

# Sprachqualität



Dipl.-Ing. José Manuel de la Rosa Govantes  
Vortrag bei der GSE Tagung Dresden, 20.09.2018



# AGENDA



## Über uns

Die semantics Kommunikationsmanagement GmbH



## Corporate Library

Forschung und Tools zur Textpflege und -optimierung



## Architektur

Unbeschränkter Einsatz durch Webservices



## CL-Sprachtool

Beispiel Realisierung



# Über uns



## Eintritt 2004

Dipl.-Ing. José de la Rosa  
(Fusion mit VisualAix®)



## Kernthema bei Gründung

Verbindung von Forschung und  
Praxis zur Textverständlichkeit  
und Texttechnologie



## Gründung 2000

als Spin-off der RWTH Aachen  
(Prof. Dr. Christian Stetter und  
Kay Heiligenhaus M.A.)



## Interdisziplinäres Team

aus z.Zt. 40 Informatikern,  
Sprachwissenschaftlern,  
Psychologen und Consultants  
Enge Kooperation mit der RWTH  
Aachen, Prof. Dr. Thomas Niehr



## Zweites Thema (seit 2005):

Digitalisierungs- und  
Recherchesysteme für  
Bibliotheken, Archive und Museen  
Marktführer D-A-CH



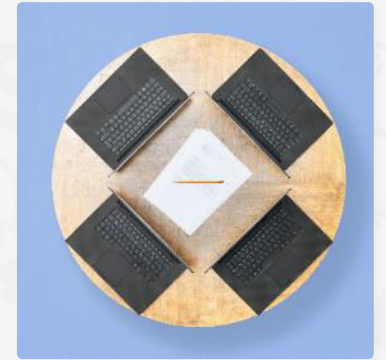
# Über uns

Methoden und Technik



## Texttechnologie

Redaktion, Kollaboration,  
Dokumentation, Recherche,  
Analyse, automatisierte  
Verarbeitung, Migration



## Sprachwissenschaft

Schriftgut-Analysen,  
Sprachrichtlinien,  
Textoptimierung,  
Sprachschulungen,  
Redaktion

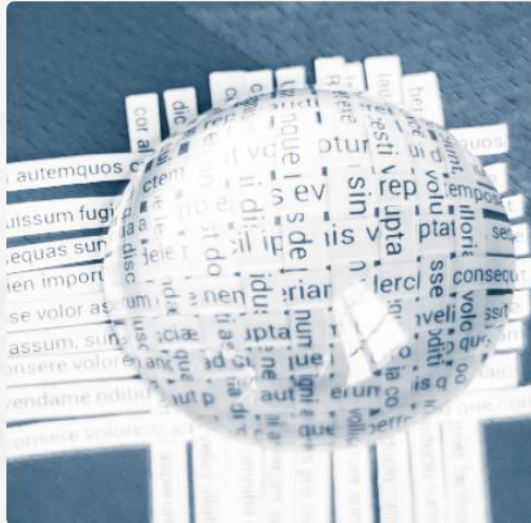


## Forschung

Empirische Studien,  
Korpuslinguistik,  
Linguistische Informatik,  
Lehre und Ausbildung



# Corporate Library



Library Manager



Redaktion

Kollaboration, Redaktion,  
Qualitätssicherung,  
Kompetenzgemeinschaften



Freies Schreiben

Sprachprüfung, Lesbarkeit,  
Corporate Wording,  
Vorschlagswesen



Empfänger

Zielgruppe, Motivation,  
Kontext



Library Textanalysis



Library Assistant

Sentiment Inbound



Grundlagen

Wissensmanagement  
Organisationstheorie, Text Mining,  
Korpuslinguistik, LSA, LDA, ...

Textlinguistik, Verständlichkeits- /  
Lesbarkeitsforschung,  
Empirische Studien

Kognitionswissenschaft,  
Mentale Modelle,  
Sprechakttheorie

# Sprache

## Relevanz und Ziel

”

Man kann nicht nicht kommunizieren.

(Paul Watzlawick)

Man kann es nur gut oder schlecht tun.

