

Computerlinguistik I

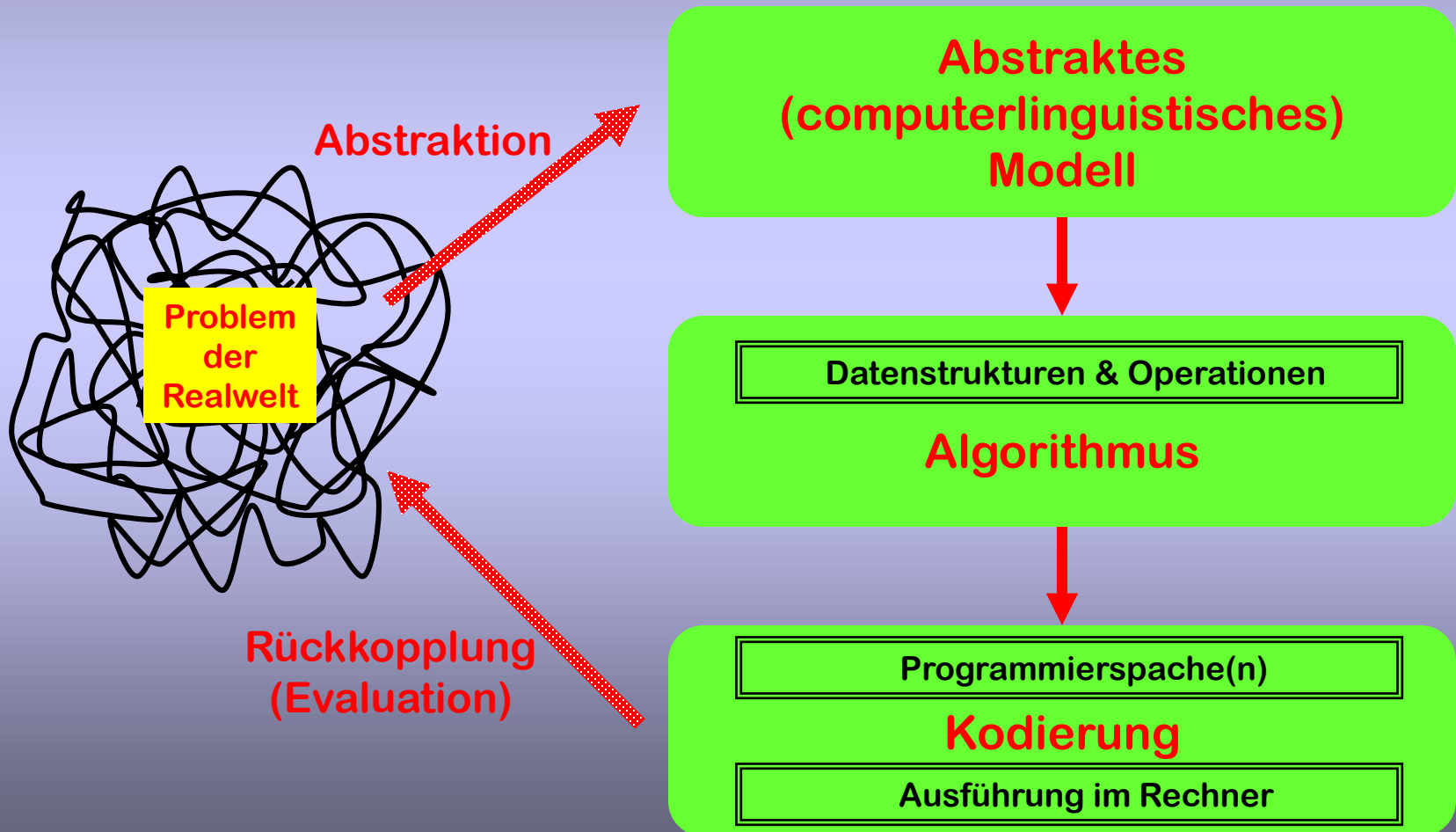
Vorlesung im WiSe 2018/19
(M-GSW-09)

Prof. Dr. Udo Hahn

Lehrstuhl für Computerlinguistik
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena

