

```
:A104.1 a      :A104 , :Motif , owl:NamedIndividual ;
  rdfs:comment  "\"Terminal motif A104.1\""@en ;|
  rdfs:label    "\"Living person becomes god.\""@en .
```

Und die Klasse, von der A104.1 eine Instanz ist, wird als Subklasse von A kodiert:

```
:A104 a      owl:Class ;
  rdfs:comment  "\"Index A104 of TMI\""@en ;
  rdfs:label    "\"The making of gods.\""@en ;
  rdfs:subClassOf :A .

:A a      owl:Class ;
  rdfs:comment  "\"Index A of TMI\""@en ;
  rdfs:label    "\"MYTHOLOGICAL MOTIFS.\""@en ;
  rdfs:subClassOf :TMI .
```

Aktuelle Arbeit besteht darin, links zu einem anderen Referenzwerk in der Folkloristik zu etablieren: Die Arne-Thompson-Uther (ATU, s. Uther 2004) Taxonomie von Märchentypen. Eine erste Ontologie Version von ATU ist von uns bereitgestellt worden, aber muss noch überprüft werden. Links von ATU zu TMI werden dann mit Hilfe von symmetrischen „Object Properties“ beschrieben. Wir arbeiten auch daran, die „labels“ der TMI Klassen und Instanzen in anderen Sprachen (Deutsch und Französisch) zu erweitern.

## Notes

1. OWL steht für „Ontology Web Language“. Siehe auch <http://www.w3.org/TR/owl-features/>
2. Siehe <http://www.ruthenia.ru/folklore/thompson/index.htm>
3. Siehe Abbildung 1 weiter unten, für einen Screenshot des von uns verwendeten Ontologie Editors “TopBraid” (cf. TopQuadrant).

## Bibliographie

**Eisenreich, Christian / Ott, Jana / Süßdorf, Tonio / Willms / Declerck, Thierry** (2014): “From Tale to Speech: Ontology-based Emotion and Dialogue Annotation of Fairy Tales with a TTS Output”, in: *Proceedings of ISWC 2014, Riva del Garda, Italy*, Springer.

**Declerck, Thierry / Lendvai, Piroska / Darányi, Sándor** (2012): “Multilingual and Semantic Extension of Folk Tale Catalogues”, in: Jan Christoph Meister (ed.): *Digital Humanities 2012. Conference Abstracts*, Hamburg, Germany. Hamburg: Hamburg University Press.

**TopQuadrant: TopBraid Live** <http://www.topquadrant.com/products/topbraid-live/> [letzter Zugriff 11. Februar 2016].

**Stith Thompson** (1955-1958): *Motif-index of folk-literature*. A classification of narrative elements in

folktales, ballads, myths, fables, medieval romances, exempla, fabliaux, jest-books, and local legends. Revised and enlarged edition. Bloomington: Indiana University Press.

**Uther, Hans-Jörg** (2004): *The Types of International Folktales. A Classification and Bibliography*. Based on the system of Antti Arne and Stith Thompson (= FF Communications 284–286). Helsinki: Suomalainen Tiedekatemia.

**W3C = World Wide Web Consortium** (2004): *OWL Web Ontology Language* <http://www.w3.org/TR/owl-features/> [letzter Zugriff 16. Februar 2016].

## Metaphern digital Auf dem Weg von der Annotation zur automatischen Detektion

### Do Dinh, Erik-Lân

dodinh@kds1.informatik.tu-darmstadt.de  
UKP Lab, TU Darmstadt

### Gerloff, Malte

gerloff@kds1.informatik.tu-darmstadt.de  
Institut für Philosophie, TU Darmstadt

### Núñez, Alexandra

nunez@kds1.informatik.tu-darmstadt.de  
Institut für Sprach- und Literaturwissenschaft, TU Darmstadt

