

UNIVERSITÄT ATHEN
ABTEILUNG FÜR DEUTSCHE SPRACHE UND LITERATUR

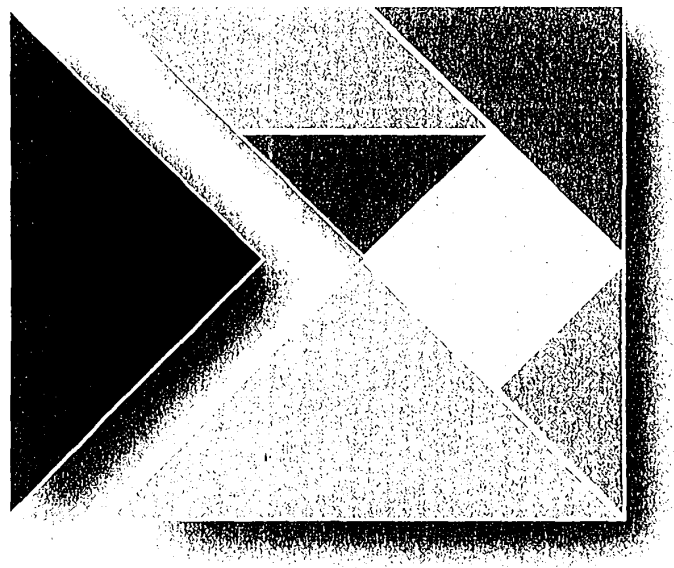
UNTERLAGEN ZU DEN KURS

MORPHOLOGIE

Christina Alexandris

b a c h e l o r - w i s s e n

b a c h e l o r - w i s s e n



Albert Busch / Oliver Stenschke

Germanistische Linguistik

Eine Einführung

2., durchgesehene und korrigierte Auflage

bachelor-wissen ist die Reihe für die modularisierten Studiengänge

- ▶ die Bände sind auf die Bedürfnisse der Studierenden abgestimmt
- ▶ das fachliche Grundwissen wird in zahlreichen Übungen vertieft
- ▶ der Stoff ist in die Unterrichtseinheiten einer Lehrveranstaltung gegliedert
- ▶ auf www.bachelor-wissen.de finden Sie begleitende und weiterführende Informationen zum Studium und zu diesem Band

gnv Gunter Narr Verlag Tübingen

Apl. Prof. Dr. Albert Busch lehrt seit 1997 an der Georg-August-Universität Göttingen Germanistische Linguistik. Von ihm stammen die Einheiten 1, 3, 7, 11, 12, 13, 14.

Dr. Oliver Stenschke lehrt seit 2000 ebenfalls Germanistische Linguistik an der Georg-August-Universität Göttingen. Er hat die Einheiten 2, 4, 5, 6, 8, 9, 10 verfasst.

Idee und Konzept der Reihe: Johannes Kabatek, Lehrstuhl für Romanische Sprachwissenschaft an der Eberhard-Karls-Universität Tübingen

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.d-nb.de>> abrufbar.

2., durchgesehene und korrigierte Auflage 2008
1. Auflage 2007

© 2008 · Narr Francke Attempto Verlag GmbH + Co. KG
Dischingerweg 5 · D-72070 Tübingen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Gedruckt auf chlorfrei gebleichtem und säurefreiem Werkdruckpapier.

Internet: <http://www.bachelor-wissen.de>
E-Mail: info@narr.de

Satz: Informationsdesign D. Fratzke, Kirchentellinsfurt
Druck und Bindung: CPI – Ebner & Spiegel, Ulm
Printed in Germany

ISSN 1864-4082
ISBN 978-3-8233-6414-6

Inhalt

Vorwort	1
---------------	---

Themenblock 1: Sprache und Zeichen

Einheit 1: Sprache und Linguistik	3
1.1 Was ist Linguistik?	4
1.2 Was ist Sprache?	5
1.3 Wie gehen Linguisten bei der Untersuchung von Sprache vor? ...	12
1.4 Übungen	14
1.5 Verwendete und weiterführende Literatur	15
Einheit 2: Semiotik	17
2.1 Semiotik – Was macht das Zeichen zum Zeichen?	18
2.2 Zeichentypen	19
2.3 Sprachliche Zeichen	21
2.4 Zeichen und Zeichenbenutzer	28
2.5 Semiotik als Wissenschaft	33
2.6 Übungen	34
2.7 Verwendete und weiterführende Literatur	34

Themenblock 2: Laut und Wort

Einheit 3: Phonetik und Phonologie	37
3.1 Phonetik und Phonologie – Wie sprechen wir?	38
3.2 Phonetik – Was tun wir, um zu sprechen?	38
3.2.1 Artikulation	40
3.2.2 Artikulation der Konsonanten	43
3.2.3 Artikulation der Vokale	47
3.3 Phonologie – Was tun wir, um Laute zu erkennen?	49
3.3.1 Das Phonemsystem	50
3.3.2 Die Silbe	54
3.4 Übungen	55
3.5 Verwendete und weiterführende Literatur	56
Einheit 4: Graphematik und Orthographie	57
4.1 Graphematik und Orthographie – Wie schreiben wir?	58
4.2 Graphematische Grundbegriffe	58

Morphologische Analyse

Inhalt

5.1	Morphologie – Was ist ein Wort?	76
5.2	Morphologische Grundbegriffe	78
5.3	Die Analyse der unmittelbaren Konstituenten (IC-Analyse)	83
5.4	Spezielle Probleme der IC-Analyse	89
5.5	Übungen	91
5.6	Verwendete und weiterführende Literatur	92

5.1 | Morphologie – Was ist ein Wort?

Morphologie

In der Phonetik-Einheit konnten Sie bereits das Bibelzitat lesen, dass das Wort am Anfang von allem war. Die linguistische Teildisziplin, die sich der Beschreibung von Wörtern widmet, heißt MORPHOLOGIE (griech.: morphé = Gestalt, Form). In dieser Einheit wird zunächst die Frage geklärt, was ein Wort eigentlich ist. Danach geht es gewissermaßen noch einen Schritt zurück, d.h., es wird die Frage gestellt, woraus ein Wort besteht. Nachdem das Wort auf diese Weise in seine Bestandteile zerlegt wurde, wird es in einer weiteren morphologischen Einheit aus umgekehrter Perspektive um die Konstruktion von Wörtern gehen, also um die Wortbildung.

Definition

Morphologie (von Goethe ursprünglich für die Lehre von der Form und Struktur lebender Organismen verwendeter Terminus): linguistische Teildisziplin, die sich mit der Gestalt, Flexion (Beugung) und Bildung von Wörtern beschäftigt.

Wort

Zuvor soll der Begriff WORT definiert werden. Das mag Ihnen banal erscheinen, aber schon bei der Pluralbildung kann man ins Schleudern kommen: Heißt es *Worte* oder *Wörter*? Bzw. wann nimmt man welchen Plural? Und auch die in Abb. 5.1 aufgeführten Alltagssprachlichen Verwendungen machen eine Definition des Begriffs „Wort“ nicht leichter.

Abb. 5.1 |

Einige Alltagssprachliche Verwendungen von „Wort“

- 1) *Ein Mann, ein Wort.*
- 2) *Das Wort Gottes.*
- 3) *Denk an meine Worte!*
- 4) *Durch die neue Rechtschreibung werden viele Wörter auseinandergerissen.*
- 5) *Die Aussprache des Wortes „China“ variiert in verschiedenen Dialekten des Deutschen.*
- 6) *Der Lebenslauf sollte nicht mehr als 500 Wörter umfassen.*

Während in den ersten drei Beispielen unter *Wort* eine Äußerung bzw. ein Text zu verstehen ist, bezieht sich *Wort* in den Beispielen (4) bis (6) eher auf kleinere, abgeschlossene Einheiten innerhalb eines Textes oder des Systems Sprache allgemein. Die Frage ist nun, anhand welcher Kriterien die Einheit Wort im linguistischen Sinne abgegrenzt werden kann. Die Definitionsansätze sind erwartungsgemäß vielfältig; die entsprechenden Einträge in linguistischen Fachlexika wie dem „Lexikon der Sprachwissenschaft“ (2002) oder dem „Metzler-Lexikon Sprache“ (2004) entsprechend umfangreich. Zu den immer wieder genannten Abgrenzungskriterien zählen die in Abb. 5.2 genannten Aspekte.

Sämtliche Kriterien haben verschiedene Haken, wie z.B. an dem Satz *Der Deutsch-Lehrer setzt mir die Aufgabe auseinander* deutlich wird. So sind hier die Einheiten *der*, *mir* und *die* nicht oder nur teilweise durch phonetische Grenzsignale markiert, tragen keinen Akzent – es sei denn, man möchte sie

- 1) *phonetisch-phonologisch*: Einheit mit einem Akzent, durch Grenzsignale wie Pause oder Knacklaut isolierbar
- 2) *orthographisch*: Einheit zwischen zwei Leerzeichen
- 3) *morphologisch*: strukturell stabile, nicht trennbare, minimale freie Einheit, mit der eine Frage beantwortet werden kann
- 4) *lexikalisch-semantisch*: kleinster, relativ selbständiger Träger von Bedeutung, der im Lexikon verankert ist
- 5) *syntaktisch*: kleinste verschlebbare Einheit im Satz

besonders betonen – und wären somit keine Wörter. Die Einheiten *setzt* und *auseinander* sind durch Leerzeichen abgetrennt und wären demnach zwei Wörter, von den aktuellen orthographischen Streitigkeiten ganz zu schweigen (bis 1996 hieß es *auseinandersetzen*, dann einige Jahre *auseinander setzen* und seit den Empfehlungen des Rechtschreibrates von 2006 wieder *auseinandersetzen*). Mit der Einheit *der* kann man kaum eine Frage beantworten – und wenn Sie meinen, das ginge doch, versuchen Sie es mal mit dem Genitiv *des*. Die Einheit *Deutsch-Lehrer* kann man in zwei Einheiten aufteilen, die sich jeweils separat als selbständige Träger von Bedeutung erweisen. Dafür lässt sich die Bedeutung einer funktionalen Einheit wie *der* für sich genommen kaum beschreiben. Außerdem kann man sie zwar im Satz verschieben, aber nur zusammen mit der Einheit *Deutsch-Lehrer*.

Ein weiteres Problem kommt hinzu, wenn man sich die folgende Auflistung aus der Duden-Grammatik (2005) anschaut (vgl. Abb. 5.3). Wie viele hervorgehobene Wörter enthält dieser Kasten? Man kann die Antwort aus verschiedenen Perspektiven formulieren. Aus der Sicht eines Wörterbuchs oder des MENTALEN LEXIKONS (gewissermaßen der im Gehirn abgespeicherten Wörter) würde man sagen, dass es sich bei den hervorgehobenen Einheiten immer um ein und dasselbe Wort handelt. Man spricht daher auch vom LEXIKALISCHEN WORT oder LEXEM.

- a) *Die Türme der Burg waren schon von weitem zu sehen.*
- b) *Auf den Türmen wehten bunte Fahnen.*
- c) *Der eine Turm war vierzig Meter hoch.*
- d) *Der andere Turm war nur etwa dreißig Meter hoch.*
- e) *Wir sind auf den Turm geklettert.*
- f) *Auf dem Turm hatten wir eine prächtige Aussicht.*
- g) *Die Mauern des Turms bestanden aus dicken Quadern.*

Stellt man die Perspektive der Sprachverwendung in den Mittelpunkt, zählt man hingegen sieben Verwendungen des lexikalischen Wortes *Turm*. Zur Unterscheidung der abstrakten Kategorie des lexikalischen Wortes *Turm*

Abb. 5.2

Abgrenzungskriterien für die linguistische Einheit „Wort“ in linguistischen Fachlexika

Mentales Lexikon

Lexikalisches Wort/Lexem

Abb. 5.3

Ein Lexem – sechs Wortformen – sieben Tokens (vgl. Duden-Grammatik 2005: 129)

Syntaktisches Wort/Wortform

(als Element der Langue) und seiner konkreten Verwendung (in der Parole) kann man das schon in der Semiotik-Einheit beschriebene Konzept der Type-Token-Relation anwenden. Demnach findet man für das Type *Turm* als lexikalische Einheit im obigen Kasten verschiedene Tokens, die sich im Hinblick auf zwei Kriterien unterscheiden: ihre äußere Form und ihre grammatischen Eigenschaften. Nimmt man beide Kriterien zusammen, lassen sich in Abb. 5.3 sechs verschiedene WORTFORMEN oder SYNTAKTISCHE WÖRTER unterscheiden, nämlich *Turm* (Nom. Sg.), *Turm* (Akk. Sg.), *Turm* (Dat. Sg.), *Turms* (Gen. Sg.), *Türme* (Nom. Pl.) und *Türmen* (Dat. Pl.). Nimmt man die hier fehlenden Formen des Akk. Pl. (*Türme*) und Gen. Pl. (*Türme*) hinzu, erhält man das vollständige Flexionsparadigma.

