

„Der Helmut Kohl unter den Brotaufstrichen“. Zur Extraktion vossianischer Antonomasien aus großen Zeitungskorpora

Jäschke, Robert

r.jaschke@sheffield.ac.uk
University of Sheffield

Strötgen, Jannik

jannik.stroetgen@mpi-inf.mpg.de
Max-Planck-Institut für Informatik, Saarbrücken

Krotova, Elena

kroelebor@gmail.com
Higher School of Economics, Moskau

Fischer, Frank

frafis@gmail.com
Higher School of Economics, Moskau

Einführung und Forschungslage

Wenn Peter Paul Rubens als »Tarantino des Barock« beschrieben wird (im *Tagesspiegel*, 2014) oder Alice Schwarzer als der »Erich Honecker des Feminismus« (in *Cicero*, 2014), dann handelt es sich um eine Vossianische Antonomasie. Diese Trope ist nach dem niederländischen Humanisten und Rhetoriklehrer Vossius benannt (und wird im Folgenden als »Vossanto« abgekürzt, in Anlehnung an den Vorschlag von Fischer/Wälzholz 2014). Generell spricht man von Antonomasie, wenn eine bestimmte Eigenschaft einer Person für diese selbst steht (z. B. »der Leimener« für Boris Becker). Beim Spezialfall der Vossanto wird einer Person über die Nennung einer anderen (bekannteren, populäreren, berühmteren) Person als Referenzgröße eine bestimmte Eigenschaft zugeschrieben. Dabei sorgt ein »untypologisches, aktualisierendes Signal« (Lausberg 1960) für den Bedeutungstransfer (in den oben genannten Beispielen wären dies der Barock und der Feminismus). Anders ausgedrückt: Die Vossanto stellt über einen »modifier« einen Zusammenhang zwischen »source« und »target« her (Bergien 2013). Entitäten können sowohl als »source« als auch als »target« auftreten, wie ebd. am Beispiel Obama demonstriert: bis 2011 trat er in Vossantos vor allem als »target« auf, danach diente er immer mehr als »source«. Die »source«-Referenz wird im Fachdiskurs im Anschluss an Lakoff 1987 auch als

»paragon« bezeichnet (»a specific example that comes close to embodying the qualities of the ideal«, ebd.).

Der Begriff »Vossianische Antonomasie« wird international kaum verwendet, stattdessen wird etwa zwischen »Antonomasia1« und »Antonomasia2« unterschieden: »metonymic« vs. »metaphorical antonomasia« (Holmqvist/P#uciennik 2010). Innerhalb dieses Klassifikationsschemas wäre unsere Vossanto ein Spezialfall von »Antonomasia2«, nämlich wenn es um »comparisons with paragons from other spheres of culture« geht: »Lyotard is a pope of postmodernism, Bush is no Demosthenes; and we can buy the Cadillac of vacuum cleaners.« (ebd.)

Dieses Stilmittel, dessen reger Gebrauch seit der Antike belegt ist, ist heute medial ubiquitär anzutreffen. Oft findet es sich schon in Überschriften, da es zugleich informativ und rätselhaft sein kann und zudem oft unterhaltsame Qualitäten bietet. Eine eigene größere Sammlung an Musterexemplaren (<http://www.umblaetterer.de/datenzentrum/vossianische-antonomasien.html>) gab den Ausschlag, dieses Phänomen systematisch zu erforschen, mit historischer Perspektive und auf Grundlage größerer englischer und deutscher Zeitungskorpora. Ziel dieser Arbeit ist eine erste methodisch-explorative Analyse des Phänomens Vossanto in der Tageszeitung *New York Times* (1987–2007) und der Wochenzeitung *Die Zeit* (1995–2011). Die Korpora wurden aufgrund ihrer Verfügbarkeit, Bedeutung und ihres Umfangs gewählt. Die Extraktion der Vossantos erfolgte jeweils korpuspezifisch, um den verschiedenen Formaten und Sprachen Rechnung zu tragen.

