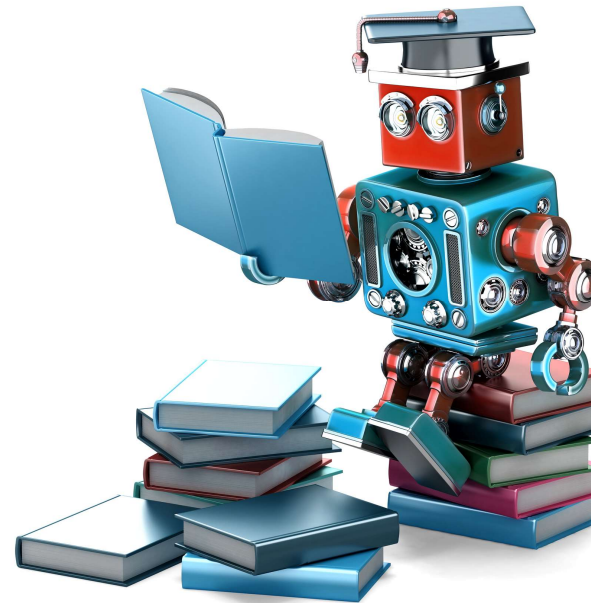


Neu in NRW!

Start im WS 2017/2018



Interdisziplinär

- Abteilung für Computerlinguistik
- Abteilung für allgemeine Sprachwissenschaft
- Institut für Informatik

## Was ist Computerlinguistik?

Computerlinguistik = Linguistik + Informatik

- Schnittstellenfach: Anwendung informatischer Methoden auf sprachwissenschaftliche Fragestellungen
- Formalisierung linguistischer Modelle
- Implementierung von Softwarewerkzeugen zur maschinellen Sprachverarbeitung:
  - Bedeutungserkennung, Wörterbucherstellung, maschinelle Übersetzung, Dialogsysteme, Marktanalysen, Plagiatserkennung, ...

## Computerlinguistik in Aktion

The screenshot shows the Google Translate web interface. At the top is the Google logo and a navigation bar with a grid icon, a red circle with the number '1', and a red circle with the letter 'E'. Below the logo is the text 'Google Übersetzer' and a link 'Sofortübersetzung deaktivieren' with a star icon. The main interface has two language selection boxes: the first with 'Englisch', 'Deutsch', 'Französisch', and 'Sprache erkennen'; the second with 'Deutsch', 'Englisch', and 'Französisch'. A blue 'Übersetzen' button is to the right. The input text on the left is 'Die Volkswirtschaftslehre (auch Nationalökonomie, Wirtschaftliche Staatswissenschaften oder Sozialökonomie, kurz VWL) ist ein Teilgebiet der Wirtschaftswissenschaft'. The output on the right is 'The economics (including economics, economics, economics, economics, economics, economics, economics, economics, economics, economics)'. Below the input text are icons for speaker, microphone, and keyboard, and a character count '164/5000'. Below the output text are icons for star, copy, speaker, and share, and a link 'Änderung vorschlagen'.

*Quelle: Google Translate.  
Ehrlich.  
Probiert es selbst aus!*

## Computerlinguistik in Aktion



Aris Toteles  
@smu1401

Folgen

Hier noch die sehr poetische Übersetzung für  
"Wir haben eine Klimaanlage"



Quelle: Tweet von @smu1401

<https://twitter.com/smu1401/status/355796756581195776>

## Computerlinguistik in Aktion

ELIZA: What are your feelings now ?  
YOU: I'm annoyed with you, to be honest  
ELIZA: Is it because you are annoyed with I that you came to  
me ?  
YOU: No????  
ELIZA: Why not ?  
YOU: You are not being very helpful right now!  
ELIZA: Do you sometimes wish you were not being very helpful  
right now ?

