

# Wie funktioniert eigentlich Google Translate?

Utz Pöhlmann

Universität Hamburg - Proseminar: Künstliche Intelligenz

*4poehlma@informatik.uni-hamburg.de*

2. August 2015

# Warum Sprachübersetzungssysteme?

# Warum Sprachübersetzungssysteme?

- mehr als 7000 Sprachen weltweit

# Warum Sprachübersetzungssysteme?

- mehr als 7000 Sprachen weltweit
- Reisen sind „IN“

# Warum Sprachübersetzungssysteme?

- mehr als 7000 Sprachen weltweit
- Reisen sind „IN“
- Übersetzung überall und für jedermann

- 1 Einleitung
- 2 Sprachübersetzung von Grund auf
- 3 Schablonentechnik
- 4 Fazit & Ausblick

# WLAN?

WLAN ist vorhanden!

WLAN ist nicht vorhanden!

# Ein Übersetzungsbeispiel



Sign in

Translate




English Spanish French English - detected ▼



German English Spanish ▼

Translate

|







Type text or a website address or [translate a document](#).






# Ein Übersetzungsbeispiel

[Sign in](#)

**Translate**

EnglishSpanishFrenchEnglish - detected



GermanEnglishSpanish

[Translate](#)

undefined

Ä 🔊 ✎

nicht definiert

☆ 📄 🔊 ✎ [Wrong?](#)

### Definitions of undefined

*adjective*  
not clear or defined.  
"undefined areas of jurisdiction"  
*synonyms:* unspecified, unexplained, unspecific, in...


### Translations of undefined

*adjective*

■ undefiniert	undefined
■ nicht definiert	undefined
■ undefinierbar	indefinable, undefinable

See also

undefine



# Ein Übersetzungsbeispiel

The screenshot shows the Google Translate web interface. At the top, the Google logo and a 'Sign in' button are visible. Below the logo, the word 'Translate' is displayed in red. The interface shows a translation from English to German. The English input is 'undefined', and the German output is 'nicht definiert'. A dropdown menu is open, showing alternative translations: 'nicht definiert', 'undefiniert', 'undefinierten', 'undefinierte', and 'undefinierbar'. A tooltip above the dropdown says 'Click to edit and see alternate translations'. Below the input field, there are icons for pronunciation and a 'Wrong?' link. The page also includes sections for 'Definitions of undefined' and 'Translations of undefined'. The 'Definitions of undefined' section lists 'adjective not clear or defined.' and 'undefined areas of jurisdiction' with synonyms. The 'Translations of undefined' section lists 'adjective' translations: 'undefiniert', 'nicht definiert', and 'undefinierbar'.

Google

Sign in

Translate

English Spanish German English - detected

German English Spanish Translate

Click to edit and see alternate translations

undefined

Ä 🔊 ✎

Definitions of undefined

*adjective*

not clear or defined.  
"undefined areas of jurisdiction"  
synonyms: unspecified, unexplained, unspecific, in...

See also

undefined

nicht definiert

nicht definiert  
undefiniert  
undefinierten  
undefinierte  
undefinierbar

Improve this translation


Translations of undefined



*adjective*

undefiniert undefined  
nicht definiert undefined  
undefinierbar indefinable, undefinable


Wrong?

# Ein Übersetzungsbeispiel

[Sign in](#)



Translate




EnglishSpanishGermanGerman - detected



GermanEnglishSpanish

Translate

Die Weisheit jagt mich, aber ich bin schneller!  


The wisdom of chasing me, but I am faster!  
 Wrong?

[Google Translate for Business: Translator Toolkit](#)[Website Translator](#)[Global Market Finder](#)

# Ein Übersetzungsbeispiel

The screenshot shows the Google Translate web interface. At the top is the Google logo and a 'Sign in' button. Below is the 'Translate' heading. The interface features two input boxes for text and two dropdown menus for selecting the source and target languages. The source language is set to 'German - detected' and the target language is 'English'. The German text 'Die Weisheit jagt mich, aber ich bin schneller!' is entered in the left box, and its English translation 'The wisdom of chasing me, but I am faster!' is shown in the right box. A tooltip with the text 'Click to edit and see alternate translations' points to the word 'wisdom' in the English text. At the bottom, there are links for 'Google Translate for Business', 'Translator Toolkit', 'Website Translator', and 'Global Market Finder'.

Google

Sign in

Translate

g+

English Spanish German German - detected

German English Spanish Translate

Click to edit and see alternate translations

Die Weisheit jagt mich, aber ich bin schneller!

The wisdom of chasing me, but I am faster!

Wrong?