Definition

Wort: als lexikalisches Wort oder Lexem abstrakte Einheit des Lexikons. Als syntaktisches Wort oder Wortform in konkreter Verwendung vorliegende Einheit einer Äußerung bzw. eines Textes.

Man könnte hinsichtlich der Type-Token-Relation noch einen Schritt weiter gehen und auf der Ebene der Wortformen ebenfalls noch einmal zwischen Types und Tokens unterscheiden. Dann wäre jede Wortform ein eigenes Type und jede Verwendung einer Wortform ein Token. Dementsprechend läge dann in Abb. 5.3 für die fünf Types *Turm* (Akk. Sg.), *Turm* (Dat. Sg.), *Turms* (Gen. Sg.), *Türme* (Nom. Pl.) und *Türmen* (Dat. Pl.) jeweils ein Token vor; das sechste Type *Turm* (Nom. Sg.) läge zweimal als Token vor (in (c) und (d)); die Types *Türme* (Akk. Pl.) und *Türme* (Gen. Pl.) wären hier hingegen nicht als Tokens realisiert.

5.2 | Morphologische Grundbegriffe

Morph Nachdem klar ist, was man unter dem Begriff „Wort“ in der Linguistik versteht, soll es nun darum gehen, wie man Wörter genauer beschreiben kann. Es gibt unterschiedlich komplexe Arten von Wörtern. Wörter wie *Fisch*, *schön*, *und*, *bald* usw. sind morphologisch gesehen relativ einfach, weswegen sie auch als **SIMPLIZIA** (lat.: simplex = einfach) bezeichnet werden. Sie bestehen aus nicht weiter zerlegbaren Einheiten, die über eine Bedeutung verfügen. Hingegen sind Wörter wie *Deutsch-Lehrer*, *auseinandersetzen* und *Turms* ganz offensichtlich aus mehreren Bestandteilen zusammengesetzt, die jeder für sich eine (nicht immer ganz leicht zu beschreibende) Bedeutung oder doch zumindest ein grammatisches Merkmal innehaben. Diese Bestandteile, die elementaren Einheiten der Morphologie, nennt man **MORPHE**.

Definition

Morph: elementare lautliche oder graphische Einheit der Sprache, der eine Bedeutung bzw. ein grammatisches Merkmal zugeschrieben werden kann. Morphe werden in geschweiften Klammern notiert, z. B. {Turm}, {s}, {lehr}, {er}.

Die Kombinationsmöglichkeiten von Morphen werden vielfach auch in der linguistischen Fachliteratur als nahezu unbegrenzt dargestellt. Das Ergebnis solcher Kombinationsprozesse wird als **WORTBILDUNGSKONSTRUKTION** oder **WORTBILDUNGSPRODUKT** bezeichnet. Aus einem der berühmtesten Beispiele für ein sehr komplexes Wortbildungsprodukt, der im Rechtschreib-Duden aufgeführten *Donau-Dampfschiffahrtsgesellschaft*, macht Donalies (2005) in ihrem Überblick über die Wortbildung des Deutschen die *Donaudampfschiffahrtsgesellschaftskapitänswitwenkompositabildungsexpertenrunde*. Obwohl die für das vorliegende Buch verwendete Version von MS-Word keinerlei Beanstandungen gegen dieses Lexem hat, geht es vielleicht auch Ihnen so, dass Sie spätestens nach *Witwen* aussteigen. Und auch im Internet mit seinen Phantastilliarden von Wortformen lässt sich dieses Lexem nicht wiederfinden. Es scheint also doch Grenzen zu geben. Gleichwohl ist dem Deutschen eine gewisse Tendenz zu relativ komplexen Wortbildungsprodukten zu eigen.

Wortbildungs-
konstruktion/Wort-
bildungsprodukt

Simplex: ein Lexem, das nur aus einem einzigen Morph besteht und daher nicht weiter zerlegt werden kann; z. B. *Fisch*, *bald*.

Wortbildungsprodukt: ein Lexem, das aus mehreren Morphen besteht und daher in verschiedene Bestandteile zerlegt werden kann; z. B. *Deutsch-Lehrer*.

Definition

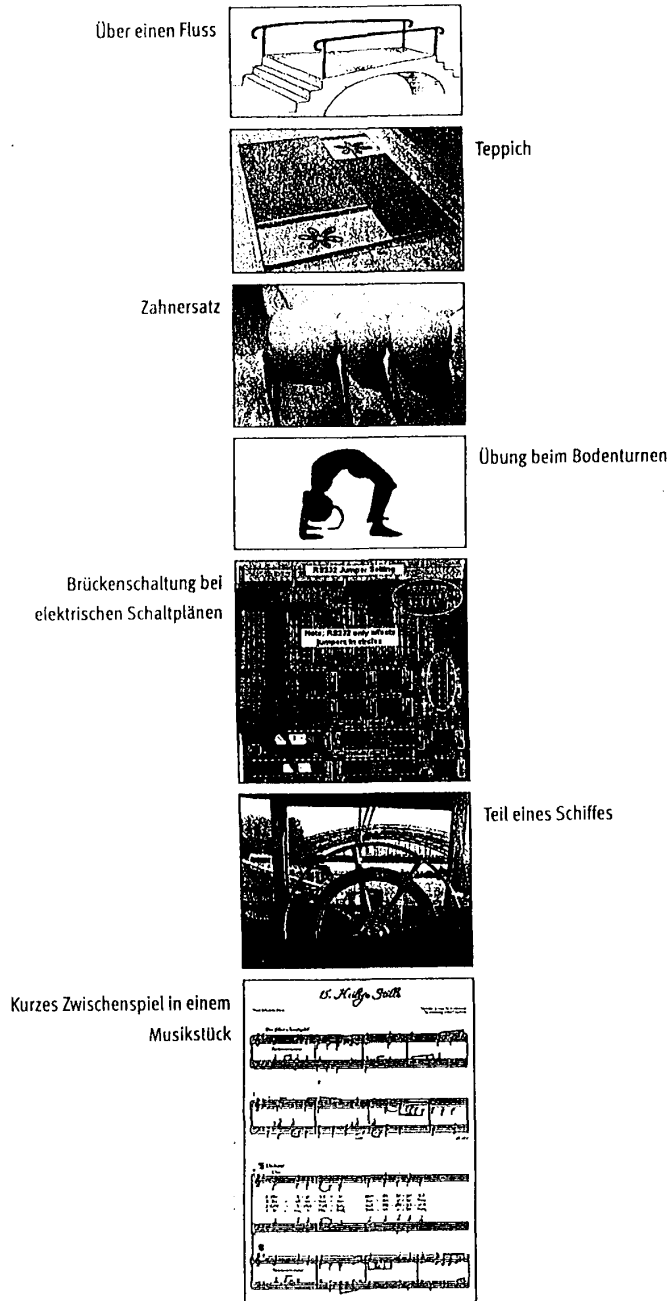
Leider verhält sich die deutsche Sprache nicht so, dass jedem Morph genau eine Bedeutung zugeordnet werden kann. Vielleicht haben Sie selbst schon einmal das Spiel „Teekesselchen“ gespielt. Dabei geht es darum, sich mehrdeutige Begriffe, also Wörter mit gleicher Ausdrucksseite, aber verschiedener Bedeutung, auszudenken und diese dann ihren verschiedenen Bedeutungen entsprechend zu erklären.

Beispielsweise kann man das Wort *Brücke* als ein Bauwerk, eine Art Teppich, eine Übung beim Bodenturnen, einen Zahnersatz, eine Schaltungsform in der Elektronik, einen Teil eines Schiffes, ein kurzes Zwischenspiel in einem Musikstück und noch einiges mehr erklären (s. Abb. 5.4). Mit anderen Worten: Ein und dieselbe Laut- bzw. Buchstabenfolge *Brücke* kommt in Kombination mit mehreren verschiedenen Bedeutungen und damit in mehreren Morphen vor. Man spricht hier auch von **POLYSEMIE** bzw. **HOMONYMIE** (vgl. Kapitel 11.5.5).

Umgekehrt gibt es eine ganze Reihe von Bedeutungen oder grammatischen Merkmalen, die durch unterschiedliche Morphe, sogenannte **ALLOMORPHE** (griech.: állos = ein anderer) realisiert werden, wie z. B. das Merkmal „Plural“ durch die Allomorphe {-n} (Löwen, Bauern), {-en} (Frauen), {-er} (Kinder), {-s} (Kinos, Muttis), {Umlaut} (Väter, Mütter, Öfen), {-ø} (Fahrer; ø = sog. Nullmorphem, vgl. Kapitel 5.3) u. a. oder das Verb *singen*, zu dem die Morphe {sing}, {sang}, {säng} und {sung} gehören (ich singe, ich sang, ich sänge, ich habe gesungen). Die Einheit aus einer bestimmten Bedeutung bzw. einem grammatischen Merkmal, also der Inhaltsseite des Zeichens, und dem dazugehörigen

Allomorph

Abb. 5.4 |
Einige Bedeutungen
des Morphs (Brücke)



Morph bzw. den dazugehörigen Allomorphen (= der Ausdrucksseite) heißt MORPHEM. Im Unterschied zum Phonem bzw. Graphem, die weiter oben als die kleinsten bedeutungsUNTERSCHIEDENDEN Einheiten der gesprochenen bzw. geschriebenen Sprache definiert wurden, handelt es sich beim Morphem um die kleinste bedeutungSTRAGENDE Einheit der Sprache.

Morphem

9

Morphem: kleinste bedeutungstragende Einheit der Sprache, bestehend aus einem Morph oder mehreren Allomorphen (Ausdrucksseite) und einer Bedeutung bzw. einem grammatischen Merkmal (Inhaltsseite). Morpheme werden ebenfalls in geschweiften Klammern notiert.

Definition

Morpheme sind also die elementaren Bestandteile der Wortbildung. Man unterscheidet zwischen BASISMORPHEMEN (auch GRUNDMORPHEME genannt) wie {Fisch}, {Kind}, {bald}, {schön}, {schreib} etc. und WORTBILDUNGSMORPHEMEN wie {heit}, {ung}, {lich}, {ver}, {un} usw. Erstere sind wortfähig, entsprechen also einem Eintrag im Lexikon und kommen daher (mit Ausnahme vieler verbaler Basismorpheme wie {helf}, die im Lexikon im Infinitiv zitiert werden) in der Regel FREI vor, also als selbständige Lexeme. Das gilt für substantivische Basismorpheme wie das oben erwähnte Wort *Brücke*, das für mehrere Morpheme stehen kann, die gleichzeitig alle über einen eigenen Lexikoneintrag verfügen. Auch adjektivische Basismorpheme {schön} sowie adverbiale Basismorpheme {bald}, konjunktionale Basismorpheme {und, weil}, pronominale Basismorpheme {jemand, dies} und präpositionale Basismorpheme {auf, wegen} kommen frei vor. Dagegen treten Wortbildungsmorpheme nur GEBUNDEN auf, also als Bestandteile von Wortbildungsprodukten. Auch die dritte Morphemklasse der FLEXIONSMORPHEME kommt ausschließlich gebunden vor. Dazu gehören z. B. die oben erwähnten Pluralmorpheme oder auch die Morpheme, mit welchen Verben in eine bestimmte Person flektiert werden, wie z. B. {t} bzw. {et} für die 3. Pers. Sg. Ind. Präs. (*er schreib-t, er arbeit-et*; dass bei Letzterem ein Schwa-Laut eingefügt wird, hat phonetische Ursachen – versuchen Sie mal, *er arbeit-t* so auszusprechen, dass man es vom verbalen Basismorphem {arbeit} unterscheiden kann). Ausnahmen sind z. B. Konjunktionen und Präpositionen, die häufig als Beispiele für freie grammatische Morpheme angeführt werden.

Basismorphem
Wortbildungs-
morphem

Freie Morpheme

Gebundene Morpheme
Flexionsmorphem

Flexionsmorphem	Ausdrucksseite	<i>schreib-{t}, arbeit-{et}</i>
„3. Pers. Sg. Ind. Präs.“ mit den beiden Allomorphen {t}, {et}	Inhaltsseite	3. Pers. Sg. Ind. Präs.