Quelle: Eliza-Bot  
<http://www.masswerk.at/elizabot/>

## Voraussetzungen

- Interesse an Sprache(n)
- Aufgeschlossenheit gegenüber Mathematik und Informatik

**Programmierkenntnisse sind keine Voraussetzung! Sie werden in den Grundlagenmodulen praxisbezogen vermittelt.**

## Studienziele

Eine breite computerlinguistische Ausbildung:

- Fähigkeit zur fundierten Analyse von Sprachdaten
- Verständnis mathematisch-informatischer Grundlagen
- Programmierkenntnisse und Kenntnisse in strategischer Softwareentwicklung
- Erfahrung in der Anwendung computerlinguistischer Modelle auf Sprachdaten (Hypothesenbildung, Implementierung, Evaluation)

## Berufsfelder

Vor allem in der IT/Softwareentwicklung:

- Abfrage, Aufbereitung, Verwaltung von Sprachdaten
  - Erstellung von Textdatenbanken, Wörterbucherstellung, ...
- Analyse von Sprachdaten, Spracherkennung
  - Suchmaschinen, Information Retrieval, automatische Textzusammenfassung, Sentimentanalyse, ...
- Generierung von Sprachdaten
  - Textgenerierung, Text-to-Speech, maschinelles Übersetzen, ...

