs, Ralph Lauren, Ralph Lauren, Ralph Lauren, Ralph Lauren |
| Karl Marx | Angela Davis, Beck, Bernie Sanders, John Jay, John Rawls, John Rawls |
| Leni Riefenstahl | DW Griffeth, David Wark Griffith, Frank Capra, Judy Garland |
| Ludwig van Beethoven | Aaron Copland, Aaron Copland, Aaron Copland, Elvis Presley, Frank Sinatra, George Gershwin, George Gershwin, Scott Joplin |
| Marlene Dietrich | Bette Davis, Clara Bow, Elizabeth Taylor, Marilyn Monroe, William Tecumseh Sherman |
| Martin Luther | Barry Goldwater, Brigham Young, Joseph Smith, Joseph Smith, Joseph Smith |
| Otto von Bismarck | Abraham Lincoln, George Washington, George Washington, George Washington, George Washington, Ulysses S. Grant |
| Pope Benedict XVI | Billy Graham, Billy Graham, Brigham Young, John Carroll , Seán Patrick O'Malley |
| Richard Wagner | Charles Ives, Frank Sinatra, Leonard Bernstein, Philip Glass |

Table 3: Veale NOC German→American adaptations.

| Entity | Human Adaptation: NOC American→German adaptations |
|---|---|
| Abraham Lincoln | Helmut Kohl, Konrad Adenauer, Wilhelm Friedrich Ludwig von Preußen, Willy Brandt, Willy Brandt |
| Al Capone | Adolf Leib, Carlos Lehder-Rivas, Jan Marsalek, Nasser Abou-Chaker, Nasser About-Chaker |
| Alfred Hitchcock | Bernd Eichinger, Bernd Eichinger, Michael Bully Herbig, Roland Emmerich, Wim Wenders |
| Benedict Arnold | Hansjoachim Tiedge, Otto von Bismarck, Otto von Bismarck, Robert Blum |
| Bill Gates | Andreas von Bechtolsheim, Carl Benz, Dietmar Hopp, Konrad Zuse |
| Britney Spears | Helene Fischer, Herbert Grönemeyer, Jeanette Biedermann, Nena, Til Schweiger |
| Charles Lindbergh | Ferdinand von Richthofen, Heinrich Horstman, Karl Wilhelm Otto Lilienthal, Ludwig Hofmann, Wernher von Braun |
| Donald Trump | Adolf Hitler, Adolf Hitler, Carsten Maschmeyer, Christian Lindner |
| Elvis Presley | Peter Kraus, Rammstein, The Scorpions, Udo Lindenberg, Udo Lindenberg |
| Ernest Hemingway | Günter Grass, Hermann Hesse, Johann Wolfgang von Goethe, Karl May, Martin Walser |
| Frank Lloyd Wright | Gerhard Richter, Hugo Häring, Karl Lagerfeld, Max Dudler, Walter Gropius |
| George Washington | Friedrich II, Heinrich I, Konrad Adenauer, Otto I. der Große, Otto von Bismarck |
| Henry Ford | Carl Benz, Carl Benz, Carl Benz, Ferdinand Porsche, Gottlieb Wilhelm Daimler |
| Hillary Clinton | Angela Merkel, Angela Merkel, Angela Merkel, Kramp-Karrenbauer, Sahra Wagenknecht |
| Homer Simpson | Alf, Heidi, Pumuckl, Werner, Werner - Beinhart! |
| Jack The Ripper | Armin Meiwes, Der Bulle von Tölz, Joachim Kroll, Karl Denke, Rudolf Pleil |
| Jay Z | Capital Bra, Marteria, Sido, Sido, Sido |
| Jimi Hendrix | Bela B., Gisbert zu Knyphausen, Herbert Grönemeyer, Rudolf Schenker, Spider Murphy Gang |
| John F. Kennedy | Hanns Martin Schleyer, Willy Brandt, Willy Brandt, Wolfgang Schäuble |
| Kim Kardashian | Carmen Geiss, Gina-Lisa Lohfink, Heidi Klum, Heidi Klum, Sarah Connor |
| Louis Armstrong | Günter Sommer, Helmut Brandt, Jan Delay, Michael Abene, Mozart |
| Marilyn Monroe | Heidi Klum, Ingrid Steeger, Marlene Dietrich, Micaela Schäfer, Uschi Glas |
| Michael Jordan | Dirk Nowitzki, Dirk Nowitzki, Dirk Nowitzki, Franz Beckenbauer, Michael Schuhmacher |
| Neil Armstrong | Alexander Gerst, Sigmund Jähn, Sigmund Jähn, Ulf Merbold, Wernher von Braun |
| Noam Chomsky | Helmut Glück, Juergen Habermas, Jürgen Habermas, Ludwig Wittgenstein, Wilhelm Röntgen |
| Oprah Winfrey | Anne Will, Arabella Kiesbauer, Maybrit Illner, Thomas Gottschalk, Thomas Gottschalk |