Tab. 5.1

Flexionsmorphem mit Allomorphen

Nicht immer lassen sich Morpheme eindeutig einer dieser drei Klassen zuordnen. Die Grenzen zwischen Basis- und Wortbildungsmorphemen sind aufgrund des Sprachwandels, im Zuge dessen lexikalische Konstituenten

Affixoid

tendenziell grammatikalisiert werden, eher fließend. Eine Zeit lang wurden Morpheme wie {zeug} in *Flug-zeug*, {werk} in *Mäh-werk*, {frei} in *sorgen-frei*, {arm} in *natrium-arm*, die sich im Übergang vom Basis- zum Wortbildungsmorphem befinden, in der Wortbildungsforschung als AFFIXOIDE bezeichnet. Auf die Problematik dieses Begriffs wird in Kapitel 6.2 näher eingegangen.

Affix Der Terminus „Affixoid“ ist abgeleitet von AFFIX (lat.: affixum = angeheftet). Dabei handelt es sich um die Oberkategorie für alle Begriffe, mit denen Morpheme hinsichtlich ihrer Position beschrieben werden. Ein PRÄFIX (lat.: praefixum = vorn angeheftet) steht immer vor dem Lexem, mit dem es ein neues Wort bildet (z. B. {auf} in *auf-essen*); ein SUFFIX (lat.: suffixum = (hinten) angeheftet) dahinter (z. B. {keit} in *Tapfer-keit*). Außerdem gibt es noch ZIRKUMFIXE (lat.: circum = ringsum, auch DISKONTINUIERLICHE MORPHEME genannt) wie z. B. {ge + e} in *Ge-birg-e* oder {be + ig} in *be-sänftig-(en)*, deren Status allerdings umstritten ist, sowie in manchen Fachbüchern das INFIX (lat.: infixum = hineingeheftet) in Wörtern wie *funktions-un-tüchtig*, über deren Zustandekommen man sich ebenfalls streiten kann; es ließe sich auch als Komposition aus *Funktion* und *untüchtig* erklären, wobei Letzteres eher selten frei vorkommt.

Diskontinuierliche Morpheme Ein weiteres Unterscheidungskriterium, das in der Definition von Morphemen als Träger von Bedeutung bzw. grammatischen Merkmalen bereits angeklungen ist, ist die Differenzierung von GRAMMATISCH VS. LEXIKALISCH. Während sich die Flexionsmorpheme wie z. B. das Plural-Morphem mit seinen diversen Allomorphen klar den grammatischen Morphemen zuordnen lassen, handelt es sich bei den meisten Basismorphemen (z. B. {Fisch}) eindeutig um lexikalische Morpheme. Sie sind im mentalen Lexikon offenbar als eigene Einheiten organisiert (wie auch immer man sich das genau physiologisch vorzustellen hat; die Forschung steckt diesbezüglich immer noch in den Anfängen; vgl. Einheit 12). Dementsprechend sind sie in einem konkreten Lexikon als eigene Einträge verzeichnet. Etwas schwieriger stellt sich die Lage bei den Wortbildungsmorphemen dar. Ein Morphem wie das Suffix {ung} hat in den allermeisten Fällen die (vorwiegend grammatische) Funktion, aus einem Verb (wie z. B. *üben*) ein Substantiv zu bilden (→ *Übung*), das dann immer Femininum ist. Hingegen verkehrt das Präfix {un} die Bedeutung eines Adjektivs (*unfrei*, *unschön*, *unwillig*, *unbeabsichtigt* usw.) oder – sehr viel seltener – eines Substantivs (z. B. *Undank*, *Untiefe*) ins Gegenteil. Es hat also gewissermaßen die Bedeutung von *nicht*, was man auch daran sieht, dass substantivierte Verben oder von Verben abgeleitete Substantive in der Regel mit {nicht} anstelle von {un} verneint werden (z. B. *Nichterscheinen*, *Nichterfüllung*; Ausnahme: *Unvermögen*). Die Klasse der Wortbildungsmorpheme schwankt also zwischen lexikalischer Bedeutung und grammatischer Funktion. Da die große Mehrheit der Wortbildungsmorpheme dazu dient, Wörter in eine

andere Wortart zu überführen, werden sie oftmals der Einfachheit halber pauschal den grammatischen Morphemen zugeordnet. In dieser Einführung wird hingegen nach einem einfachen Kriterium unterschieden: Überführt das Wortbildungsmorphem ein Basismorphem in eine andere Wortart (z. B. *Umleit-ung*), wird es zu den grammatischen Morphemen gerechnet, ansonsten (z. B. *un-frei*) zu den lexikalischen. In Fällen, wo man nicht entscheiden kann, welcher Wortart das Basismorphem zuzuordnen ist (kommt z. B. *be-antwort(en)* vom Substantiv *Antwort* oder vom Verb *antwort(en)*?), werden einfach beide Kategorien angegeben. In Tab. 5.2 sind die verschiedenen Klassifikationsmöglichkeiten für Morpheme noch einmal überblicksartig dargestellt.

Klassifikationsmöglichkeiten der Morpheme

Morphemtyp			
frei		gebunden	
lexikalisch	grammatisch	lexikalisch	grammatisch
verbales Basismorphem (identisch mit Imperativ): {komm}	pronominales Basismorphem: {jemand}	verbales Basismorphem (nicht identisch mit Imperativ): {helf}	Flexionsaffixe: Zirkumfix: {ge} + {t} (<i>gelegt</i>)
substantivisches Basismorphem: {Fisch}	präpositionales Basismorphem: {auf}	Konfixe (vgl. Kap. 5.4): {bio}, {thek}	Suffix: {t}/(et) (<i>arbeitet</i> , <i>schreibt</i>)
adjektivisches Basismorphem: {schön}	konjunktionales Basismorphem: {und}	Wortbildungsaffixe:	
adverbiales Basismorphem: {bald}		Präfix: {un} (<i>unfrei</i>) {ver} (<i>vergleichen</i>)	
		Zirkumfix: {ge} + {e} (<i>Gebirge</i>) {be} + {t} (<i>berühmt</i>)	
		Suffix {ei} (<i>Bücherei</i>) {ung} (<i>Leitung</i>)	

Tab. 5.2 Die Klassifikation von Morphemtypen

Die Analyse der unmittelbaren Konstituenten (IC-Analyse)

5.3

Die im vorherigen Kapitel erwähnten Kategorien bilden das Grundinstrumentarium, um Wörter morphologisch zu analysieren. Wie Sie anhand dieser Formulierung schon richtig errahnen, reicht das allerdings noch nicht ganz aus, weswegen Sie in diesem Kapitel das Verfahren der sogenannten IC-ANALYSE kennen lernen werden. IC steht für „Immediate Constituents“ (= unmittelbare

IC-Analyse

Bestandteile), so dass man auch von Konstituentenanalyse spricht. Ziel dieses Verfahrens ist es, einen sprachlichen Ausdruck – z. B. einen Satz oder wie hier ein Wort – in seine Bestandteile zu zerlegen. Die morphologisch kleinsten Bestandteile eines Wortes sind die Morpheme, die wie oben beschrieben mit einem Morph oder verschiedenen Allomorphen in der Sprache vorkommen. In dieser Einführung geht es um die Analyse konkreter Wortformen in Texten. Zu diesem Zweck sei ein kurzer Text zitiert – ein sogenanntes Dialogfenster, also eine Nachricht, mit der gewissermaßen ein Computerprogramm Kontakt zum Benutzer aufnehmen möchte.



Abb. 5.5 |
Windows-Dialogfenster
(Quelle: <http://www.swr3.de/dialoge/>;
Stand: September 2006)

Dieser kurze Text, wenn man ihn denn als solchen bezeichnen kann, enthält inkl. Überschrift und Schaltflächen zehn Wörter.

Im Folgenden soll es um die beiden Wörter *Änderungen* und *Schriftartgleichungs-Ausnahmedatei* gehen. Um mit dem Einfacheren anzufangen, wird zunächst das Wort *Änderungen* analysiert.

Segmentierung

Prinzip der Binarität

Im ersten Teil einer morphologischen IC-Analyse geht es darum, die vorliegende Wortform zu segmentieren, also in ihre Bestandteile zu zerlegen. Dabei gelten zwei Prinzipien. Das erste, sogenannte PRINZIP DER BINARITÄT (lat.: bini = je zwei) besagt, dass die Wortform immer in zwei Konstituenten aufgeteilt wird und diese dann, wenn möglich, wiederum in zwei Konstituenten, bis irgendwann nichts mehr aufzuteilen ist. Das zweite Prinzip lautet: Wenn die Wortform flektiert ist, spricht: ein Flexionsaffix enthält, so ist dieses zuerst abzutrennen. Die Wortform *Änderungen* enthält ein solches Affix, nämlich das Morphem {en}, das hier den Nominativ Plural anzeigt. Der erste Analyse-schritt sieht also wie folgt aus:

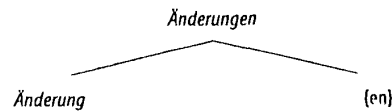


Abb. 5.6 |
IC-Analyse der Wortform *Änderungen*,
1. Schritt

Der Einfachheit halber wird in dieser Einführung davon ausgegangen, dass bei jeder Wortform höchstens ein Flexionsmorphem vorliegt. Beispielsweise bestünde bei der Verbform (*wir*) *machten* die Möglichkeit, zwei verschiedene Flexionsmorpheme anzunehmen: einmal {t} für Präteritum und einmal {en} für 1. Pers. Pl. Allerdings hat die deutsche Sprache die Eigenheit, dass nicht jedes grammatische Merkmal immer mit einem eigenen Morphem dargestellt wird; so verfügt das Substantiv *Änderung* zwar über ein Morphem für das Merkmal „Plural“, aber weder im Singular noch im Plural

über irgendwelche Morpheme, die den Kasus (hier: Nominativ) anzeigen. Das hätte für die Darstellung die Konsequenz, dass man für das Merkmal „Nominativ“ bei der Wortform *Änderungen* ein sogenanntes NULLMORPHEM ansetzen müsste, worauf hier aus Gründen der Überschaubarkeit verzichtet wird.

Nullmorphem

∞

Nullmorphem (∅): Ein grammatisches Merkmal wird ausdrucksseitig nicht durch ein Morphem realisiert, wie z. B. die verschiedenen Kasus im Plural der meisten Substantive oder der Plural ganz allgemein in Substantiven wie „Fahrer“. Genau gesagt handelt es sich hier um Nullallomorphe.

Definition

Die Endung {en} ist nicht weiter segmentierbar, mit anderen Worten: Die Analyse ist hier bereits auf der Morphemebene angekommen, weswegen die Endung in geschweiften Klammern notiert wird. Das sieht beim übrig gebliebenen Rest *Änderung* anders aus. Dieser lässt sich weiter segmentieren, wie der folgende Strukturbaum zeigt:

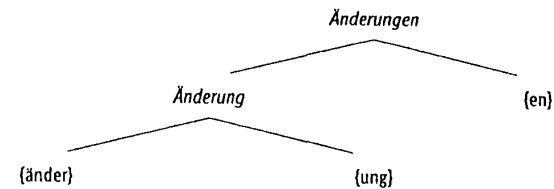


Abb. 5.7
IC-Analyse der Wortform *Änderungen*,
vollständig segmentiert

Damit ist die Wortform vollständig SEGMENTIERT. Die einzelnen Morpheme können nun im zweiten Teil der Analyse KLASSIFIZIERT werden. Bei {en} handelt es sich um ein gebundenes, grammatisches Flexionssuffix, bei {ung} um ein gebundenes, grammatisches Wortbildungssuffix. Schwieriger ist der Fall bei {änder}. Dieses könnte man als verbales (*änder-n*) oder adjektivisches Basismorphem (*eine ander-e Gelegenheit*) einstufen. Im letzteren Fall würde das bedeuten, dass hier mit {änder} ein Allomorph zu {ander} vorliegt. Allerdings besteht das Problem, dass adjektivische Basismorpheme normalerweise frei vorkommen (*eine gut-e Gelegenheit – die Gelegenheit ist gut*, aber: *eine ander-e Gelegenheit – die Gelegenheit ist *ander*). Deswegen folgen wir hier der Einschätzung von Fleischer/Barz (1995) und dem Großen Duden, die {ander} als pronominales Basismorphem einordnen, ähnlich wie z. B. {solch} oder {irgendein}, die allerdings beide frei vorkommen. Aus dem pronominalen {ander} ließe sich dann mittels Flexion ein Pronomen konstruieren oder aber im Zuge eines Wortbildungsprozesses ein Adverb (*ander-s*; ähnlich wie *besonder-s*, *stet-s*, *link-s*, die auch alle ohne das {s} nicht frei vorkommen, aber mittels Flexion ein Adjektiv bilden können; z. B. *eine besonder-e Eigenschaft*, *stet-er Tropfen*, *auf der linken Seite*) bzw. der Verbstamm {änder} bilden. Wie solche Wortbildungspro-

Klassifizierung

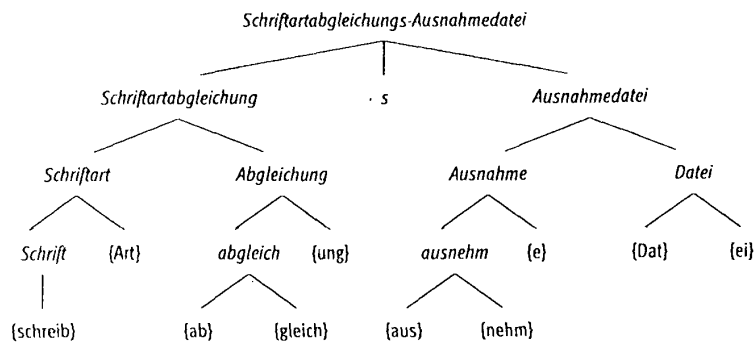
zesse genau funktionieren, ist Gegenstand der folgenden Einheit. Im Ergebnis wäre {änder} also als gebundenes, lexikalisches verbales Basismorphem oder als Allomorph des gebundenen, grammatischen, pronominalen Basismorphems {ander} zu klassifizieren.

