

# Morphologie

---

Vertiefung der Grundlagen der CL

WiSe 2020/21

Marcelina Wisniewska, Karolina Spiel

09.02.2021

# Überblick

---

1. Grundbegriffe
2. Flexion und Wortbildung
3. Sprachvergleich
4. Morphologie in der CL

# Grundbegriffe

---

- Morphologie:
  - *morphe* (Gestalt, Form) + *logos* (Wort, Lehre)
  - Teildisziplin der Linguistik, die sich mit der Struktur von Wörtern befasst
- Morphem
  - Die kleinste sprachliche Einheit, die eine Bedeutung oder eine grammatische Funktion hat
  - Frei oder gebunden: (*schön* + *es*) *Bild*
- Wurzel vs. Affix
  - *Tür* → *Türschlüssel*, *Haustür*
  - *ver* → *verkaufen*
- Zwischen Wurzeln und Affixen → Konfix
  - Philosemit
  - Bibliophil

# Affixe

---

- Präfix - missverstehen
- Suffix - Spieler
- Zirkumfix - gemachtt
- Infix
  - Tagalog: *sulat* „schreib-“ → *sumulat* „schreiben / jemand schreibt“
- Interfix
  - Tannenubaum
  - Besprechungsu- und Konferenzraum

# Grundbegriffe

---

- Basis - bezogen auf einen morphologischen Prozess
  - *Fußball*
  - **Fußball**tor
- Stamm - die Form, an die Flexionsaffixe angehängt werden
  - eiern → Wurzel: *ei*, Stamm: *eier*
- Unikale Morpheme - treten in einem einzigen Wort auf, haben keine eigene Bedeutung
  - Brombeere, Schornstein, Nachtigall
- Nullmorphem - ein Morphem wird aufgrund einer allgemeinen Tendenz in einer Sprache erwartet, erscheint aber nicht
  - *Frau* → *Frauen*, *Kind* → *Kinder*
  - *Meister* → *Meister-0*, *Messer* → *Messer-0*
- Allomorphie - ein Morphem wird durch lautlich unterschiedliche Formen repräsentiert
  - *ich* halte - *du* hältst

# Flexion und Wortbildung

---

## Flexion (Wortgrammatik)

- Schafft keine neuen Wörter → nur grammatische Formen ein und desselben Wortes
- *Land* → *Landes, Länder, Ländern*
- (Kern-)Bedeutung und Wortart bleiben gleich

## Wortbildung (lexikalischer Bereich)

- Es entstehen neue Lexeme durch Modifikation von bereits vorhandenen Morphemen
- *un + schön* → *unschön*

# Flexion

---

- Deutsch:
  - Flektierbare Wortarten:
    - deklinierbar (nach **Kasus**, Numerus und Genus): Substantive, Artikelwörter, Pronomen und Adjektive
    - konjugierbar (nach **Tempus**, Person, Numerus, Modus und Genus): Verben
  - nicht-flektierbare Wortarten: Adverbien, Präpositionen, Konjunktionen, Subjunktionen, Partikeln
- Flexionsparadigmen - enthalten alle Wortformen des jeweiligen Lexems

	Singular	Plural
Nominativ	Stift	Stifte
Akkusativ	Stift	Stifte
Dativ	Stift	Stiften
Genitiv	Stiftes	Stifte

**Tabelle 1:** Flexionsparadigma für STIFT

# Wortbildung

---

- Komposition - mehrere Wurzeln/Stämme werden zu einem komplexen Wort zusammengefügt
  - *Donaudampfschiffahrtsgesellschaftskapitän*
- Derivation - ein Affix wird an eine Wurzel bzw. einen Stamm angehängt
  - *lesbar*, *unschön*, *Spieler*
- Konversion - Lexeme sind lautlich identisch, unterscheiden sich nur in der Wortkategorie
  - *tief* → *(das) Tief*
- Kürzung: *Universität* → *Uni*
- Akronym: *AusZUBildender* → *Azubi*
- Abkürzung: *Lastkraftwagen* → *LKW*



# Sprachvergleich

---



# Sprachvergleich

## Sprachtypologie



“Wie lassen sich Sprachen / sprachl. Erscheinungen typisieren bzw. klassifizieren?”

z.B. durch

- Wortstellungstypologie
- Phonologische Sprachtypologie
- ***Morphologische Sprachtypologie:***

“Wie werden einzelne lexikalische Konzepte miteinander in Sätzen verbunden, um eine Satzbedeutung zu erzeugen?”

# Sprachvergleich

---

## Klassische morphologische Typologie

- F. v. Schlegel (1808):
  1. ***Flektierende***, fusionierende Sprache ("Sprachen durch Flexion")
  2. ***Agglutinierende*** Sprachen ("Sprachen durch Affixa")
- A.W. Schlegel (1818):
  3. ***Isolierende*** Sprachen
- W. v. Humboldt (1836):
  4. ***Polysynthetische***, "einverleibende", inkorporierende Sprachen

# Sprachvergleich

---

## 1. Flektierende Sprachen

- z.B. (fast alle) indogermanische Sprachen
- Grammatische Rolle des Wortes im Satz wird durch Beugung markiert
- “Fusionierend”:  
*Eine* nicht mehr zerlegbare Ausdruckseinheit kann *mehrere* grammatische Informationen tragen (vice versa)
- Dt.        *Kind - ern*  
→ 1 Morphem (-ern) trägt 2 Informationen (Numerus, Kasus)

