

# Einführung

Katja Markert

Institut für Computerlinguistik  
Uni Heidelberg

`markert@cl.uni-heidelberg.de`  
mit Folien von Yannick Versley und Anette Frank

October 18, 2022

- 1 Was ist Computerlinguistik?
- 2 Anwendungsbeispiele
- 3 Linguistik und CL: Fragen und Teilgebiete
- 4 Was passiert in diesem Kurs?

- 1 Was ist Computerlinguistik?
- 2 Anwendungsbeispiele
- 3 Linguistik und CL: Fragen und Teilgebiete
- 4 Was passiert in diesem Kurs?

- Computerlinguistik beschäftigt sich mit der maschinellen Verarbeitung natürlicher Sprache: Algorithmen zum Verstehen und Generierung natürlicher Sprache
- Modelle zur Erklärung von strukturellen Eigenschaften und Verarbeitungsmechanismen natürlicher Sprache.

- Studium der formalen Eigenschaften von Sprache: Wie funktioniert Sprache? Wie kann ich dies formal repräsentieren und modellieren?

Beispielfrage: Welche Ausdrücke sind korrektes Deutsch/Englisch?

- Erklärung / Simulation von Sprachverständnis
  - Analyse
  - Generierung
  - Übersetzung
  - Von natürlicher Sprache zu strukturierter Information und zurück

Beispielfrage: Was sagen Menschen? Welche Muster im Sprachgebrauch gibt es?

- 1 Was ist Computerlinguistik?
- 2 Anwendungsbeispiele**
- 3 Linguistik und CL: Fragen und Teilgebiete
- 4 Was passiert in diesem Kurs?

- Extraktion von Meinungen und Emotionen aus Texten
- Typisch: Vorhersage von Verkäufen oder Filmerfolgen oder Verfolgung von Markenreputation
- Social Media Monitoring

## Demos:

- **Lexalytics Demo** <https://www.lexalytics.com/nlp-demo>  
(Example Hotel, Sample 2)
- **Sentiment Treebank Demo** <http://nlp.stanford.edu:8080/sentiment/rntnDemo.html> (auch Teil der CoreNLP pipeline)  
<https://stanfordnlp.github.io/CoreNLP/index.html>)

# Lexalytics Input (Screenshot 2020)

Semantria Web Demo | Lexalytics - Mozilla Firefox

Semantria Web Dem... x Semantria API End U... x Spectre Reviews & Ra... x How to make screens... x Font size and zoom - ... x

https://www.lexalytics.com/demo

LEXALYTICS Semantria ▾ Salience ▾ Tech ▾ Demo Price

A dull pointless movie.

I was finding it incredibly hard to stay awake through this movie.  
Even the action scenes were quite lacklustre.  
I think it had nothing to do with the big budget special effects but more to do with the emotionless performance of Daniel Craig's James Bond.

I just cannot watch a mindless film about fighting, explosions and getting laid when there's no passion.

Speaking of passion the "bond girls" were a big disappointment and had no connection with Bond. It was not believable and super awkward.

Current Character Count: 0 / 16384

English ▾

Start Analysis

← Paste your  
Or try with the def

← Select lang

# Lexalytics Output (Screenshot 2020)

The screenshot shows the Lexalytics web application interface. At the top, there's a navigation bar with the Lexalytics logo and links for Semantria, Salience, Tech, Demo, Price, Blog, Support, and Login. Below the navigation bar, there are two tabs: 'Detailed' and 'Discovery (one per line)'. The 'Detailed' tab is active. The main content area displays a sample text input field with the URL 'https://www.lexalytics.com/demo' and a 'Go' button. Below the input field, the analyzed text is shown with sentiment scores and labels. The text is: 'A dull pointless movie. I was finding it boring hard to stay awake through this movie. Even the action scenes were quite lacklustre. I think it had nothing to do with the big budget special effects but more to do with the emotionless performance of Daniel Craig's James Bond. I just cannot watch a mindless film about fighting, explosions and getting laid when there's no passion. Speaking of passion the 'bond girls' were a big disappointment and had no connection with Bond. It was not believable and super awkward.' The sentiment scores are: -0.174 (negative). The sentiment labels are: fighting, awkward, passion, pointless, dull, disappointment, super, mindless, incredibly, lacklustre, believable. A scroll down for full report button is visible.

Semantria Web Demo | Lexalytics - Mozilla Firefox

https://www.lexalytics.com/demo

Lexalytics Semantria Salience Tech Demo Price Blog Support Login

Detailed Discovery (one per line)

This mode takes a single document, categorizes it, extracts entities, determines sentiment and creates a summary.

Analyze URL Go

A dull pointless movie.

I was finding it boring hard to stay awake through this movie. Even the action scenes were quite lacklustre. I think it had nothing to do with the big budget special effects but more to do with the emotionless performance of Daniel Craig's James Bond.