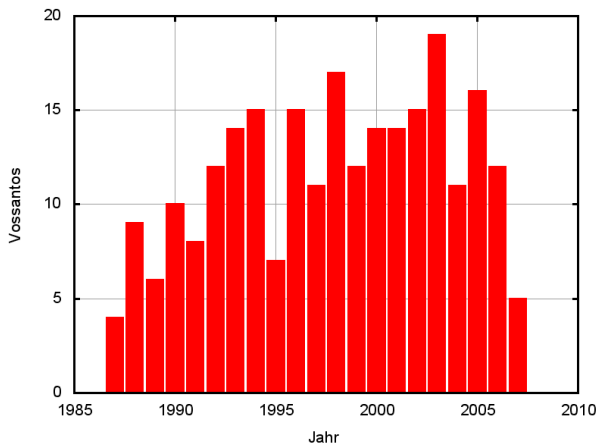
Englischsprachiges Korpus

Aus den XML-Daten des *New York Times*-Korpus (Sandhaus 2008) wurde für jeden der 1.854.726 Artikel der Volltext extrahiert. Anschließend wurde der Text mit Hilfe des NLTK (Bird/Loper/Klein 2009) in Sätze zerlegt und die Wörter jedes Satzes mit dem Part-of-Speech-Tagger des NLTK annotiert. Zusätzlich wurden Named Entities (Personen, Orte, Organisationen) mit dem NLTK-eigenen Named-Entity-Extraktor annotiert. Die so annotierten Sätze wurden mit einer Liste von Vossantotypischen Mustern (in Form von regulären Ausdrücken) abgeglichen. Eine vereinfachte Darstellung eines solchen Musters ist beispielsweise:

```
\((PERSON|ORGANIZATION|GPE) *\\) (is|has)
(often|sometimes)? (been)? (called)? the \((PERSON|
ORGANIZATION|GPE) *\\) (of|among|from)
\((PERSON|ORGANIZATION|GPE) *\)
```

Die zu findenden drei Elemente einer Vossanto sind darin durch * gekennzeichnet. Passte ein Satz auf eines der Muster, so wurden diese drei Teile extrahiert und tabellarisch ausgegeben. Anschließend wurden die extrahierten 10.744 Kandidaten manuell überprüft. Nicht-Vossantos und Vossantos mit Städten und Firmen wurden

entfernt (Treffer der Art »Algarve, the Riviera of Portugal« oder »Pepsi is the Nike of soft drinks« sind eine eigene Untersuchung wert) und der Fokus auf Vossantos gelegt, in denen Individuen (Personen, Tiere, fiktive Figuren) als ›source‹ oder ›target‹ dienen. 246 Vossantos blieben dabei übrig (Übersicht in unserem Arbeitsrepo, siehe Bibliografie), die sich wie folgt über das Korpus verteilen:



vossantos_nyt.png

Am häufigsten als ›source‹ verwendet wurden folgende Namen:

Anzahl	source
6	Michael Jordan
5	Michelangelo
4	Babe Ruth
3	Zelig
3	Rodney Dangerfield
3	Neil Young
3	Elvis
3	Don Quixote

Als Beispiel für Treffer seien diejenigen für Michael Jordan genannt:

- »Romario is the *Michael Jordan* of soccer and Bebeto is the Magic Johnson of soccer« (1994)
- »Bonfire, the *Michael Jordan* of dressage horses« (1998)
- »Brian Foster, the *Michael Jordan* of BMX racing« (1998)
- »The stunt biker Dave Mirra, the *Michael Jordan* of the dirt set« (2000)
- »Cynthia Cooper is the *Michael Jordan*, the Larry Bird, the Magic Johnson of this league« (2000)
- »McNabb has been called the *Michael Jordan* of the National Football League« (2001)

Trotz der zeitlichen Einschränkung des Korpus lassen sich bereits einige vielversprechende Beobachtungen anstellen und Thesen bilden: 1. Produktive Referenzgrößen einer Vossanto sind sowohl reale als auch fiktionale Figuren (Bsp. für letztere aus der obigen Liste: Woody Allens »Zelig«, Cervantes' »Don Quixote«). 2. Öffentliche Personen oder bekannte fiktionale Charaktere haben bestimmte Eigenschaften, die sie für die Verwendung als Referenzgröße einer Vossanto prädestinieren oder nicht (es bleibt etwa zu erforschen, warum gerade Michael Jordan und Michelangelo sich so gut eignen und nicht andere Sportler bzw. Künstler). 3. Es gibt historisch stabile Referenzgrößen, deren Bekanntheit vorausgesetzt werden kann (z. B. Michelangelo), und es gibt ephemere Referenzgrößen, die ab irgendeinem Zeitpunkt nicht mehr als Bezugspunkt taugen (für das benutzte zeitgenössische Korpus eher noch nicht relevant).

Deutschsprachiges Korpus

Das deutsche Datenset besteht aus einer Sammlung des Archivs der Wochenzeitung *Die Zeit* und enthält die Artikel aus den Jahren 1995 bis 2011. Insgesamt umfasst das Korpus 126.702 Dokumente.