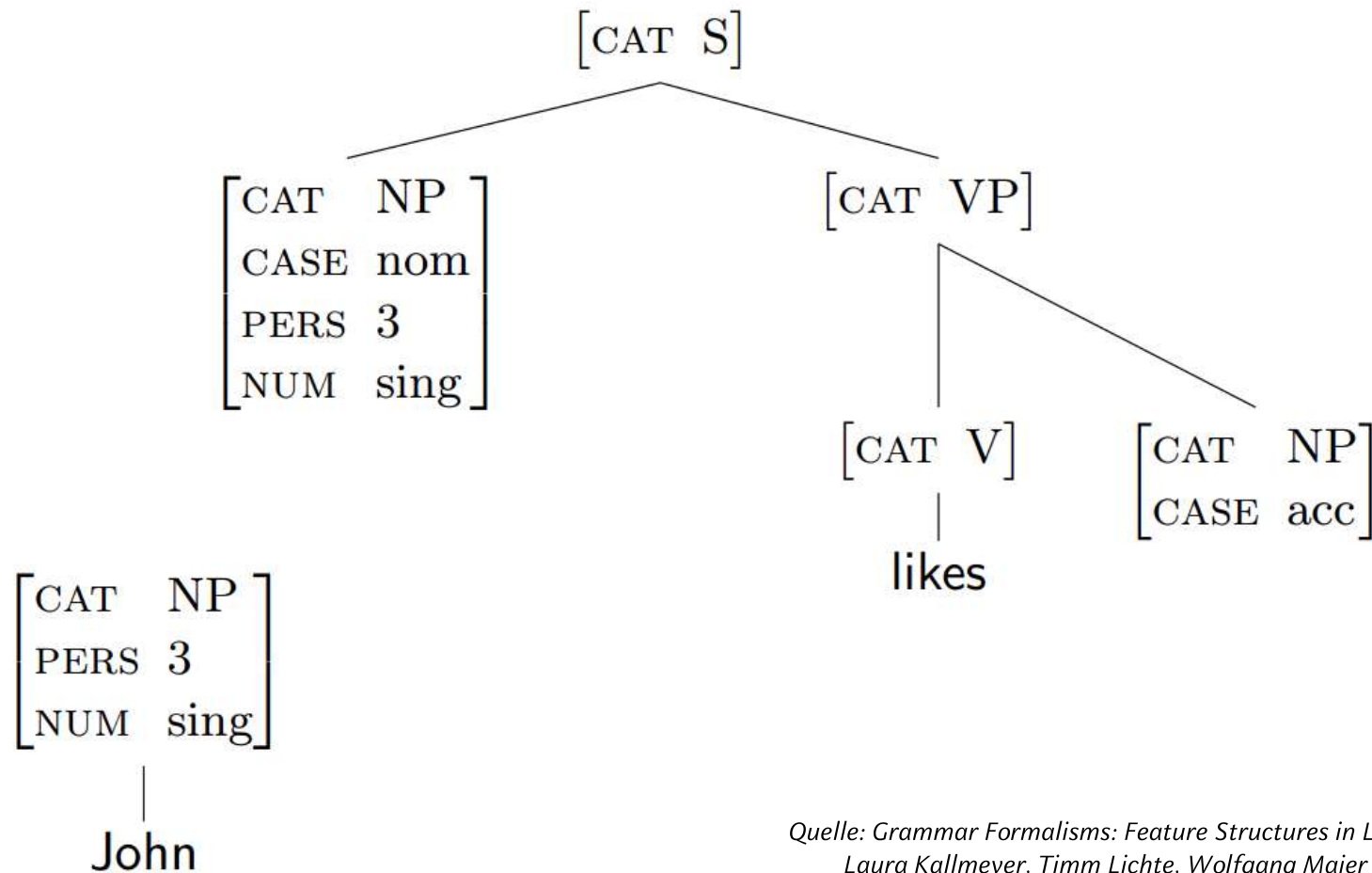


## Studienverlaufsplan

FS	Module						SWS	CP
1	<b>P: Propädeutik der CL</b> (16 SWS, 20 CP, 1 AP) Mathematische Grundlagen (4 SWS, 4 CP) Einführung in die Logik (4 SWS, 4 CP) Grundkurs Linguistik (4 SWS, 4 CP) CL Programmierung I: Python (4 SWS, 4 CP)			<b>L1: Phonetik &amp; Phonologie</b> (4 SWS, 4 CP) Einführung in die Phonetik (2 SWS, 2 CP)	<b>L2: Morphologie &amp; Syntax</b> (4 SWS, 8 CP, 1 AP) Einführung in die Morphologie (2 SWS, 2 CP)	<b>L3: Semantik &amp; Pragmatik</b> (4 SWS, 8 CP, 1 AP) Einführung in die Semantik (2 SWS, 2 CP)	22	26 +4
2	<b>CL1: Basismodul</b> (8 SWS, 12 CP, 1 AP) Einführung in die Computerlinguistik (4 SWS, 4 CP)	<b>CL2: Quantitative Methoden in der Sprachverarbeitung</b> (14 SWS, 20 CP, 1 AP) Grundlagen (2 SWS, 2 CP) CL-Programmierung II (4 SWS, 4 CP)		Einführung in die Phonologie (2 SWS, 2 CP)	Einführung in die Syntax (2 SWS, 2 CP)	Einführung in die Pragmatik (2 SWS, 2 CP)	16	24 +6
3	Grammatikformalismen (4 SWS, 4 CP)	Quantitative Methoden (4 SWS, 4 CP)	<b>CL3: Mathematische Linguistik</b> (8 SWS, 14 CP, 1 AP) Automatentheorie und formale Sprachen (4 SWS, 4 CP)	<b>CL4: Computationale Semantik</b> (12 SWS, 18CP, 1 AP) CL-Programmierung III: funktionale oder logische Programmiersprache (4 SWS, 4 CP)	<b>I: Informatik</b> (8 SWS, 10 CP, 1 AP) Programmierung (8 SWS, 10 CP)		24	30
4		Thematisches Seminar (z.B. Machine Learning) (4 SWS, 6 CP)	Thematisches Seminar (z.B. Berechenbarkeit) (4 SWS, 6 CP)	Computerlinguistische Semantik (4 SWS, 4 CP)	<b>CL5: Morphologische und syntaktische Sprachverarbeitung</b> (8 SWS, 14 CP, 1 AP) Parsing (4 SWS, 4 CP)		16	28 +2
5				Thematisches Seminar (z.B. DRT, Distributional Semantics) (4 SWS, 6 CP)	Thematisches Seminar (z.B. Tagging, Anaphor Resolution) (4 SWS, 6 CP)	<b>CL6: Vertiefungsmodul: Wahlbereich</b> (10 SWS, 14 AP, 1 AP) Linguistik, CL oder Informatik (6 SWS, 9CP)	14	29 +1
6	<b>Berufsfeldpraktikum oder Projektarbeit</b> (8CP)		<b>Abschlussarbeit</b> (2 SWS, 12 CP)			Informatik (4 SWS, 5 CP)	6	25 +5

Fächerübergreifender Wahlbereich (+18 CP)

## CL1: Einführung, Grammatikformalismen



Quelle: Grammar Formalisms: Feature Structures in LTAG  
Laura Kallmeyer, Timm Lichte, Wolfgang Maier  
<http://www.sfs.uni-tuebingen.de/emmy/gf/slides/020507-ftag.pdf>