Definition

Morphologische IC-Analyse: Verfahren zur Analyse von Wortformen. Dabei werden die Wortformen zunächst Schritt für Schritt jeweils in ihre unmittelbaren Konstituenten (immediate Constituents) segmentiert, wobei das **binäre Prinzip** anzuwenden ist. Nicht weiter zerlegbare Konstituenten (= Morpheme) werden schließlich klassifiziert.

Damit sind Sie theoretisch in der Lage, die folgende IC-Analyse der Wortform *Schriftartabgleichungs-Ausnahmedatei* nachzuvollziehen, die in Abb. 5.8 dargestellt ist.

Abb. 5.8 | Vollständige IC-Analyse der Wortform *Schriftartabgleichungs-Ausnahmedatei*



Detaillierte IC-Analyse

Wenn Sie diesen Strukturbaum ohne fremde Hilfe selbständig genauso gezeichnet hätten, können Sie getrost zum nächsten Kapitel weiterblättern. Wahrscheinlich ergeben sich aber einige Fragen. Das fängt damit an, dass im ersten Schritt vom Prinzip der Binarität abgewichen wurde. Das ist immer dann legitim, wenn es sich bei der Wortform um eine Komposition (vgl. dazu Einheit 6) aus zwei Bestandteilen handelt, welche mit einem sogenannten FUGENELEMENT (alternativ auch: INTERFIX) verbunden sind. Das hier vorliegende Fugen-s ist keine Genitiv-Endung, wie oft vermutet wird (auch wenn das Fugen-s historisch aus einem Flexionssuffix entstanden ist), denn der Genitiv von *Schriftartabgleichung* besitzt kein s. Auch bei anderen Fugenelementen wie -er- in *Kind-er-spielplatz* oder -n- in *Blume-n-vase* liegen nicht etwa Pluralmorpheme vor. Der Status der vielfältigen Formen von Fugenelementen (vgl. auch *Glaube-ns-frage*, *Trag-e-tasche*, *Held-en-mut*, *Sieg-es-zug*, *Schmerz-ens-schrei*, *Kontinent-al-verschiebung*, *Gas-o-meter*, *Agr-i-kultur*, *Woll-(ø)-decke*, *lexik-al-isch*, *öff-en-t-lich*) ist umstritten und lässt sich kaum pauschal beschreiben. Grundsätzlich wird man der Fuge wohl am ehesten gerecht, wenn man

sagt, dass sie gegenwartssprachlich betrachtet der phonetischen Strukturierung komplexer Komposita und der besseren Aussprechbarkeit dient. Da das Fugenelement weder eine eigene Bedeutung noch ein grammatisches Merkmal trägt, wird es ohne geschweifte Klammern notiert.

Definition

Fugenelement: aus phonetischen Gründen eingefügter Laut oder Lautkette in komplexen Wortbildungsprodukten wie *Umleitung-s-empfehlung*, *Kontinent-al-verschiebung*, *Kind-er-garten* oder *brasil-l-an-isch*. Fugen sind keine Morpheme, da sie zwar bedeutungsUNTERSCHIEDEND (*Land-es-polizei* vs. *Länd-ø-polizei* vs. *Länd-er-pölizei*) sein können, nicht aber bedeutungsSTRAGEND.

Weitere Segmentierung

Die weitere Segmentierung der linken Konstituente *Schriftartabgleichung* dürfte weitgehend nachvollziehbar sein. Geschweifte Klammern finden sich immer erst dann, wenn eine Konstituente nicht weiter segmentierbar ist, also Morph(em)status besitzt. Eine Besonderheit stellt hier die Konstituente *Schrift* dar. Man könnte an dieser Stelle auch aufhören und {Schrift} als freies, lexikalisches substantivisches Basismorphem werten. Allerdings kann man bei längerem Grübeln auf den Gedanken kommen, dass {Schrift} etwas mit {schreib} zu tun hat. Noch längeres Nachdenken führt irgendwann zum ähnlichen Fall {Gift}, dessen Bezug zu {geb} aus synchroner Sicht noch in *Mitgift* deutlich ist. Allerdings lassen sich aus Sicht der Gegenwartssprache {Gift} und {Schrift} kaum segmentieren (obwohl einige wenige weitere Beispiele wie *Vernunft*, *Auskunft*, evtl. auch *Kunst*, *Brunst*, *Gunst* den Verdacht nahelegen, dass es in der Vergangenheit mal ein Morphem {t} gegeben haben könnte; tatsächlich handelt es sich um eine Spur des in althochdeutscher Zeit produktiven Morphems {ti}, das zur Bildung von Abstrakta diente). Erst recht nicht kann man aus synchroner Perspektive entscheiden, was eher da war. Deswegen wird an dieser Stelle erstens erneut vom Prinzip der Binarität abgewichen und die Beziehung von {Schrift} zu {schreib} lediglich mit einer senkrechten Linie markiert; zweitens ist dieser Schritt insgesamt ohne Hilfsmittel (wie etymologische Wörterbücher, in denen die Wortgeschichte von Lexemen verzeichnet ist, und sogenannte rückläufige Wörterbücher, in denen der Wortschatz alphabetisch geordnet vom Wortende her verzeichnet ist) bestenfalls als optional zu betrachten.

Auch wären Sie wohl nicht unbedingt darauf gekommen, dass man bei der rechten Konstituente *Ausnahmedatei* die Konstituente *Datei* noch weiter segmentieren kann. Ein Blick z. B. in Fleischer/Barz (1995) zeigt aber, dass das Wortbildungssuffix {ei} einen Ort bezeichnet, an dem sich jemand befindet, etwas hergestellt oder aufbewahrt wird o. Ä., wie man an den analogen Fällen *Kartei*, *Kantorei* oder *Ziegelei* erkennen kann. Eine *Datei* ist also ein Aufbewahrungsort für Daten.

Bemerkenswert ist noch, dass man bei der Analyse der Konstituente *Ausnahme* von Allomorphie auszugehen hat. Der Verbstamm *ausnehm*, der sich

seinerseits noch weiter segmentieren lässt, wird nicht nur mit dem Wortbildungssuffix {e} versehen, sondern verändert auf der betonten Silbe auch noch den Vokal. Dieses Phänomen tritt ebenfalls sehr häufig auf, z. B. bei {back} und {Bäck} + {er} oder {geb} und {Gab} + {e}. Auch bei diesen Fällen kann man sich natürlich immer fragen, wie man darauf kommt, dass das Substantiv vom Verb abgeleitet wurde und nicht umgekehrt. Gerade das Beispiel *Ausnahme* ist aber ein gutes Indiz dafür, dass das Substantiv nach dem Verb gebildet wurde, da man das Substantiv nicht in {aus} und {nahme} segmentieren kann – es sei denn, Sie können erklären, was {nahme} sein soll. In der Tendenz kann man davon ausgehen, dass das Verb vor dem Substantiv da war. Ausnahmen wie *fernsehen* von *Fernseher* bestätigen diese Regel.

Schließlich ist an der gezeigten IC-Analyse gut zu erkennen, warum von „unmittelbaren Bestandteilen“ gesprochen wird. Damit ist die schrittweise Analyse benannt. So ist beispielsweise das Morphem {aus} zwar auch eine Konstituente von *Schriftartabgleichungs-Ausnahmedatei*, aber eben erst eine mittelbare, d. h. in mehreren Schritten isolierbare. Die unmittelbaren Konstituenten findet man immer in der Zeile direkt unter dem gerade analysierten Bestandteil.

Wenn die Segmentierung abgeschlossen ist, müssen die erhaltenen Morpheme noch klassifiziert werden. Dies geschieht in Tab. 5.3:

Klassifizierung der Morpheme

Tab. 5.3 | Klassifizierung der aus der Segmentierung von *Schriftartabgleichungs-Ausnahmedatei* hervorgegangenen Konstituenten

{schreib}	frei, lexikalisch, verbales Basismorphem
{Art}	frei, lexikalisch, substantivisches Basismorphem
{ab}	gebunden, grammatisch, Wortbildungspräfix
{gleich}	frei, lexikalisch, adjektivisches Basismorphem
{ung}	gebunden, grammatisch, Wortbildungssuffix
s	Fugenelement
{aus}	gebunden, lexikalisch, Wortbildungspräfix
{nehm}	gebunden, lexikalisch, verbales Basismorphem
{e}	gebunden, grammatisch, Wortbildungssuffix
{Dat}	gebunden, lexikalisch, substantivisches Basismorphem
{ei}	gebunden, lexikalisch, Wortbildungssuffix

Der Verbstamm {schreib} wird hier als frei gewertet, weil er als Imperativ selbständig ist. Hingegen unterscheidet sich der Verbstamm {nehm} – jedenfalls im Moment noch – vom Imperativ *nimm!*. Das Lexem *Daten* tritt in der vorliegenden Bedeutung im Deutschen nur im Plural auf, weswegen das substantivische Basismorphem {Dat} nicht frei vorkommt bzw. unklar ist, wie der Singular heißen würde. Der vom Duden vorgeschlagene Singular

Datum ist zumindest alltagssprachlich wohl eher auf Zeitangaben beschränkt. Schließlich werden die Morpheme {aus} und {ab} als gebunden klassifiziert, da sie nicht dieselbe Bedeutung wie die frei vorkommenden präpositionalen Basismorpheme haben. Andererseits verfügen sie, ähnlich wie das Suffix {ei}, noch über eine gewisse Bedeutung, so dass eine eindeutige Zuordnung entweder zu den grammatischen oder lexikalischen Morphemen nicht möglich ist.

Spezielle Probleme der IC-Analyse

Auch wenn es bei den bisherigen Analysen schon einige wenige knifflige Fälle gab, so waren diese doch alle noch relativ eindeutig zu lösen. Es gibt jedoch auch Fälle, in denen Fragen offen bleiben, wie z. B. beim Wort *Unbescheidenheit*. Hier ergibt sich ein Problem bei der Segmentierungs-Reihenfolge. Die beiden Alternativen sind in Abb. 5.9 zu sehen; sie gehen einher mit der Frage, ob man dieses Wort eher als die Eigenschaft des Unbescheidenseins (links) oder als das Gegenteil von Bescheidenheit (rechts) auffasst. Beide Auffassungen sind gleichermaßen vertretbar.



5-4

Segmentierungsprobleme

Abb. 5.9

Varianten der Segmentierung

Auch die Analyse der oben erwähnten Zirkumfixe bereitet notationstechnisch gewisse Probleme, kollidiert doch das Prinzip der Binarität hier mit der Tatsache, dass drei unmittelbare Konstituenten vorliegen. Meistens wird dieses Problem so gelöst wie in Abb. 5.10 dargestellt. Zu beachten ist dabei, dass der Infinitiv ein Flexionssuffix ist und daher zuerst abgetrennt werden muss. Außerdem enthält der Verbstamm *besänftig* neben dem Zirkumfix ein umgelautetes Allomorph von {sanft}.

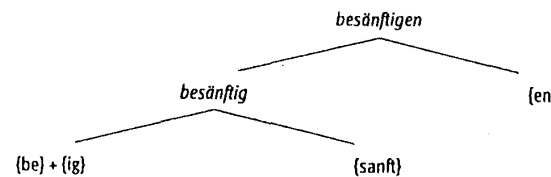


Abb. 5.10

IC-Analyse von Zirkumfixen

Ein weiteres häufig auftretendes Problem ist, ob man eine Konstituente noch weiter segmentieren kann oder nicht. Das zeigte sich schon andeutungsweise bei den Konstituenten *Schrift* und *Datei*. Dabei wurde das Morphem {ei} als im Deutschen zwar noch vorkommendes und damit aktives, aber nicht mehr produktives Wortbildungsmorphem identifiziert. Was genau ist mit **PRODUKTIV** und **AKTIV** gemeint?

