“The Michael Jordan of Greatness”:
Extracting Vossian Antonomasias from Two Decades of
the New York Times, 1987–2007

Frank Fischer
National Research University
Higher School of Economics, Moscow
ffischer@hse.ru

Robert Jäschke
Humboldt-Universität zu Berlin
robert.jaeschke@hu-berlin.de

Abstract

Vossian Antonomasias is a prolific stylistic device, in use since antiquity. It can compress
the introduction or description of a person or another named entity into a terse, poignant
formulation and can best be explained by an example: When Norwegian world champion
Magnus Carlsen is described as “the Mozart of chess”, it is Vossian Antonomasias we are
dealing with. The pattern is simple: A source (Mozart) is used to describe a target (Magnus
Carlsen), the transfer of meaning is reached via a modifier (“of chess”). This phenomenon
has been discussed before (as ‘metaphorical antonomasias’ or, with special focus on the
source object, as ‘paragons’), but no corpus-based approach has been undertaken as yet to
explore its breadth and variety. We are looking into a full-text newspaper corpus (The New
York Times, 1987–2007) and describe a new method for the automatic extraction of Vossian
Antonomasias based on Wikidata entities. Our analysis offers new insights into the occurrence
of popular paragons and their distribution.

1 Introduction and Related Works

When Peter Paul Rubens is described as “the Tarantino of Baroque painting” (in Tagesspiegel,
2014) or Slavoj Žižek as “the Elvis of cultural theory” (in Slate, 2014), we deal with a phenomenon
called Vossian Antonomasias (VA). This trope is named after Gerardus Vossius (1577–1649), the
Dutch humanist and professor of rhetoric, who first described VA as a specific sub-phenomenon
of antonomasias. Generally speaking, an antonomasias incorporates the idea of a particular prop-
erty of a person standing in for this person (e.g., “The Voice” standing in for Frank Sinatra). In the
case of Vossian Antonomasias, a person is attributed a specific property by reference to another (more
well-known, more popular, more notorious) person. An “atypological, modifying signal (pronoun, adject-
ive, genitive)” provides for the transfer of meaning (Lausberg, 1960). “Baroque painting” and “cultural
theory” work as such in the above-mentioned examples. In other words, VA uses a ‘modifier’ to
establish a relationship between a ‘source’ and a ‘target’ (Bergien, 2013). In our paper, we will study
this popular variant, although VA generally works without such modifier (as in “Go and denounce me,
you Judas!”, where Judas stands for “traitor” and no modifier is presented). Entities can occur both
as source and as target, as is shown by the example of Obama in Bergien (2013): until 2011 he mostly
appeared as target, but increasingly as source thereafter. Following Lakoff (1987), the entity serving
as ‘source’ is also referred to as ‘paragon’, “a specific example that comes close to embodying
the qualities of the ideal” (Bergien, 2013).

Internationally, the term ‘Vossian Antonomasias’ is rarely used; instead, we encounter a distinc-
tion between ‘Antonomasias1’ and ‘Antonomasias2’: ‘metonymic’ vs. ‘metaphorical antonomasias’
(Holmqvist and Pluciennik, 2010). In this classification scheme, Vossian Antonomasias would occur
within ‘Antonomasias2’, in the form of “comparisons with paragons from other spheres of culture: Ly-
otard is a pope of postmodernism, Bush is no Demosthenes; and we can buy the Cadillac of vacuum
cleaners.” (Holmqvist and Pluciennik, 2010)

The use of this stylistic device dates back to classical antiquity. Nowadays, it is a welcome element
in many journalistic genres and often occurs in headings, since it can be both informative and enigmatic,
and is notorious for its unique entertaining qualities. A larger collection of examples1 gave the
impetus to research this phenomenon more system-

1See https://www.umblaetterer.de/datenzentrum/
vossianische-antonomasien.html and Fischer and Wälzholz
(2014).
atically, with historical perspective and on the basis of larger English and German newspaper corpora (Jäschke et al., 2017). The aim of this paper is an exploratory automated analysis of VA in the New York Times (1987–2007). The object of study was chosen because we were looking for a corpus of a general contemporary English-language newspaper and found the NYT Annotated Corpus a perfect fit being readily available as it was released to the research community (Sandhaus, 2008).