Zunächst wurden die Volltexte (inklusive Überschriften) aller Dokumente extrahiert. Diese wurden dann mit Hilfe des Part-of-Speech-Taggers und Named-Entity-Recognition-Tools des Stanford CoreNLP Package verarbeitet. Für die Analyse deutschsprachiger Texte enthält Stanford CoreNLP speziell für das Deutsche trainierte Modelle (Faruqui und Pado 2010). Somit können alle Texte auf drei Ebenen untersucht werden: auf der Wortebene, der Part-of-Speech-Ebene sowie der Named-Entity-Ebene. Mithilfe von regulären Ausdrücken, die auf den verschiedenen Ebenen angewandt werden können, wurde dann nach Vossanto-Mustern gesucht. Im Gegensatz zur Verarbeitung des englischsprachigen Korpus wurde jedoch noch nicht versucht, auch das ›target‹ einer Vossanto zu extrahieren. Stattdessen wurden Muster entworfen, die das ›source‹-Objekt sowie das »aktualisierende Signal« matchen. Ausschlaggebend für diese Herangehensweise waren die in einem Testdurchlauf beobachtete hohe Anzahl an Vossantos ohne unmittelbaren Verweis auf das ›target‹ sowie eine große Vielfalt an möglichen Formulierungen, die auf die Relation zum ›target‹ hinweisen können. Mithilfe relativ strikter Regeln konnte die Anzahl an falschen Extraktionen im Rahmen gehalten werden. Ein vereinfachtes Beispiel für eine Extraktionsregel lautet etwa: »eine Art PERSON (der/des) (ADJECTIVE)? NOUN«.

Die Produktivität der beiden häufigsten Referenznamen des NYT-Korpus bestätigt sich im verwendeten deutschen Korpus, etwa wenn vom »Michael Jordan der analytischen Philosophie« die Rede ist (*Die Zeit* 44/1999) oder vom »bulgarischen Michelangelo« (*Die Zeit* 14/2001). Ansonsten scheint es sprachen- bzw. kulturspezifische Präferenzen zu geben. Die häufigsten ›sources‹ sind:

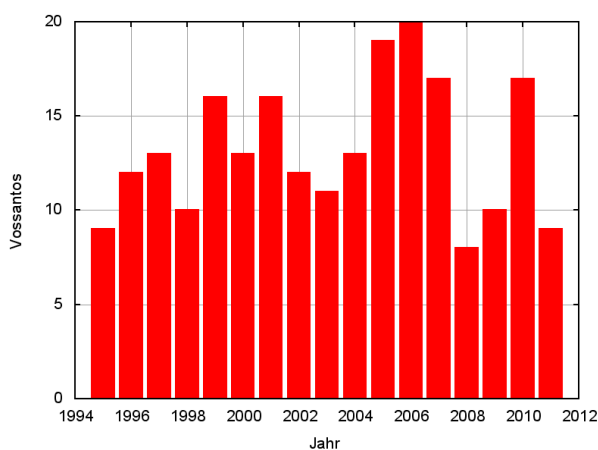
Anzahl	source
9	Robin Hood
6	Bill Gates
4	Franz Beckenbauer
3	Daniel Düsentrieb
3	Heinz Rühmann
3	James Dean
3	Jesus Christus
3	Norbert Blüm
3	Willy Brandt

Ähnlich wie im NYT-Korpus ist erkennbar, wie stark typisierend mythische bzw. fiktive Figuren sind (Robin Hood, Daniel Düsentrieb). Daneben zeigt sich, dass »Bill Gates«, der im NYT-Korpus nur zweimal als »source« einer Vossanto vorkommt, im *Zeit*-Korpus sechs Mal als Referenz vertreten ist:

- »eine Art Bill Gates des Stolperns« (1998)
- »Der Bill Gates von Aurich« (2001)
- »der Bill Gates von Ostfriesland« (2001)
- »der Bill Gates von Aurich« (2002)
- »der britische Bill Gates« (2008)
- »der Bill Gates von Estland« (2010)

Die wiederholte Verwendung des »Bill Gates von Aurich« zeigt, wie stark ein »target« mit einer »source« verwachsen kann. (Paradebeispiel hierfür ist im Übrigen Vittorio Hösle, »der Boris Becker der Philosophie«, eine Bezeichnung, die es bis in den Wikipedia-Artikel zu Hösle geschafft hat.) Am Beispiel Bill Gates' lässt sich wie zuvor am Beispiel Obama demonstrieren, dass ein Name sowohl als »target« als auch als »source« vorkommen kann. Bevor Bill Gates selbst als Referenz verwendet wird, wird er in einem Artikel von 1995 noch durch eine andere Person beschrieben: »Bill Gates ist der Henry Ford des Computerzeitalters«.

Insgesamt wurden aus 1.456 Vossanto-Kandidaten 225 manuell als Vossantos markiert, die sich wie folgt über die im Korpus vorhandenen Jahre verteilen:



vossantos_zeit.png

Zu den fälschlich extrahierten Named Entities gehören »der Berliner Klaus Wowereit«, »der deutsche Michel« oder »der Anton aus Tirol«, stehende Wendungen, die grammatisch unseren definierten Vossanto-Mustern entsprechen.