Das interdisziplinäre Forschungsteam *Natur & Staat* hat sich zum Ziel gesetzt, ein innovatives Computerprogramm für die (semi-)automatische Metapherndetektion zu entwickeln. Der didaktische Nutzen und wissenschaftliche Mehrwert des Tools für die geisteswissenschaftliche Forschung und Lehre lassen sich folgendermaßen umreißen: Synchrone und diachrone Textanalysen des kontextuellen Metapherngebrauchs können z. B. bei vorliegenden großen Textkorpora rascher durchgeführt und erste wissenschaftliche Hypothesen evaluiert und modifiziert werden. Es können zudem auch textsortenübergreifende Analysen qualitativer bis quantitativer Art durchgeführt werden. Des Weiteren soll das Tool sowohl auf verschiedene Metaphertheorien als auch auf die jeweiligen Forschungsabsichten anpassbar sein. Der Fokus des interdisziplinären Vortrags liegt auf der geisteswissenschaftlichen Methodik, der didaktischen Vermittlung der statistischen Modellierung des Tools und insbesondere auf den bis jetzt erzielten Zwischenergebnissen unserer Arbeit.

## Der Untersuchungsgegenstand

Den namensgebenden Untersuchungsgegenstand bildet das neunbändige Volltextkorpus, das unter dem programmatischen Titel „Natur und Staat. Beiträge zur naturwissenschaftlichen Gesellschaftslehre“ (im Folgenden: *Natur & Staat*) u. a. von dem Zoologen Ernst Haeckel herausgegeben und im Zeitraum von 1903-1911 publiziert wurde. Die Bände liegen als Volltextdigitalisate in Frakturschrift vor.

Vorausgegangen war ein national und international stark rezipiertes Preisausschreiben, das 1900-1901 ausgeschrieben und vom bekannten Großindustriellen Friedrich Alfred Krupp (1854-1902) anonym finanziert wurde. Die Preisfrage lautete: „Was lernen wir aus den Prinzipien der Descendenztheorie in Beziehung auf die innerpolitische Entwicklung und Gesetzgebung der Staaten?“. Diskursanalytisch betrachtet, ist das Korpus u. a. eine Antwort auf die in verschiedenen wissenschaftlichen Disziplinen vorangegangene Rezeption der Theorien Charles Darwins im 19. Jahrhundert. Zusammen mit der sozialdarwinistischen Bewegung wurde ein neues Deutungsmodell des Menschen und seiner Stellung in der Welt vorbereitet und schließlich etabliert.

*Natur & Staat* bildet besonders aufgrund der Textsorte (vgl. Fix 2011) einen geeigneten Ausgangspunkt für die Metapherndetektion: Es handelt es sich um (populär-)wissenschaftliche Texte (vgl. Polenz 1999, 1981), die primär das Ziel verfolgen, Sachverhalte, Ideen, Theoreme auf der Basis einer sozialdarwinistischen Agenda zu reflektieren, argumentativ aufzubereiten und einem breiten Adressatenkreis zu vermitteln. Das Preisausschreiben öffnete den Weg für einen tiefgreifenden Wandel „ethisch“ genannter, auf eine Einflussnahme kollektiver „Entwicklungen“ abzielender Handlungsmaximen. Die Abhandlungen entwarfen Szenarien einer sozialdarwinistischen Governance von Bildung: einer Sitten- und Wertepolitik für das „Leben“ – mitsamt biotechnischen und eugenischen Implikationen (vgl. Gehring 2009).

## Interaktionale Metaphernmodelle

Innerhalb des aufgezeigten Kontextes sollen Ausdrucksgestalt und semantische Funktion von sprachlichen Metaphern analysiert werden. Interaktionale Modelle mit ihrer binären Übertragungsstruktur fokussieren u. a. konventionalisierte, ubiquitäre Alltagsmetaphern. Eine wissenschaftliche Herausforderung in diesem Forschungsparadigma bilden jedoch die kühnen Metaphern.

## Kühne Metapherntheorien

Metapherntheorien, die die kühne Metapher ins Auge fassen, verorten sich innerhalb des hermeneutischen Paradigmas und legen das Primat auf das Besondere.