| | |
|---|---|
| Orville Wright | Carl Benz, Gustav Otto, Gustav Weißkopf, Otto Lilienthal, Wernher von Braun |
| Richard Nixon | Franz Josef Strauss, Helmut Kohl, Ludwig Erhard, Ludwig Erhard, Richard von Weizsäcker |
| Rosa Parks | Anne Wizorek, Marie Juchacz, Sophie Scholl, Sophie Scholl, Vera Lengsfeld |
| Serena Williams | Andrea Petkovic, Boris Becker, Sabine Lisicki, Steffi Graf, boris becker |
| Steve Jobs | Carl Benz, Dietmar Hopp, Dietmar Hopp, Karl Lagerfeld |
| Steven Spielberg | Michael Bully Herbig, Roland Emmerich, Roland Emmerich, Roland Emmerich, Wim Wenders |
| Superman | Bibi Blocksberg, Fix and Foxi, Maverick, Superman, Till Eulenspiegel |
| Tiger Woods | Boris Becker, Martin Kaymer, Martin Kaymer, Michael Schumacher, Serge Gnabry |
| Walt Disney | Axel Springer, Christian Becker, Franz Mack, Gerhard Hahn, Rötger Feldmann |

Table 4: Veale NOC American→German adaptations.

| Entity | Human Adaptation: Wikipedia German→American |
|---|---|
| ARD | NPR, PBS, PBS |
| Adolf Hitler | Donald Trump, Donald Trump, Franklin D. Roosevelt, Franklin D. Roosevelt, Franklin D. Roosevelt |
| Airbus | Boeing, Boeing, Boeing, Boeing, Lockheed Martin |
| Albert Einstein | Carl Sagan, J. Robert Oppenheimer, J. Robert Oppenheimer, John Forbes Nash Jr., Thomas Edison |
| Alice Merton | Ariana Grande, Elle King, K.T. Tunstall, P!NK, Vanessa Carlton |
| Alternative für Deutschland | Libertarian Party , Republican Party, Tea Party movement |
| Andrea Nahles | Elizabeth Warren, Hillary Clinton, Nancy Pelosi, Tammy Duckworth |
| Andrej Mangold | Kawhi Leonard, Kevin Durant, Kris Humphries, Yao Ming |
| Annalena Baerbock | Al Gore, Al Gore, Alexandria Ocasio-Cortez, Bernie Sanders, Jill Stein |
| Anne Frank | Anna Green Winslow, Clara Barton, Emmett Till, Kunta Kinte |
| Annegret Kramp-Karrenbauer | Condoleezza Rice, Hillary Clinton |
| AnnenMayKantereit | Guns N' Roses, Milky Chance, Polar Bear Club, Red Hot Chili Peppers |
| Apache 207 | Fetty Wap, Tekashi 69, XXXTentacion, Zayn Malik |
| Arnold Schwarzenegger | Chuck Norris, Dwayne Johnson, Ronnie Coleman, Sylvester Stallone, Sylvester Stallone |
| BMW | Cadillac, Cadillac, Chevrolet, Chrysler |
| Babylon Berlin | Game of Thrones, Man From U.N.C.L.E., Peaky Blinders , The Americans, Turn |
| Baden-Württemberg | California, Chicago metropolitan area, San Diego, Southern United States, Texas |
| Bastian Yotta | Chad Johnson, Colton Underwood, Dan Bilzerian |
| Bauhaus | Frank Lloyd Wright |
| Bayerischer Rundfunk | NPR, National Public Radio, National Public Radio, national public ra |
| Bayern | Florida, New York, The Confederacy |
| Benjamin Piwko | Bruce Lee, Colton Underwood, Derek Hough |
| Berlin | New York City, Portland Oregon, Washington D.C., Washington D.C., Washington D.C. |
| Berliner Mauer | Border Patrol Police, Mason–Dixon line, Mason–Dixon line, US-Mexican border |
| Bertolt Brecht | Tennessee Williams, Tennessee Williams |
| Björn Höcke | Lindsey Graham, Mike Pence |
| Borussia Dortmund | Golden State Warriors, New England Patriots, New England Patriots |
| Brandenburg | Maryland, New York, Northeastern United States, Richmond Virginia, Virginia |
| Bruno Ganz | Clint Eastwood, Ethan Hawke, Marlon Brando, Robert De Niro, Robert De Niro |
| Bundespräsident | First Lady, President of the United States, Speaker of the House |
| Bundeswehr | Department of Defense , US military, United States Armed Forces, United States Army |
| Capital Bra | Drake, Eminem, Eminem, Kanye West, Kendrick Lamar |
| Carola Rackete | American Civil Liberties Union, Dawn Wooten, Rosa Parks, Whale Wars |
| Carolin Kebekus | Amy Schumer, Sarah Silverman, Tina Fey, Tina Fey |

