

Einführung in die Computerlinguistik

Morphologie

Robert Zangenfeind

Center for Information and Language Processing

2023-10-30

Die Grundfassung dieses Foliensatzes wurde von Dr. Benjamin Roth unter Zuhilfenahme von Materialien (insbesondere von Beispielsätzen) aus Vorlesungen von Prof. Dr. Tania Avgustinova und von mir erstellt. Fehler und Mängel liegen ausschließlich in meiner Verantwortung.

Outline

- 1 Intro
- 2 Morpheme
- 3 Wortstruktur
- 4 Flexion
- 5 Derivation
- 6 Komposition
- 7 Morphologische Prozesse
- 8 Automaten

Der Begriff der Morphologie

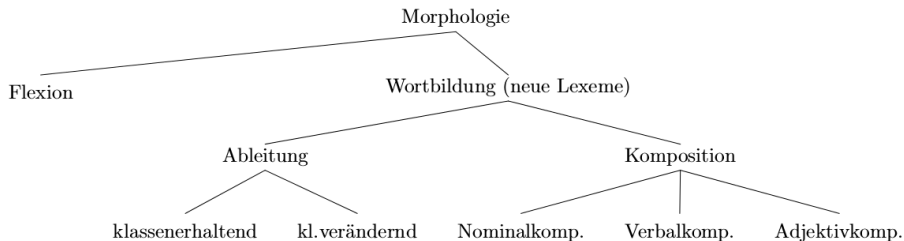
- *griech.* morphé (Form, Gestalt) + logos (Sinn, Lehre) \Rightarrow Formenlehre
- Aufbau von Wörtern aus kleinsten bedeutungstragenden Einheiten
- interne Struktur der Wörter



Die zwei Bereiche der Morphologie

- **Flexionsmorphologie** (Wortformbildung):
 - Markierung von Tempus, Person, Kasus, Numerus, ...
 - Aufbau von Wortformen aus Basis und Flexionsendung (Wort als Flexionsparadigma):
der Mann – des Mannes
 - aber auch:
geht – ging
- **Wortbildungslehre:**
 - **Derivationsmorphologie:** Bedeutungsverändernde Bildung von Wörtern aus einer Basis und einem Derivationsmorphem.
klar → unklar
Sache → sächlich / sachlich
 - **Komposita:** Zusammensetzung von mehreren Teilen (“Stämmen”)
Bauer + Hof → Bauernhof
Sonne + baden → sonnenbaden

Morphologie: Übersicht



Begriffe “Morph” und “Morphem”

- die elementaren Einheiten des Wortes
- die kleinsten sprachlichen Einheiten, die Bedeutung haben
- Morphem: abstrakte Einheit, die durch konkrete Repräsentationseinheiten (Morphe) realisiert wird, und zwar
 - in der gesprochenen Sprache als **Phonemfolge**
 - in der Schrift als **Graphemfolge**
- Morph: Realisierung (im Text) eines bestimmten Morphems

Beispiel: zwei Formen des Morphems {Stuhl} (entspricht: /ʃtu:l/)

- *Stuhl*

-s

-es

-e

- *Stühl*

-e

-en

- Wortformen: {Stuhl, Stuhls, Stuhles, Stuhle, Stühle, Stühlen}

- Prinzip:

Kombination von wenigen Elementen ergibt viele neue Elemente

stuhl_morphem + plural_morphem (*Stühle*)

hoch_morphem + haus_morphem (*Hochhaus*)

Wilhelm von Humboldt (1767–1835)

Das Verfahren der Sprache ist aber nicht bloß ein solches, wodurch eine einzelne Erscheinung zu Stande kommt; es muss derselben zugleich die Möglichkeit eröffnen, eine unbestimmbare Menge solcher Erscheinungen, und unter allen, ihr von dem Gedanken gestellten Bedingungen hervorzubringen. Denn sie steht ganz eigentlich einem unendlichen und wahrhaft gränzenlosen Gebiete, dem Inbegriff alles Denkbaren, gegenüber. **Sie muss daher von endlichen Mitteln einen unendlichen Gebrauch machen**, und vermag dies durch die Identität der Gedanken und Sprache erzeugenden Kraft.