Nur Metaphern, die besonders seien, seien laut Max Black (1954) für die Philosophie relevant, da nur sie eine erkenntnistheoretische Funktion hätten. Des Weiteren stellt er in der Folge von I. A. Richards (1936) fest, dass die Metapher zwei Bestandteile aufweise: Fokus und Rahmen. Beide bestünden mindestens aus einem Wort und sie seien interaktional, derart, dass ein System von Implikationen, welches dem Rahmen unterliege, auf das Implikationssystem des Fokus‘ rückbezüglich wirke. Die Übertragung der Implikationen geschehe auf der Basis von Ähnlichkeiten. Sie speise sich somit aus der immanenten Semantik des Implikationssystems, welches durch statistische Verfahren z. B. über selektionale Präferenzen für den Computer abbildbar gemacht werden kann.

Gehring (2011) erweiterte dann Blacks Interaktionstheorie, insofern sie den Kontextbruch als eine notwendige Bedingung der Metapher einführt, da aus diesem und dem Fehlen des wörtlichen Sinns der Interaktion die Metapher entstehe. Da aber nicht jede semantische Übertragung eine Metapher sei, weil man ansonsten in die Beliebigkeit abdriften würde, ist die Interaktion der beiden Entitäten, auch eine Interaktion der besonderen Art. Die Größe des Rahmens respektive des Kontexts der Metapher ist sowohl bei Gehring als auch bei Black variabel, könne allerdings durch eine Weglassprobe evaluiert werden. Einigkeit besteht überdies auch darüber, dass die Metapher nur in der Gesamtheit von Fokus und Rahmen bestehe. Weder hinreichende noch notwendige Bedingungen der Metapher seien hingegen, laut Gehring (2011), sowohl Bildlichkeit als auch lexikalische sowie grammatikalische Indikatoren, weil die Metapher deutungsoffen sei; auch schließt sie aus, dass es ein Kontinuum zwischen Begriff und Metapher gebe.

## Ubiquitäre Metapherntheorie

*Metapher* bezeichnet im Rahmen der konzeptuellen Metapherntheorie (Lakoff / Johnson 1980; Lakoff 1993; Goatly 2007; Kövecses 2015) zunächst ein zentrales kognitives Vermögen. Die Kernthese der konzeptuellen Metapherntheorie bildet die Annahme, dass eine unbekannte Erfahrung ( *target domain*) in Analogie zu einer bereits bekannten Erfahrung ( *source domain*) sprachlich konzeptualisiert wird. Dies betrifft beispielsweise vage Konzepte wie Emotionen (Kövecses 2003), Theorien und andere abstrakte Sachverhalte, Relationen (Johnson 1987) und Prozesse (Núñez 2014). Diese metaphorischen Übertragungsmuster lassen sich auch im Textkorpus *Natur & Staat* indexikalisch auf der Sprachoberfläche mit dem Fokus auf Lexemen, z. B. Genitivkonstruktionen ([NP] der/ des [NP]), und weiteren usuellen Konstruktionen korpuslinguistisch eruieren und annotieren. Die Verteilung der metaphorischen Sprachphänomene ist dabei im Vergleich zu den bereits erwähnten kühnen Metaphern

als ubiquitär (Paul 1909, Bühler 1934, Paprotte / Dirven 1985) einzustufen. Die entlehnten konzeptuellen und sprachlich umgesetzten Domänen können schließlich in einem zweiten Schritt hinsichtlich regelmäßiger Konzeptübertragungen zwischen den beiden Domänen systematisiert werden.