| | |
|---|---|
| Charité | Call the Midwife, Grey’s Anatomy, Grey’s Anatomy, The Queen’s Gambit |
| Chris Töpperwien | Gordon Ramsey , Guy Fieri, Jeff Probst |
| Christoph Waltz | Anthony Hopkins, Christoph Waltz, Denzel Washington |
| Dark | Stranger Things, Stranger Things |
| Deutsche Bahn | Amtrack, Norfolk Southern Railway, Union Pacific Corporation |
| Deutsche Demokratische Republik | Confederate States of America, Confederate States of America, Texas, The Confederacy, The Confederate States of America |
| Deutsche Nationalhymne | Born in the U.S.A., Lazy Eye , Star Spangled Banner, The Star Spangled Banner |
| Deutschland | America, America, Continental United States, USA, United States, United States |
| Dieter Bohlen | Billy Joel, Blake Shelton, Daryl Hall, Paula Abdul, Ryan Seacrest |
| Dirk Nowitzki | LeBron James, Michael Jordan, Shaquille O’Neal |
| Doreen Dietel | Jessica Alba, Lisa Kudrow, Warrick Brown |
| Dreißigjähriger Krieg | American Civil War, American Civil War, American Indian Wars, Civil war |
| Elisabeth von Österreich-Ungarn | Edith Roosevelt, Hillary Clinton, Jackie Kennedy |
| Elyas M’Barek | Adam Sandler, Adam Sandler, Chris Pine |
| Europawahl in Deutschland 2019 | 2018 United States elections, American presidential election 2020, Us election 2018 |
| Europäisches Parlament | North Atlantic Council, Representative of the United States of America to the European Union, United Nations, United States Congress |
| Evelyn Burdecki | Hannah Brown, Kaitlyn Bristowe, Kim Kardashian, Kim Kardashian |
| FC Bayern München | Dallas Cowboys, Dc United, New York Yankees, New York Yankees, New York Yankees |
| Falco | David Bowie, Frederick William Schneider III, MC Hammer, Michael Jackson |
| Ferdinand Sauerbruch | Ben Carson, Ben Carson, Cornelius P. Rhoads, Jonas Salk, Virginia Apgar |
| Flughafen Berlin Brandenburg | Cincinnati Subway, DCA , John F. Kennedy International Airport, LaGuardia Airport |
| Frankfurt am Main | Chicago, Los Angeles, Los Angeles, New York City, Washington D.C. |
| Fritz Honka | Ted Bundy, Ted Bundy, Ted Bundy, Zodiac |
| Hamburg | Chicago, Chicago, Los Angeles, New York, Philadelphia |
| Hannelore Elsner | Elizabeth Taylor, Jane Lynch, Julia Roberts |
| Heidi Klum | Chrissy Teigen, Cindy Crawford, Gigi Hadid, Karlie Kloss, Tyra Banks |
| Heinz-Christian Strache | Anthony Weiner, Ben Carson, Donald J. Trump, Rob Ford, Roger Stone |
| Helene Fischer | Beyoncé, Kelly Clarkson, Taylor Swift, Taylor Swift |
| Hessen | Arizona, Illinois, Mid-Atlantic , Napa County California |
| Holocaust | Chattel Slavery, Japanese interned in American camps, Slavery in the United States |
| Ich bin ein Star – Holt mich hier raus! | Survivor, Survivor |
| Jürgen Klopp | Bill Belichick, Bill Belichick, John Wooden |
| Kevin Kühnert | Bernie Sanders, Bernie Sanders, Bernie Sanders, Pete Buttigieg |

| | |
|---|---|
| Klaus Kinski | Christopher Lee, Clark Gable, John Wayne, Robert Pattinson, Robert Pattinson |
| Kontra K | 50 Cent, Eminem, Eminem, Jesus Is King, Travis Scott |
| Köln | Boston, Chicago, Chicago, Houston |
| Leila Lowfire | Paris Hilton, Sasha Grey, Zendaya |
| Leipzig | Denver, Detroit, Miami, San Diego |
| Lena Meyer-Landrut | Ariana Grande, Kelly Clarkson, Kelly Clarkson, Meghan Trainor, Selena Gomez |
| Liechtenstein | Connecticut, Mexico, Philippines, Victoria British Columbia |
| Lisa Martinek | Julie Benz, Katherine Heigl, Mandy Moore, Meryl Streep |
| Ludwig van Beethoven | Aaron Copland, Aaron Copland, Aaron Copland, Aaron Copland, Elvis Presley, Frank Sinatra, George Gershwin, George Gershwin, Scott Joplin |
| Lufthansa | Delta, United, United Airlines, United Airlines |
| Luxemburg | Canada, Connecticut, Mexico, Victoria British Columbia |
| Mark Forster | Bruno Mars, Post Malone |
| Mero | DaBaby, Fetty Wap, Lil Nas X, Lil Nas X, Post Malone |
| Michael Schumacher | Dale Earnhardt, Dale Earnhardt, James Gordon, Jeff Gordon, Tiger Woods |
| München | Chicago, Los Angeles, New York City, New York City, Washington D.C. |
| Nico Santos | Harry Styles, Justin Bieber, Shawn Mendes |
| Niki Lauda | Dale Earnhardt, Dale Earnhardt Jr., Jeff Gordon, Jeff Gordon, Tiger Woods |
| Norddeutscher Rundfunk | NPR, NPR, National Public Radio, PBS, Sirius XM |
| Nordrhein-Westfalen | California, California |
| Philipp Amthor | Alexandria Ocasio-Cortez, Ben Shapiro |
| RAF Camora | Bad Bunny, Drake, Drake , Eminem, Future |
| Rammstein | Green Day, Metallica, Metallica, Metallica, Sum 41 |
| Rhein | Mississippi, Mississippi River, Mississippi River |
| Robert Habeck | Al Gore, Bernie Sanders, Jill Stein, Ralph Nader |
| Rudi Assauer | Dave Roberts, Gregg Berhalter, Tom Flores, Vince Lombardi, Vince Lombardi |
| Sahra Wagenknecht | Alexandria Ocasio-Cortez, Elizabeth Warren, Elizabeth Warren, Elizabeth Warren, Nancy Pelosi |
| Sarah Connor | Beyoncé, Britney Spears, Mariah Carey |
| Schweiz | Canada, Canada, Iowa, Mexico, United States |
| Sebastian Kurz | Alexandria Ocasio-Cortez, Greg Abbott, Justin Trudeau, Justin Trudeau, Mitch McConnell |
| Serge Gnabry | Clint Dempsey, JuJu Smith-Schuster, Phillip Rivers, Stephen Curry, Zion Williamson |
| Sido | Eminem, Eminem, Macklemore |
| The Cratez | DJ Khaled, Drake , Twenty One Pilots |
| Thüringen | Iowa, Midwestern United States, Tennessee, Tennessee |
| Till Lindemann | James Hetfield, James Hetfield, James Hetfield, Ozzy Osbourne |
| Tom Kaulitz | Adam Levine, Blink-182, Chris Martin, Green Day, Maroon 5 |
| UEFA Champions League | Major League Soccer, NFC, NFL, National Football League, Ncaa |
| Udo Jürgens | Aretha Franklin, Billy Joel, Elton John, Michael Jackson, Rolling Stone, Tom Lehrer |
| Udo Lindenberg | Johnny Cash, Mick Jagger, Roger Taylor , Travis Barker |