2 Extraction and Evaluation

The method described in Jäschke et al. (2017) was based on NLTK’s NER module and complex handcrafted regular expressions. This resulted in low precision (few of the resulting candidates were VA) and recall (only few VA were detected). Reason enough to work on a novel method for the extraction of VA. We did so by restricting our search to the clear-cut pattern “the SOURCE of”. This pattern limitation lead to the extraction of a much larger set of VA than before (by a factor of 10) with just minimal additional effort. It has to be noted that this strategy works well for our English-language corpus, but would have to be adjusted for languages that do not use articles or prepositions to determine cases.

To overcome the shortcomings of current NER implementations we resorted to Wikidata, the crowd-sourced fact database (Vrandečić and Krötzsch, 2014), from which we extracted a list of possible sources of a VA expression. In most other use-case scenarios, using Wikidata for named-entity matching would lead to a deterioration or limitation of the results. But in our very special case, this simple idea was the breakthrough to achieve clean extractions and to immensely increase precision. Because one prerequisite for being a source of a VA is to be famous or popular or notorious for something, and this arguably goes along with being present on at least one of the hundreds of Wikipedia versions and thus having a Wikidata entry.

By help of the Wikidata Toolkit we downloaded and filtered the Wikidata dump from August 14, 2017. We used the instanceOf property to retrieve all instances of the class ‘human’. For each individual we retrieved the English representation of their name and potential alias names. The resulting list comprises 2,801,931 individuals with 2,955,761 unique names.

From the XML-encoded New York Times corpus (Sandhaus, 2008), we extracted the full text for each of the 1,854,726 articles and split it into sentences using the Natural Language Toolkit, NLTK (Bird et al., 2009). Within each sentence we then searched for character sequences matching the regular expression `\bthe\s*([\w,.-]+\s+){1,5}\sof\b`. It matches all phrases beginning with “the” and ending with “of”, with up to five words in-between.

We had to increase this margin to be able to include dots, commas, apostrophes and hyphens, allowing us to match sources like “Robert Downey, Jr.”, “Shaquille O’Neal”, or “Jean-Luc Godard”. Each sentence containing a matched phrase was stored together with the extracted source candidate.

We then tried to match each source candidate against the Wikidata list of individuals. This effectively reduced the number of candidate sentences from 11,452,615 to 96,731, but still included one-word source candidates like “Church” or “Hall”, caused by ambiguous nicknames or aliases of some persons stored in Wikidata, or incomplete data.

We decided to filter out such cases with a curated blacklist, which mainly concerned one-word aliases (of the 1,289 entries in the blacklist, only 62 consist of more than one word). After filtering the source candidates by help of our blacklist, we were down to 3,753 sentences from 20 years of New York Times containing possible VA.

While we cannot provide numbers for recall due to missing gold annotations for the VA phenomenon, we can present numbers for the precision of our extraction method. Compared to other stylistic devices, VA are a notably rare phenomenon – we were able to evaluate a candidate list of this magnitude manually.

The final result is a list of 2,775 sentences containing VA, making for a 73.9% precision score. (While these are all cases of valid VA, we have eliminated duplicates for the rankings in Section 3 if an expression reoccurred in a text or if an article was republished, trimming the list to 2,646 unique occurrences of VA.) Figure 1a shows these num-

---

2See https://www.mediawiki.org/wiki/Wikidata_Toolkit.
3See https://www.wikidata.org/wiki/Property:P31.
4See https://www.wikidata.org/wiki/Q5.
5E. g., https://www.wikidata.org/wiki/Q21508608.
6E. g., missing first names, see https://www.wikidata.org/wiki/Q5116540.
7The blacklist and all other data can be found in our working repository at https://github.com/weltliteratur/vossanto and via the project website at https://vossanto.weltliteratur.net/.
bers on a per-year basis (the corpus ends early in 2007, which explains the abrupt decrease in that year). The slight increase of the curves in Figure 1b (relative frequency) suggests that the use of VA increased over the years covered by our corpus. Since VA are typically built around a popular person, a paragon, it is unlikely that Wikipedia/Wikidata has significant gaps for the time covered by our corpus, but we have to at least consider that Wikipedia, and as a consequence, Wikidata, too, might have a contemporary bias.