## CL2: Quantitative Methoden

### Sentiment analysis

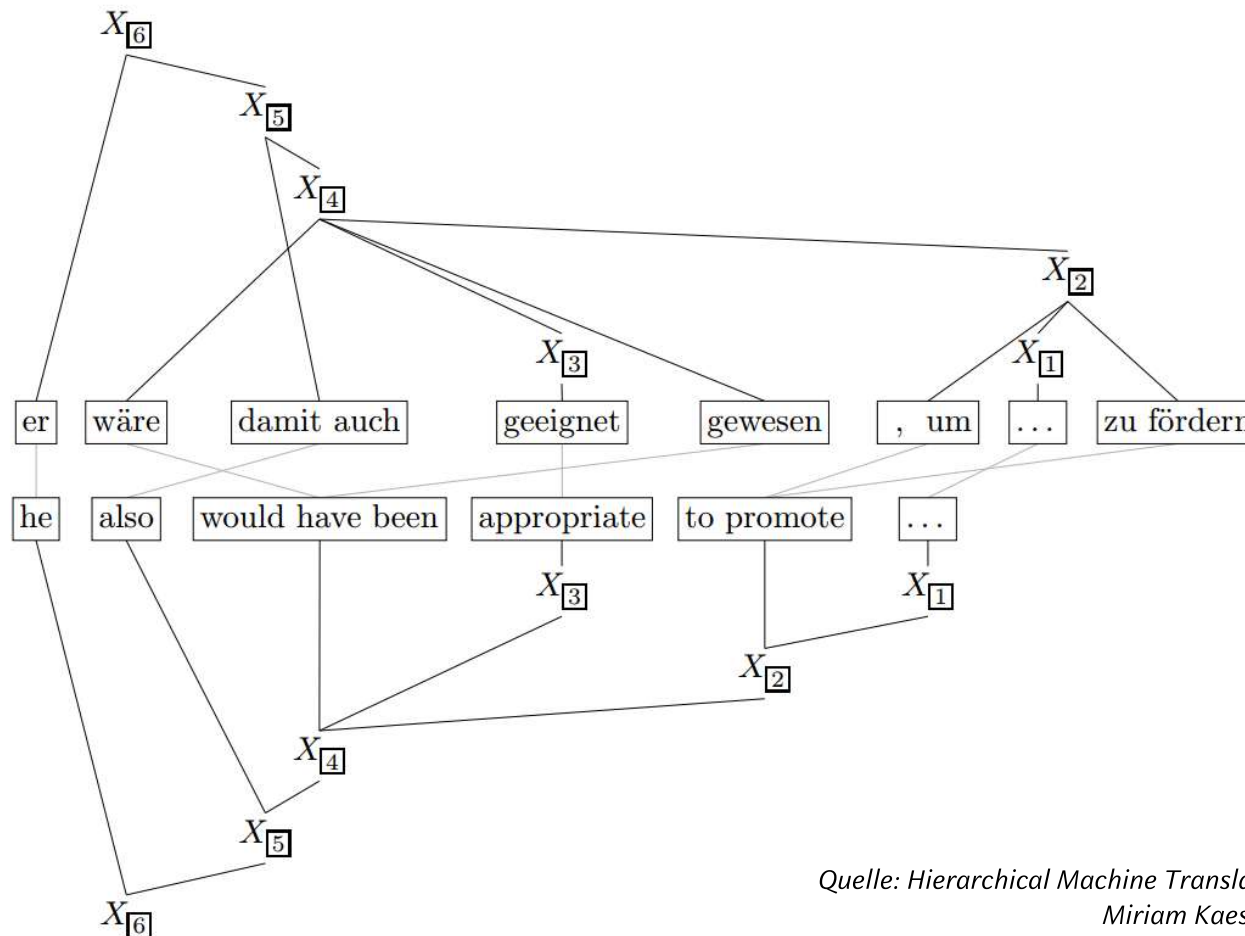
Goal: Build a classifier that assigns a class  $c \in \{\text{positive, negative, neutral}\}$  to documents, for instance to comments on films:

$c$	document
negative	“total langweilig”
negative	“auf den Film kann man echt verzichten”
negative	“ich habe selten so etwas schlechtes gesehen”
neutral	“ich habe schon bessere Filme gesehen”
neutral	“kann man angucken”
positive	“super Film”
positive	“sicher einer der besten Filme dieses Sommers”
positive	“gute Schauspieler und eine gute Story”