<e>



### Ansprüche Kunden steigen

Multi- / Omni-Channel:  
Infos, Zeit, Umfang und Kanal –  
und dabei bitte den richtigen  
Ton



### Planung der Kommunikation

Brief, Email, Portale, Apps, SMS, ...  
unterschiedlich zu bedienen,  
Aufwand steigt



### Bindeglied *Sprache*

Wiedererkennung,  
Zufriedenheit,  
Textprozesse  
*automatisieren*



# Sprache

Beispiel aus der Praxis

„Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Fall, dass er wegen eines während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses, das den Tod, die Verletzung oder Gesundheitsschädigung von Menschen (Personenschaden) oder die Beschädigung oder Vernichtung von Sachen (Sachschaden) zur Folge hatte, für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadenersatz in Anspruch genommen wird.“

„Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Fall, dass er wegen eines während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses, das den Tod, die Verletzung oder Gesundheitsschädigung von Menschen (Personenschaden) oder die Beschädigung oder Vernichtung von Sachen (Sachschaden) zur Folge hatte, für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadenersatz in Anspruch genommen wird.“

Quelle: „Zur Verständlichkeit von Geschäftstexten am Beispiel Allgemeiner Versicherungsbedingungen“, Magisterarbeit Björn Eßer, RWTH Aachen, 2012; §1 Nr. 1 AHB, zit. nach Nitschke (2002), S.122.



# Sprache

Grammatische Analyse

„Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Fall, dass er während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses, das den Tod, die Verletzung oder Gesundheitsschädigung von Menschen (Personenschaden) oder die Beschädigung oder Vernichtung von Sachen (Sachschaden) zur Folge hatte, für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadenersatz in Anspruch genommen wird.“

*Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Fall, dass er während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses, das den Tod, die Verletzung oder Gesundheitsschädigung von Menschen (Personenschaden) oder die Beschädigung oder Vernichtung von Sachen (Sachschaden) zur Folge hatte, für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadenersatz in Anspruch genommen wird.*

Hauptaussage

konditionaler Nebensatz

kausale Bestimmung

Linksattribut

Relativsatz

Rückbezug

Verweis auf rechtliche Basis



# Sprache

Überarbeitung

„Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Schaden, der während der Wirksamkeit der Versicherung eingetretenen Schadenereignis, Verletzung oder Gesundheitsschaden (Personenschaden) oder die Vernichtung von Sachen (Sachschaden) für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen r

*Der Versicherer gewährt dem Versicherungsnehmer Deckung für Haftpflichtansprüche, die gegen ihn geltend gemacht werden.*

*Haftpflichtansprüche unterliegen gesetzlichen Haftpflichtbestimmungen privatrechtlichen Inhalts.*

*Sie bestehen dann, wenn der Versicherungsnehmer für Personen- oder Sachschäden verantwortlich gemacht werden kann.*

*Dieser Schaden muss während der Versicherungszeit entstanden sein, damit er unter die gewährte Deckung fällt.*

**Hauptaussage**

**Verweis auf rechtliche Basis**

**Kausale Bestimmung**

**Bedingung**



# Sprache

Thematische Kohärenz zur  
Unterstützung mentales  
Modell

(Thema/Rhema Paradigma,  
anaphorische Bezüge,  
Rekurrenzen)

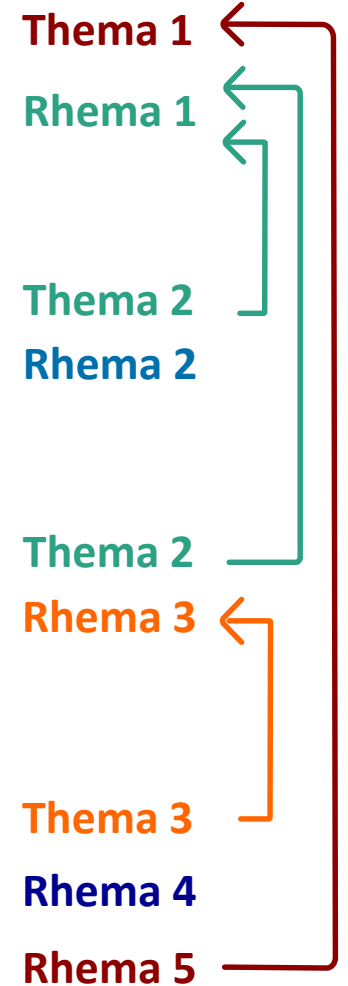
„Der Versicherer gewährt a  
Versicherungsschutz für de  
während der Wirksamkeit c  
eingetretenen Schadenereig  
Verletzung oder Gesundheit  
(Personenschaden) oder die  
Vernichtung von Sachen (S  
für diese Folge aufgrund ge  
Haftpflichtbestimmungen r

Der *Versicherer gewährt* dem Versicherungsnehmer *Deckung*  
*für Haftpflichtansprüche*, die gegen ihn geltend  
gemacht werden.

*Haftpflichtansprüche* unterliegen gesetzlichen  
*Haftpflichtbestimmungen* privatrechtlichen Inhalts.