3 Results

3.1 Most Frequent Sources

The 40 most frequent sources are shown in Table 1. They include artists, people from sports, business, the culture industry, historical figures, and scientists. While especially the upper segment of the table is dominated by (US-American) men, Madonna is the most frequent female paragon in our corpus, followed by Martha Stewart, Greta Garbo, Mother Teresa, Oprah Winfrey, Rosa Parks, Susan Lucci, and Julia Child.

The top source, Michael Jordan, is followed by Rodney Dangerfield, a stand-up comedian preeminently known for his catchphrase “I don’t get no respect!” He thus finally gets some respect, at least as a highly productive source for VA. Next in the ranking are Babe Ruth, an American baseball player, and Elvis Presley. Altogether, Table 1 assembles international celebrities, some of which are prone to appear beyond the New York Times. Yet, many references are not likely as sources for VA in other nations’ newspapers (just take all the baseball references), which confirms the notion that VA is a language- and culture-dependent phenomenon (Kövecses, 2005; Jäschke et al., 2017).

3.2 High-Frequent Modifiers

Looking at the most frequent modifiers, we find proof for one of the main functions of VA, which is inculturation (Holmqvist and Płuciennik, 2010). Next to unspecific modifiers such as “of his [or her] day”, “of his [or her] time”, etc., we can find toponyms such as Japan, China, and Brazil (Table 2). This becomes even clearer if we limit the output to just countries (Table 3).

The same can be said about other kinds of modifiers, such as sports (baseball, tennis, basketball, hockey, football). They can transfer meaning between different disciplines, like in this sentence: “It’s one thing to win a gold medal, but the way he [Andre Agassi] dominated, he was the Michael Johnson of the tennis field today.” (NYT, 1996) But given the right source, they can also inherit meaning from other domains, as shows this example: “[MBL pitcher Greg] Maddux is the Picasso of baseball.” (NYT, 2001)
3.3 Distribution of VA

We also gained insights into the distribution of VA across our corpus. VA as a phenomenon is particu-
larly popular in the cultural pages (“Arts”, “Books”, “Movies”, “Theater”, “Magazine”) and the “Sports” section, as indicated in Table 4. In contrast to that, authors in the “Business” section of the New York Times (the section containing the most articles in our corpus, see Table 5) clearly think twice before operating a VA expression into their texts. The popularity of VA especially on the cultural pages is confirmed by the example of German weekly Die Zeit studied in (Jäschke et al., 2017) where the paper’s “Feuilleton” section exhibits the highest frequency of VA. Our quantitative approach also backs up findings of qualitative research on antonomasia in the “Sports” section of Croatian media (Grgić and Nikolić, 2013).

3.4 Authors with a Thing for VA

One next step was to see if there are certain authors with a thing for VA. If so, we could assume that at least some authors would use VA very consciously being well aware of its stylistic impact. Table 6

Table 2: Most frequent modifiers.

| count | modifier    |
|-------|-------------|
| 55    | his day     |
| 33    | his time    |
| 29    | Japan       |
| 16    | tennis      |
| 16    | his generation |
| 16    | baseball    |
| 15    | China       |
| 13    | her time    |
| 13    | her day     |
| 12    | our time    |
| 11    | the 1990’s  |
| 10    | the Zulus   |
| 10    | the 90’s    |
| 10    | politics    |
| 10    | hockey      |
| 10    | Brazil      |
| 10    | basketball  |
| 10    | ballet      |
| 9     | jazz        |
| 9     | fashion     |
| 8     | today       |
| 8     | Israel      |
| 8     | his era     |
| 8     | hip-hop     |
| 8     | golf        |
| 8     | dance       |
| 7     | the art world |
| 7     | the 19th century |
| 7     | Long Island |
| 7     | Iran        |

Table 3: Selected countries serving as modifiers.

| count | country |
|-------|---------|
| 29    | Japan   |
| 15    | China   |
| 10    | Brazil  |
| 8     | Israel  |
| 7     | Iran    |
| 4     | South Africa |
| 4     | Mexico  |
| 3     | Spain   |
| 3     | South Korea |
| 3     | Russia  |
| 3     | Poland  |
| 3     | Pakistan |

(a) Absolute numbers

Figure 1: VA candidates and true VA extracted per year. The right vertical axes measure the resulting precision (percentage of true VA among the candidates).