## Forschungslage

Seit einigen Jahren lässt sich eine verstärkte Entwicklung (zumeist überwachter) automatisierter Verfahren für die Identifikation und Interpretation von Metaphern beobachten. Diese Verfahren nutzen überwiegend als Grundlage für die Modellierung des Untersuchungsgegenstandes *Metapher* die bereits vorgestellte konzeptuelle Metaphertheorie von Lakoff und Johnson (1980). Dabei gibt es Unterschiede in der Zielsetzung einzelner Verfahren: Während einige Ansätze lediglich die metaphorische Verwendung bestimmter Konstruktionen (z. B. Subjekt-Verb-Objekt oder Adjektiv-Nomen) bewerten (Turney et al. 2011; Tsvetkov et al. 2013; Shutova 2013), weiten andere Verfahren eine solche Klassifizierung auf alle Inhaltswörter aus (Beigmann Klebanov et al. 2014; Dunn 2013). Darüber hinaus existieren Unterschiede in der Tiefe der Analyse. So wird bei einem Großteil der Verfahren die konzeptuelle Ebene ausgeblendet und Metaphern werden ausschließlich als Realisierungen auf der Sprachoberfläche identifiziert, wenn auch mit Methoden, die sich auf das Vorhandensein einer konzeptuellen Ebene stützen oder diese voraussetzen. Wenige Verfahren versuchen, metaphorische Abbildungen auf der konzeptuellen Ebene zu erkennen (Mason 2004; Shutova et al. 2013).

Neben der Wahl der untersuchten Metaphertheorie und der Ebene, auf der Metaphern erkannt werden, ist ein weiteres Merkmal bestehender automatischer Verfahren die Sprache der behandelten Texte. Diese ist – nicht zuletzt aus praktischen Gründen, wie dem Vorhandensein annotierter Korpora und weiterer Ressourcen wie vordefinierten Datenbanken zu Abstraktheits- und Konkretheitsbewertungen von Wörtern – üblicherweise Englisch. Für einige Verfahren existieren überwachte maschinelle Lernverfahren, die es ermöglichen, auf annotierten, englischen Daten ein Modell zu trainieren, welches durch zweisprachige Wörterbücher für die automatische Identifikation auf Texten anderer Sprachen anwendbar ist (Tsvetkov et al. 2013).

## Annotationstool, Vorgehen und Ziele

Für die Annotation sowohl kühner als auch ubiquitärer Metaphern verwenden wir zunächst WebAnno (Yimam et al. 2013), ein Web-Annotationsprogramm für mehrere

Benutzer mit frei definierbaren Annotationsarten. Dazu setzen wir jeweils unterschiedliche Annotationsebenen ein. Zwei Benutzer mit geisteswissenschaftlicher Expertise im Bereich der Metaphertheorien annotieren dieselben vereinbarten Textabschnitte, um Vergleichswerte für ein Inter-Annotator Agreement zu erhalten. Dabei nutzen wir die Exportmöglichkeiten von WebAnno, um die Annotationsdaten der Nutzer sowie weitere linguistische Merkmale wie Wortarten zu exportieren, mit denen dann ein Inter-Annotator Agreement berechnet, sowie die automatische Weiterverarbeitung für die Identifizierung von Metaphern ausgestaltet werden können.

Hierfür testen wir zunächst in einer Pilotstudie, wie bestehende state-of-the-art Verfahren für ubiquitäre Metaphern (Tsvetkov et al. 2013; Beigmann Klebanov et al. 2014) auf deutschen, insbesondere historischen, Texten abschneiden. Dafür benötigte Ressourcen werden erstellt oder erweitert. Außerdem wird die Weiterentwicklung bestehender Methoden sowie das Entwickeln neuer Methoden vorangetrieben, um auch kühne Metaphern zu identifizieren. Die Identifikation solcher Metaphern stand bislang nicht im Vordergrund automatischer Systeme. Durch die Wahl des Korpus‘ ergeben sich weitere Herausforderungen, zum Beispiel eine ungenügende Abdeckung von *Natur & Staat* durch bestehende manuell erstellte Ressourcen wie etwa GermaNet (Hamp / Feldweg, 1997; Henrich / Hinrichs 2010), sowie ein Mangel an Vergleichskorpora für statistische Verfahren, die beispielsweise mittels Kookkurrenzen oder selektionaler Präferenzen Unterschiede zur „Standardsprache“ feststellen können.

