

Einführung in die Computerlinguistik

Morphologie I

Hinrich Schütze & Robert Zangenfeind

Centrum für Informations- und Sprachverarbeitung, LMU München

26.10.2015

Take-away

- Einleitung
- Morpheme
- Prozesse und Prozeduren
- Das Wort als Lemma

Begriff der Morphologie

- *griech.* morphé (Form, Gestalt) → Formenlehre
- Aufbau von Wörtern aus kleinsten bedeutungstragenden Einheiten (Morpheme)
- interne Struktur der Wörter

Die zwei Bereiche der Morphologie

- Flexionslehre (Wortformbildung): Aufbau von Wortformen aus Wortstamm und Flexionsendung (Wort als Flexionsparadigma)
- Wortbildungslehre (Wortstammbildungslehre): Komposition, Präfigierung, Suffigierung (Derivation): Aufbau komplexer Formen aus Stammmorphem und Wortbildungsaffix

Begriff des Morphems

- sprachliche Zeichen
- elementare Einheiten des Wortes
- kleinste sprachliche Einheiten, die Bedeutung haben
- bestehen aus Silben
- mehr als 20 000 Morpheme im Dt.
- System: Kombination von wenigen Elementen ergibt viele neue Elemente
- z.B.: Graphemkette *Gartenzwerg*
- z.B. {mann} u. {männ}
- z.B. {er}, {e}, {en}
- Einheiten gleicher Bedeutung mit komplementärer Verteilung

Arten von Morphemen

- lexikalisches Morphem: trägt lexikalische Bedeutung, z.B. {sag}

vs.

- grammatisches Morphem (Grammem, funktionales Morphem): rein grammatische Funktion, z.B. {te} (→ *sagte*)

bzw.

- freies (ungebundenes) Morphem: kann ohne Vorhandensein anderer Morpheme ein Wort bilden; z.B. {Garten}, {Zwerg}

vs.

- gebundenes Morphem: kann nicht selbständig ein Wort bilden, z.B. {s} als Genitiv Singular

Wurzel

- Morphem, das den größten Teil der Syntaktik einer Wortform beinhaltet
- → Kookurrenz der Wortform mit anderen Wortformen (z.B. "Die *Ziele* erreichen")
- Großteil der Morpheme einer Sprache

vs.

Affix

- Morphem, das keine Wurzel ist
- gebundenes Morphem
- reihenbildend: modifiziert versch. Wurzeln in analoger Weise
- Suffixe; z.B. {sam} (z.B. *aufmerksam*)
- Präfixe; z.B. {auf}
- Infixe; (z.B. engl. *fan-bloody-tastic*)
- Zirkumfixe (diskontinuierlich); z.B. {ge- -t} (→ *gesagt*)

vs.

Stamm

- Bestandteil einer Wortform, der die Wurzel enthält
- und eventuell weitere Teile, die keine Flexionsmorpheme darstellen
- z.B. *verkauf*

Definition Morphem

(vgl. Def. von Mel'čuk, I.A.: Aspects of the Theory of Morphology. Berlin 2006, S. 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 Signifikate von allen m_i sind 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)

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

Grundsätzliches

- interne Struktur
- Sprachökonomie: größter kommunikativer Erfolg bei minimalem Aufwand
- → Gleichgewicht zw. Umfang des Lexikons und Wortbildung

1. Segmentieren (1)

- Verstehen: Zerlegen der Äußerung in kleinere Einheiten
- → Regeln finden!
- ohne Rest
- meist intuitiv möglich

1. Segmentieren (2)

Beispiele

- *Straßenbahnfahrerin*
→ Straße-n-bahn-fahr-er-in
- Klammerung zur Darstellung von Strukturunterschieden:
→ [Straßenbahn][fahrerin]
- *Mädchenhandelsschule*
→ [Mädchen][handelsschule] vs. [Mädchenhandel][s][schule]?
- falsche Trennung erschwert Segmentieren beim Lesen, z.B.:
Talent-wässerung, Gebirg-stier, Wachs-tube, Tau-schwert,
Mais-turm

2. Grammatikalisierung und Lexikalisierung (1)

- “Arbeitsteilung”: Inventar von Elementen und Regeln zur Kombination

Apresjan: “Unserer linguistischen Beschreibung liegt die Auffassung von einer prinzipiellen Parität des Wörterbuchs und der Grammatik zugrunde. Alle wirklich allgemeinen, lexikalisch nicht eingeschränkten oder nicht zu sehr eingeschränkten Gesetzmäßigkeiten möchten wir gerade durch grammatische Regeln beschreiben.”

Mel’čuk: “Eine Grammatik wird als etwas betrachtet, das nicht mehr ist als eine Reihe von Verallgemeinerungen über ein gutes Wörterbuch.”