(b) Relative frequency (one per thousand)
Table 4: Occurrences of VA by newspaper section. Categories are as extracted from corpus, this includes empty category fields.

| VA    | 2,646 | category          | articles  | 1,854,726 |
|-------|-------|-------------------|-----------|------------|
| 336   | 12.7% | Sports            | 160,888   | 8.7%       |
| 334   | 12.6% | Arts              | 88,460    | 4.8%       |
| 290   | 11.0% | New York and Region | 221,897  | 12.0%      |
| 237   | 9.0%  | Arts; Books       | 35,475    | 1.9%       |
| 158   | 6.0%  | Movies; Arts      | 27,759    | 1.5%       |
| 109   | 4.1%  | Business          | 291,982   | 15.7%      |
| 102   | 3.9%  | Opinion           | 134,428   | 7.2%       |
| 96    | 3.6%  | U.S.              | 89,389    | 4.8%       |
| 95    | 3.6%  | Magazine          | 11,464    | 0.6%       |
| 62    | 2.3%  | Style             | 65,071    | 3.5%       |
| 61    | 2.3%  | Arts; Theater     | 13,283    | 0.7%       |
| 46    | 1.7%  | World             | 79,786    | 4.3%       |
| 39    | 1.5%  | Home and Garden; Style | 13,978 | 0.8%       |
| 32    | 1.2%  | Travel            | 10,440    | 0.6%       |
| 31    | 1.2%  | Technology; Business | 23,283  | 1.3%       |
| 27    | 1.0%  |                    | 42,157    | 2.3%       |
| 25    | 0.9%  | Week in Review    | 17,107    | 0.9%       |
| 25    | 0.9%  | Home and Garden   | 5,546     | 0.3%       |
| 17    | 0.6%  | World; Washington | 24,817    | 1.3%       |
| 17    | 0.6%  | Style; Magazine   | 1,519     | 0.1%       |

Table 5: Number of articles by newspaper category.

| articles   | 1,854,726 |
|------------|-----------|
| categories | 1,580     |
| Business   | 291,982   |
| Sports     | 160,888   |
| Opinion    | 134,428   |
| U.S.       | 89,389    |
| Arts       | 88,460    |
| World      | 79,786    |
| Style      | 65,071    |
| Obituaries | 19,430    |
| Magazine   | 11,464    |
| Travel     | 10,440    |

presents some authors with an inclination to use VA.

3.5 Anti-VA

As we have shown in Table 1, Michael Jordan is a universal super ‘source’ for VA, a meme in its own right (Lovinger, 1999). The Wall Street Journal runs a website collecting all kinds of occurrences of “Michael Jordans of . . .”, counting 1,505 occurrences in different media to date (Cohen et al., 2015). On top of that, Barack Obama, in his honorific speech for Michael Jordan when awarding him the Medal of Freedom in November 2016, said the following: “There is a reason you call someone the Michael Jordan of [something]. They know what you’re talking about because Michael Jordan is the Michael Jordan of greatness. He is the definition of somebody so good at what they do that everybody recognizes it. That’s pretty rare.” (Payne, 2016)

Obama stretches VA past its breaking point by implementing recursivity, using an identical source and target (“Michael Jordan is the Michael Jordan of greatness”). Another way of doing that would be to use a non-modifying modifier, as has been acted out by German author Joachim Lottmann, who ironically referred to himself as “the German version of Rainald Goetz” (Seidl, 2004). Since Rainald Goetz, the ‘source’ for this VA, is a German author himself, the whole formulation stops making sense as the modifier wasn’t used to transfer the meaning into a different field, which lets this VA expression explode in irony. Another example for this sub-phenomenon will demonstrate that we are not talking about a single occurrence here: “I am [right now] in Norway, the Sweden of Scandinavia,” said Swiss satirist Gabriel Vetter in a recent interview, also leaving the modifier un-modified (Fopp, 2016).