I just cannot watch a mindless film about fighting, explosions and getting laid when there's no passion.

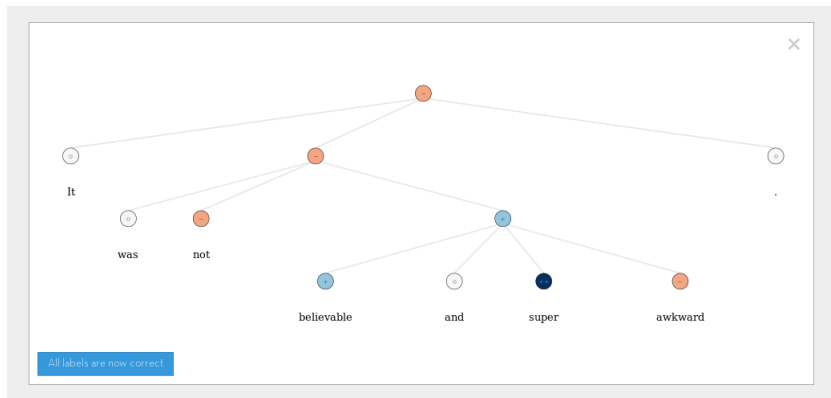
Speaking of passion the 'bond girls' were a big disappointment and had no connection with Bond. It was not believable and super awkward.

This document is: **negative**  
**(-0.174)**

fighting awkward passion  
pointless dull disappointment super  
mindless incredibly lacklustre  
believable

Scroll down for full report

# Stanford sentiment treebank (Screenshot Oktober 2021)



Suche nach Antworten:

- Faktoid: Wann wurde Leonardo Da Vinci geboren?
  - Komplex: Wie alt war Leonardo da Vinci als Michelangelo geboren wurde?
  - Schlussfolgerung: Leonardo da Vinci war älter als Michelangelo.
  - Allgemeinwissen oder aus vorgegebenem Text
- 1 Linguistische Analyse der Frage: Fragetyp, Thema → Suche nach passender Information (Kandidaten), Evaluation von Hypothesen → Aufbereitung der Antwort
  - 2 Große Sprachmodelle wie GPT-2 und deren Vorhersagen

Demo: Wolfram Alpha <http://www.wolframalpha.com/> (accessed 15.10.2022)

- When was Leonardo da Vinci born? → Saturday, April 15, 1452
- How old was Leonardo da Vinci when Michelangelo was born?  
→ Result: 22 years 10 months 18 days
- Has Elvis died → Yes
- Has Elvis kicked the bucket? → Yes
- Has Elvis given up the ghost? → Interpreting as "color ghost".
- When did Shakespeare die? → Tuesday, April 23, 1616
- Shakespeare wrote many plays. When did he die? → Input interpretation "wrote" (English word).

Demo: Wolfram Alpha <http://www.wolframalpha.com/> (accessed 15.10.2022)

- When was Leonardo da Vinci born? → Saturday, April 15, 1452
- How old was Leonardo da Vinci when Michelangelo was born?  
→ Result: 22 years 10 months 18 days
- Has Elvis died → Yes
- Has Elvis kicked the bucket? → Yes
- Has Elvis given up the ghost? → Interpreting as “color ghost”.
- When did Shakespeare die? → Tuesday, April 23, 1616
- Shakespeare wrote many plays. When did he die? → Input interpretation “wrote” (English word).

Demo: Wolfram Alpha <http://www.wolframalpha.com/> (accessed 15.10.2022)

- When was Leonardo da Vinci born? → Saturday, April 15, 1452
- How old was Leonardo da Vinci when Michelangelo was born?  
→ Result: 22 years 10 months 18 days
- Has Elvis died → Yes
- Has Elvis kicked the bucket? → Yes
- Has Elvis given up the ghost? → Interpreting as “color ghost”.
- When did Shakespeare die? → Tuesday, April 23, 1616
- Shakespeare wrote many plays. When did he die? → Input interpretation “wrote” (English word).

## Einige mögliche Demos mit großen Sprachmodellen:

- `https://visbert.demo.dataxis.com/`
- `https://www.pragnakalp.com/demos/BERT-NLP-QnA-Demo/`
- **Auf Deutsch:** `https://www.pragnakalp.com/demos/BERT-NLP-QnA-Demo/german\_qna.php`

- Sicher eine der bekanntesten und wichtigsten Anwendungen
- Unterschiedliche linguistische Eigenschaften verschiedener Sprachen
- Interpretation, Wissen und Übersetzung
- Varianten: Vollübersetzung (wissensbasiert – statistisch – neuronal ), Unterstützte Systeme (HAMT) – human-aided MT, Unterstützende Systeme (MAHT) – machine-aided HT

## Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.
- The doctor went home Der Arzt ging nach Hause.
- The secretary went home Die Sekretärin ging nach Hause.
- The doctor washed her hands Der Arzt wusch ihr die Hände.
- The nurse/secretary washed her hands Die Krankenschwester/Sekretärin wusch sich die Hände

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

Google Translate (15.10.2022)

- Im Wesentlichen handelt Star Wars vom ständig andauernden Kampf zwischen Gut und Böse. **Essentially, Star Wars is about the ongoing battle between good and evil.**
- Die Kaffeemaschine ist kaputt. Ich lasse sie reparieren. **The coffee maker is broken. I'll have them repaired.**
- The coffee machine is broken. I'll have it repaired. **Die Kaffeemaschine ist kaputt. Ich lasse es reparieren.**
- The doctor went home **Der Arzt ging nach Hause.**
- The secretary went home **Die Sekretärin ging nach Hause.**
- The doctor washed her hands **Der Arzt wusch ihr die Hände.**
- The nurse/secretary washed her hands **Die Krankenschwester/Sekretärin wusch sich die Hände**

- Alle diese Anwendungen müssen Sprache “verstehen” → Nutzen von Linguistik (?)
- Alle nicht perfekt: → Empirische Methoden
- **Linguistik:** Analyse der gesprochenen und geschriebenen natürlichen Sprache in Bezug auf Form, Bedeutung und Kontext

- 1 Was ist Computerlinguistik?
- 2 Anwendungsbeispiele
- 3 Linguistik und CL: Fragen und Teilgebiete**
- 4 Was passiert in diesem Kurs?

- 1 Welchem deutschen Wort entspricht [ra:t]
- 2 Wie wird das Wort *rasten* ausgesprochen?
- 3 Sprechen Sie das Wort *ja* aus.
- 4 Wie verstehen Sie die folgende (deutsche) Sprachsequenz: [hasm mo'men'tsaɪt]

- 1 Welchem deutschen Wort entspricht [ra:t]?
- 2 Wie wird das Wort *rasten* ausgesprochen?
- 3 Sprechen Sie das Wort *ja* aus.
- 3 Wie verstehen Sie die folgende (deutsche) Sprachsequenz:  
[hasm mo'men'tsa:t]

- 1 Welchem deutschen Wort entspricht [ra:t]?
- 2 Wie wird das Wort *rasten* ausgesprochen?
- 3 Sprechen Sie das Wort *ja* aus.
- 3 Wie verstehen Sie die folgende (deutsche) Sprachsequenz:  
[hasm mo'men'tsai]

- 1 Welchem deutschen Wort entspricht [ra:t]?
- 2 Wie wird das Wort *rasten* ausgesprochen?
- 3 Sprechen Sie das Wort *ja* aus.
- 4 Wie verstehen Sie die folgende (deutsche) Sprachsequenz: [hasm mo'mɛn'tsaɪt]

- Phonetik und Phonologie
  - Artikulatorische Merkmale und Lautstruktur
  - Laute einer bestimmten Sprache
  - Wortsegmentierung, Aussprache, Prosodie
- **Phonem:** Kleinste Spracheinheit, die Bedeutungsunterschiede ausmacht
- **Homophone:** verschiedene Worte mit gleicher Aussprache
- Variation in Aussprache
  - **lexikalisch:** durch Ambiguität; oft soziolinguistisch
  - **allophonisch:** meist kontextuelle Variation
  - **Prosodie:** Betonung, Intonation, Rhythmus

- Phonetik und Phonologie
  - Artikulatorische Merkmale und Lautstruktur
  - Laute einer bestimmten Sprache
  - Wortsegmentierung, Aussprache, Prosodie
- **Phonem:** Kleinste Spracheinheit, die Bedeutungsunterschiede ausmacht
- **Homophone:** verschiedene Worte mit gleicher Aussprache
- Variation in Aussprache
  - **lexikalisch:** durch Ambiguität; oft soziolinguistisch
  - **allophonisch:** meist kontextuelle Variation
  - **Prosodie:** Betonung, Intonation, Rhythmus

Spracherkennung: Übersetzung von gesprochener Sprache in Text

Sprachsynthese/Text-to-Speech: künstliche Erzeugung der menschlichen Sprechstimme

- 1 Ist *riche* ein deutsches Wort? Wie ist es mit *freche*?
- 2 Is *un* ein Wort? Hat es eine Bedeutung?
- 3 Wie viele "Bedeutungseinheiten" hat *Frechheiten*? Wie ist es mit *Staubbecken*?

- 1 Ist *riche* ein deutsches Wort? Wie ist es mit *freche*?
- 2 Is *un* ein Wort? Hat es eine Bedeutung?
- 3 Wie viele "Bedeutungseinheiten" hat *Frechheiten*? Wie ist es mit *Staubbecken*?