| | |
|---|---|
| Ursula von der Leyen | Condoleezza Rice, Hillary Clinton, Mike Pence, Sarah Palin, Susan Rice |
| Volkswagen AG | Ford Motor Company, Ford Motor Company, Ford Motor Company, Ford Motor Company, Ford Motor Company |
| Walter Lübcke | Harvey Milk, John F. Kennedy, John Roll, Steve Scalise |
| Weimarer Republik | America, Confederation Period, Congress of the Confederation, Counterculture of the 1960s, The Confederate States of America |
| Westdeutscher Rundfunk Köln | ABC News, NBC, NPR |
| Wien | Austin Texas, Richmond Virginia, Toronto, Washington D.C. |
| Wilhelm II. | William Howard Taft, Woodrow Wilson, Woodrow Wilson |
| Wolfgang Amadeus Mozart | Alan Menken, Elvis Presley, Leonard Bernstein |
| ZDF | NPR, NPR, National Public Radio, PBS, PBS |
| Österreich | Canada, Mexico, Texas, Texas, United States |
| Ötzi | Spirit Cave mummy, Spirit Cave mummy, Spirit Cave mummy, Sue |

Table 5: Top Wikipedia German→American adaptations.

| Entity | Human Adaptation: Wikipedia American→German |
|---|---|
| 13 Reasons Why | Club der roten Bänder, Gute Zeiten schlechte Zeiten, Lammbock, Türkisch für Anfänger |
| Albert Einstein | Albert Einstein, Albert Einstein, Albert Einstein, Max Planck, Max Planck |
| Alexander Hamilton | Konrad Adenauer, Max Weber, Otto von Bismarck, Otto von Bismarck |
| American Civil War | Deutscher Krieg, Dreißigjähriger Krieg, German Revolution of 1918–1919, German revolutions of 1848–1849 |
| American Horror Story | Dark, Der goldene Handschuh, Good Bye Lenin!, Tintenherz |
| Angelina Jolie | Barbara Schöneberger, Franka Potente, Marlene Dietrich, Romy Schneider, Veronica Maria Cäcilia Ferres |
| Apple Inc. | BMW, Fujitsu, SAP, Siemens |
| Ariana Grande | Lena Meyer-Landrut, Lena Meyer-Landrut, Lena Meyer-Landrut, Sarah Connor, Sarah Connor |
| Arnold Schwarzenegger | Arnold Schwarzenegger, Karl Lauterbach, Matthias Steiner, Peter Maffay, Ralf Rudolf Möller |
| Ashton Kutcher | Florian David Fitz, Matthias Schweighöfer, Til Schweiger, Til Schweiger |
| Australia | Australia, Russia, Schweiz, South Africa, Österreich |
| Avengers Infinity War | Das Arche Noah Prinzip, Fack ju Göhte, Fantastic Four, Who Am I |
| Barack Obama | Angela Merkel, Angela Merkel, Angela Merkel, Helmut Schmidt, Helmut Schmidt |
| Beyoncé | Helene Fischer, Sarah Connor, Veronica Ferres, Xavier Naidoo, Yvonne Catterfeld |
| Black Mirror | Dark, Dark, Die kommenden Tage, Krabat |
| Blake Lively | Josefine Preuß, Maria Furtwängler, Maria Furtwängler, Til Schweiger |
| Brad Pitt | Florian David Fitz, Frederick Lau, Til Schweiger, Til Schweiger, Til Schweiger |
| Bruce Lee | Götz Georg, Henry Maske, Julian Jacobi, Max Schmeling, no one is like Bruce Lee |
| Caitlyn Jenner | Kristin Otto, Magdalena Neuner, Magdalena Neuner, Niklas Kaul, Ulrike Meyfarth |
| California | Bavaria, Bavaria, Bayern, Bayern |
| Camila Cabello | Helene Fischer, Lena Meyer-Landrut, Lena Meyer-Landrut, Nadja Benaissa |
| Canada | Austria, Italy, Schweiz, Sweden, Österreich |
| Cardi B | Ace Tee, Pamela Reif, Sabrina Setlur, Sarah Connor, Schwester Ewa |
| Charles Manson | Andreas Baader, Issa Rammo, Papst benedikt xvi, Paul Schäfer |
| Charlize Theron | Baran bo Odar, Josefine Preuß, Josefine Preuß, Veronica Ferres, Veronica Maria Cäcilia Ferres |
| Cher | Marlene Dietrich, Nena, Nena, Nena |
| Chris Pratt | Elyas M'Barek, Jan Josef Liefers, Matthias Schweighöfer, Ralf Moeller, Til Schweiger |
| Clint Eastwood | Heinz Erhardt, Klaus Kinski, Mario Adorf, Til Schweiger, Wim Wenders |
| Darth Vader | Adolf Hitler, Belzebub, Hagen von Tronje, Jens Maul |
| Donald Glover | Elyas M'Barek, Helge Schneider, Money Boy, Stefan Raab |