<http://www.julielab.de>

Informatischer Problemlösungszyklus



Informatischer Problemlösungszyklus

- **Modellbildung**
 - **Abstraktion** von allen unwesentlichen Details der Problemstellung im Hinblick auf die algorithmische Lösung
 - Spezifikation der **logischen** Abhängigkeiten zwischen problemlösungsrelevanten Objekten
 - **(computer)linguistisches Wissen**

Informatischer Problemlösungszyklus

- **Algorithmisierung**
 - Übersetzung der modellbezogenen Spezifikation in
 - eine Menge von **Objekten** (Datenstrukturen) mit bestimmten Eigenschaften und Beziehungen zueinander
 - die erlaubten **Operationen** auf diesen Objekten
 - **Algorithmus**: (möglichst präzise) Beschreibung einer Folge zulässiger Operationen auf den Objekten, um das Problem zu lösen
 - **Computerlinguistische Kernexpertise**

Informatischer Problemlösungszyklus

- **Kodierung (Programmierung)**
 - Übersetzung der algorithmischen Spezifikation in Konstrukte einer (geeigneten) Programmiersprache
- **Ausführung des Programms**
 - Hier erst Bezug auf konkrete Maschinen (Datenstrukturen und Algorithmen sind abstrakte Konstruktionen)
 - Test-Modifikationszyklus ... Dokumentation !
 - **Informatisches Know-How**

Morphologische Prozesse: Flexion - Deflexion

- Kombination von **Grundformen** mit **Flexionsaffixen** (Kasus, Numerus, Tempus)
 - Deklination
 - **Land**: Land, Land**es**, Land**e**, L**ä**nder, L**ä**nder**n**
 - Konjugation
 - **landen**: land**e**, land**est**, land**et**, land**eten**, **gelandet**
- primär syntaktische, nur minimale semantische Information, keine grundlegenden Wortartwechsel

Morphologische Prozesse: Derivation - Dederivation

- Kombination von **Grundformen** mit **Derivationsaffixen**
 - **Land**: landen, verlanden, anlanden,
 - **Land**: Landung, Verlandung, Anlandung
 - **Land**: ländlich, verländlichen, Verländlichung
- modifizierende semantische Information, häufig mit Wortartwechsel verbunden

Morphologische Prozesse: Komposition - Dekomposition

- Kombination von Grundformen mit Grundformen (mittels Fugeninfixen)
 - Land: Landnahme, Landflucht, Landgang
 - Land: Heimatland, Ausland, Bauland
 - Land: Landesrekord, Landesverrat, Landsmann
 - Land: Inlandsflug, Landesratspräsidentengattin
- starke semantische Modifikation, fast keine Wortartwechsel
 - ... aber: Rotkehlchen, Weichteile

Lemmatisierung vs. Wort-Parsing

Eingabe	Lemma	Wort-Parse
Töchtern	Tochter	
Houses	Haus	
sagte	sagen	
Spiegelungen	Spiegelung	
leichter	leicht	
verlängerte	verlängert	
	verlängern	

Lemmatisierung vs. Wort-Parsing

Eingabe	Lemma	Wort-Parse
Töchtern	Tochter	Tochter [+N, +FEM, +PL, +DAT]
Houses	Haus	Haus [+N, +NEU, +SG, +GEN]
sagte	sagen	sagen [+V, +SG, {1P,3P}, +PAST]
Spiegelungen	Spiegelung	[Spiegel] _N [ung] _{ds} [+N, +FEM, +PL, {NOM,GEN,DAT,AKK}]
leichter	leicht	leicht [+Adj, +POS, +MAS, +SG, +NOM] [+Adj, +KOM]
verlängerte	verlängert	[ver] _{dp} [[lang] _{Adj} [er] _{ds} Adj[t] _{ds} [+Part, {MAS,FEM,NEU}, +SG, + DEF, +NOM] [+Part, {FEM,NEU}, +SG, + DEF, +AKK]
	verlängern	[ver] _{dp} [[lang] _{Adj} [er] _{ds} Adj[n] _{ds} [+V, +SG, {1P,3P}, +PAST]