- 1 Ist *riche* ein deutsches Wort? Wie ist es mit *freche*?
- 2 Is *un* ein Wort? Hat es eine Bedeutung?
- 3 Wie viele "Bedeutungseinheiten" hat *Frechheiten*? Wie ist es mit *Staubbecken*?

- 1 Ist *riche* ein deutsches Wort? Wie ist es mit *freche*?
- 2 Is *un* ein Wort? Hat es eine Bedeutung?
- 3 Wie viele "Bedeutungseinheiten" hat *Frechheiten*? Wie ist es mit *Staubbecken*?

- 1 Ist *riche* ein deutsches Wort? Wie ist es mit *freche*?
- 2 Is *un* ein Wort? Hat es eine Bedeutung?
- 3 Wie viele “Bedeutungseinheiten” hat *Frechheiten*? Wie ist es mit *Staubecken*?

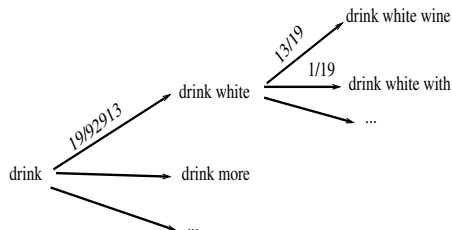
- **Morphologie:** Beschreibung von Bildung und Struktur von Wörtern
- **Wort:** Kleinste alleinstehende Einheit
- **Lemma:** Lexikoneintrag eines Wortes
- **Wortform:** Voll flektierte Oberflächenform eines Wortes

- **Morphem:** kleinste bedeutungstragende Einheit
- Systematische Beziehungen zwischen Wörtern und Wortform:  
Flexion, Derivation, Komposition (*frech* - *Frechheit* - *Frechdachs*)
- Prozesse/Regeln zur Erzeugung von Wortformen
- **Morphologisches Parsing:** Finde die Morpheme eines Wortes  
→ endliche Automaten, formale Sprachen

- 1 Welche Worte folgen wahrscheinlich der (englischen) Sequenz  
“white ...” ?
- 2 Welche Worte folgen wahrscheinlich der englischen Sequenz  
“drink white ...” ?

- 1 Welche Worte folgen wahrscheinlich der (englischen) Sequenz  
“white ...” ?
- 2 Welche Worte folgen wahrscheinlich der englischen Sequenz  
“drink white ...” ?

**ngram modelling:** Ordne Wortsequenzen Wahrscheinlichkeiten zu  
→ Wahrscheinlichkeitstheorie und Schätzung

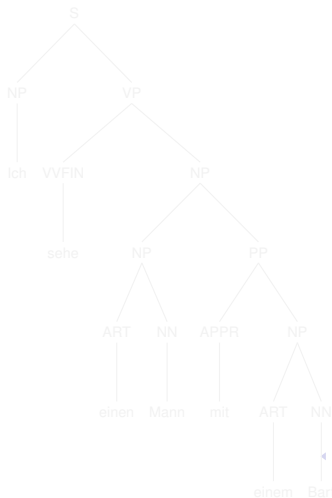


$$P(w_i | w_1 \dots w_{i-1}) = \frac{P(w_1 \dots w_i)}{P(w_1 \dots w_{i-1})} \approx \frac{f(w_1 \dots w_i)}{f(w_1 \dots w_{i-1})}$$

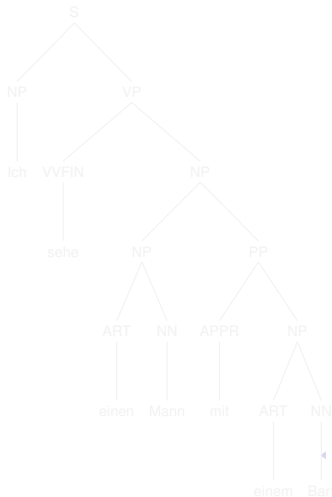
Alternativ: ngram modelling mit neuronalen Netzen

- Ist *content* in “a content person” ein **Nomen oder Adjektiv**?
- Ist “Ich sehe einen Mann mit meinem Fernglas” **grammatikalisch richtig**? Hat der Mann mein Fernglas oder sehe ich ihn?

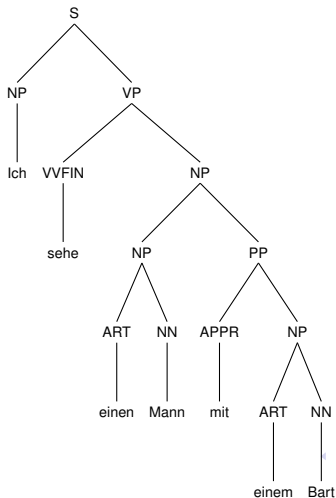
Vergleich mit “Ich sehe einen Mann mit einem Bart”?