Quelle: *Machine Learning for Natural Language Processing*  
Laura Kallmeyer

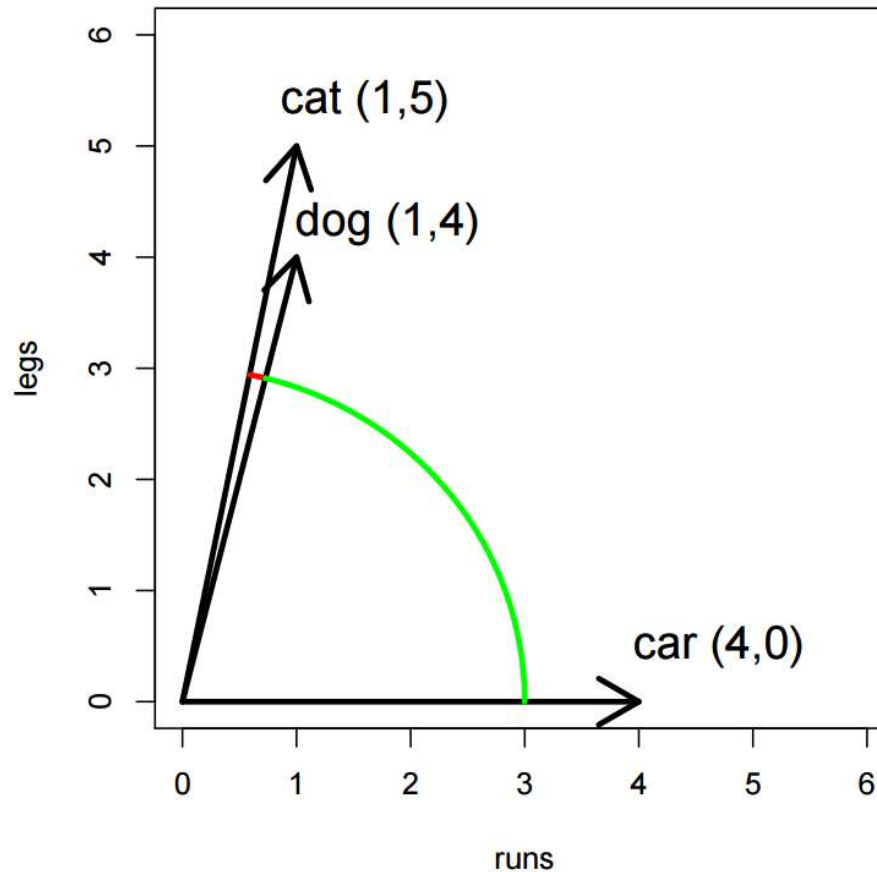
<https://user.phil.hhu.de/~kallmeyer/MachineLearning/intro.pdf>

## CL3: Mathematische Linguistik



Quelle: Hierarchical Machine Translation With Discontinuous Phrases  
Miriam Kaeshammer  
<http://www.statmt.org/wmt15/pdf/WMT28.pdf>