# Sprachvergleich

---

## 2. Agglutierende Sprachen

- z.B. Türkisch, Ungarisch, Japanisch
- **Eine** nicht mehr zerlegbare Ausdruckseinheit (meist Suffix) trägt **eine** bestimmte grammatische Information und umgekehrt:

Türk.:	<i>ev</i>	“Haus”
	<i>ev - ler</i>	“Häuser”
	<i>ev - ler - im</i>	“meine Häuser”
	<i>ev - ler - im - de</i>	“in meinen Häusern”

# Sprachvergleich

---

## 3. Isolierende Sprachen

- z.B. Mandarin, Vietnamesisch
- Keine Flexion, Position im Satz drückt grammatische Funktion aus
- **Alle** Wörter bestehen aus nur *einem* Morphem (ideal):

Man.:    *tā*    *zài*    *túshūguǎn*    *kàn*    *bào*

*he*    *at*    *library*            *read*    *newspaper*

“He’s at the library reading a newspaper.”

# Sprachvergleich

- z.B. (fast alle) indigene nordamerikanische Sprachen
- Extremer Gebrauch von Affixen
- Grammatische Informationen an einem Stamm + **mehrere** Stämme mit grammatischen Morphemen zu einem “Wort” **zusammengebunden** (Phrasenkonstruktion, nicht Wortbildung!):

# Morphologie in der CL

---



# Morphologie in der CL

Morph. Analyse / korrekte Generierung von abgeleiteten Wörtern + Wortformen  
= **Notwendige Voraussetzung** für (fast alle) Anwendungen in der Computerlinguistik!

## Aufgaben

- Erkennung von (falschen) **Flexionsformen**:  
*\*Pinguins sollten die Gäste kein Futter geben*
- **Lemmatisierung**:  
*Die Flügel des [Pinguin]s dienen ...*
- **Analyse** abgeleiteter Wortformen + Komposita:  
*[pinguin]artig – Königs[pinguin]*
- **Wortformengenerierung**:  

<NOMEN,nom,sing>	liebt	<NOMEN,akk,plur>
Maria	liebt	Pinguine



# Morphologie in der CL

---

## Problemstellungen

- **Regularitäten** / Ausnahmen erkennen und kodieren:  
*lachen – lache – lacht – lachte*  
*gehen – gehe – geht – \*gehte*
- **Kompositasegmentierung:**  
*Wachstube – [Wach][stube] – [Wachs][tube]*
- **Bedeutungsregularitäten** in der Wortbildung:  
*König <HERRSCHER> – Königin <HERRSCHER&FEM>*
- (Morph.) **Ambige** Wortformen  
*Frau (Kasus?) – Treffen (Numerus?) – Treffen (Wortart?)*

# Morphologie in der CL

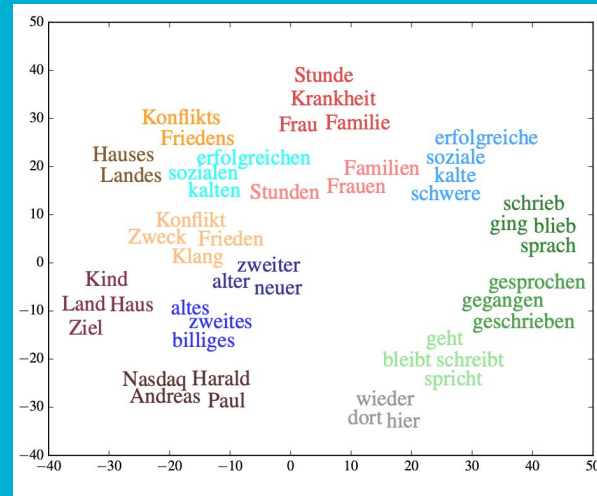
---

## Anwendungen

- Maschinelle Übersetzung
- Textzusammenfassung
- Textgenerierung
- Suchmaschinen
- Rechtschreibkorrektur
- Grammatikkorrektur
- ...

# Morphologie und Word Embeddings

- Word Embeddings - (semantisch) ähnliche Worte kommen im ähnlichen Kontext vor
  - morphologisch verwandte Worte?
- Word2Vec - Addition von Vektoren
- Morphological Word Embeddings
  - Embeddings werden auf einem mit morphologischen Tags annotierten Korpus trainiert
  - Trainingsziel: das nächste Wort und morphologische Tag vorhersagen
- DagoBERT - Derivationally and generatively optimized BERT)
  - generiert morphologisch komplexe Englische Worte
  - (this jacket is \_) + (wear) -> unwearable



# Quellen

---

- Langer, S. (2008): Morphologie in der Computerlinguistik. LMU München. Online unter: [https://www.cis.uni-muenchen.de/~stef/seminare/morphologie/skript\\_morphologie1.pdf](https://www.cis.uni-muenchen.de/~stef/seminare/morphologie/skript_morphologie1.pdf) (Letzter Zugriff: 02.02.21)
- Vogel, R. & Sahel, S. (2013): Einführung in die Morphologie des Deutschen. Darmstadt: WBG.
- Cotterell, R., Schütze H. (2015): Morphological Word-Embeddings
- Hofmann, V., Pierrehumbert, J., Schütze, H. (2020): DagoBERT: Generating Derivational Morphology with a Pretrained Language Model
- Zifonun, G. (2003): Sprachtypologie und Sprachvergleich. Universität Mannheim. Online unter: <https://www1.ids-mannheim.de/fileadmin/gra/texte/zif4.pdf> (Letzter Zugriff: 02.02.21)
- Bilder: <https://www.hiclipart.com/> (Letzter Zugriff: 02.02.21)