- Ist *content* in “a content person” ein **Nomen oder Adjektiv**?
- Ist “Ich sehe einen Mann mit meinem Fernglas” **grammatikalisch richtig**? Hat der Mann mein Fernglas oder sehe ich ihn?  
Vergleich mit “Ich sehe einen Mann mit einem Bart”?



- Ist *content* in “a content person” ein **Nomen oder Adjektiv**?
- Ist “Ich sehe einen Mann mit meinem Fernglas” **grammatikalisch richtig**? Hat der Mann mein Fernglas oder sehe ich ihn?  
Vergleich mit “Ich sehe einen Mann mit einem Bart”?



- **Part-of-Speech (POS)**: Lexikalische Kategorie/Wortart eines Wortes
- **Syntax**: beschreibt strukturelle Beziehung zwischen Wörtern
- Typische Fragen: Grammatikalität? Parse eines Satzes
- **Syntaktische Regeln**: Prozesse, die Sätze generieren können
  - VP → VVFIN NP
  - VP → VP PP
- **Parsing**: Zähle die syntaktischen Strukturen (parses) von Sätzen auf und entscheide Dich für eine präferierte Struktur; benutzt syntaktische Regeln
- Präferenzen: oft auch wieder statistisch
  - VP → VVFIN NP | 0.7
  - VP → VP PP | 0.3

- **Part-of-Speech (POS)**: Lexikalische Kategorie/Wortart eines Wortes
- **Syntax**: beschreibt strukturelle Beziehung zwischen Wörtern
- Typische Fragen: Grammatikalität? Parse eines Satzes
- **Syntaktische Regeln**: Prozesse, die Sätze generieren können
  - VP → VVFIN NP
  - VP → VP PP
- **Parsing**: Zähle die syntaktischen Strukturen (parses) von Sätzen auf und entscheide Dich für eine präferierte Struktur; benutzt syntaktische Regeln
- Präferenzen: oft auch wieder statistisch


|    |   |          |  |     |
|----|---|----------|--|-----|
| VP | → | VVFIN NP |  | 0.7 |
| VP | → | VP PP    |  | 0.3 |

- 1 Kann *grün* in *ein grüner Junge* durch *unerfahren* ersetzt werden?  
Wie steht es bei *ein grüner Baum*? Wieviele Bedeutungen von *grün* gibt es?
- 2 Welches Wort fällt Ihnen als erstes ein, wenn Sie hören: *Apfel*
- 3 Welches Wort passt nicht: *Hut, Hose, Rock, Papier, Pullover*
- 4 Was bedeutet *Jeder Holländer besitzt einen Wohnwagen*?

- 1 Kann *grün* in *ein grüner Junge* durch *unerfahren* ersetzt werden? Wie steht es bei *ein grüner Baum*? Wieviele Bedeutungen von *grün* gibt es?
- 2 Welches Wort fällt Ihnen als erstes ein, wenn Sie hören: *Apfel*?
- 3 Welches Wort passt nicht: *Hut, Hose, Rock, Papier, Pullover*?
- 4 Was bedeutet *Jeder Holländer besitzt einen Wohnwagen*?

- 1 Kann *grün* in *ein grüner Junge* durch *unerfahren* ersetzt werden? Wie steht es bei *ein grüner Baum*? Wieviele Bedeutungen von *grün* gibt es?
- 2 Welches Wort fällt Ihnen als erstes ein, wenn Sie hören: *Apfel*
- 3 Welches Wort passt nicht: *Hut, Hose, Rock, Papier, Pullover*
- 4 Was bedeutet *Jeder Holländer besitzt einen Wohnwagen*?

- ① Kann *grün* in *ein grüner Junge* durch *unerfahren* ersetzt werden? Wie steht es bei *ein grüner Baum*? Wieviele Bedeutungen von *grün* gibt es?
- ② Welches Wort fällt Ihnen als erstes ein, wenn Sie hören: *Apfel*
- ③ Welches Wort passt nicht: *Hut, Hose, Rock, Papier, Pullover*
- ④ Was bedeutet *Jeder Holländer besitzt einen Wohnwagen*?

- 1 Kann *grün* in *ein grüner Junge* durch *unerfahren* ersetzt werden? Wie steht es bei *ein grüner Baum*? Wieviele Bedeutungen von *grün* gibt es?
- 2 Welches Wort fällt Ihnen als erstes ein, wenn Sie hören: *Apfel*
- 3 Welches Wort passt nicht: *Hut, Hose, Rock, Papier, Pullover*
- 4 Was bedeutet *Jeder Holländer besitzt einen Wohnwagen*?