## Ausblick/Fazit

Während mit dem ubiquitären Metaphernmodell somit eher implizite Mechanismen der sprachlichen Sachverhaltsperspektivierung, kurz: konventionalisierte Metaphern, in den Fokus rücken, wird das kühne Metaphernmodell besonders den strukturellen Rezeptions- und Gestaltungsprinzipien des Textes gerecht und vermag auf der Basis des vorgestellten Fokus- und Rahmenkonzepts die besonders markanten Sinnbezirke im Text in den Vordergrund zu rücken. Kühne Metaphern in theoretischen, argumentativen Texten heben sich deutlich von der diskursiv omnipräsenten Sprachstruktur ab: Sie weisen zugleich dadurch, dass sie einen Bruch mit der stilistisch homogenen Sprachstruktur erzeugen, erst auf eben diesen textuell gegebenen Sprachstandard hin und weisen zugleich aufgrund ihres semantischen Verdichtungspotenzials über diesen hinaus (Gehring 2011). Als Sinnbezirke auf der Sprachoberfläche rhetorisch inszeniert und umgesetzt, besitzen kühne Metaphern in *Natur & Staat* insbesondere epistemologisches Potenzial, indem sie neue Theoreme und Idee erst eine besondere sprachliche Gestalt zu geben vermögen.

Hinsichtlich der Annotation sowohl kühner als auch ubiquitärer Metaphern ist ein hohes Inter-Annotator Agreement notwendig; einerseits als Bestätigung für eine hinreichend gute Operationalisierung und Modellierung der verwendeten Theorie, andererseits um einen verlässlichen Goldstandard für das Training und die Evaluation automatischer Verfahren bereit zu stellen. Diesen Zwischenschritt und die daraus resultierenden Möglichkeiten für Geisteswissenschaften als auch Informatik werden wir in unserem Vortrag vorstellen.

## Bibliography

**Black, Max** (1954): „Metaphor“, in: *Proceedings of the Aristotelian Society. New Series* 55: 273-294.

**Black, Max** (1977): „More about Metaphor“, in: *Dialectica* 31: 431-457.

**Beigman Klebanov, Beata / Leong, Ben / Heilman, Michael / Flor, Michael** (2014): „Different Texts, Same Metaphors: Unigrams and Beyond“, in: *Proceedings of the Second Workshop on Metaphor in NLP*, Baltimore, MD, USA 11-17.

**Bühler, Karl** (1934 / 1982): *Sprachtheorie*. Die Darstellungsfunktion der Sprache. Stuttgart / New York: Gustav Fischer.

**Dunn, Jonathan** (2013): „Evaluating the Premises and Results of Four Metaphor Identification Systems“, in: *Proceedings of CICLing 2013*, Samos, Griechenland: 471-486.

**Fix, Ulla** (2011): *Texte und Textsorten - sprachliche, kommunikative und kulturelle Phänomene*. Berlin: Frank & Timme.

**Gehring, Petra** (2006): „Vom Begriff zur Metapher. Elemente einer Methode der historischen Metaphernforschung“, in: Günter Abel (ed.): *Kreativität*. Kolloquiumsbeiträge des XX. Kongresses der Allgemeinen Gesellschaft für Philosophie in Deutschland. Hamburg: Meiner 800-815.

**Gehring, Petra** (2009): „Biologische Politik um 1900 – Reform, Theorie, Experiment?“, in: Griesecke, Birgit / Krause, Marcus / Pethes, Nicolas / Sabisch, Katja (eds.): *Kulturgeschichte des Menschenversuchs im 20. Jahrhundert*. Frankfurt am Main: Suhrkamp 48-76.

**Gehring, Petra** (2010): „Erkenntnis durch Metaphern? Methodologische Bemerkungen zur Metapherntheorie“, in: Junge, Matthias (ed.): *Metaphern in Wissenskulturen*. Wiesbaden: VS Verlag für Sozialwissenschaften 203-220.

**Gehring, Petra** (2011): „Metaphertheoretischer Visualismus – Ist die Metapher »Bild«?“, in: Kroß, Matthias / Zill, Rüdiger (eds.): *Metapherngeschichte – Perspektiven einer Theorie der Unbegrifflichkeit*. Berlin: Parerga Verlag 15-31.

**Gehring, Petra / Gurevych, Iryna** (2014): „Suchen als Methode? Zu einigen Problemen digitaler Metapherndetektion“, in: *Journal für Phänomenologie*

*Schwerpunkt: Metaphern als strenge Wissenschaft* 41: 99-109.