Arten von Morphemen

- **Freies Morphem**: Morphem, das ohne Vorhandensein anderer Morpheme ein Wort bilden kann.

Beispiele:

{Garten}, {Zwerg}, {book}, {sing}

- **Gebundenes Morphem**: Morphem, das nicht selbständig ein Wort bilden kann.

Beispiele:

{ver-} (*versprechen*)

{-s} als Genitiv Singular (*Peters*)

- **Fugenelemente** seien hier nicht als Morpheme gesehen, weil sie keine identifizierbare Bedeutung tragen.

vgl. Beispiele: (*Krankheit*)-s-(*zeichen*), (*Schwan*)-en-(*hals*)

Morphem: Mel'čuks formale Definition (vereinfacht!)

(Mel'čuk, I.A.: Aspects of the Theory of Morphology. Berlin 2006:388)

Ein Morphem ist ein nicht-leeres Set von allen Morphen

$m_1, m_2, \dots, m_n = \{m_i\}$, die folgende drei Bedingungen erfüllen:

- Die Bedeutung von allen m_i ist identisch.
- Alle m_i gehören zur gleichen Klasse von Morphen, d.h. alle m_i sind entweder Wurzeln oder Affixe.
- Alle m_i stehen in komplementärer Distribution, die durch allgemeine Regeln beschrieben werden kann. Affixe können auch identische Distribution haben.
- Beispiel: "Stuhl-", "Stühl-"

Allomorphe des Morphems $\{M\}$: alle Morphe m_i , die zu $\{M\}$ gehören

Komplementäre Distribution

Zwei sprachliche Elemente kommen nie in gleicher Umgebung vor, d.h. ihre Vorkommen schließen sich in einer gegebenen Umgebung gegenseitig aus.

Morphem: Mel'čuks formale Definition (vereinfacht!)

(Mel'čuk, I.A.: Aspects of the Theory of Morphology. Berlin 2006:388)

Ein Morphem ist ein nicht-leeres Set von allen Morphen $m_1, m_2, \dots, m_n = \{m_i\}$, die folgende drei Bedingungen erfüllen:

- Die Bedeutung von allen m_i ist identisch.
- Alle m_i gehören zur gleichen Klasse von Morphen, d.h. alle m_i sind entweder Wurzeln oder Affixe.
- Alle m_i stehen in komplementärer Distribution, die durch allgemeine Regeln beschrieben werden kann. Affixe können auch identische Distribution haben.
- Beispiel: "Stuhl-", "Stühl-"

Allomorphe des Morphems $\{M\}$: alle Morphe m_i , die zu $\{M\}$ gehören

- **Derivation und Flexion**: Eine analysierbare Wortform kann rekursiv als Kombination eines Affixes und eines “Restes”, der **Basis** (die ggf. weiter analysierbar ist), aufgefasst werden.
- **Komposition**: Eine analysierbare Wortform kann rekursiv als Kombination zweier Teile (“Stämme”) aufgefasst werden (die ggf. weiter analysierbar sind).

- **Affix**: gebundenes Morphem, das verschiedene Basen in analoger Weise modifiziert (d.h. reihenbildend)

Unterkategorien:

- Flexionsaffix
- Derivationsaffix

- **Basis**: jede Form, an die ein Affix angefügt werden kann

Unterkategorien:

- die meisten Wurzeln sind Basen: *Tisch* → *Tisch-e*
- wenige Wurzeln sind jedoch keine Basen: *kunter* (in *kunterbunt*)
- Basen sind sehr oft keine Wurzeln: *touchable*, *unklar*, *sachlich*

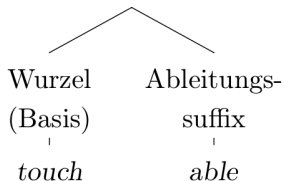
- **Flexionsaffix**: geht eine Verbindung mit einer **Basis** ein; markiert grammatikalische Funktionen wie Tempus, Person, Kasus oder Numerus
- **Derivationsaffix**: geht ebenfalls eine Verbindung mit einer **Basis** ein; verändert die Bedeutung der Basis
- **Wurzel(n)**: Teil, der übrig bleibt, nachdem alle Affixe einer Basis entfernt wurden (mit lexikalischer Bedeutung)