- ① Kann *grün* in *ein grüner Junge* durch *unerfahren* ersetzt werden? Wie steht es bei *ein grüner Baum*? Wieviele Bedeutungen von *grün* gibt es?
- ② Welches Wort fällt Ihnen als erstes ein, wenn Sie hören: *Apfel*
- ③ Welches Wort passt nicht: *Hut, Hose, Rock, Papier, Pullover*
- ④ Was bedeutet *Jeder Holländer besitzt einen Wohnwagen*?

- **Semantik:** Wissenschaft der Bedeutung
- **Lexikalische Semantik:** Wortbedeutung
- Distributionelle Semantik: Worte, die in gleichen Kontexten auftauchen teilen, semantische Bedeutung
- Vektorsemantik, lineare Algebra, neuronale Netze

- **Satzsemantik und das Kompositionalitätsprinzip:** Bedeutung eines Satzes ergibt sich aus der Bedeutung seiner Teile.  
Ausnahmen?
- **Semantische Analyse:** Bilde linguistischen Input auf formale Bedeutungsrepräsentationen ab. Semantische Relationen.
- Prädikatenlogik, Lambda-Kalkül (siehe *Einführung in die Logik*)

- ❶ Sie gehen zum Bahnhof und sagen am Verkaufsschalter: "Ich möchte nach Hamburg". Die Antwort ist: "Ich auch. Letztes Jahr war ich in der Elbphilharmonie und das Jahr davor...". Wurde die Bedeutung Ihres Satzes richtig verstanden?
- ❷ Sie reden mit einer Freundin und sagen: "Ich möchte nach Hamburg". Die Antwort ist: "Das macht 60.50 Euro. Du musst zweimal umsteigen". Wurde die Bedeutung Ihres Satzes richtig verstanden?
- ❸ Warum macht der folgende Dialog Sinn? A: *Gehst Du zu Mias Geburtstag?* B: *Ich muss arbeiten.*

- ① Sie gehen zum Bahnhof und sagen am Verkaufsschalter: “Ich möchte nach Hamburg”. Die Antwort ist: “Ich auch. Letztes Jahr war ich in der Elbphilharmonie und das Jahr davor...”. Wurde die Bedeutung Ihres Satzes richtig verstanden?
- ② Sie reden mit einer Freundin und sagen: “Ich möchte nach Hamburg”. Die Antwort ist: “Das macht 60.50 Euro. Du musst zweimal umsteigen”. Wurde die Bedeutung Ihres Satzes richtig verstanden?
- ③ Warum macht der folgende Dialog Sinn? A: *Gehst Du zu Mias Geburtstag?* B: *Ich muss arbeiten.*

- ① Sie gehen zum Bahnhof und sagen am Verkaufsschalter: “Ich möchte nach Hamburg”. Die Antwort ist: “Ich auch. Letztes Jahr war ich in der Elbphilharmonie und das Jahr davor...”. Wurde die Bedeutung Ihres Satzes richtig verstanden?
- ② Sie reden mit einer Freundin und sagen: “Ich möchte nach Hamburg”. Die Antwort ist: “Das macht 60.50 Euro. Du musst zweimal umsteigen”. Wurde die Bedeutung Ihres Satzes richtig verstanden?
- ③ Warum macht der folgende Dialog Sinn? *A: Gehst Du zu Mias Geburtstag? B: Ich muss arbeiten.*

- **Pragmatik:** Wie wird Sprache gebraucht, um Ziele zu erreichen
- **Sprechakt:** Äußerung hat eine performative Funktion zum Beispiel Bitte, Warnung, Frage
- Kommunikative Ziele folgen (kulturspezifischen) Konventionen!
- **Grice Principles of Conversation:** Quantität, Qualität, Relevanz, Manner (brief, non-ambiguous)

- 1 *Shakespeare war einer der produktivsten Schriftsteller des Elisabethanischen Zeitalters und unter König James immer noch beliebt. **Er** schrieb 38 Stücke. Auf wen referiert er?.*
- 2 *Wie sind die folgenden Satzteile relationiert? Anna arbeitete an ihren ECL Übungen während sie Musik hörte.*
- 3 *Wie sind die folgenden Satzteile relationiert? Anna arbeitete an ihren ECL Übungen während Mia auf eine Party ging.*

- 1 *Shakespeare war einer der produktivsten Schriftsteller des Elisabethanischen Zeitalters und unter König James immer noch beliebt. **Er** schrieb 38 Stücke. Auf wen referiert er?.*
- 2 *Wie sind die folgenden Satzteile relationiert? Anna arbeitete an ihren ECL Übungen während sie Musik hörte.*
- 3 *Wie sind die folgenden Satzteile relationiert? Anna arbeitete an ihren ECL Übungen während Mia auf eine Party ging.*