We will conclude this parade of examples with
Table 6: Ranking of VA-embracing authors (absolute number of VA per author – percentage among all VA extracted – total number of articles per author – percentage among all articles in the corpus). Some author fields in the corpus were empty, so from the corpus itself we do not know for 15.5% of extracted VA who authored them.

| VA  | 2,646 | author articles | 1,854,726 |
|-----|-------|----------------|-----------|
| 411 | 15.5% | 961,052        | 51.8%     |
| 30  | 1.1%  | 5,098          | 0.3%      |
| 29  | 1.1%  | 2,874          | 0.2%      |
| 26  | 1.0%  | 2,739          | 0.1%      |
| 23  | 0.9%  | 3,140          | 0.2%      |
| 22  | 0.8%  | 717            | 0.0%      |
| 20  | 0.8%  | 2,661          | 0.1%      |
| 19  | 0.7%  | 1,647          | 0.1%      |
| 19  | 0.7%  | 1,704          | 0.1%      |
| 18  | 0.7%  | 1,515          | 0.1%      |
| 17  | 0.6%  | 568            | 0.0%      |
| 16  | 0.6%  | 4,090          | 0.2%      |
| 16  | 0.6%  | 4,544          | 0.2%      |
| 15  | 0.6%  | 2,497          | 0.1%      |
| 15  | 0.6%  | 817            | 0.0%      |
| 15  | 0.6%  | 1,368          | 0.1%      |
| 15  | 0.6%  | 2,188          | 0.1%      |
| 15  | 0.6%  | 2,735          | 0.1%      |
| 14  | 0.5%  | 1,437          | 0.1%      |
| 14  | 0.5%  | 2,492          | 0.1%      |

a third class of anti-VA, quoting a passage which was returned when processing the NYT corpus: “During their decade-long marriage, Elizabeth Taylor quipped that Burton was ‘the Frank Sinatra of Shakespeare.’” (NYT, 12 March 1989) The use of another human as modifier (here standing pars pro toto for Shakespeare’s oeuvre, too) is another attempt to wrap VA around itself. The same pattern is found in a well-known German example, Eckhard Henscheid’s satirical opera guide “Verdi ist der Mozart Wagners” (1979). The description of these three ways to direct VA against itself was only possible because our exploratory large-scale analysis sharpened our senses for these kinds of absurd antonomasia.

4 Summary and Outlook

We started this project with the notion that Vossian Antonomasia is a stylistic device with a simple structure. It turned out, however, that it is a far richer phenomenon than we thought. Writing extraction rules for this phenomenon is anything but trivial. By resorting to Wikidata, we circumvented shortcomings of available NER tools. By discussing the most frequent VA extracted from the New York Times corpus we added a quantifying approach to the discourse around paragons. We also presented first insights into the distribution of VA and found that certain journalistic genres have an inclination towards using this device, whereas others avoid it (“Sports” and “Arts” sections vs. “Business”). We also confirmed that VA is a language- and culture-dependent phenomenon, much like other stylistic devices (Kövecses, 2005).

Our approach comes with known limitations. For one thing, we have focused our attention on just one grammatical variant of VA. Although our pragmatic limitation to the pattern “the . . . of” aimed at the most common shape of VA, we thereby missed expressions like “the American Oscar Wilde”, “Harlem’s Mozart”, or “the Japanese Nolan Ryan” (all taken from the NYT corpus), which certainly leaves room for further work. The same can be said about our limitation to persons as source entities (living or dead, but in any case non-fictional and non-mythological). While working with our corpus we made some observations, though, that we don’t want to withhold:

- Productive ‘sources’ for VA can be real persons as well as mythological or fictional characters (productive examples for the latter include Woody Allen’s “Zelig”, Cervantes’ “Don
Quixote” or James Hilton’s “Mr. Chips”, but also sources like “the [Santa Claus, Satan, Midas, Godzilla, King Kong, Cinderella, Pied Piper, Robin Hood, Energizer Bunny, Cassandra, Jupiter, Icarus] of –”.

- Animals can be targets, even if the source is a human, like in “Sea Hero is the Bobo Holloman of racing” (NYT, 1993) or “Bonfire, the Michael Jordan of dressage horses” (NYT, 1998).