Google Translate for Business: [Translator Toolkit](#) [Website Translator](#) [Global Market Finder](#)

# Ein Übersetzungsbeispiel

The screenshot shows the Google Translate web interface. At the top, the Google logo is on the left, and a 'Sign in' button is on the right. Below the logo, the word 'Translate' is displayed in red. To the right of 'Translate' are the Google+ logo and a bookmark icon. The main interface features two language selection dropdowns: the first is set to 'German - detected' and the second to 'English'. A blue 'Translate' button is positioned to the right of these dropdowns. The input text box on the left contains the German sentence 'Die Weisheit jagt mich, aber ich bin schneller!'. The output text box on the right shows the English translation 'The wisdom of chasing me, but I'. A dropdown menu is open under the word 'The wisdom of', listing suggestions: 'The wisdom of', 'The wisdom', 'Wisdom', 'Die Weisheit', and 'Improve this translation'. A small tooltip with the word 'Weisheit' is visible next to the 'Wisdom' suggestion. Below the output text, there is a 'Wrong?' link with a pencil icon. At the bottom of the page, there are links for 'Google Translate for Business: Translator Toolkit', 'Website Translator', and 'Global Market Finder'.

Google

Sign in

Translate

g+

English Spanish German German - detected

German English Spanish

Translate

Die Weisheit jagt mich, aber ich bin schneller!

The wisdom of chasing me, but I



The wisdom of  
The wisdom  
Wisdom  
Die Weisheit  
Improve this translation



Weisheit

Wrong?


Google Translate for Business: [Translator Toolkit](#) [Website Translator](#) [Global Market Finder](#)

# Ein Übersetzungsbeispiel

[Sign in](#)

Translate



EnglishSpanishGermanGerman - detected







GermanEnglishSpanish

Translate

Die Weisheit jagt mich, aber ich bin schneller!



Wisdom chasing me, but I am faster!

 Wrong?  Undo edits

Your contribution will be used to improve translation quality and may be shown to users anonymously

ContributeClose



chasing 



< > x

See also


chase

# Ein Übersetzungsbeispiel

[Sign in](#)

Translate



EnglishSpanishGermanGerman - detected






GermanEnglishSpanish

Translate

Die Weisheit jagt mich, aber ich bin schneller!




Wisdom chasing me, but I am faster!




chasing  
chases  
hunts  
hunting  
hunt

Verfolgungsjagden

Wrong?  Undo edits

Your contribution will be used to improve translation quality and may be shown to users anonymously



ContributeClose



chasing 

< > x


See also  
chase

# Ein Übersetzungsbeispiel

[Sign in](#)



Translate






EnglishSpanishGermanGerman - detected



GermanEnglishSpanish


Translate

Die Weisheit jagt mich, aber ich bin schneller!  
 

Wisdom **chases** me, but I am faster!  
    Wrong?  Undo edits

Your contribution will be used to improve translation quality and may be shown to users anonymously

ContributeClose

chases 



< > x



See also

chase




# Ein Übersetzungsbeispiel


[Sign in](#)



Translate

EnglishSpanishGermanGerman - detected ▾








GermanEnglishSpanish ▾[Translate](#)





Die Weisheit jagt mich, aber ich bin schneller! 

Wisdom **chases** me, but I am faster!

    Wrong?  Undo edits

Thank you for your submission.

chases    

See also  
chase

# Arten von Übersetzungssystemen

- Wort für Wort

# Arten von Übersetzungssystemen

- Wort für Wort
- Schablonentechnik

# Arten von Übersetzungssystemen

- Wort für Wort
  - Übersetzen Marke Lexikon
  - Sinnabschnitte und Wahrscheinlichkeiten
- Schablonentechnik

- Gehe zum ersten Wort

Vorsicht ist die Mutter der  
Porzellankiste.

- Gehe zum ersten Wort
- Schau ins Lexikon

*die Vorsicht – attention*

*die Vorsicht – caution*

*die Vorsicht – wariness*

*Vorsicht! – Watch out!*

[...]

- Gehe zum ersten Wort
- Schau ins Lexikon
- Nimm die erste Wortbedeutung

**attention** ist die Mutter der Porzellankiste.

- Gehe zum ersten Wort
- Schau ins Lexikon
- Nimm die erste Wortbedeutung
- Gehe zum nächsten Wort

attention **ist** die Mutter der  
Porzellankiste.



- Gehe zum ersten Wort
- Schau ins Lexikon
- Nimm die erste Wortbedeutung
- Gehe zum nächsten Wort
- Fahre fort mit Schritt 2

attention actual amount that  
mother that Porzellankiste.