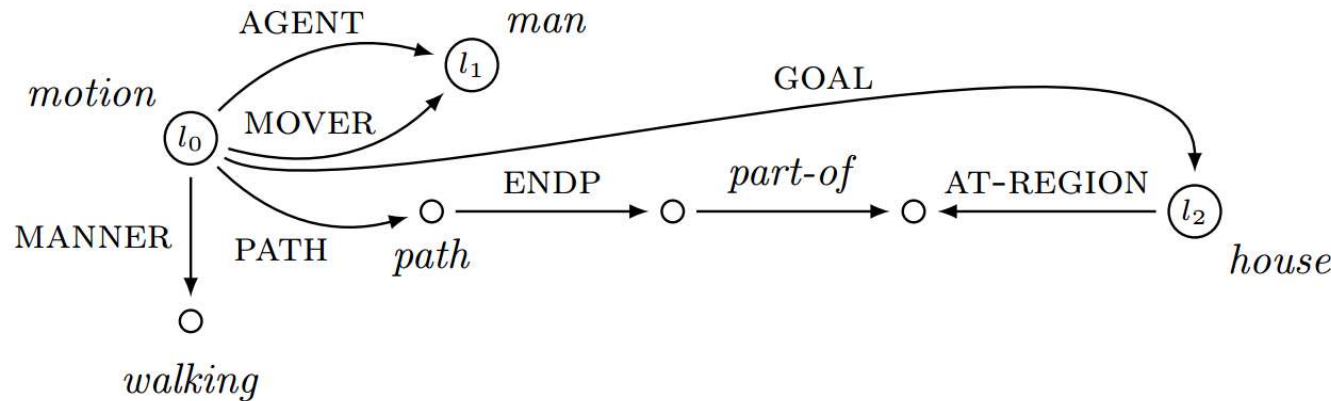
## CL4: Computationelle Semantik



Quelle: *Distributional Semantics*  
Gemma Boleda, Marco Baroni

<https://www.cs.utexas.edu/~mooney/cs388/slides/dist-sem-intro-NLP-class-UT.pdf>

## CL4: Computationelle Semantik



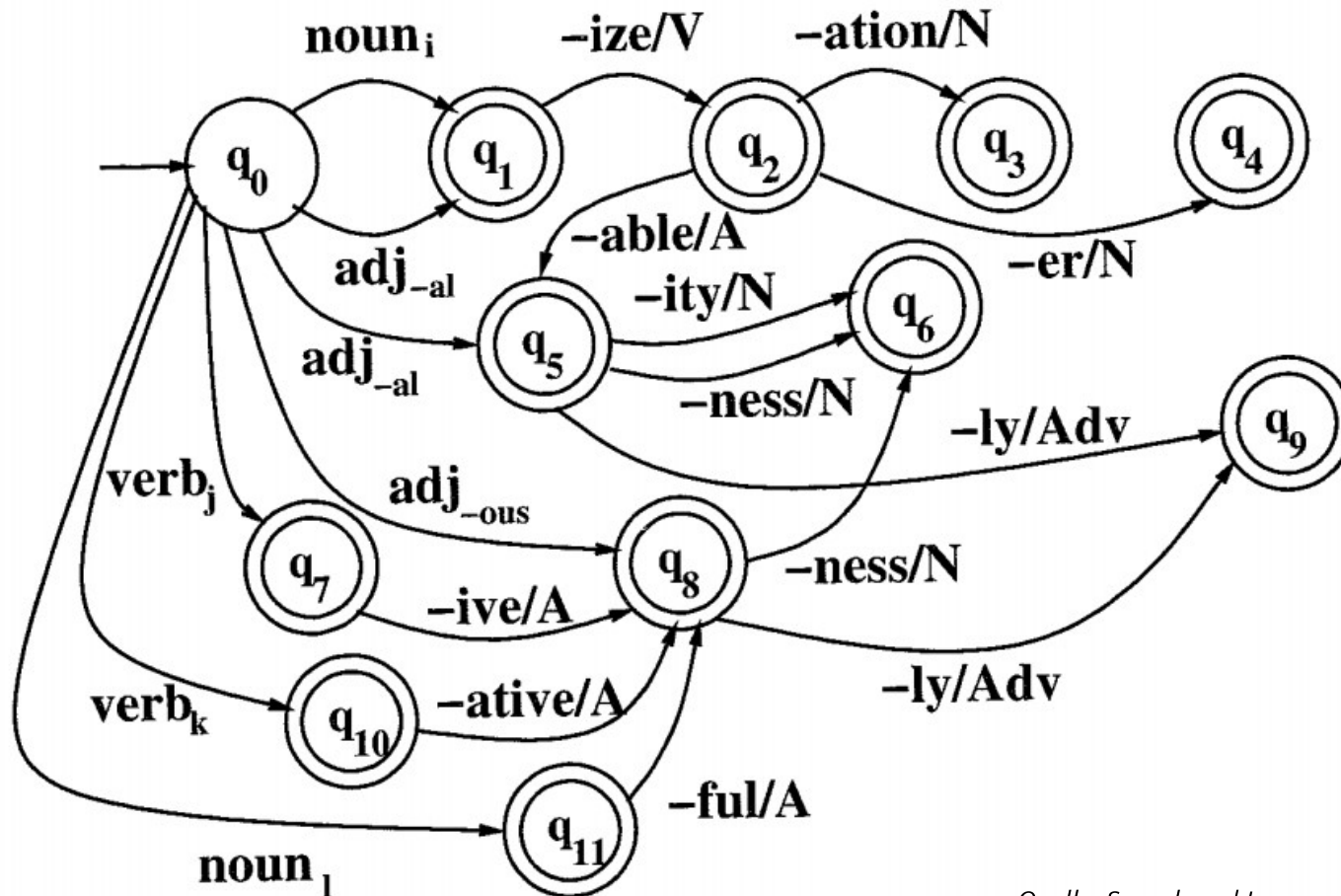
**Fig. 1.** Frame for the meaning of *the man walked to the house*