Produktive und aktive Morpheme

Wie jede natürliche Sprache ist auch das Deutsche einem permanenten Sprachwandel unterworfen. Im Zuge dieses Wandels kommt es auch vor, dass z. B. Wortbildungsmorpheme nicht mehr zur Wortbildung verwendet werden, also ihre **PRODUKTIVITÄT** verlieren. Neben dem erwähnten {ti} könnte man hier das von Eisenberg beschriebene {ling} anführen, das einstmals vor allem zur Bildung von Personenbezeichnungen aus Verben (*Lehrling, Säugling, Prüfling, Feigling, Erstling* (ursprünglich „der erste Sohn“) usw.), aber offenbar auch allgemeiner für Objekte verbaler Handlungen (*Bratling, Setzling*) verwendet wurde. Heutzutage ist es, wenn überhaupt, nur noch für pejorative (= abwertende) Personenbezeichnungen produktiv (*Schönling, Weichling*). Insgesamt kann man solchen Morphemen wohl eher den Status der **AKTIVITÄT** zuweisen, d. h. sie kommen zwar noch in Wörtern vor, werden aber nicht mehr zur Bildung neuer Wörter verwendet. Dieses Phänomen beschränkt sich nicht auf Wortbildungsmorpheme. Auch ein Basismorphem wie {heisch} kommt zwar noch aktiv im Deutschen vor (*er-heisch-en* und *beifallheischend/Beifall heischend*), ist aber deutlich auf dem Rückzug, während ein Basismorphem wie {Computer} hoch produktiv ist und in nahezu unendlichen Neubildungen auftritt. Manche (Basis-)Morpheme haben sich bereits so weit zurückgezogen, dass sie nur noch ein einziges Mal im Deutschen zu finden sind (z. B. {Him} in *Himbeere*, {Schorn} in *Schornstein*). Diese bezeichnet man als **UNIKALE MORPHEME** (lat.: unicus = einzigartig). Weitere Beispiele sind {lier} in *verlier(en)*, {ginn} in *Beginn* oder {winn} in *Gewinn*.

Unikale Morpheme

Definition

Produktives Morphem: Morphem, das in der Gegenwartssprache zur Wortbildung verwendet wird, wie {Computer}, {ung} etc.
Aktives Morphem: Morphem, das in der Gegenwartssprache noch vorkommt, aber nicht mehr zur Wortbildung verwendet wird, wie {ling}, {heisch} etc.
Unikales Morphem: Morphem, das nur noch restartig in einem Lexem vorkommt, wie z. B. {Him} in *Himbeere*.

Segmentierungsregel

Als Regel ist festzuhalten: Wörter sind dann noch segmentierbar, wenn beide Morpheme nicht nur aktiv sind, sondern ihnen auch ein Anteil an der Bedeutung der übergeordneten Konstituente zugewiesen werden kann. Bei {ei} in *Datei* war das noch möglich; bei *Himbeere* und *verlieren* kann man {Him} und {lier} zwar einen Anteil an der Bedeutung der gesamten Konstituente zuschreiben, aber es fällt schwer, diesen zu beschreiben. Sie stellen also Grenzfälle dar. Bei Wörtern wie *Schrift* ist es aus synchroner Sicht schon sehr schwierig zu erkennen, wo man sie segmentieren sollte. Ähnliches gilt prinzipiell für alle entlehnten Morpheme, wobei man allerdings bei dem Beispiel *online* argumentieren kann, dass in der deutschen Sprache das Wort *offline* ja ebenfalls recht weit verbreitet sei, was für eine Segmentierung spräche. Nicht aufs Glatteis führen lassen darf man sich auch von Wörtern wie *Hammer*, die auf

den ersten Blick wie eine Bildung aus dem Wortbildungssuffix {er} und einem Verbstamm {*hamm} wirken, den es allerdings nicht gibt.

Das Gegenstück zu den unikalenen Morphemen sind die sogenannten **KONFIXE**. Während sich Erstere auf dem sprachlichen Rückzug befinden, finden Letztere – in der Regel nachdem sie über ein oder mehrere Fremdwörter ins Deutsche gelangt sind – zunehmende Verbreitung durch ihre Verwendung in der Wortbildung. Typische Beispiele hierfür sind {bio}, {thek}, {polit}, {elektr} etc. Konfixe kommen nicht frei vor, sondern nur in Verbindung mit anderen Morphemen (lat.: confixum = zusammengeheftet). Man könnte einwenden, dass das bei Verbstämmen oft auch so ist; allerdings müssen Verbstämme nur mit einem Flexionsmorphem kombiniert werden, um eine Wortform zu bilden, während Konfixe in Verbindung mit einem Wortbildungsaffix oder einem weiteren Basismorphem überhaupt erst ein Wort bilden können. Konfixe unterscheiden sich auch dadurch von Basismorphemen, dass es praktisch unmöglich ist, sie einer bestimmten Wortart zuzuordnen. So kommt beispielsweise das Konfix {elektr} in Adjektiven (*elektr-isch*), Substantiven (*Elektr-ik*) und Verben (*elektr-isieren*) vor. Konfixe treten nicht mehr nur in Fremdwörtern (*Biologie, biometrisch, Biodynamik*), sondern auch in Mischungen aus Fremdwörtern und einheimischen Wörtern auf (*Bioladen, Biogemüse*).

Konfix

Konfix: gebundenes lexikalisches Morphem, das aus einer Fremdsprache stammt und in der Wortbildung verwendet wird.

Definition

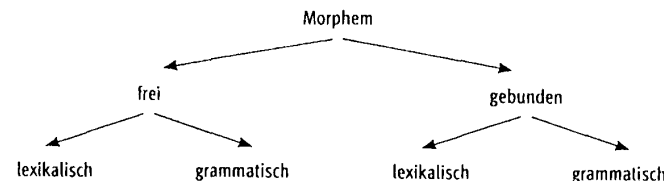
Übungen

15.5

1 Wie viele lexikalische bzw. syntaktische Wörter enthalten die folgenden Zeilen aus einem bekannten Kinderlied:

Auf der Mauer, auf der Lauer
 sitzt 'ne kleine Wanze.
 Seht euch nur die Wanze an,
 wie die Wanze tanzen kann!

2 Finden Sie für jede der Kategorien drei Beispiele. Was ist jeweils der Normalfall (unmarkierter Fall) und was der Ausnahmefall (markierter Fall)?



- 3 Segmentieren Sie die folgenden Lexeme und klassifizieren Sie die erhaltenen Morpheme nach den Kategorien Basis-, Wortbildungs- und Flexionsmorphem:

Rührei; Eimer; Mogelei; Rentner; Salbei; Rechner.

- 4 Fertigen Sie eine IC-Analyse der Lexeme *bedachen*, *Unabhängigkeit* und *Geschirrspülmaschine* an. Welche Probleme gibt es?

5.6 | Verwendete und weiterführende Literatur

- Best, Karl-Heinz (2005):** Linguistik in Kürze. Mit einem Ausblick auf die Quantitative Linguistik. 3., überarb. Aufl. Göttingen: Skript.
- Donalies, Elke (2005):** Die Wortbildung des Deutschen. 2., überarb. Aufl. Tübingen: Narr.
- Duden (2005) – Die Grammatik.** 7., völlig neu bearb. u. erw. Aufl. Hrsg. v. d. Dudenredaktionen. Mannheim et al.: Dudenverlag.
- Duden (2000) – Das große Wörterbuch der deutschen Sprache.** Hrsg. v. Wissenschaftlichen Rat der Dudenredaktion. CD-ROM-Ausgabe auf Basis der 3., völlig neu bearb. u. erw. Aufl. der Buchausgabe in 10 Bänden von 1999. Mannheim: Bibliographisches Institut.
- Eisenberg, Peter (2004):** Grundriß der deutschen Grammatik. Bd. 1. Das Wort. 2., überarb. u. aktual. Aufl. Stuttgart, Weimar: Metzler.
- Fleischer, Wolfgang; Barz, Irmhild (1995):** Wortbildung der deutschen Gegenwarts-
- sprache. 2., durchges. und erg. Aufl. Tübingen: Niemeyer.
- Kluge (2002) – Etymologisches Wörterbuch der deutschen Sprache.** Bearb. v. Elmar Seebold. 24., durchges. und erw. Aufl. Berlin, New York: de Gruyter.
- Lexikon der Sprachwissenschaft (2002).** Hrsg. v. Hadumod Bußmann. 3., aktual. und erw. Aufl. Stuttgart: Kröner.
- Lohde, Michael (2006):** Wortbildung des modernen Deutschen. Ein Lehr- und Übungsbuch. Tübingen: Narr.
- Metzler-Lexikon Sprache (2004).** Hrsg. v. Helmut Glück. CD-ROM-Ausgabe. Berlin: Directmedia Publishing.
- Motsch, Wolfgang (1999):** Deutsche Wortbildung in Grundzügen. Berlin, New York: de Gruyter.
- Muthmann, Gustav (2001):** Rückläufiges deutsches Wörterbuch: Handbuch der Wortausgänge im Deutschen mit Beachtung der Wort- und Lautstruktur. 3., überarb. und erw. Aufl. Tübingen: Niemeyer.

Wortbildung und Flexion

Inhalt

6.1	Wortbildung und Flexion – Wie wird ein Wort gebildet?	94
6.2	Komposition	95
6.3	Derivation	99
6.4	Kurzwortbildung	102
6.5	Sonstige Verfahren der Wortschatzerweiterung	104
6.6	Flexion	110
6.7	Übungen	113
6.8	Verwendete und weiterführende Literatur	114

6.1 | Wortbildung und Flexion – Wie wird ein Wort gebildet?

Wortbildung In den bisherigen Teilkapiteln zur Morphologie ging es darum, wie man Wörter bzw. Wortformen segmentieren und dann die nicht mehr weiter segmentierbaren Bestandteile klassifizieren kann. In den folgenden Teilkapiteln wird die umgekehrte Sichtweise eingenommen: Mit Hilfe welcher Prozesse werden aus Morphemen Wörter bzw. Wortformen? Dementsprechend wird es zunächst um die verschiedenen Typen der WORTBILDUNG und in Kapitel 6.6 dann um die Flexion gehen.

Definition **Wortbildung:** Prozess der Bildung neuer Wörter, welcher der Erweiterung des Wortschatzes einer Sprache dient. Dabei werden auf verschiedene Weise Morpheme miteinander kombiniert oder verändert.

Neologismus Zunächst einmal spricht man, wenn ein neues Wortbildungsprodukt in eine Sprache aufgenommen wird, von einem NEOLOGISMUS (griech.: néos = neu, lógos = Wort). Als solches wird es allerdings erst bezeichnet, wenn es sich dauerhaft im Wortschatz etabliert hat, also LEXIKALISIERT wurde. Der einfachste Fall eines Wortbildungsprodukts ist das bereits in Einheit 5 beschriebene

Simplex SIMPLEX, bei dem ein Morphem entweder mit einem Lexem identisch ist, wie z. B. bei den Adjektiven *frei, schön, rot, krank*, den Substantiven *Schrank, Wand, Haus, Ast, Zimmer, Wein*, oder, wie zum Beispiel bei den verbalen Basismorphemen {schlag}, {helf}, {schreib}, im Lexikon immer noch das Infinitivsuffix angefügt wird.

Morphologische Prozesse Es fällt auf, dass die meisten Simplizia ein- oder höchstens zweisilbig sind. Das ist kein Zufall. Sobald ein Wort mehr als zwei Silben hat, kann man in aller Regel davon ausgehen, dass es sich entweder um ein komplexes Wortbildungsprodukt oder um eine ENTLEHNUNG aus einer anderen Sprache handelt, die sich ohne Fremdsprachenkenntnisse nicht segmentieren lässt. Obwohl Entlehnungen einen beträchtlichen Anteil am Wachstum des Wortschatzes einer Sprache allgemein und des Deutschen im Besonderen haben, bilden den zentralen Gegenstand dieses Kapitels Wortbildungsprodukte, die entstehen, wenn indigene (lat.: indiges = einheimisch; vgl. Einheit 3) Morpheme des Deutschen miteinander kombiniert werden. Dabei sind verschiedene Verfahren grundsätzlich zu unterscheiden: zunächst die KOMPOSITION und die DERIVATION, die oft auch als ABLEITUNG bezeichnet werden. Auf beide Verfahren wird in den Kapiteln 6.2 und 6.3 eingegangen. Andere Verfahren zur Erweiterung des Wortschatzes wie Kurzwortbildung, Konversion, Entlehnung u. a. werden dann in den Kapiteln 6.4 und 6.5 beschrieben, bevor dann in Kapitel 6.6 die wichtigsten Aspekte des Verfahrens angesprochen werden, mittels dessen aus lexikalischen Wörtern syntaktische Wörter werden: der FLEXION.