| | |
|---|---|
| Drake | Bushido, Cro, Falco, Fler |
| Dwayne Johnson | Alexander Wolfe, Arnold Schwarzenegger, Peter Alexander, Tim Wiese, Tim Wiese |
| Elon Musk | Alexander Samwer, August Horch, Carl Benz, Herbert Diess, Werner von Siemens |
| Eminem | Bushido, Kollegah, Sido, Sido, Sido |
| Facebook | Das Erste, Lokalisten, Lokalisten, Schüler VZ, StudiVZ, StudiVZ |
| Friends | Gute Zeiten schlechte Zeiten, Gzsz, Lindenstraße, Stromberg |
| Game of Thrones | Babylon Berlin, Babylon Berlin, Babylon Berlin, Die unendliche Geschichte, Krabat |
| Google | Ecosia, Fastbot, SAP, SAP, i.d.k. |
| Harry Potter | Die Unendliche Geschichte, Die unendliche Geschichte, Harry Potter und ein Stein, Meggie Folchart |
| Heath Ledger | Christoph Waltz, Florian David Fitz, Henry Blanke, Matthias Schweighöfer, Tilman Valentin Schweiger |
| It | Dark, Der goldene Handschuh, Die Wolke, Pandorum |
| Jason Momoa | Arnold Schwarzenegger, Benno Fürmann, Christoph Waltz, Elyas M'Barek, Elyas M'Barek, Elyas M'Barek |
| Jeff Bezos | Alexander Samwer, Beate Heister, Martin Winterkorn, Oliver Samwer |
| Jeffrey Dahmer | Armin Meiwes, Fritz Haarmann, Joachim Kroll, Karl Denke, Karl Denke |
| Jennifer Aniston | Barbara Schöneberger, Diane Kruger, Diane Kruger, Franka Potente, Iris Berben |
| Jennifer Lawrence | Iris Berben, Josefine Preuß, Karoline Herfurth, Ruby O. Fee |
| Jennifer Lopez | Heidi Klum, Helene Fischer, Jeanette Biedermann, Mandy Capristo, Sarah Connor |
| John Cena | Arnold Schwarzenegger, Max Schmeling, Max Schmeling, Ralf Möller |
| Johnny Cash | Fantastischen vier, Helge Schneider, Peter Maffay, Peter Maffay |
| Johnny Depp | Christoph Maria Herbst, Christoph Waltz, Cro, Til Schweiger, Xavier Naidoo |
| Julia Roberts | Karoline Herfurth, Maria Furtwängler, Marlene Dietrich, Marlene Dietrich |
| Justin Bieber | Cro, Felix Jaehn, Lukas Rieger, McFittie, Mike Singer |
| Keanu Reeves | Daniel Brühl, Mario Adorf, Til Schweiger, til schweiger |
| Kylie Jenner | Barbara Schöneberger, Heidi Klum, Karoline Einhoff, Sarah Connor, Stefanie Giesinger |
| Lady Gaga | Helene Fischer, Nena, Nena, Nina Hagen, Sarah Lombardi |
| LeBron James | Dirk Nowitzki, Dirk Nowitzki, Dirk Nowitzki, Dirk Nowitzki, Toni Kroos |
| Leonardo DiCaprio | Matthias Schweighöfer, Moritz Bleibtreu, Til Schweiger, Til Schweiger, Til Schweiger |
| Lisa Bonet | Franka Potente, Iris Berben, Karoline Herfurth, Maria Furtwängler |
| Madonna | Blümchen, Helene Fischer, Helene Fischer, Helene Fischer, Sarah Connor |
| Mark Wahlberg | Florian David Fitz, Til Schweiger, Tilman Valentin Schweiger, Alexei Alexejewitsch |
| Martin Luther King Jr. | Hans Scholl, Hans Scholl, Helmut Palmer, Robert Blum, Sophie Scholl |