**Goatly, Andrew** (2007): *Washing the Brain – Metaphor and Hidden Ideology*. Amsterdam/Philadelphia: John Benjamin Publishing Company.

**Hamp, Birgit / Feldweg, Helmut** (1997): „GermaNet - a Lexical-Semantic Net for German“, in: *Proceedings of the ACL Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, Madrid, Spanien 9-15.

**Henrich, Verena / Hinrichs, Erhard** (2010): „GernEiT - The GermaNet Editing Tool“, in: *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, Valletta, Malta 2228-2235.

**Johnson, Mark** (1987): *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: Chicago University Press.

**Kövecses, Zóltan** (2003): *Metaphor and Emotion. Language, Culture, and the Body in Human Feeling*. Cambridge: Cambridge University Press.

**Kövecses, Zóltan** (2015): *Where Metaphor Come From. Reconsidering Context in Metaphor*. Oxford: University Press.

**Kohl, Katrin** (2007): *Metapher*. Stuttgart: Sammlung Metzler.

**Lakoff, George** (2006): „Conceptual Metaphor. The Contemporary Theory of Metaphor [1993]“, in: Geeraerts, Dirk (ed.): *Cognitive Linguistics: Basic Readings*. Berlin: Mouton de Gruyter 185-238.

**Lakoff, George / Johnson, Mark** (1980): *Metaphors We Live by*. Chicago: Chicago University Press.

**Mason, Zachary** (2004): „Cormet: A Computational, Corpus-based Conventional Metaphor Extraction System“, in: *Computational Linguistics*, 30, 1: 23-44.

**Núñez, Alexandra** (2014): „Wenn das 'Embodiment' politisch wird: Das Image-Schema PATH und seine Realisierung im Mediendiskurs zum 'Arabischen Frühling'“, in: Polzenhagen, Frank / Kleinke, Sonja / Kövecses, Zoltán / Vogelbacher, Stefanie (eds.): *Cognitive Explorations into Metaphor and Metonymy*. Bern / Berlin / Frankfurt am Main / New York / Paris / Wien: Peter Lang 149-164.

**Paprotté, Wolf / Dirven, René** (ed.) (1985): *The Ubiquity of Metaphor. Metaphor in Language and Thought*. Amsterdam / Philadelphia: John Benjamin Publishing Company.

**Paul, Hermann** (1909): *Prinzipien der Sprachgeschichte*. Halle a. S.: Niemeyer.

**Polenz, Peter von** (1981): „Über die Jargonisierung von Wissenschaftssprache und wider die Deagentivierung“, in: Bungarten, Theo (ed.): *Wissenschaftssprache- Beiträge zur Methodologie, theoretische Fundierung und Deskription*. München: Wilhelm Fink Verlag 85-110.

**Polenz, Peter von** (1999): *Deutsche Sprachgeschichte. Vom Spätmittelalter bis zur Gegenwart*.

Bd. III. 19. und 20. Jahrhundert. Berlin / New York:  
Walter de Gruyter.

**Ricoeur, Paul** (1972): „La métaphore et le problème central de l'herméneutique“, in: *Revue philosophique de Louvain* 70: 93-112.

**Shutova, Ekaterina** (2013): „Metaphor Identification as Interpretation“, in: *Proceedings of \*SEM 2013*, Atlanta, GA, USA 276-285.

**Shutova, Ekaterina / Sun, Lin** (2013): „Unsupervised Metaphor Identification using Hierarchical Graph Factorization Clustering“, in: *Proceedings of NAACL*, Atlanta, GA, USA 978-988.

**Tsvetkov, Yulia / Mukomel, Elena / Gershman, Anatole** (2013): „Cross-lingual Metaphor Detection using Common Semantic Features“, in: *Proceedings of the First Workshop on Metaphor in NLP*, Atlanta, GA, USA 45-51.

**Turney, Peter D. / Neuman, Yair / Assaf, Dan / Cohen, Yohai** (2011): „Literal and Metaphorical Sense Identification through Concrete and Abstract Context“, in: *Proceedings of EMNLP 2011*, Stroudsburg, PA, USA 680-690.