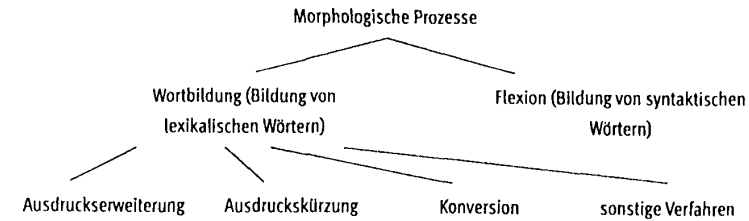


Abb. 6.1 Die morphologischen Prozesse im Überblick

Komposition

Bei der KOMPOSITION (lat.: compositio = Zusammensetzung) werden zwei lexikalische Morpheme miteinander verbunden. Im gängigsten Fall entsteht daraus ein DETERMINATIVKOMPOSITUM (lat.: determinare = bestimmen), bei dem der erste Teil (DETERMINANS) den zweiten (DETERMINATUM) näher bestimmt. Ein *Mülleimer* ist ein Eimer für Müll, eine *Haustür* ist die Tür, durch die man ins Haus kommt, eine *Waschmaschine* ist eine Maschine zum Waschen, eine *Bio-Kartoffel* ist eine Kartoffel aus biologischem Anbau usw. Man erkennt an diesen Beispielen schon zwei Dinge: Zum einen gilt für Determinativkomposita immer die Formel: „Ein AB ist ein B.“ Ein *Foto-Kalender* ist also ein Kalender, ein *Kalender-Foto* ist hingegen entweder ein Foto in einem Kalender oder von einem Kalender, auf jeden Fall ein Foto.

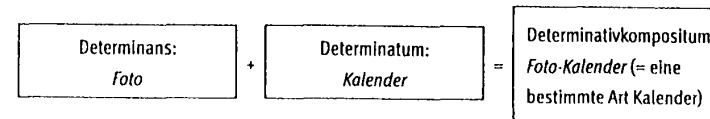


Abb. 6.2 Das Determinativkompositum

Am letzten Beispiel wird zudem das Phänomen der UNTERDETERMINIERTHEIT – mit anderen Worten: Unbestimmtheit – deutlich. Die meisten Komposita kann man nur verstehen, wenn man sein allgemeines Weltwissen zurate zieht. Bei einem nicht besonders gebräuchlichen Kompositum wie *Kalender-Foto* ist ohne Kontext schon nicht mehr ganz klar, was es eigentlich konkret bedeutet, weil dieses Wort zwar möglich, aber nicht lexikalisiert, d. h. nicht in unserem allgemeinen Sprachgebrauch verankert (und dementsprechend auch nicht im Duden verzeichnet) ist. Man spricht auch von AD-HOC-BILDUNGEN (lat.: ad hoc = [nur] zu diesem [Zweck]) oder OKKASIONALISMEN (lat.: occasio = Gelegenheit), GELEGENHEITS- oder AUGENBLICKSBILDUNGEN. Sie sind häufig durch Bindestrich-Schreibung markiert. Was z. B. die Erdbeer-Gruppe sein könnte, fällt Ihnen vielleicht am ehesten ein, wenn Sie an einen Kindergarten denken. Sehr viele solcher Bildungen findet man naturgemäß in der Literatur-

Unterdeterminiertheit

Ad-hoc-Bildung/ Okkasionalismus

sprache, wie das in Abb. 6.3 abgedruckte Gedicht von Paul Celan beispielhaft illustriert. Aber auch die Alltagssprache hat manchmal poetische Anwendungen. Eine *Schrankwand* beispielsweise ist nur in metaphorischer Hinsicht eine Wand, eigentlich aber ein Schrank.

Abb. 6.3 | Paul Celan: Anabasis. In: Ders. (1986): Gesammelte Werke. Bd. 1. Frankfurt am Main, 256f.

ANABASIS

Dieses	mit den
schmal zwischen Mauern geschriebne	sekundenschön hüpfenden
unwegsam-wahre	Atemreflexen -: Leucht-
Hinauf und Zurück	glockentöne (dum-,
in die herzhelle Zukunft.	dun-, un-,
	<i>unde suspirat</i>
Dort.	<i>cor</i> ,
Silben-	aus-
mole, meer-	gelöst, ein-
farben, weit	gelöst, unser.
ins Unbefahrne hinaus.	
	Sichtbares, Hörbares, das
Dann:	frei-
Bojen-,	werdende Zeltwort:
Kummerbojen-Spalier	Mitsammen.

Definition

Ad-hoc-Bildungen (auch: **Okkasionalismen, Gelegenheits- oder Augenblicksbildungen**): Wortbildungsprodukte, die situationsgebunden gebildet, aber nicht dauerhaft in den Wortschatz einer Sprache übernommen, also nicht lexikalisiert werden.

Rektionskompositum

Als Spezialfall des Determinativkompositums kann man das REKTIONSKOMPOSITUM (lat.: regere (Partizip Perfekt Passiv: rectum) = regieren) betrachten. Dessen Besonderheit besteht darin, dass das (von einem Verb abgeleitete) Zweitglied (REGENS) eine Ergänzung (REKTUM) fordert, so dass für das Kompositum nur eine Lesart möglich ist. Während also z. B. bei einer *Holz-kiste* allein aus dem Wortbildungsprodukt nicht deutlich wird, ob es sich um eine Kiste aus Holz oder zur Aufbewahrung von Holz handelt, herrschen bei *Deutsch-Lehrer* (*lehren* fordert eine Akkusativ-Ergänzung) oder *Terroristen-Fahndung* (*fahnden* fordert eine Präpositional-Ergänzung) keinerlei Zweifel, dass es sich um jemanden handelt, der Deutsch (Akk.) lehrt, bzw. um die Fahndung nach (Präp.) einem oder mehreren Terroristen. Allerdings existieren dieselben Zweitglieder auch in Determinativkomposita, die keine Rektionskomposita sind, wie die Beispiele *Junglehrer* oder *Rasterfahndung* belegen.

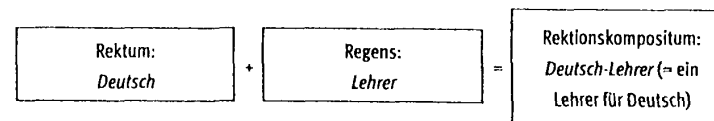


Abb. 6.4
Das Rektionskompositum

Die bisher genannten Beispiele für Determinativkomposita waren noch relativ überschaubar, da sie aus nur zwei Konstituenten bestanden. Bei Wortbildungsprodukten wie *Altherrenmannschaft* oder *Mehrzweckwaffe* wird es schon komplizierter. In beiden Fällen kommt die erste Konstituente nicht frei vor, sondern immer nur in Verbindung mit anderen Konstituenten (Letztere z. B. laut Rechtschreib-Duden (2006) in *Mehrzweckgerät, -halle, -maschine, -möbel, -raum, -tisch*, aber es sind noch weitere Komposita denkbar). Es handelt sich hierbei also um ein Kompositum aus einer bestimmenden Wortgruppe aus Adjektiv und Substantiv (*alte Herren, mehrere Zwecke*) plus einem dadurch näher bestimmten Substantiv (*Mannschaft, Waffe*). Ebenfalls mit Wortgruppen, die allerdings aus anderen Wortarten bestehen, sind Komposita wie *Geschirrspülmaschine, Zwanzig-Euro-Schein, Arzt-Patienten-Verhältnis, Nacht-und-Nebel-Aktion, Nachkriegsgeneration, Gewinn-und-Verlust-Rechnung, Fünf-vor-zwölf-Stimmung, Blut-Schweiß-und-Tränen-Rede* gebildet. Auch ganze Sätze findet man als Erstkonstituenten von Determinativkomposita wieder: *Jetzt-komm-ich-Ausstrahlung, Leck-mich-am-Arsch-Haltung*. Im Rahmen einer IC-Analyse würde man solche Wortbildungsprodukte zunächst ganz normal in zwei Komponenten segmentieren. Während sich die zweite Konstituente (also z. B. *Ausstrahlung*) in der Regel unproblematisch weiteranalysieren lässt, kann man die komplexe Erstkonstituente als Konvertat (lat.: convertere = umwandeln) aus einer Wortgruppe (*jetzt komme ich*) betrachten; was genau Konversion bedeutet, wird in Kapitel 6.5 beschrieben.

Kopulativkompositum

Neben den Determinativkomposita gibt es eine Reihe von Komposita, auf die die Formel „Ein AB ist ein B“ nicht zutrifft. Beispiele dafür sind *Hosenrock, Radiowecker, Affenmensch, süß-sauer, nasskalt*, bei denen man nicht sagen kann, ob ein AB mehr ein A oder ein B ist. Solche Wortbildungsprodukte werden als KOPULATIVKOMPOSITA (lat.: copulatio = Verknüpfung) bezeichnet; beide Konstituenten sind hier gleichrangig, ohne dass das Erstglied das Zweitglied determiniert.

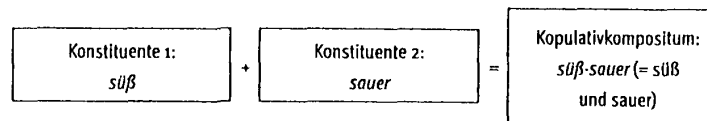


Abb. 6.5
Das Kopulativkompositum

Bei *Glatzkopf*, *Rotkehlchen* und *Zwölfzylinder* ist das AB weder ein A noch ein B, sondern etwas, das einen Glatzkopf, eine rote Kehle oder eben zwölf Zylinder besitzt, weswegen man auch von POSSESSIVKOMPOSITA (lat.: possidere = besitzen) spricht. Weil die Eigenschaften der Menschen, Vögel oder Autos (bzw. Motoren), die so benannt werden, gewissermaßen außerhalb des Zentrums dessen liegen, was die Komposita bezeichnen, werden diese als EXOZENTRISCHE (griech.-lat.: exo = außerhalb, centrum = Mittelpunkt) KOMPOSITA bezeichnet. Ähnliches gilt auch für *Nachmittag* oder *Untertasse*, die auch PRÄPOSITIONALE REKTIONSKOMPOSITA genannt werden, wobei allerdings hier, anders als bei den oben beschriebenen Rektionskomposita, das Erstglied ursprünglich das Zweitglied regiert. Allerdings haben sich bei diesen beiden Spezialkategorien der exozentrischen Komposita – wohl auch aufgrund ihrer Seltenheit – die Bezeichnungen bisher nicht allgemein durchgesetzt, weshalb solche Sonderfälle hier keiner eigenen Kategorie zugeordnet werden.

Ein letztes hier zu erwähnendes Phänomen ist der allmähliche Übergang von scheinbaren Basismorphemen zu Wortbildungsmorphemen in Wortbildungsprodukten wie *Riesenüberraschung*, *wirkungsvoll*, *abgasarm*. Die Konstituenten {Riesen}, {voll} und {arm} erinnern hier zwar an gleichlautende freie Morpheme, sind mit diesen aber nicht mehr bedeutungsgleich, sondern verfügen nur noch über eine reduzierte Bedeutung. Gleichzeitig wirken sie reihenbildend, d.h. sie kommen in mehreren Wortbildungsprodukten zur Anwendung. Eine *Riesenüberraschung* ist keine Überraschung von einem oder für einen Riesen, sondern eine große Überraschung. Eine *wirkungsvolle Verpackung* ist keine Verpackung, in der sich Wirkung befindet, sondern die Verpackung selbst besitzt eine gewisse Wirkung. Und Sie können mit den genannten Konstituenten nahezu beliebig viele neue Wörter bilden. Der 10-bändige Duden verzeichnet beispielsweise für {arm} die folgenden Bedeutungen und Beispiele:

Abb. 6.6 | Die Bedeutung des Suffixes {arm} nach Duden (2000)

1. drückt in Bildungen mit Substantiven aus, dass etw. nur in äußerst geringem Umfang vorhanden ist: emotions-, fleisch-, handlungsarm.
2. drückt in Bildungen mit Substantiven aus, dass sich etw. nur in äußerst geringem Umfang entwickelt, dass etw. nur in äußerst geringem Grad hervorgerufen wird: austausch-, emissions-, schadstoffarm.
3. a) drückt in Bildungen mit Verben (Verbstämmen) aus, dass die beschriebene Sache etw. nur in äußerst geringem Grad macht: knitter-, klirr-, rausch-arm; b) drückt in Bildungen mit Verben (Verbstämmen) aus, dass etw. nur in äußerst geringem Grad gemacht zu werden braucht: bedien-, bügel-, pflegearm.