| | |
|---|---|
| Marvel Cinematic Universe | Bavaria Film, Havelstudios, Phantásien, Rat Pack Filmproduktion, Tatort |
| Michael Jackson | Herbert Grönemeyer, Nena, Udo Jürgens, Xavier Naidoo, Xavier Naidoo |
| Mila Kunis | Josefine Preuß, Matthias Schweighöfer, Vanessa Mai |
| Miley Cyrus | Lena Meyer-Landrut, Lukas Rieger, Nena, Sarah Connor, Yvonne Catterfeld |
| Muhammad Ali | Alexander Abraham, Boris Becker, Max Schmeling, Max Schmeling, Sven Ottke |
| Natalie Portman | Barbara Schöneberger, Diane Kruger, Franka Potente, Iris Berben |
| New York City | Berlin, Berlin, Berlin, Berlin, Frankfurt |
| Nicole Kidman | Evelyn Hamann, Franka Potente, Senta Berger, iris berben |
| Peaky Blinders | Dark, Dieter Schwarz, Im Westen Nichts Neues, Tatort, Tatort |
| Philippines | Greece, Griechenland, Mallorca, Mallorca |
| Post Malone | Bushido, Bushido, Cro, Cro, Kollegah |
| Rihanna | Helene Fischer, Lena Meyer-Landrut, Lena Meyer-Landrut, Nena |
| Riverdale | Babylon Berlin, Berlin Tag und Nacht, Neues vom Süderhof, Türkisch für Anfänger |
| Robert Downey Jr. | Christoph Waltz, Günter Strack, Martin Semmelrogge, Moritz Bleibtreu, Til Schweiger |
| Robin Williams | Hape Kerkeling, Heinz Erhardt, Peter Maffay, Silvia Seidel, Tim Bendzko |
| Ronald Reagan | Helmut Schmidt, Konrad Adenauer, Konrad Adenauer, Konrad Adenauer |
| Ryan Reynolds | Daniel Brühl, Florian David Fitz, Matthias Schweighöfer, Til Schweiger, Til Schweiger |
| Scarlett Johansson | Lena Gercke, Romy Schneider, Sarah Connor, Sarah Connor, Veronica Ferres |
| Selena Gomez | Lena Meyer-Landrut, Lena Meyer-Landrut, Nena, Nora Tschirner |
| September 11 attacks | Anschlag im OEZ, Dresden Bombing, Mauerfall, RAF-Attentate, Terroranschlag Olympia 1972 |
| Shaquille O'Neal | Dirk Nowitzki, Dirk Nowitzki, Mehmet Scholl, Niklas Süle |
| Star Wars | Dark, Metropolis, Traumschiff Surprise – Periode 1, Who Am I?, i.d.k |
| Stephen Curry | Dirk Nowitzki, Dirk Nowitzki, Dirk Nowitzki, Dirk Nowitzki, Manuel Neuer |
| Stranger Things | 8 Tage, Babylon Berlin, Dark, Tatort, Tatort |
| Sylvester Stallone | Henry Blanke, Jan Josef Liefers, Michael Bully Herbig, Michael Fassbender, Til Schweiger |
| Taylor Swift | Lena Meyer-Landrut, Lena Meyer-Landrut, Sarah Connor, Sarah Connor, Yvonne Catterfeld |
| Ted Bundy | Joachim Kroll, Josef Fritzl, Niels Högel, Rudolf Pleil, Rudolf Pleil |
| The Big Bang Theory | Doctor's Diary, Stromberg, Stromberg, der Tatortreiniger |
| The Crown | Babylon Berlin, Deutschland 83, Die Deutschen, Karl der Große |
| The Handmaid's Tale | Dark, Dark, Der Pass, Die Wanderhure, Er ist wieder da |
| The Walking Dead | Dark, Dark, Der goldene Handschuh, Zombies From Outer Space |
| Tom Brady | Franz Beckenbauer, Michael Ballack, Oliver Kahn, Thomas Müller, Uli Stein |
| Tom Cruise | Benno Fürmann, Benno Fürmann, Christoph Waltz, Elyas M'Barek, Matthias Schweighöfer |
| Tom Hanks | Christoph Waltz, Christoph Waltz, Daniel Brühl, Til Schweiger |

| | |
|---|---|
| Tom Hardy | Bruno Ganz, Michael Herbig, Til Schweiger, Wotan Wilke Möhring |
| Tom Holland | Daniel Brühl, Frederick Lau, Matthias Schweighöfer, Matthias Schweighöfer, Til Schweiger |
| Tupac Shakur | Farid Bang, Haftbefehl, Kollegah, Kristoffer Klauß, Peter Fox |
| United States | BRD, Bundesrepublik Deutschland, Deutschland, Germany, Germany |
| Vietnam War | Berlin Wall, First world war, Kosovokrieg, World War II |
| Wikipedia | Brockhaus, Brockhaus Enzyklopädie, Brockhaus Enzyklopädie, Duden, dict.cc |
| Will Smith | Daniel Brühl, Elyas M'Barek, Sascha Reimann, Sido, Til Schweiger |
| X-Men | Abwärts, Fantastic Four, Freaks, Krabat, Who Am I |
| YouTube | Lokalisten, MyVideo, MyVideo, ProSieben, lokalisten |
| Zac Efron | Frederick Lau, Lukas Rieger, Peter Kraus, Walter Sedlmayr |
| Zendaya | Franka Potente, Iris Berben, Lena Meyer-Landrut, Lena Meyer-Landrut, Yvonne Catterfeld |