Affixe können in vier Klassen eingeteilt werden, je nach ihrer Position bezüglich der Basis:

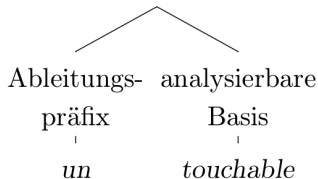
- Präfix; Beispiele:
 {auf-} (*aufpassen*)
 {un-} (*unglücklich*)
- Suffix; Beispiele:
 {-sam} (*unterhaltsam*)
 {-keit} (*Heiterkeit*)
 engl. {-ize} (*nationalize*)
- Infix (selten in europ. Sprachen); Beispiel:
 engl. {-bloody-} (*fan-bloody-tastic*)
- Zirkumfix (selten in europ. Sprachen); Beispiel:
 {Ge-e} (*Ge-zerr-e*)

Beispiel: Teilanalysen

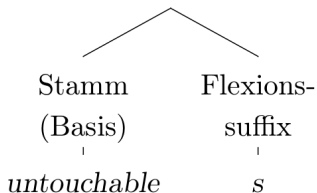
1.



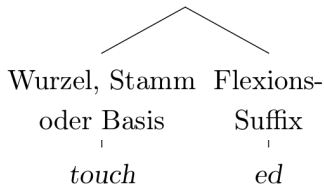
2.



3.



4.

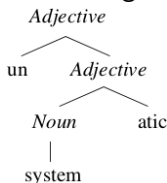


(vgl. auch: The Hierarchical Structure of Words. in: Fromkin et al. 2011)

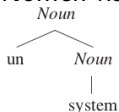
- Wörter haben eine interne Struktur, die durch Regeln der Zusammensetzung festgelegt wird.
- Beispielregel: Die meisten Affixe können nur mit Basen bestimmter Wortarten kombiniert werden.

Beispiel: Bestimmung der Wortstruktur von *unsystematic*

- Das Suffix “-atic” kann mit Nominalbasen kombiniert werden.
- Das Präfix “un-” kann mit Verb- und Adjektivbasen kombiniert werden, aber nicht mit Nominalbasen.
- Daraus ergibt sich folgende Analyse (Fromkin et al. 2011):



- Der folgende Teilbaum ist nicht möglich, weil “un-” nicht mit Nomen kombiniert werden kann:



Bestimmung der Wortstruktur: Wie?

- Testen, ob die Teilanalysen für sich genommen mögliche Basen ergeben
- Wenn mehrere Analysen möglich sind, entscheidet man sich für diejenige, deren Bedeutung am plausibelsten ist (d.h. bei der die internen Teile sinnvoll interpretiert werden können).
- Flexionsaffixe werden immer als letztes angehängt.

Flexionsmorphologie: Konstruktion von Paradigmen (1)

- z.B. *Tag*:

	sg	pl
nom	–	e
gen	es	e
dat	–	en
akk	–	e

- Generalisierung der Paradigmen → lexikalische Kategorie

Flexionsmorphologie: Konstruktion von Paradigmen (2)

- Flexionsendungen haben unspezifisches Bedeutungspotential, vgl. *Mensch*:

	sg	pl
nom	–	en
gen	en	en
dat	en	en
akk	en	en

- Häufige Erscheinung in flektierenden Sprachen
- **Synkretismus**: Innerhalb eines Paradigmas (und auch in verschiedenen Paradigmen) werden verschiedene Werte eines Merkmals von grammatischen Kategorien durch die gleiche Form repräsentiert. Die Formen “fallen zusammen”.
- z.B.: [*He*] *tried* vs. [*He has*] *tried* (simple past vs. past participle im present perfect simple) → Synkretismus (vgl. dagegen *He came* vs. *He has come*)

Flexionsmorphologie: Beispiel *klein*

- Starke (pronominale, determinierende) Deklination des Adjektivs (Wortformen bei Verwendung ohne Artikel u.a.)