- Übergang zur Grammatik: Grammatikalisierung
- Übergang zum Lexikon: Lexikalisierung

2. Grammatikalisierung und Lexikalisierung (2)

Grammatikalisierung:

- Wort geht in Richtung grammatisches Morphem
- Grammatik entsteht im Sprachwandel:
- vgl. dazu: pragmatic mode: autonome Elemente werden aneinandergereiht
- vs. syntactic mode: (strengere grammatische Regulierung → normierte Standardsprache)
- 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*

2. Grammatikalisierung und Lexikalisierung (3)

Beispiele:

- 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'
- Nomen auf *-er*: oft Ableitungen von Verben: *Bohrer, Spieler, Seufzer*
- auch abgeleitet von Nomen: *Fußballer, Berliner, Politiker*
- vgl. aber: *Mutter, Schulter, Fenster* → lexikalisiert
- anderer Weg: *das Buch von Peter* ← *Peters Buch*

Das Wörterbuch/Lexikon

- versch. Arten von Wörterbüchern
- meist: alphabetische Anordnung
- Lemma: Grundform eines Wortes (Zitierform): “Überschrift” im Eintrag eines Wörterbuchs
- Unterpunkte: Kriterien zur Unterscheidung oft unklar
- Bsp. aus Duden:

Wort, das; -[e]s, Wörter u. Worte [mhd., ahd. Wort, eigtl. = feierlich Gesprochenes]:

1.a) (Pl. Wörter; gelegtl. auch: Worte) kleinste selbständige sprachliche Einheit von Lautung (2) u. Inhalt (2a) bzw. Bedeutung:

ein [...]kurzes, zusammengesetztes, deutsches, [...] W.; dieses W. ist ein Substantiv [...]

b) (Pl. Worte) Wort (1a) in speziellem Hinblick auf seinen bestimmten Inhalt, Sinn; Ausdruck, Begriff:

Liebe ist ein großes W.; Angst ist ein zu hartes W.

2. (Pl. Worte) etw., was man als Ausdruck seiner Gedanken, Gefühle o.ä. zusammenhängend äußert; Äußerung:

(2a) Worte des Dankes, Trostes; aufmunternde [...] überflüssige -e [...]

mir fehlen die -e, ich habefinde keine -e [dafür]! (ich bin vor Entrüstung o.ä. sprachlos) [...]

auf ein W.! (ich möchte Sie/dich kurz sprechen)

du sprichst ein großes W. gelassen aus (so einfach ist das nicht; nach Goethe, „Iphigenie“, I. 3);

3. <Pl. Worte> Ausspruch:

*ein wahres, [...] geflügeltes W. (bekannter, viel zitierter Ausspruch; LÜ von griech. *épea pteróenta* [Homer]).*

4. <Pl. Worte> (geh.) Text von, zu etw.:

W. und Weise;

man weiß nicht, wer die -e zu dieser Melodie schrieb [...]

5. <o. Pl.> förmliches Versprechen; Versicherung:

jmdm. das W. abnehmen zu schweigen; sein W. einlösen, halten [...]

ich gebe Ihnen mein W. (Ehrenwort) darauf; auf mein W. (dafür verbürge ich mich!)

a) (Rel.) Kanon, Sammlung heiliger Schriften, bes. die darin enthaltene Glaubenslehre:

das W. Gottes (Gottes Offenbarung im Wort der Heiligen Schrift) [...] Im Zweifelsfall galt ihm das W. des Korans (Stern, Mann I 40);

b) (christl. Theol.) Logos (4):

Und das W. ward Fleisch (Ioh. I. 14)

Das Wörterbuch/Lexikon

andere Wörterbücher, z.B.:

- Synonym-WB (Verwendung)
- Etymologische WB
- Rechtschreib-WB
- Aussprache-WB
- Valenz-WB (Basis für wissenschaftliche Untersuchungen, für Fremdsprachenlerner)
- Kollokations-WB
- Rückläufige WB (Suffixe – Wortbildung)
- Bild-WB (Problem: Ontologie)
- Frequenz-WB

Das Wörterbuch/Lexikon

Wichtige Unterscheidung:

Lemmatisierung

- Zuordnung der Wortformen zu Lemmata
- → schnellere Suche möglich
- nicht immer eindeutig; Kontext!
- Programme hierzu:
CST Lemmatiser (Demo):
<http://cst.dk/online/lemmatiser/uk/>
SMOR – dt. Morphologie (Download):
<https://code.google.com/p/cistern/wiki/SMOR>

vs.

Tokenisierung (Aufteilung in Token)

Das mentale Lexikon

3 wichtige Aspekte für Verwendung und Zugriff auf ein Wort:

- phonologische Form
- Bedeutung
- Wortart

zur Form: “Badewanneneffekt”:

luat enier sidtue an eienr uvrnsäiet ist es eagl, in wcheler rhnfgeeloie die bstuchbaen in eniem wrot sheetn, das eniizg whictgie ist, dsas der etrse und der lztete bstuchbae am rtigeinchn paltz snid. Der rset knan tatol deiuranchnedr sien und man knan es iemmr ncoh onhe porbelm lseen. Das legit daarn, dsas wir nhcit jeedn bstuchbaen aeilln lseen, srednon das Wrot als Gzanes.

→ Input fällt auf sprachliches Wissen; Netzwerke

Literatur

- Heringer, H.-J.: Morphologie. Paderborn 2009.
- Mel'čuk, I.A.: Aspects of the Theory of Morphology. Berlin 2006.

Take-away

- Einleitung
- Morpheme
- Prozesse und Prozeduren
- Das Wort als Lemma