Table 6: Top Wikipedia American→German adaptations.

| Entity | Top Five WikiData Adaptations |
|---|---|
| Abraham Lincoln | Victor Adler, Johann Joachim Christoph Bode, Willem Barentsz, Hermann Wagener, Robert von Mohl |
| Al Capone | Hans H. Zerlett, Fritz Thyssen, Adam Rainer, Franz Winkelmeier, Christian Louis, Duke of Brunswick-Lüneburg |
| Alfred Hitchcock | Edgar Reitz, Jan Josef Liefers, Mario Adorf, Max Frisch, Armin Mueller-Stahl |
| Benedict Arnold | Hans-Georg Hess, Isabelle Eberhardt, Günther Heydemann, Max Schreck, Louis Blenker |
| Bill Gates | Ferdinand von Zeppelin, Günther Jauch, Nikolaus Harnoncourt, Sepp Blatter, Alfred Grosser |
| Britney Spears | Herta Müller, Günter Grass, Joachim Gauck, Hans-Dietrich Genscher, Koča Popović |
| Donald Trump | Max Frisch, Thomas Gottschalk, Jan Josef Liefers, Rainer Werner Fassbinder, Christa Wolf |
| Elvis Presley | Reinhard Lakomy, James Last, Herbert Achternbusch, Fritz Hauser, Hans-Peter Pfammatter |
| Ernest Hemingway | Karlheinz Böhm, Ricarda Huch, Michael Ballhaus, Arnold Zweig, Michael Fassbender |
| Frank Lloyd Wright | Ferdinand Hodler, Johan Zoffany, Hans Thoma, Arne Jacobsen, Lucas Cranach the Younger |
| George Washington | Friedrich Wilhelm von Seydlitz, Dagobert Sigmund von Wurmser, Heinz Guderian, Ernst Gideon von Laudon, George Olivier, count of Wallis |
| Henry Ford | Heinz Sielmann, Wieland Schmied, Manfred Krug, Paul Maar, Armin Mueller-Stahl |
| Hillary Clinton | Pope Benedict XVI, Willy Brandt, Angela Merkel, Helmut Schmidt, Kurt Biedenkopf |
| Homer Simpson | Elizabeth Lavenza, Hans Fugger, Baron Strucker, Herbert of Wetterau, Prince Johannes of Liechtenstein |
| Jimi Hendrix | Marius Müller-Westernhagen, Karl Richter, Reinhard Lakomy, Michael Cretu, Paul van Dyk |
| Kim Kardashian | Erika Mann, Frank Wedekind, Til Schweiger, Fritz von Opel, Carmen Electra |
| Marilyn Monroe | Gerhart M. Riegner, Viktor de Kowa, Otto Sander, Hans Hass, Dorothee Sölle |
| Michael Jordan | Jean-Claude Juncker, Richard von Weizsäcker, Herta Müller, Konrad Adenauer, Helmut Kohl |
| Louis Armstrong | Herbert Prikopa, Till Lindemann, Nico, Klaus Voormann, Jakob Adlung |
| Neil Armstrong | Stefan Hell, Franz-Ulrich Hartl, Reinhard Genzel, Charles Weissmann, Harald zur Hausen |
| Noam Chomsky | Günter Grass, Herta Müller, Heinrich Böll, Peter Handke, Juli Zeh |
| Oprah Winfrey | Günter Grass, Peter Scholl-Latour, Elfriede Jelinek, Juli Zeh, Christa Wolf |
| Orville Wright | Frank Thiess, Jessica Hausner, Elmar Wepper, Wolf Jobst Siedler, Marc Rothemund |
| Richard Nixon | Heinrich von Brentano, Ernst Benda, Gustav Heinemann, Heiner Geißler, Heinrich Albertz |
| Superman | Magneto, Nightcrawler, Sinterklaas, Silent Night, Victor Frankenstein |

| | |
|---|---|
| Steve Jobs | Victor Klemperer, Joschka Fischer, Jürgen Kuczynski, Joachim Fest, Dieter Hallervorden |
| Steven Spielberg | Herta Müller, Jean-Claude Juncker, Hans-Dietrich Genscher, Joachim Gauck, Koča Popović |
| Tiger Woods | Charles Dutoit, Shania Twain, Lise Meitner, Michael Haneke, Otto Hahn |
| Walt Disney | Shania Twain, Charles Dutoit, Lise Meitner, Otto Hahn, Michael Haneke |
| John F. Kennedy | Bernhard von Bülow, Otto von Habsburg, Hans-Jochen Vogel, Prince Henry of Prussia, Frederick Augustus III of Saxony |
| Charles Lindbergh | Pina Bausch, Ferdinand von Zeppelin, Nikolaus Harnoncourt, Jan Josef Liefers, Wolf Biermann |
| Rosa Parks | Hermann Lenz, Wilhelm Feldberg, Horst Tappert, Peter Stein, Gert Jonke |
| Serena Williams | Charles Dutoit, Lise Meitner, Michael Haneke, Richard von Coudenhove-Kalergi, Klaus Clusius |