*Sie bestehen dann*, wenn der Versicherungsnehmer  
*für Personen- oder Sachschäden verantwortlich*  
*gemacht werden kann*.

*Dieser Schaden* muss während der Versicherungszeit  
*entstanden sein*, damit er unter die gewährte *Deckung* fällt.





# Forschung

Stärkerer Angang seitens  
Geisteswissenschaftler  
(Sprachwissenschaft kombiniert  
mit Text Mining etc.)



113

Thomas Niehr / Eva Dickmeis / Bianka Trevisan / Eva-Maria Jakobs

## Neue Wege der linguistischen Diskursforschung

### Computerbasierte Verfahren der Argumentanalyse

**Zusammenfassung:** Der vorliegende Beitrag beschreibt und diskutiert eine neuartige Verbindung quantitativer und qualitativer Verfahren für die Analyse von Big Data in der linguistischen Diskursforschung. Der vorgestellte Ansatz kombiniert Methoden der diskurslinguistischen Argumentationsanalyse mit Methoden des Linguistischen Text Mining. Das Ziel der Methodenentwicklung ist ein computergestütztes Verfahren für die semi-automatisierte Identifizierung und Analyse von Argumenten in großen Textkorpora. Erprobt wird das Verfahren an einem Diskurs über Infrastrukturmaßnahmen. Im Beitrag werden sprachliche Mittel vorgestellt, die im Korpus gemeinsam auftreten und damit als Merkmale von Argumentmustern betrachtet werden können. Solche Argumentmuster können das Vorkommen von Argumenten und ihren Verwendungsweisen in Texten indizieren.

**Schlagwörter:** Linguistische Diskursanalyse, Argumentationsanalyse, Text Mining, Argumentation Mining, Argument

**Abstract:** The following paper presents and discusses an innovative approach for the automated detection and analysis of argument patterns in large corpora. The approach combines methods of discourse linguistics and linguistic text mining. The goal of the method development is a computer-supported procedure for the semi-automated identification and analysis of arguments in large corpora. The procedure is tested on a discourse about infrastructure measures. In the paper, linguistic means are presented that co-occur in the corpus and can thus be considered as features of argument patterns. Such argument patterns can indicate the occurrence of arguments and their uses in texts.

6.1 Text Mining

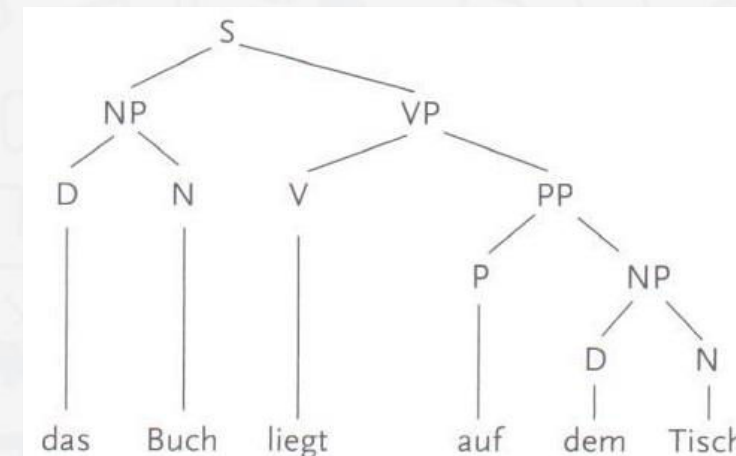
Text Mining ist eine Methodik, die aus der Informatik stammt. Sie wurde erstmals von Feldman und Dagan (1995) als *Knowledge Discovery in Textual Database (KDD)* eingeführt. Gegenstand der Methodik ist die (semi-)automatische Textanalyse schwach strukturierten Texten.

**Keywords:** Discourse, Argument

## 1. Einleitung

Der Beitrag diskutiert eine neuartige Verbindung quantitativer und qualitativer Verfahren für die Analyse von Big Data in der linguistischen Diskursforschung. Der vorgestellte Ansatz kombiniert Methoden der diskurslinguistischen Argumentationsanalyse mit Methoden des Linguistischen Text Mining. Das Ziel der Methodenentwicklung ist ein computergestütztes Verfahren für die semi-automatisierte Identifizierung und Analyse von Argumenten in großen Textkorpora. Erprobt wird das Verfahren an einem Diskurs über Infrastrukturmaßnahmen. Im Beitrag werden sprachliche Mittel vorgestellt, die im Korpus gemeinsam auftreten und damit als Merkmale von Argumentmustern betrachtet werden können. Solche Argumentmuster können das Vorkommen von Argumenten und ihren Verwendungsweisen in Texten indizieren.