- Personifications can be part of a VA expression, that is, the use of individual persons/characters as sources for companies, clubs, bands or places as ‘targets’, like in “Sturm, Ruger is the Benedict Arnold of the gun industry” (NYT, 1989), “Aerosmith, the Dorian Gray of rock bands” (NYT, 1993), “the Hudson has been the John Barrymore of rivers, noble in profile but a sorry wreck” (NYT, 1996), “the National Collegiate Athletic Association, the Kenneth Starr of sports” (NYT, 1998).

In conclusion, we would like to propose more holistic approaches to the phenomenon rather than just building on Lakoff’s ‘paragons’, that is, the ‘sources’. With quantitative explorations of Vossian Antonomasie like the one presented here, we can capture the phenomenon as a whole, including source, target and (if provided) modifier. We will find a network of people, connected by a diversity of modifiers, with nodes that can be sources or targets (or both, as shown by the Obama example), a network that can help us understand the hidden patterns of role models and how they differ from country to country, from language to language.

5 Acknowledgement

We would like to thank the Information School at the University of Sheffield for granting one of the authors (F. F.) a visiting scholarship in spring 2017 which helped to finish our research.

References

Angelika Bergien. 2013. Names as frames in current-day media discourse. In Name and Naming. Proceedings of the second international conference on onomastics, pages 19–27, Cluj-Napoca. Editura Mega.

Steven Bird, Ewan Klein, and Edward Loper. 2009. Natural Language Processing with Python, 1st edition. O’Reilly Media, Inc.

Ben Cohen, Geoff Foster, Matthew Oshinsky, Jon Keegan, and Tom McGinty. 2015. The Michael Jordan of . . . . Website.

Frank Fischer and Joseph Wälzholz. 2014. Jeder kann Napoleon sein. Vossianische Antonomasie: Eine Stilkunde. Frankfurter Allgemeine Sonntagszeitung, page 34. Dec 21, 2014.

Andrea Fopp. 2016. “Die Wurst ist Geheimnis, ist Gefahr, ist Geisterbahn” [interview with Gabriel Vetter]. TagesWoche. Sep 1, 2016.

Ana Grgić and Davor Nikolić. 2013. The Cowboys, the Poets, the Professor . . . – Antonomasia in Croatian Sports Discourse. In G. Kišiček and I.Z. Žagar, editors, What Do We Know about the World? Rhetorical and Argumentative Perspectives, pages 408–428. University of Windsor; Pedagoški inštitut, Windsor; Ljubljana.

Kenneth Holmqvist and Jaroslav Plucennik. 2010. Princess Antonomasia and the Truth: Two Types of Metonymic Relations. In Armin Burkhardt and Brigitte Nerlich, editors, Tropical Truth(s). The Epistemology of Metaphor and Other Tropes, pages 373–381. De Gruyter, Berlin/New York.

Robert Jäschke, Jannik Strötgen, Elena Krotova, and Frank Fischer. 2017. “Der Helmut Kohl unter den Brotaufstrichen”. Zur Extraktion Vossianischer Antonomasien aus großen Zeitungskorpora. In Proceedings of DHd 2017, DHd ’17, pages 120–124. Digital Humanities im deutschsprachigen Raum.

Zoltán Kövecses. 2005. Metaphor in Culture: Universality and Variation. Cambridge University Press, Cambridge.

George Lakoff. 1987. Women, Fire, and Dangerous Things. What Categories Reveal about the Mind. The University of Chicago Press, Chicago.

Heinrich Lausberg. 1960. Handbuch der literarischen Rhetorik. Eine Grundlegung der Literaturwissenschaft. Hueber, München.

Caitlin Lovinger. 1999. They’re Sort of Like Mike. New York Times. Jan 17, 1999.

Marissa Payne. 2016. Obama tells Michael Jordan he’s ‘more than just an Internet meme,’ then makes MJ tear up. Washington Post. Nov 22, 2016.

Evan Sandhaus. 2008. The New York Times Annotated Corpus LDC2008T19. DVD, Linguistic Data Consortium, Philadelphia.

Claudius Seidl. 2004. Jugend ohne Sex. Ein Manifest gegen die Wirklichkeit: Joachim Lottmanns Roman “Die Jugend von heute”. Frankfurter Allgemeine Sonntagszeitung, page 34. Oct 17, 2004.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A Free Collaborative Knowledgebase. Communications of the ACM, 57(10):78–85.