Affixoid | Aufgrund dieses besonderen Verhaltens wurden derartige Morpheme zeitweise in der Forschungsliteratur zur Klasse der AFFIXOIDE zusammengefasst. Diese Einstufung verschleiert aber nur den Blick darauf, dass die Entwicklung von einem Wort zu einem Affix in der deutschen Sprachgeschichte nichts

Ungewöhnliches ist; z. B. entwickelte sich das Suffix {heit} aus dem mhd. Wort *heit*, das etwa „Art und Weise“, „Eigenschaft“, „Person“ bedeutete. Abgesehen davon löst die Einführung einer Zwischenkategorie nicht die Abgrenzungsproblematik zwischen Affix und Lexem, sondern verdoppelt sie noch, da sie drei Kategorien mit zwei Abgrenzungen zur Folge hat. Im Einzelfall kann man sich genauso gut zwischen Noch-Wörtern und Schon-Affixen entscheiden.

Derivation

In Einheit 5, die die morphologische Analyse von Wörtern behandelt, ist Ihnen bereits der Terminus „Wortbildungsmorpheme“ begegnet. Diese werden auch als Derivationsmorpheme oder Derivateme bezeichnet. Das bedeutet nun nicht, dass Wortbildung und DERIVATION (lat.: derivatio = Ableitung) das Gleiche sind. Vielmehr handelt es sich bei Letzterem – genauer gesagt: bei der EXPLIZITEN DERIVATION – um den neben der Komposition zweiten wichtigen Prozess zur Bildung neuer Wörter. Dabei werden, anders als bei der Komposition, keine freien, lexikalischen Morpheme miteinander kombiniert, sondern an in der Regel freie, lexikalische Morpheme (z. B. {schön}, {freund}) werden Morpheme (z. B. {heit}, {lich}) angehängt (affigiert), die als gebunden und tendenziell eher grammatisch einzustufen sind, wobei auch hier die Ausnahme die Regel bestätigt. Denn wie bereits in Einheit 5 ausgeführt, dient die große Mehrheit der Wortbildungsmorpheme zwar dazu, Wörter in eine andere Wortart zu überführen; das trifft z. B. auf Morpheme wie {er}, {heit} oder {ung} zu, die dazu dienen, beispielsweise aus einem Verb oder Adjektiv ein Substantiv abzuleiten (*Lehrer*, *Dunkelheit*, *Erinnerung*). Auch Verben (z. B. mit {be} in *beantwort(en)*), Adjektive (z. B. mit {lich} in *erfreulich*) oder Adverbien (z. B. mit {ens} in *höchstens*, *bestens*) lassen sich so bilden. Neben dieser Funktion, z. B. aus einer verbalen Basis ein substantivisches Wortbildungsprodukt zu gewinnen, tragen diese Morpheme aber immer auch eine Bedeutung. Das Wortbildungssuffix {er} z. B. bezeichnet in den meisten Fällen eine Person, die die durch das Verb bezeichnete Tätigkeit ausübt. Ein *Lehrer* ist also „jemand, der lehrt“; allerdings ist ein *Jauchzer* keine Person, sondern eine Handlung. Noch stärker in den Vordergrund tritt die lexikalische Komponente von Wortbildungsmorphemen, wenn mit ihnen beispielsweise aus einem Substantiv ein anderes Substantiv gebildet wird, wie es bei {er} in *Musiker* oder *Chemiker* der Fall ist. Auch bei einem Suffix wie {bar} lässt sich die Bedeutung oft problemlos paraphrasieren: Wenn etwas *erlernbar* ist, dann kann es erlernt werden. Das Suffix {bar} bezeichnet in Zusammenhang mit einem verbalen Basismorphem die Tatsache, dass die dadurch bezeichnete Handlung auf einen bestimmten Gegenstand angewendet werden kann, mit anderen Worten: *anwendbar* ist. Ein Morphem wie das bereits erwähnte {un} hat sogar fast ausschließlich lexikalischen Charakter, indem es prinzipiell der Negation von Substantiven

16.3

Explizite Derivation

(*Undank, Unordnung*) oder Adjektiven (*unfrei, unvorsichtig*) dient, ohne deren Wortart zu beeinflussen. Allerdings lässt es sich nicht immer einfach mit *nicht* paraphrasieren, wie das Beispiel *Unfall* zeigt (hier bedeutet {un} soviel wie „vom Normalen abweichend“).

Wie die bisherigen Beispiele für Derivation gezeigt haben, lassen sich verschiedene Typen differenzieren, je nachdem, wo etwas angehängt wird und was dabei als Wortbildungsprodukt entsteht. Es wird unterschieden zwischen nominalen (z. B. *Erz-feind, Un-tat*), adjektivischen (*ur-alt, un-frei*) und verbalen (*ver-antwort(en), be-lüg(en), miss-acht(en)*) PRÄFIXDERIVATEN. Bei den SUFFIXDERIVATEN findet man nominale (*Einsam-keit, Finster-nis, Anhörung*), adjektivische (*brenn-bar, ekel-haft, vorsicht-ig*) sowie wenige verbale (*halb-ier(en), brumm-el(n)*) und adverbiale (*ander-s, probe-halber, ab-wärts, rück-lings*). Relativ selten schließlich sind nominale (*Ge-birg-e*), adjektivische (*ge-lehr-ig*) und verbale (*be-sänft-ig(en)*) ZIRKUMFIXDERIVATE, wobei beim letzten Beispiel die Zirkumfixierung mit einer Umlautung der adjektivischen Basis einhergeht ({sanft} → {sänft}). Die Zirkumfixderivation wird auch KOMBINATORISCHE DERIVATION genannt, weil hier gewissermaßen ein Präfix und ein Suffix in Kombination angehängt werden, was man auch bei der IC-Analyse solcher Wörter berücksichtigen muss (vgl. Kapitel 5.4). Wie Sie sehen, sind manche Affixe auf eine Wortart beschränkt (z. B. {lich} auf die adjektivische Derivation), während z. B. die Präfixe {un}, {ur} oder {erz} jeweils sowohl beim Adjektiv als auch beim Substantiv vorkommen, was auch mit ihrem tendenziell lexikalischen Charakter zusammenhängt. Adverbiale Präfix- und Zirkumfixderivate gibt es im Deutschen nicht; falls Sie eins finden, können Sie (jedenfalls in der Linguistik) berühmt werden.

Einen Spezialfall der expliziten Derivation stellen die sogenannten ZUSAMMENBILDUNGEN dar. Auf den ersten Blick könnte man z. B. *Geldgeber* für ein Determinativkompositum und *blauläugig* für ein herkömmliches Derivat halten. Allerdings kommen weder *blauaug* noch *Geber* in der Standardsprache vor; Letzteres werden Sie vielleicht bestreiten, wenn Sie Skat spielen, aber in diesem Fall sei noch der *Dickhäuter* angeführt, wo wohl niemand behaupten kann, er habe schon mal einen „Häuter“ getroffen. Hier wird vielmehr, ähnlich wie bei den komplexen Determinativkomposita, aus einer Wortgruppe als Basis (*Geld + geb(en), blau + Aug(e), dick + Haut*) mit Hilfe eines Derivationsuffixes ({er}, {ig}) ein Derivat abgeleitet. Auch hierbei wird bisweilen begleitend der Vokal der substantivischen Basis umgelautet.

Wenn eingangs dieses Kapitels von expliziter Derivation gesprochen wurde, stellt sich automatisch die Frage, wozu das Attribut *explizit* (lat.: explicare = auseinanderfalten, ausdehnen) notwendig ist. Die Lösung erkennt man leicht, wenn man sich Wörter wie *setz(en), Trank* oder *Schritt* anschaut. Dort wird nichts an die ursprünglich verbale Basis ({sitz}, {trink}, {schreit}) angehängt, sondern das Derivat wird durch eine Veränderung des Basismorphems gewonnen. Man spricht daher von IMPLIZITER DERIVATION (lat.: implicare; Part. Perf.: implicatum/implicitum = verflechten, einwickeln; eingewickelt). Die nominalen Derivate wie *Wurf* und *Flug* werden allerdings in der Regel diachron erklärt, indem sie allgemein historisch vor dem Hintergrund des mhd. Ablauts oder spezieller als Konvertate (vgl. Kapitel 6.5) von mhd. Präteritumstämmen (vgl. z. B. 3. Pers. Pl. Prät.: *sie wurfen, sie flugen*) betrachtet werden. Somit blieben für den Wortbildungsprozess der impliziten Derivation nur noch die von einer verbalen Basis abgeleiteten impliziten verbalen Derivate wie *tränk(en)* und *setz(en)* übrig. In jedem Fall handelt es sich hierbei um ein nicht mehr produktives Muster der Wortbildung.

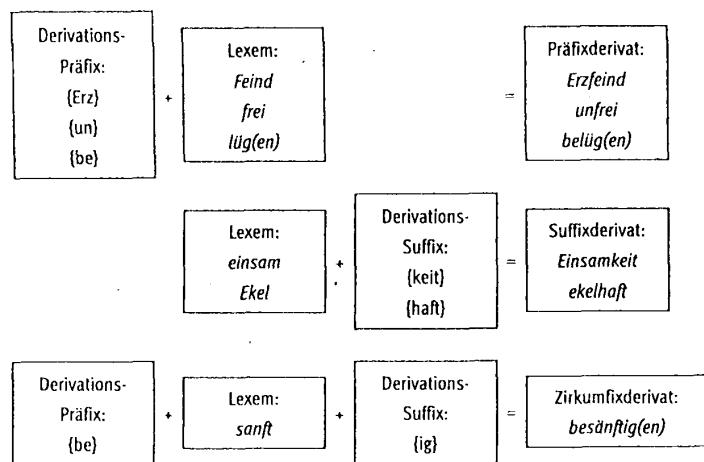
Aus einer diachronen Perspektive werden häufig auch Lexeme wie *Sanftmut* und *notlanden* betrachtet. In beiden Fällen wird dann davon ausgegangen, dass es sich bei *sanftmütig* und *Notlandung* nicht, wie man eigentlich vermuten würde, um Derivationsprodukte handelt, weil Letztere älter seien als Erstere. Wenn man also der Auffassung folgt, dass das verbale Basismorphem {notland} aus dem substantivischen Basismorphem *Notlandung* und das substantivische Basismorphem {sanftmut} aus dem Adjektiv *sanftmütig* abgeleitet wurden, hat man im Ergebnis keine explizite Derivation, sondern genau das Gegenteil einer Affigierung. Dieser Prozess wird dann als RÜCKBILDUNG bezeichnet. In einigen Publikationen (vgl. z. B. Fleischer/Barz 1995) zählen hierzu auch solche Fälle, in denen das rückgebildete Basismorphem dann wiederum mit einem Suffix versehen wird. Beispiele wären *Häm-e* aus *hämisch*, *Emanz-e* aus *Emanzipation* sowie die zahlreichen Spitznamen auf *-i* in diversen Varianten (*Klins-i* ← *Klinsmann*, *Ol(l)-i* ← *Oliver*, *Ang-ie* ← *Angela*). Dieses Muster, bei dem eigentlich zwei Wortbildungsprozesse – Kürzung und Derivation – gleichzei-

Zusammenbildung

Implizite Derivation

Rückbildung

Abb. 6.7 | Verschiedene explizite Derivationstypen



tig ablaufen, ist in der Gegenwartssprache sehr produktiv. Insgesamt bewegt sich die Rückbildung, wenn man sie denn als Wortbildungsprozess akzeptiert, im Grenzbereich zwischen der Derivation und der Kurzwortbildung, auf die als nächstes eingegangen wird.

6.4 | Kurzwortbildung

Ausdrucks Kürzung Die bisher beschriebenen Wortbildungstypen hatten (mit Ausnahme der impliziten Derivation und mit Einschränkungen bei der Rückbildung) gemeinsam, dass dabei die Ausdrucksseite der jeweiligen Basis erweitert wurde, weswegen sie auch unter dem Terminus **AUSDRUCKSERWEITERUNG** zusammengefasst werden. Diesem Verfahren gegenüber steht die sogenannte **AUSDRUCKSKÜRZUNG**, bei der entsprechend die jeweilige Basis verkürzt wird. Zu beachten ist dabei zunächst, dass nur dann von Wortbildung gesprochen wird, wenn sowohl auf graphematischer als auch auf phonologischer Ebene ein neues Wort entsteht. Rein schriftsprachliche Abkürzungen wie z. B., *d. h.*, *etc.*, *usw.* gehören also nicht dazu.