$$\begin{aligned}
 &l_0 \wedge \text{motion} \wedge \langle \text{AGENT} \rangle (l_1 \wedge \text{man}) \wedge \langle \text{MOVER} \rangle l_1 \wedge \langle \text{GOAL} \rangle (l_2 \wedge \text{house}) \wedge \\
 &\quad \langle \text{MANNER} \rangle \text{walking} \wedge (\exists v \ w. \langle \text{PATH} \rangle (\text{path} \wedge \langle \text{ENDP} \rangle v) \wedge \\
 &\quad \quad @_{l_2} ((\langle \text{AT-REGION} \rangle w) \wedge @_v (\langle \text{part-of} \rangle w))
 \end{aligned}$$

Quelle: Quantification in Frame Semantics with Hybrid Logic  
 Laura Kallmeyer, Timm Lichte, Rainer Osswald, Sylvain Pogodalla, Christian Wurm  
<https://hal.inria.fr/hal-01151641/document>



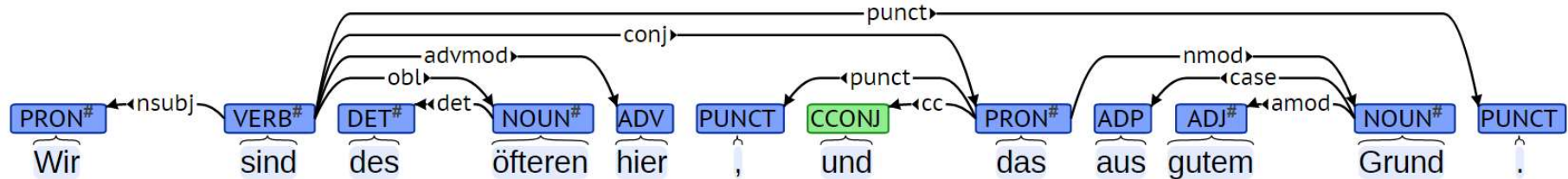
## CL5: Morphologische Sprachverarbeitung



Quelle: *Speech and Language Processing*  
Dan Jurafsky, James Martin

<http://ling.umd.edu/~idsardi/620/jurafsky/jurafsky2000-3.pdf>

## CL5: Syntaktische Sprachverarbeitung



Quelle: Universal Dependencies treebank  
Joakim Nivre and others  
<http://universaldependencies.org/>



## Lehrformen

- Vorlesung:
  - Wissensvermittlung Professor\*in → Studierende
  - Vorwiegend Frontalunterricht
  - Prüfungsform: meist Klausur
- Seminar:
  - Interaktive Wissensvermittlung
  - Teils Frontalunterricht, teils Diskussion, oft mit freiwilligen Hausaufgaben
  - Prüfungsform: Klausur, Hausarbeit oder mündliche Prüfung
- Übung:
  - Interaktive Wiederholung oder praktische Anwendung von Stoff aus anderen Lehrveranstaltungen
  - Oft von Studierenden unterrichtet
  - Keine Prüfung

## Ist Computerlinguistik das Richtige für mich?

- Bei weiteren Fragen helfen wir euch am Infostand gerne weiter!
- Bei komplizierten Anliegen könnt ihr uns auch per Mail erreichen:
  - Studiengangwechsel: Timm Lichte ([lichte@phil.hhu.de](mailto:lichte@phil.hhu.de)), Wiebke Petersen ([petersen@phil.hhu.de](mailto:petersen@phil.hhu.de))
  - Organisatorisches zum Studienbeginn: Fachschaftsrat Linguistik ([fsasw@phil.hhu.de](mailto:fsasw@phil.hhu.de))
  - Inhaltliche Fragen zum Studiengang: Timm Lichte, Wiebke Petersen, Laura Kallmeyer ([kallmeyer@phil.hhu.de](mailto:kallmeyer@phil.hhu.de))