Befindet sich ein Wort nicht im Lexikon, führe folgende Schritte durch:

*Porzellankiste* — ?

Befindet sich ein Wort nicht im Lexikon, führe folgende Schritte durch:

- Benutze ein einsprachiges Wörterbuch



**Abbildung:** Ein beispielhaftes einsprachiges Wörterbuch

**Quelle:** Google

Befindet sich ein Wort nicht im Lexikon, führe folgende Schritte durch:

*Porzellankiste = Porzellan + Kiste*

- Benutze ein einsprachiges Wörterbuch
- Zerlege das Wort in Teilworte die im Wörterbuch stehen

Befindet sich ein Wort nicht im Lexikon, führe folgende Schritte durch:

- Benutze ein einsprachiges Wörterbuch
- Zerlege das Wort in Teilworte die im Wörterbuch stehen
- Übersetze diese Teilworte

*das Porzellan – china*

[...]

*die Kiste – box*

[...]

Befindet sich ein Wort nicht im Lexikon, führe folgende Schritte durch:

attention actual amount that mother that china box.

- Benutze ein einsprachiges Wörterbuch
- Zerlege das Wort in Teilworte die im Wörterbuch stehen
- Übersetze diese Teilworte
- Fahre fort mit dem nächsten Wort

- Parallelen
  - Englisch und Deutsch beides germanische Sprachen
  - also ähnliche Grammatik und ähnlicher Wortschatz

- Parallelen
- Worttypen

*Vorsicht – Substantiv*

*ist – Verb*

*die – Artikel*

*Mutter – Substantiv*

*der – Artikel*

*Porzellankiste – Substantiv*



- Parallelen
- Worttypen
- Voraussagen

Beziehungen:

*Vorsicht – keine erkennbar*

*ist – zu Substantiv*

*die – zum nächsten Wort*

*Mutter – keine erkennbar*

*der – zum nächsten Wort*

*Porzellankiste – keine erkennbar*

Bisher keine sinnlosen Daten

- Parallelen
- Worttypen
- Voraussagen
- Filterung

- Parallelen
- Worttypen
- Voraussagen
- Filterung
- Auftrennen in Wortstamm und Affix

*Vorsicht – Stamm*

*ist – Stamm*

*die – Stamm*

*Mutter – Stamm*

*der – Stamm*

*Porzellankiste – Stamm + Stamm*

- Parallelen
- Worttypen
- Voraussagen
- Filterung
- Auftrennen in Wortstamm und Affix
- Übersetzen unter Beachtung der Wortstellung und der Fälle

Carefulness is the crate of porcelain's mother.

- Textabschnitte in anderen Texten wiederfinden
- Vorsicht ist die Mutter der Porzellankiste.
- Vorsicht ist die Mutter der Porzellankiste.
- Vorsicht ist die Mutter der Porzellankiste.
- [...] Vorsicht ist die Mutter der Porzellankiste [...]

- Textabschnitte in anderen Texten wiederfinden
- Übersetzung aus dem anderen Text nehmen
- Discretion is the better part of valor
- Prudence is the better part of valor.
- Prudence is the better part of valour.
- [...] caution is the mother of wisdom  
[...]

- Textabschnitte in anderen Texten wiederfinden
- Übersetzung aus dem anderen Text nehmen
- Als Schablone behalten

1 = *Discretion is the better part of valor*  
2 = *Prudence is the better part of valor.*  
3 = *Prudence is the better part of valour.*  
4 = *caution is the mother of wisdom*

- Textabschnitte in anderen Texten wiederfinden
- Übersetzung aus dem anderen Text nehmen
- Als Schablone behalten
- Wenn Schablone ohne gegenseitigen Konflikt den ganzen Text abdecken

Die Schablone steht in Konflikt zu:

*Schablone 1 : 2, 3, 4*

*Schablone 2 : 1, 3, 4*

*Schablone 3 : 1, 2, 4*

*Schablone 4 : 1, 2, 3*

Schablonen die ohne Konflikt den gesamten Text abdecken:

- 1
- 2
- 3
- 4



- Textabschnitte in anderen Texten wiederfinden
- Übersetzung aus dem anderen Text nehmen
- Als Schablone behalten
- Wenn Schablone ohne gegenseitigen Konflikt den ganzen Text abdecken
  - Diese Schablonen als aktuelle Übersetzung ausgeben
- Sonst kleinere Schablonen wählen

Schablone 2 kommt am häufigsten in anderen Schablonen vor.

Deshalb wird sie ausgegeben:

Prudence is the better part of valor.