**Weinrich, Harald** (1976): *Sprache in Texten*. Stuttgart: Ernst Klett Verlag.

**Yimam, Seid Muhie / Gurevych, Iryna / Eckart de Castilho, Richard / Biemann, Chris** (2013): „WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations“, in: *Proceedings of ACL 2013, Demo Session*, Sofia, Bulgaria 1-6.

## CLARIN-D: Ressourcen gesprochener Sprache und Webservices des Bayerischen Archivs für Sprachsignale

### Draxler, Christoph

draxler@phonetik.uni-muenchen.de  
Bayerisches Archiv für Sprachsignale, Institut für  
Phonetik und Sprachverarbeitung, Deutschland

### Schiel, Florian

schiel@phonetik.uni-muenchen.de  
Bayerisches Archiv für Sprachsignale, Institut für  
Phonetik und Sprachverarbeitung, Deutschland

### Reichel, Uwe

reichelu@phonetik.uni-muenchen.de  
Bayerisches Archiv für Sprachsignale, Institut für  
Phonetik und Sprachverarbeitung, Deutschland

### Kisler, Thomas

kisler@phonetik.uni-muenchen.de  
Bayerisches Archiv für Sprachsignale, Institut für  
Phonetik und Sprachverarbeitung, Deutschland

## Das BAS Repository

Das Repository des BAS ist in einen öffentlich zugänglichen und einen Zugangsgeschützten Bereich unterteilt. Der öffentliche Bereich enthält die Metadaten der im Repository enthaltenen Datenbanken gesprochener Sprache, der geschützte Bereich die Annotationen und Sprachsignaldaten.

Aktuell (Stand 31.12.2015) umfasst das Repository 33 Korpora mit den Schwerpunkten Sprachtechnologie, regionale Variation, Sprechermerkmale und Grundlagenforschung. Auf der obersten Repository-Ebene werden die wichtigsten Angaben prominent platziert: Name und Eigentümer der Ressource, Audio bzw. Video, verwendete Sprache, und die Zugangsrestriktionen (s. Kap. 6).

Von den 33 Korpora enthält eines arabische Sprachaufnahmen, und enthalten zwei italienische, drei englische und 26 deutsche Sprachaufnahmen; dazu kommt ein Korpus in Deutscher Gebärdensprache. 26 Korpora enthalten nur gesprochene Sprache, 6 sowohl gesprochene Sprache als auch Video, eines nur Video. Ein Korpus ist allgemein verfügbar, 30 sind für akademische Nutzer\_innen frei zugänglich, zwei nur als lizenzierte Korpora.

Der Datenumfang beträgt mehr als 2,86 TB an Audio-, Video- und Sensorsignaldaten, sowie ca. 17,4 GB an Annotations- und Metadaten.

The screenshot shows the website interface for the BAS CLARIN Repository. At the top, there is a navigation menu with options: Repository, Search, Web Services, and Links. The main content area is titled 'BAS CLARIN Repository' and includes a 'Metadatum' section with the following details:

- Metadatum:**
  - PID:** 11858/00-1779-0000-0006-BF00-E
  - CMCI:** | text/xml | (Distance: Usage)
- Korpora:**
  - Gender:**
    - Owner: Bavarian Archive for Speech Signals, Ludwig-Maximilians-Universität München
    - Title: sGender
    - Modality: Spoken
    - Subject language(s): German
    - Access: free for science
  - ALG:**
    - Owner: Bavarian Archive for Speech Signals, Ludwig-Maximilians-Universität München
    - Title: BAS Alcohol Language Corpus
    - Modality: Spoken
    - Subject language(s): German
    - Access: free for science
  - AsiCa:**
    - Title: LMU AsiCa
    - Modality: Spoken
    - Subject language(s): Italian
    - Access: restricted (contact bas@bas.uni-muenchen.de to obtain a user license)

**Abb. 1:** Startseite des BAS Repository mit der Liste verfügbarer Korpora.