- 1 *Shakespeare war einer der produktivsten Schriftsteller des Elisabethanischen Zeitalters und unter König James immer noch beliebt. **Er** schrieb 38 Stücke. Auf wen referiert er?.*
- 2 *Wie sind die folgenden Satzteile relationiert? *Anna arbeitete an ihren ECL Übungen während sie Musik hörte.**
- 3 *Wie sind die folgenden Satzteile relationiert? *Anna arbeitete an ihren ECL Übungen während Mia auf eine Party ging.**

- **Diskurs:** Linguistik über die Satzgrenze hinaus.
- Hauptfrage: was macht einen Text kohärent?
- **Referenzresolution:** Welche Entitäten in einem Text sind koreferent?
- **Rhetorische Relationen/Diskursrelationen/Kohärenzrelationen:**  
Diskursrelationen zwischen Sätzen (z.B. temporal, Kausal etc).

- Phonetisch: [ra:t]
- Betonung: *Ich habe nie gesagt, dass er in meine Wohnung eingebrochen ist.*
- Lexikalische Bedeutung: *grün, Bank*
- POS: *engl.: duck*
- syntaktische Anbindung: *I saw a man with a beard/telescope*
- Kompositionelle Bedeutung: *Jeder Holländer hat einen Wohnwagen.*
- Sprechakt: *Ich will nach Hamburg.*

*Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.*

Wieviele Lesarten besitzt dieser Satz?

$$2 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 2 \cdot 4 \cdot 2 \cdot 4 \cdot 2 \cdot 2 \cdot 7 \cdot 2 = 258.048$$

Quelle: Hans Uszkoreit

*Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.*

- *Früher* kann eigenständiges Adverb oder Komparativ von *früh* sein (2);
- die Verbform *stellten* ist ambig zwischen Präteritum und Konjunktiv (2)
- die Nominalphrase *die Frauen* kann Subjekt oder Objekt des Satzes sein (2)
- *am Wochenende* kann *die Insel*, *die Frauen* oder das Verb modifizieren (3);
- *mit Blumenmotiven* kann sich auf *die Kopftücher* beziehen, ein Instrument der Herstellung sein oder ein Adjunkt im Sinne von gemeinsam mit Blumenmotiven(3);
- *Her* hat auch eine direktionale Bedeutung (2)

Quelle: Hans Uszkoreit

*Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.*

- der Relativsatz könnte jede der vier Nominalphrasen im Plural modifizieren (4);
- *die* als auch *ihre Männer* kann Subjekt des Relativsatzes sein (2);
- das Possessivpronomen *ihre* kann auf jede der Nominalphrasen referieren (4);
- *Montagen* hat eine zweite Lesart als Nominalisierung von montieren (2);
- *Hauptinsel* kann im Genitiv zu der vorangegangenen NP gehören oder im Dativ die Käuferin bezeichnen (2);
- die drei Präpositionalphrasen des Relativsatzes können sich in insgesamt sieben Kombinationen mit den jeweils vorhergehenden NPs oder mit dem Verb verbinden (7);
- *Verkauften* zeigt wieder die Ambiguität zwischen Präteritum und Konjunktiv auf (2).

- 1 Was ist Computerlinguistik?
- 2 Anwendungsbeispiele
- 3 Linguistik und CL: Fragen und Teilgebiete
- 4 Was passiert in diesem Kurs?**



- Viele Anwendungen brauchen nicht alle diese Schritte, sondern gehen direkt von Input zu Output
- Beispiel in ECL: Textklassifikation
- Beispiele in späteren Seminaren und Vorlesungen: Textzusammenfassung, MT
- Alternativen zur End-to-end Architektur: Gemeinsame Modelle, neuronale Netzwerke (arbeiten oft direkt auf Inputtext)

## Aufgabenarten:

- Klassifikation: Jede Instanz bekommt ein vordefiniertes Klassenlabel
- Regression: Jede Instanz bekommt ein numerisches Label
- Clustering: Eine Menge von Inputinstanzen wird in nicht vorherdefinierte Klassen gruppiert

## Ansätze:

- Überwacht: Trainingsinstanzen mit Lösung werden in Algorithmusentwicklung benutzt
- Unüberwacht: Trainingsinstanzen mit Lösung werden in Algorithmen nicht benutzt

- Was macht Sprache als Kommunikationssystem besonders?
- Vergleich mit anderen Kommunikationssystemen?