Table 7: We show top-5 predictions out of the top-100 for American→German adaptations on the Veale NOC subset using **WikiData**. These are compared to our human annotations in our results.

| Entity | Top Five Embedding Adaptations: American→German adaptations on the Veale NOC |
|---|---|
| Abraham Lincoln | Bismarcks, Hindenburg, Theodor Heuss, Reichskanzler Otto, Bismarck |
| Al Capone | SA-Mann, Antifaschisten, Spitzel, Adolf Eichmann, Al Capone |
| Alfred Hitchcock | Die Feuerzangenbowle, Johannes Mario Simmel, Gustav Ucicky, Alfred Hitchcock, Fritz Lang |
| Benedict Arnold | Leipziger Völkerschlacht, Carl Gustav Wrangel, Hotze, Jean-Victor Moreau, Reichstruppen |
| Bill Gates | Axel Springer, T-Online, Credit Suisse, Deutsche Telekom, WirtschaftsWoche |
| Britney Spears | Xavier Naidoo, Glasperlenspiel, Rihanna, Rosenstolz, Kool Savas |
| Donald Trump | Christoph Blocher, Angela Merkel, Wahlkampf, Frauke Petry, Donald Trump |
| Elvis Presley | Rio Reiser, Elvis Presley, Udo Jürgens, Reinhard Mey, Bob Dylan |
| Ernest Hemingway | Hermann Hesse, Robert Walser, Robert Musil, Erich Kästner, Erich Maria Remarque |
| Frank Lloyd Wright | Le Corbusier, Erich Mendelsohn, Neuen Bauens, Walter Gropius, Joseph Maria Olbrich |
| George Washington:Washington | Bern, Zürich, Genf, Aarau, Basel |
| Henry Ford:Ford | Opel, Ford, BMW, Porsche, Mercedes-Benz |
| Hillary Clinton | Angela Merkel, Wahlkampf, Christoph Blocher, Nicolas Sarkozy, Bundeskanzlerin Angela Merkel |
| Homer Simpson | Wachtmeister Studer, Schlusssequenz, Lotte Jäger, Ulrike Folkerts, Helen Dorn |
| Jack The Ripper:Ripper | Mörders, Sherlock Holmes, Killer, Jack the Ripper, Verbrecher |
| Jay Z | Xavier Naidoo, Sido, Kool Savas, Glasperlenspiel, Bass Sultan Hengzt |
| Jimi Hendrix | Rio Reiser, Ton Steine Scherben, Joy Division, James Last, Lou Reed |
| Kim Kardashian | Oliver Pocher, Miley Cyrus, Sarah Connor, Anke Engelke, Rihanna |
| Marilyn Monroe | Marlene Dietrich, Willy Fritsch, Romy Schneider, Marika Rökk, Lilian Harvey |
| Michael Jordan:Jordan | Israel, Jordanien, Syrien, Ägypten, Libanon |
| Louis Armstrong | Django Reinhardt, Michael Jary, Louis Armstrong, James Last, Ella Fitzgerald |
| Neil Armstrong | Planeten Merkur, Astronauten, Apollo, Neil Armstrong, Mond |
| Noam Chomsky | Jürgen Habermas, Georg Lukács, Erich Fromm, Carl Schmitt, Hannah Arendt |
| Oprah Winfrey | Harald Schmidt, Oliver Pocher, Frank Elstner, Hape Kerkeling, Florian Silbereisen |
| Orville Wright | Junkers, Gottlieb Daimler, Hanna Reitsch, Hugo Junkers, Louis Blériot |
| Richard Nixon | Willy Brandt, Gustav Heinemann, Helmut Kohl, Adenauer, Helmut Schmidt |
| Superman | Batman, Superman, Spider-Man, Monster, Dracula |
| Steve Jobs | Microsoft, Steve Jobs, T-Online, Google, AOL |
| Steven Spielberg | Tom Tykwer, Die Feuerzangenbowle, Sönke Wortmann, Volker Schlöndorff, Peter Schamoni |

| | |
|---|---|
| Tiger Woods | Timo Boll, Jörg Roßkopf, Steffi Graf, Conny Freundorfer, Eberhard Schöler |
| Walt Disney | Unterwegs, UFA, MGM, Tobis, Disney |
| John F. Kennedy | Willy Brandt, Konrad Adenauer, Adolf Hitler, Winston Churchill, Theodor Heuss |
| Charles Lindbergh | Ernst Udet, Hanna Reitsch, Wilhelm Gustloff, Gustloff, Max Valier |
| Rosa Parks | Käthe Kollwitz, Köpenicker Blutwoche, Ernst Barlach, Weiße Rose, Theodor Heuss |
| Serena Williams | Angelique Kerber, Roger Federer, Agnieszka Radwańska, Timo Boll, Runde eins |

Table 8: We show top-5 predictions out of the top-100 for American→German adaptations on the Veale NOC subset using **3CosAdd**. These are compared to our human annotations in our results.