Erkenntnisse und Ausblick

Die Vossanto ist als Stilmittel nur scheinbar einfach strukturiert, das Erstellen von Extraktionsregeln daher alles andere als trivial. Die vorliegenden Skripte weisen bekannte Lücken auf, die Qualität hängt v. a. von der Verlässlichkeit der benutzten NER-Tools und der Präzision der definierten Muster ab. Fehlende Goldannotationen für dieses Phänomen erschweren zudem eine Evaluierung. Die vorliegende Arbeit hat daher explorativen Charakter, die Optimierung von Precision und Recall lag noch nicht in deren Fokus, ist aber das nächste Ziel dieses Projekts.

Trotz der genannten Einschränkungen konnten durch diesen korpusbasierten Ansatz neue Erkenntnisse zur Vielgestaltigkeit des Phänomens »Vossianische Antonomasie« gewonnen werden. So lassen sich zahlreiche Spezialfälle unterscheiden und systematisch untersuchen (vgl. auch Fischer/Wälzholz 2014), beispielhaft genannt seien:

- Tiere als »target« (» *Sea Hero* is the Bobo Holloman of racing«, NYT, 1993; » *Bonfire*, the Michael Jordan of dressage horses«, NYT, 1998),
- Feminisierungen (Adele Schopenhauer, »eine Art *Donna Quichotta* des Weimarer Museenvereins«, *Die Zeit* 18/2002; »Tracey [Emin], die *Donna Giovanna* der britischen Gegenwartskunst«, *Die Zeit* 9/2006; »Kati Witt ist jetzt eine *Franziska Beckenbauer* der Münchner Olympiabewerbung.«, *Die Zeit* 39/2010),
- nicht individualisierbare »sources«: »the [God, King, Queen, Satan, Emperor, Oracle, Shogun, Czar, Sultan, Buddha] of«,
- mythologische und fiktive Figuren als »sources«: »the [Santa Claus, Midas, Godzilla, Pied Piper, Energizer Bunny, Jupiter, Icarus] of«,
- Personifizierungen, also der Einsatz individueller Personen/Figuren als »source« für Firmen, Vereine, Bands oder Orte als »target« (» *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).

Zu letzteren Beispielen gehört nun endlich auch der titelgebende »Helmut Kohl unter den Brotaufstrichen« (*der Freitag* 35/2011).

Auch zur Distribution der Vossantos innerhalb der beiden Zeitungskorpora ließen sich belastbare Ergebnisse gewinnen. Demnach sind Vossantos besonders im Kultur- und Sport-Ressort beliebt (Vorkommen in der Sektion »Arts« der NYT: 78; in der Sektion »Sports«: 57; auf dem nächsten Rang mit großem Abstand »New York and Region«: 28 – im »Feuilleton + Literatur«-Ressort der Zeit: 76, »Politik«: 54, nächstrangig ist weit entfernt »Wirtschaft« mit 23 Vorkommen; »Sport« hat hier keine Treffer, denn die gedruckte *Zeit* hat kein dediziertes Sport-Ressort).

Bibliographie

Bergien, Angelika (2013): „Names as frames in current-day media discourse“, in: Felecan, Oliviu (ed.): *Name and Naming*. Proceedings of the second international conference on onomastics. Cluj-Napoca: Editura Mega 2013: 19–27.

Bird, Steven / Loper, Edward / Klein, Ewan (2009): *Natural Language Processing with Python*. O’Reilly Media Inc.

Faruqui, Manaal / Pado, Sebastian (2010): „Training and Evaluating a German Named Entity Recognizer with Semantic Generalization“, in: *Proceedings of Konvens 2010*.

Fischer, Frank / Wälzholz, Joseph (2014): „Jeder kann Napoleon sein: Vossianische Antonomasie: Eine Stilkunde“, in: *Frankfurter Allgemeine Sonntagszeitung* 51 (21. Dezember 2014): 34 http://www.umblaetterer.de/wp-content/uploads/2014/12/vossanto_fas.png .

Holmqvist Kenneth / P#uciennik Jaros#aw (2010): „Princess antonomasia and the truth: Two types of metonymic relations“, in: Burkhardt, Armin / Nerlich, Brigitte (eds.): *Tropical Truth(s): The Epistemology of Metaphor and Other Tropes*. Berlin/New York: De Gruyter 373–381 10.1515/9783110230215.

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

Lausberg, Heinrich (1960): *Handbuch der literarischen Rhetorik. Eine Grundlegung der Literaturwissenschaft 2*. München: Hueber.

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

Arbeitsrepositorium: <https://github.com/weltliteratur/vossanto>

Folien zum Vortrag: <https://lehkost.github.io/slides/2017-bern/>