- Google Translate arbeitet mit diesem Muster

- Google Translate arbeitet mit diesem Muster
- andere Systeme sind schlechter im Übersetzen von ganzen Texten als Google Translate

- Google Translate arbeitet mit diesem Muster
- andere Systeme sind schlechter im Übersetzen von ganzen Texten als Google Translate
  - Babelfish
  - Leo
  - Linguee (Gibt nur Schablonen aus, wagt keinen Übersetzungsversuch)
  - kleine, die es nicht ins Internet geschafft haben
  - und Weitere

- Google Translate bisher beste Lösung

- Google Translate bisher beste Lösung
- Eine gänzlich neue Lösung wurde bisher noch nicht gefunden

- Google Translate bisher beste Lösung
- Eine gänzlich neue Lösung wurde bisher noch nicht gefunden
- Kleinere Bugfixes bestimmt

- Google Translate bisher beste Lösung
- Eine gänzlich neue Lösung wurde bisher noch nicht gefunden
- Kleinere Bugfixes bestimmt
- Wer selber Übersetzen will, nimmt Linguee





## Utz Pöhlmann (2015)

Wie funktioniert eigentlich Google Translate?

*Entwicklungsstadium* Stand Juli 2015.



## Front Runner (2015)

Gemeinsamkeiten und Unterschiede der deutschen und englischen Sprache

<http://front-runner.de/de/gemeinsamkeiten-unterschiede-englisch-deutsch.html>

Absatz 2, Juli 2015.

# Ende

# Ende

Noch Fragen?

- Gemeinsamkeiten Ausgangs-, Zielsprache

- Gemeinsamkeiten Ausgangs-, Zielsprache
- Suchen: z.B. Google

- Gemeinsamkeiten Ausgangs-, Zielsprache
- Suchen: z.B. Google
- Besser: vorher einprogrammiert

- Substantiv

- Substantiv
  - immer groß geschrieben
  - haben oft Artikel
  - deklinierbar



- Substantiv
  - immer groß geschrieben
  - haben oft Artikel
  - deklinierbar
- Verb

- Substantiv
  - immer groß geschrieben
  - haben oft Artikel
  - deklinierbar
- Verb
  - meist klein geschrieben
  - konjugierbar

- Substantiv
  - immer groß geschrieben
  - haben oft Artikel
  - deklinierbar
- Verb
  - meist klein geschrieben
  - konjugierbar
- Adjektiv

- Substantiv

- immer groß geschrieben
- haben oft Artikel
- deklinierbar

- Verb

- meist klein geschrieben
- konjugierbar

- Adjektiv

- meist klein geschrieben
- steigerbar

- und weitere

- Beziehung von 2 Wörtern

- Beziehung von 2 Wörtern
  - zusammengehörig (z.B. Artikel + Substantiv)
  - unabhängig

- Beziehung von 2 Wörtern
  - zusammengehörig (z.B. Artikel + Substantiv)
  - unabhängig
- Beziehung von 2 Sätzen

- Beziehung von 2 Wörtern
  - zusammengehörig (z.B. Artikel + Substantiv)
  - unabhängig
- Beziehung von 2 Sätzen
  - Nebensatz
  - implizite Referenzen (z.B. Thema)
  - unabhängig



- Berechnung aller Daten

- Berechnung aller Daten
  - Wortanzahl
  - Abstände von Wörtern
  - etc.

- Berechnung aller Daten
  - Wortanzahl
  - Abstände von Wörtern
  - etc.
- Schlechte Ergebnisse verwerfen

- Auftrennung einzelner Wörter

- Auftrennung einzelner Wörter
  - 2 Bäume bauen
  - Wort von hinten und vorne durch gehen
  - gefundene Affixe entfernen

- Auftrennung einzelner Wörter
  - 2 Bäume bauen
  - Wort von hinten und vorne durch gehen
  - gefundene Affixe entfernen
- gefundene Wortstämme auf Existenz prüfen (Lexikon)
- gefundene Affixe auf Existenz prüfen (Datenbank)

- erstes Wort nehmen

- erstes Wort nehmen
- Wortstamm im Lexikon suchen
  - bei mehreren Möglichkeiten Voraussage beachten



- erstes Wort nehmen
- Wortstamm im Lexikon suchen
  - bei mehreren Möglichkeiten Voraussage beachten
- mit Affixen richtige Form finden

- erstes Wort nehmen
- Wortstamm im Lexikon suchen
  - bei mehreren Möglichkeiten Voraussage beachten
- mit Affixen richtige Form finden
- nächstes Wort übersetzen

Krasser Stuff der mir eventuell noch einfällt.