	Singular			Plural
	Maskulinum	Neutrum	Femininum	
Nominativ	-er	-es	-e	-e
Akkusativ	-en			
Dativ	-em		-er	-en
Genitiv	-en			-er

klein+er	klein+e	klein+es	klein+e
klein+es/en	klein+er	klein+es/en	klein+er
klein+em	klein+er	klein+em	klein+en
klein+en	klein+e	klein+es	klein+e
klein+er+er	klein+er+e	klein+er+es	klein+er+e
klein+er+es/en	klein+er+er	klein+er+es/en	klein+er+er
klein+er+em	klein+er+er	klein+er+em	klein+er+en
klein+er+en	klein+er+e	klein+er+es	klein+er+e
klein+st+ er	klein+st+e	klein+st+ es	klein+st+e
klein+st+es/en	klein+st+er	klein+st+es/en	klein+st+er
klein+st+em	klein+st+er	klein+st+em	klein+st+en
klein+st+en	klein+st+e	klein+st+es	klein+st+e

- Im Deutschen insgesamt sechs Phoneme für (genuin dt.) Flexions-Suffixe:
/e, m, n, r, s, t/
- Durch die Morphologie markierte Merkmale (sprachabhängig):
 - Numerus-Systeme: auch z.B. zusätzlich Dual
 - Genus-Systeme: auch z.B. belebt-unbelebt; nur mask-fem
 - Kasus: große Differenzen in Sprachen
 - Person: Sprecher, Angesprochener, Besprochenes
 - Tempus: sprachspezifisch (Anzahl und Arten)

Traditionell unterscheidet man folgende Flexionsprozesse:

- Konjugation: bezeichnet die morphologische Kennzeichnung nach Person, Tempus, “Genus verbi” (Aktiv/Passiv), Numerus, Modus (Indikativ/Konjunktiv/Imperativ), ...
⇒ Verben
- Deklination: bezeichnet Flexion nach Kasus, Genus, Numerus
⇒ Substantiv, Pronomen, Adjektiv, Artikel
- Komparation: Graduierung und Vergleich
⇒ Adjektiv, (Adverb)

Übung: Abweichungen von diesem Paradigma?

klein+er	klein+e	klein+es	klein+e
klein+es/en	klein+er	klein+es/en	klein+er
klein+em	klein+er	klein+em	klein+en
klein+en	klein+e	klein+es	klein+e
klein+er+er	klein+er+e	klein+er+es	klein+er+e
klein+er+es/en	klein+er+er	klein+er+es/en	klein+er+er
klein+er+em	klein+er+er	klein+er+em	klein+er+en
klein+er+en	klein+er+e	klein+er+es	klein+er+e
klein+st+ er	klein+st+e	klein+st+ es	klein+st+e
klein+st+es/en	klein+st+er	klein+st+es/en	klein+st+er
klein+st+em	klein+st+er	klein+st+em	klein+st+en
klein+st+en	klein+st+e	klein+st+es	klein+st+e

- Neues Wort aus Basis + Derivationsaffix
- bei Änderung der Wortart → **klassenenverändernde** Derivation
- Derivationsuffix:
 - → neue Bedeutung des Wortes
 - reihenbildend (modifiziert in analoger Weise viele Wörter)

Inflection-derivation continuum (1)

Inflection	Derivation
(i) relevant to the syntax	not relevant to the syntax
(ii) obligatory expression of feature	not obligatory expression
(iii) unlimited applicability	possibly limited applicability
(iv) same concept as base	new concept
(v) relatively abstract meaning	relatively concrete meaning
(vi) compositional meaning	possibly non-compositional meaning
(vii) expression at word periphery	expression close to the base
(viii) less base allomorphy	more base allomorphy
(ix) no change of word-class	sometimes changes word-class
(x) cumulative expression possible	no cumulative expression
(xi) not iterable	possibly iterable

Table 5.5 A list of properties of inflection and derivation

aus: M. Haspelmath, A.D. Sims: Understanding Morphology. London 2010.

Inflection-derivation continuum (2)

Language	Formation	Example	cum	obl	[not] new	unl	cm
English	3rd singular	<i>walk/walks</i>	I	I	I	I	I
English	noun plural	<i>song/songs</i>	D	I	I	I	I
Spanish	diminutive	<i>gato/gatito</i>	D	D	I	I	I
English	repetitive	<i>write/rewrite</i>	D	D	D	I	I
English	female noun	<i>poet/poetess</i>	D	D	D	D	I
English	action noun	<i>resent/resentment</i>	D	D	D	D	D

Note: cum= cumulative expression; obl = obligatory; new = new concept;
unl = unlimited applicability; cm = compositional meaning.