- Text (und einzelne Worte) als Sequenz von Zeichen
- **Reguläre Ausdrücke:** Muster für Text, mit denen man (mit jeder Programmiersprache) suchen kann
- Beispiel: `Baa+!`  $\longrightarrow$  `Baa!`, `Baaa!` . . .
- Meist gebraucht um Textteile zu finden, die klaren sequentiellen Mustern folgen
- Endliche Automaten und reguläre Grammatiken als formaler Mechanismus hinter regulären Ausdrücken

- **Tokenisierung:** wo fangen Wörter und Sätze an, wo hören sie auf?
- **Regelbasierte Algorithmen:** Wenn der jetzige Buchstabe ein Punkt ist, und vorher keine bekannte Abkürzung, Zahl oder URL steht, dann ist der Satz zu Ende
- **Normalisierung:** z.B Stemming *misclassified* → *classify*
- Effiziente Algorithmen für Stringvergleich (Edit-Distanz zwischen *read* und *lad*): dynamic programming

- Wie sehen Wortverteilungen in Texten aus?
- Plausibilität von Wort- oder Buchstabenfolgen: *drink white ...*
- **Smoothing**: Was mache ich mit ungesehenen Wortfolgen?
- Anwendung: Language Identification
- Hintergrund: Wiederholung Wahrscheinlichkeitstheorie sowie Einführung in Informationstheorie

- Was kann man schon mit nur Worten und ein paar Wortfolgen tun? **Textklassifikation!**
- Klassifiziere Texte nach Inhalt, Genre, Autor, Meinung
- Hintergrund: maschinelles Lernen

Just came out of the theater and I'm literally blowing away! As a moviegoer and movie lover looking for a good entertaining is simply irresistible not to like this movie even just a little.

I love james bond, I've seen all the films, and i can say this is the worst one, dull , meandering script, at times i had no idea what the plot was. lots of confusiion

## Maschinelles Lernen

- Algorithmus, der aus Erfahrung lernt in Bezug auf eine Aufgabe und eine Performanzmetrik
- Überwacht oder unüberwacht
- Zielfunktion soll gelernt werden, die einen Input auf einen Output matched

Wir lernen 3 überwachte Klassifikationsmodelle kennen: Naive Bayes, logistische Regression und das Perzeptron

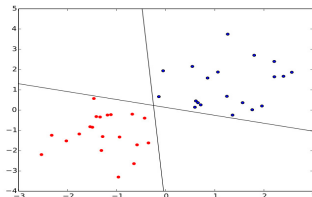
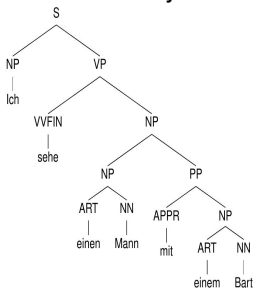


Bild von Qwertyus - Own work, CC BY-SA 4.0,

<https://commons.wikimedia.org/w/index.php?curid=41351528>

- **Distributionelle lexikalische Semantik:** Kann ich die Vorkommen eines Wortes in einem Textkorpus nutzen, um etwas über dessen Bedeutung zu erfahren? Um Ähnlichkeiten zwischen Worten zu berechnen?
- Hintergrund: lineare Algebra, sparse and dense word embeddings, clustering

- In welche Klassen (z.B. Nomen, Verben, Adjektive) kann ich Wörter einteilen? Wie kann ich diese Klassen bei Worten in einem Text automatisch zuordnen?
- Wie kreiire ich Syntaxbäume, automatisch und manuell?



- Wie behandle ich syntaktische Ambiguitäten?
- Methoden: formale Sprachen, Grammatiken, Suchalgorithmen, Hidden-Markov-Modelle, probabilistische Grammatiken, dynamische Programmierung

Möglichkeiten:

- **Koreferenzresolution:** *Peter's car is in the garage. It is red.*
  - Hintergrund: linguistische Constraints, Suchalgorithmen, Graphen
  - Weitere Klassifikationsalgorithmen: Feed-forward Neural Networks
  - Clustering: unüberwachtes Gruppieren von Instanzen in Gruppen
- Beispieldemo:

<https://projector.tensorflow.org/>

- NLP Modelle werden auf **Textkorpora** entwickelt und meist auch evaluiert
- Korpora: Sammeln und Aufbereiten sowie Sampling
- **Evaluationskriteria:** Effektivität/Akkuratheit; Effizienz; Robustheit
- Typische Effektivitätsmaße im Vergleich mit Ground Truth sowie Baselines
- Wie weiß ich, ob ich besser geworden bin? Hypothesentests

# Zusammenfassung: Was macht NLP schwierig?

- Effektivität: Ambiguität, Fehlender Kontext, Fehlerakkumulierung in Pipeline, Mangelnde Daten
- Effizienz: große Datenmengen müssen gesammelt, gespeichert und verarbeitet werden
- Robustheit: Datensätze können domänenspezifisch sein, oder haben Bias

- Mindestens eine der Anwendungen oder Demos ausprobieren
- Vorbereitung für nächstes Mal: Wie definieren Sie ein Kommunikationssystem?
- Vorbereitung für nächstes Mal: Was unterscheidet Sprache von anderen Kommunikationssystemen?