Abb. 1: Mehrebenen-Annotationsmodell von Trevisan (2014, S. 134)



# Architektur

Sprachprüfung



Redaktion



CL-Sprachtool



Editor



Content  
Management



Library Assistant

Webservice  
Nahtlose Integration,  
Anbindung Drittlösungen



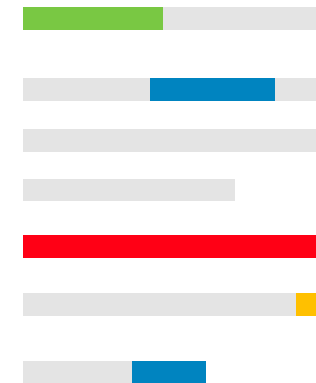
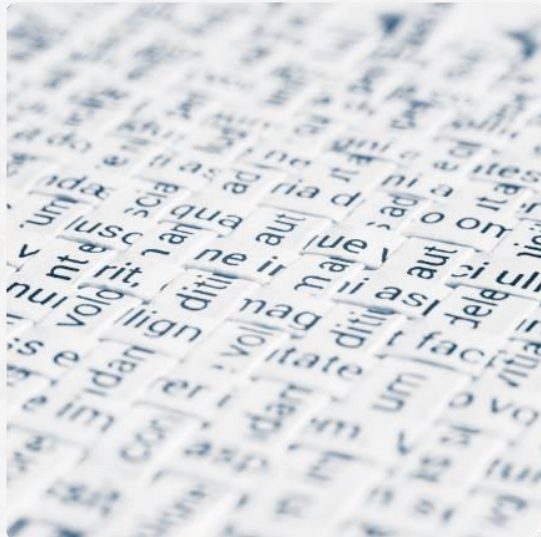
Sentiment,  
Big Data



Keyword- / Topic  
Extraction



CL-Textanalysis



# Architektur

Vorschlagswesen



Redaktion



CL-Sprachtool



Editor



Content  
Management



Library Assistant

Webservice  
Nahtlose Integration,  
Anbindung Drittlösungen



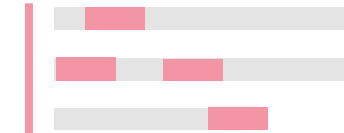
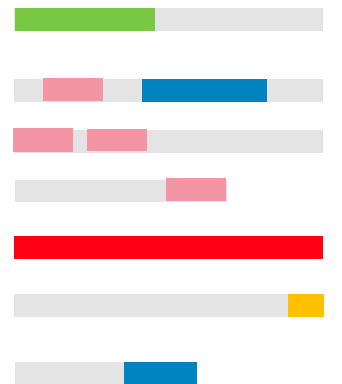
Sentiment,  
Big Data



Keyword- / Topic  
Extraction



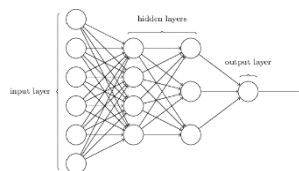
CL-Textanalysis





# Corporate Library

Sprachtool



## Künstliche Intelligenz

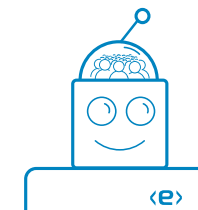
Segmentierung (99,98%),  
morphosyntaktische  
Annotationen,  
Semantik, Pragmatik

The screenshot shows the s<e>mantics web interface. At the top, the logo 's<e>mantics Kommunikationsmanagement GmbH' is visible. The main content area displays a document snippet with German text. Below the text, there is a section with a '49 Score' and 'Sentiment negativ'. The interface is divided into several panels, each showing a sentence from the document and its corresponding sentiment analysis results. For example, one panel shows the sentence 'Wir hatten für Sie eine solche Best [...] n Kennzeichen M B 1234 angefordert.' with a sentiment of 'negativ'. Another panel shows 'Bitte sehen Sie doch in den Unterla [...] amtliche Kennzeichen richtig sind.' with a sentiment of 'negativ'. The interface also includes a sidebar with icons for document, list, and settings.

## Sprachprojekt bei SV



> 13.000 User  
Feedbacksystem  
Lernende Algorithmen  
Sentiment-Analysen





Dipl.-Ing. José de la Rosa

semantics  
Kommunikationsmanagement  
GmbH  
Viktoriaallee 45  
52066 Aachen

+49 241 89 49 89 29  
j.delarosa@semantics.de  
www.semantics.de



Besten Dank für Ihre Aufmerksamkeit!