Table 5.6 A continuum from inflection to derivation

aus: M. Haspelmath, A.D. Sims: Understanding Morphology. London 2010.

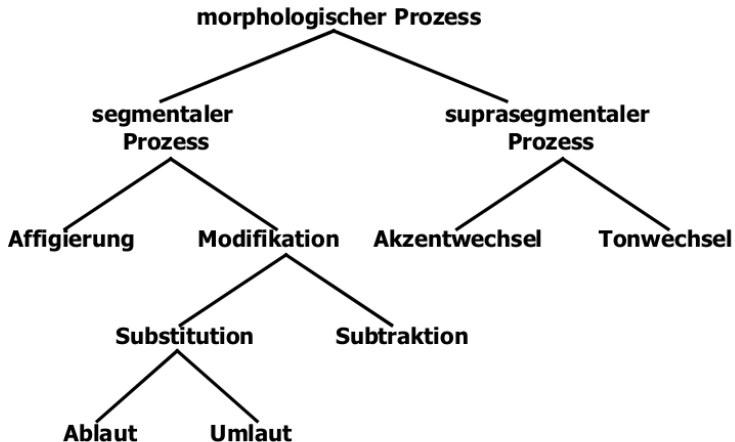
- Rekursive Kombination von Wurzeln bzw. Basen
- Fugenelemente sind keine Morpheme
- Nominalkomposition: [Straße]n[bahn], [Sprech][übung]
- Verbalkomposition: [press][schweißen], [stand][halten]
- Adjektivkomposition: [alt][ehrwürdig], [taub][stumm], [treff][sicher]
- Klammerung zur Darstellung der rekursiven Struktur (“Verschachtelung”):
→ [[Straße]n[bahn]][fahrerin]
- *Mädchenhandelsschule*
→ [Mädchen][[handel]s[schule]] vs.
[[Mädchen][handel]]s[schule]?
vgl.a. [Porzellan][[eier][korb]] vs. [[Porzellan][eier]][korb]?
• falsche Trennung erschwert Segmentieren beim Lesen, z.B.:
Gebirg-stier, Wachs-tube, Tau-schwert, Mais-turm,
Rohr-ohr-zucker; ungünstige Trennung: Talent-wässerung

(Weitere) morphologische Prozesse

Verschiedene Prozesse, mit denen neue Wortformen aus grundlegenden Elementen (z.B. Morphemen) konstruiert werden:

- Affigierung (s.o.)
- Modifikation durch Substitution:
 - Ablaut: *sing* – *sang* – *gesungen*;
 - Umlaut: *Maus* – *Mäuse*
- Modifikation durch Subtraktion (Tilgung) von Segmenten oder Merkmalen:
 - *Omnibus* \Rightarrow *Bus*
 - phonologisch im Französischen: *gris* – /gʁi/ ['grau']
(maskuline Form durch Tilgung des /z/ gebildet)
- Suprasegmental (nicht auf orthographischer/phonemischer Ebene):
 - Akzentwechsel (*pro'duce* (V) vs. 'produce (N), *per'mit* (V) vs. 'permit (N); analog: *import*, *insult*, *discount* ...)
 - Tonwechsel (in tonalen Sprachen, z.B. Kanuri [saharanische Sprache])

Morphologische Prozesse



Lehnwörter fremder Sprachen werden oft durch Kombination ähnlichklingender (und teils bedeutungsähnlicher) Morpheme nachgebildet:

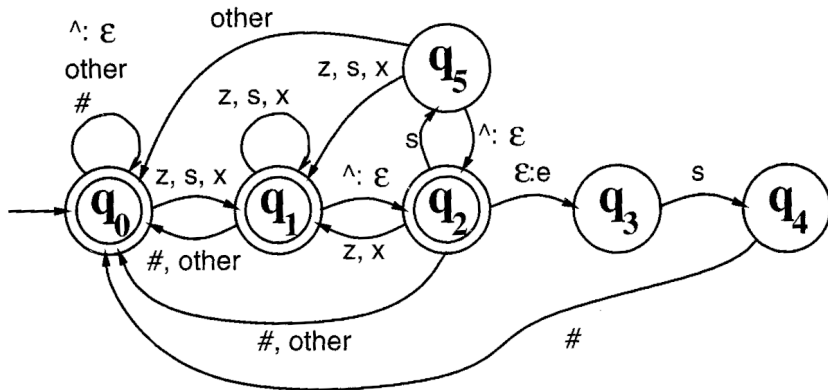
- Hängematte: von Taino/haitianisch *hamaka* (“Schlafnetz”)
- Vielfraß: von altnorwegisch *fjeldfross* (“Gebirgskater”)
- Mes[s]ner: von lateinisch *mansionarius* (“Aufseher des [Gottes]hauses”)
- Quäntchen: von lateinisch *quintus* (“ein Fünftel”)
- Tollpatsch: von ungarisch *talpas* (“Fußsoldat”)

- Morphem verliert lexikalische Bedeutung und Freiheit in der Stellung
- Inhaltswörter werden zu Funktionswörtern
- freie Morpheme zu gebundenen
- z.B. dt. Präteritum-Affix *-t-*: urspr. *sagen-tat* → *sagte*
- Suffixe *-keit*, *-heit*, *-tum*, *-lich*: urspr. eigenständige Wörter:
 - *-lich*: ahd. 'Körper, Gestalt'
 - *-keit*, *-heit*: 'Art, Weise'
 - *-tum*: 'Würde, Stand'
- derzeit im Übergang: *Weise*, *frei*, *voll*, *mäßig*, *Zeug*, *Werk*
- ähnlich: *bekommen*, vgl. *Er bekommt etwas geschenkt*, aber auch *Er bekam den Zahn gezogen*
- franz. *ne - pas*: (nicht) ← 'keinen Schritt'

Anwendung: Stemming (Stemmatisierung)

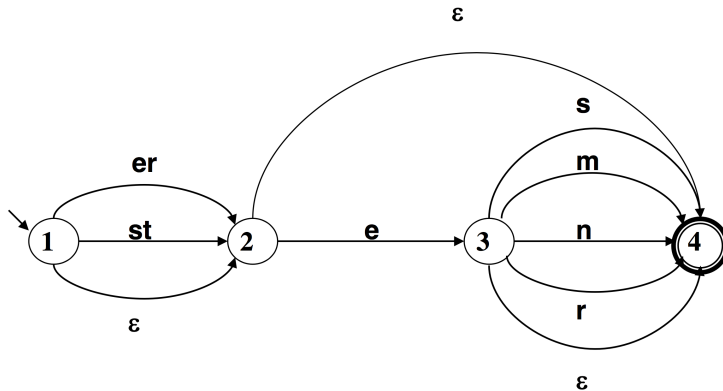
- Alternatives Verfahren zum Lemmatisieren (= Zuordnen einer Wortform zu seinem Lemma [Zitierform eines Wortes])
- Flexionsmorpheme einer Wortform werden beseitigt
⇒ Wort-“Stamm” (wird der Wortform zugeordnet)
- z.B. engl. Wortformen *process*, *processing*, *processed*
⇒ “Stamm” *process*
- Problem: sinnvolle Unterscheidungen können verloren gehen:
- z.B. *stocks* (Aktien etc.) und *stockings* (‘Strümpfe’ etc.)
⇒ “Stamm” *stock*
- Programm in Netz z.B. Porter Stemmer:
Porter-Stemmer (Demo und download):
<https://snowballstem.org>

Finite state automaton (Transducer): e-insertion engl. Plural



aus: D. Jurafsky, J.H. Martin: Speech and Language Processing. Prentice Hall 2000:78.

Transducer: Endungen dt. Adjektive



- Implementierung einer Finite-state-Morphologie

<https://www.cis.uni-muenchen.de/~schmid/tools/SMOR/>

Zum Schluss: Besonders klausurrelevant

- Morphem
- Flexion
- Derivation
- Komposition
- Morphologische Baumanalyse
- Wurzel, Basis, Affix, Fugenelement, Wortform
- Flexionsparadigma
- Synkretismus