Kurzwortbildung Den einfachsten Fall der **KURZWORTBILDUNG** stellt das sogenannte **UNISEGMENTALE KURZWORT** (lat.: unus = einer; segmentum = Abschnitt) dar. Hierbei wird gewissermaßen einfach ein Teil der jeweiligen Basis abgeschnitten. Der übrig bleibende Abschnitt wird dann nach seiner ursprünglichen Position in der zugehörigen Langform bezeichnet. Ein *Auto* (← *Automobil*) oder eine *Lok* (← *Lokomotive*) gehören demnach in die Klasse der **ANFANGSWÖRTER**, ein *Rad* (← *Fahrrad*) oder die *Bahn* (← *Eisenbahn*) hingegen zu den **ENDWÖRTERN**. Manchmal ändert sich zusätzlich das Genus beim Kurzwort (z. B. *der Frust* ← *die Frustration*) oder das Kurzwort ist ohne Kontext mehrdeutig (*Platte* ← *Schallplatte*, *Herdplatte*, *Tischplatte* etc.). Besonders der Kurzwortbildungstyp des Anfangsworts ist in der Umgangssprache produktiv bzw. aktiv, was man an unisegmentalen Kurzwörtern wie *Abi*, *Demo*, *Uni*, *Prof* usw. sieht.

Rumpfwort Relativ selten sind sogenannte **RUMPFWÖRTER** wie *Lisa* (← *Elisabeth*), die meistens bei Namen vorkommen. Umstritten ist, ob Wörter wie *Bierdeckel* oder *Kirschblüte* als **KLAMMERFORM** (zu *Bierglasdeckel*, *Kirschbaumblüte*) aufzufassen oder nicht doch eher direkt entstanden sind; ein Bierdeckel wäre demnach ein Deckel, auf den man sein Bier stellt, während bei der Kirschblüte das Morphem {Kirsch} als *Pars pro toto* (lat.: Teil anstelle des Ganzen) für *Kirschbaum* verwendet würde; das zehnbändige Duden-Wörterbuch der deutschen Sprache kennt beispielsweise die Formulierung *Die Kirschen blühen schon*. In jedem Fall wären sie dann als **MULTISEGMENTALE KURZWÖRTER** (lat.: multi = viele) zu werten, auf die im Folgenden näher eingegangen wird.

Multisegmentales Kurzwort/Akronym Multisegmental ist ein Kurzwort dann, wenn sich nicht nur ein Ausschnitt, sondern mehrere aus der Langform im Kurzwort wiederfinden. Solche Kurzwörter werden bisweilen auch als **AKRONYME** bezeichnet. Der häufigste Fall ist

das **INITIALABKÜRZUNGSWORT** (*PKW* ← *Personenkraftwagen*, *ICE* ← *InterCity-Express*). Dieser Wortbildungsprozess gehört zu den aktuell produktivsten und lässt sich beispielsweise im Computerwortschatz mit zahlreichen Beispielen belegen (*EDV*, *WWW*, *URL*), wobei viele dieser Initialabkürzungswörter aus dem Englischen entlehnt sind, was dazu führt, dass die Langformen häufig zumindest dem Laien nicht präsent sind – die meisten von Ihnen wussten bisher vermutlich nicht, dass die Langform zu *URL Uniform Resource Locator* lautet, geschweige denn was das genau bedeutet, so dass die semantisch eigentlich redundante Bildung *URL-Adresse*, die man beim Googeln in sechsstelliger Anzahl antrifft, Ihnen auf den ersten Blick vielleicht nicht besonders auffällig erscheint. Insofern ist die verbreitete These, dass ein Kurzwort immer parallel zu seiner Langform im Wortschatz einer Sprache existiert, mit Vorsicht zu genießen.

Oftmals scheinen Kurzwörter bewusst so gebildet zu werden, dass man sie nicht nach ihrem Buchstabenwert ausspricht (also z. B. *WWW* [ve:ve:ve:]), sondern nach ihrem Lautwert wie z. B. bei *UFO*, *UNO*, *TÜV*. Auch Mischformen kommen vor (*WLAN*: [ˈve:la:n]). Ebenfalls offenbar aufgrund ihrer besseren Aussprechbarkeit werden **SILBENKURZWÖRTER** gebildet. Reine Silbenkurzwörter sind *Kripo* (← *Kriminalpolizei*) und mit Einschränkungen noch *Hanuta* (← *Haselnusstafel*), während bei *Azubi* (← *Auszubildender*) und *BAföG* (*Bundesausbildungsförderungsgesetz*) wiederum Mischformen aus Initialabkürzungswort und Silbenkurzwort vorliegen; bei Letzterem schlägt sich das auch in der etwas seltsamen Orthographie nieder.

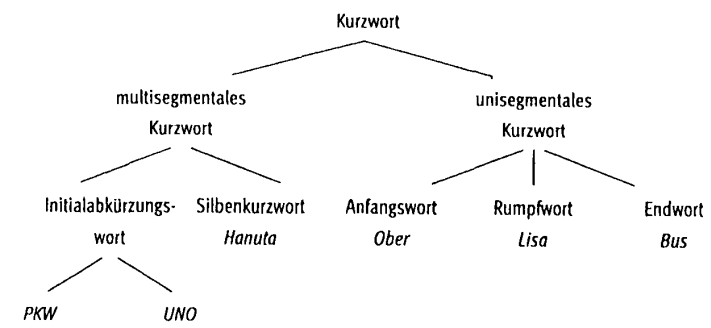


Abb. 6.8 Die wichtigsten Kurzworttypen

Schließlich gibt es auch noch den Spezialfall des **PARTIELLEN KURZWORTES**, bei dem nur ein Teil der Langform abgekürzt wird (*U-Bahn* ← *Untergrundbahn*). Dieser Wortbildungstyp ist eher selten und zu unterscheiden von dem Phänomen, dass frei vorkommende Kurzwörter ihrerseits wieder als Konstituenten von komplexeren Wortbildungsprodukten auftreten können (*UNO-Mitarbeiter*, *EDV-Seminar*). Überhaupt scheinen der Kreativität bei der

Kurzwortbildung kaum Grenzen gesetzt zu sein, was Sie u. a. in den 1980er Jahren am damals modernen *BTX* (← *Bildschirmtext*) und in Zukunft an Ihrer Steuererklärung (*Pauschbetrag* ← *Pauschalbetrag*) ablesen können. Während das erste Beispiel sich wohl am ehesten den Initialabkürzungswörtern zurechnen lässt, handelt es sich bei Letzterem um einen Grenzfall zwischen Klammerbildung und partiellem Kurzwort. Von einer gewissen Kreativität zeugen auch die ebenfalls noch im Bereich der Ausdruckskürzung anzusiedelnden Phänomene der **KONTAMINATION**, bei denen zwei Wörter entweder gekreuzt werden – d. h., zumindest eines der beiden ursprünglichen Wörter verliert einen Teil seines phonetischen und/oder orthographischen Materials wie in *jein*, *Teuro* oder *Demokratur* (← *Demokratie* + *Diktatur*) – oder miteinander verschmelzen, ohne dass eines der ursprünglichen Wörter etwas aufgeben muss (*verschlimmbessern*), wobei besonders Letztere es nur selten ins Lexikon schaffen. Alltagssprachlich relevant sind eigentlich nur die bisweilen unter dem Terminus **KONTRAKTION** zusammengefassten Wortkreuzungen aus einer Präposition und einem Artikel (*am*, *beim*, *zur*, *ins*).

Kontamination

Kontraktion

6.5 | Sonstige Verfahren der Wortschatzerweiterung

Während bei den bisherigen Wortbildungstypen entweder etwas hinzugefügt bzw. kombiniert, abgeschnitten oder doch zumindest verändert wurde, gibt es auch eine Möglichkeit, den Wortschatz zu erweitern, ohne rein äußerlich etwas an der jeweiligen Wortbasis zu verändern. Diesen Prozess nennt man **KONVERSION**. Diese geht immer mit einem Wechsel der Wortart einher. Der häufigste Fall ist die Substantivierung. Anhand von Verben, die entweder als reine Verbstämme ({*treff*} → *Treff*) oder inkl. ihrer Infinitivendung (→ *das Treffen*) konvertieren, lässt sich zeigen, dass es zwei verschiedene Typen von Konversion gibt. Bei der **SYNTAKTISCHEN KONVERSION** wird das Konvertat so, wie es in der Basis auftritt, in eine andere Wortart übertragen; im Fall von *Treffen* z. B. vom Infinitiv des Verbs in ein Substantiv. Ähnliches gilt für die aus dem verbalen Partizip II konvertierten Adjektive (*sie hat die Frage beantwortet*, *er hat die Wahl gewonnen* → *die beantwortete Frage*, *die gewonnene Wahl*), die in bestimmten Formulierungen in derselben Form wie das Partizip II vorkommen (*die Frage ist beantwortet*, *die Wahl gilt als gewonnen*), sowie die aus Adjektiven konvertierten Substantive (*ein guter Wein* → *ein Guter*) und die seltenen, aus Substantiven konvertierten Adjektive (*es ist mir ernst*). Auch bei substantivierten Adverbien (*das Gegenüber*), Konjunktionen (*ohne Wenn und Aber*), Satzäquivalenten (*ein klares Nein*) und Wortgruppen (*das So-tun-als-ob*) kann man von syntaktischen Konversionen sprechen. Letztere können dann, wie in Kapitel 6.2 beschrieben, wiederum als komplexe Erstkonstituenten von Determinativkomposita auftreten (*So-tun-als-ob-Haltung*).

Konversion

Syntaktische Konversion

Demgegenüber führt die **MORPHOLOGISCHE KONVERSION** zu Wortbildungsprodukten, die sich morphologisch in allen Verwendungen von ihrer Basis unterscheiden. Dies sieht man bei Verbstämmen wie *Treff*, die nur als substantiviertes Konvertat frei vorkommen, und bei aus Adjektiven konvertierten Verben wie *grünen*. Ansonsten ist dieser Wortbildungstyp gegenwärtig noch vereinzelt bei aus Substantiven konvertierten Verben produktiv, besonders im Bereich des Computerwortschatzes: Wenn Sie sich die Verben *chatten*, *googeln* und *computern* anschauen, können Sie hoffentlich zustimmen, dass *chatten* im Deutschen aus der im Deutschen zunächst nur als Substantiv vorhandenen Entlehnung *Chat* konvertiert wurde; ebenso verhält es sich mit *googeln*, bei dem ein Eigenname *Pate* stand. Bei *computern* fragen Sie sich vielleicht, wer das benutzt, aber da es im Duden steht, wird es hier als eindeutiger Beleg einer morphologischen Konversion eines Substantivs zu einem Verb mit aufgeführt. Anzumerken ist an dieser Stelle, dass sich bei syntaktischen Konversionen, bei denen ein Substantiv Ausgangspunkt oder Ergebnis ist, natürlich doch etwas ändert, nämlich die Orthographie (*Ernst* → *ernst*). Der orthographische Aspekt wird aber von der Morphologie vernachlässigt.

Morphologische Konversion

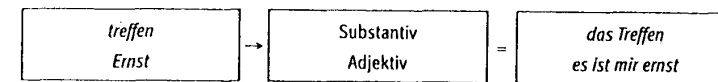


Abb. 6.9 Syntaktische Konversion

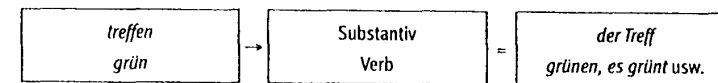


Abb. 6.10 Morphologische Konversion

In den bisherigen Ausführungen dieser Einheit ging es darum, wie der Wortschatz dadurch erweitert wird, dass Morpheme mit anderen Morphemen kombiniert, verändert, verkürzt oder in eine andere Wortart übertragen werden, mit anderen Worten: dass ihre Ausdrucksseite oder zumindest ihre Zugehörigkeit zu einer grammatischen Kategorie verändert wurde. Es gibt jedoch noch zahlreiche weitere Möglichkeiten, den Wortschatz einer Sprache zu vergrößern, von denen hier nur die wichtigsten kurz angesprochen werden können.

Das in der Gegenwartssprache am häufigsten genutzte Verfahren hierzu ist die **ENTLEHNUNG**. Dabei wird ein Wort aus einer Fremdsprache übernommen, wobei es in punkto Aussprache, Flexion und teilweise auch Orthographie an das Deutsche angepasst wird. Ein Bereich, in dem dieses Verfahren gegenwärtig besonders produktiv ist, ist der Computerwortschatz. In erster Linie werden so Substantive aus dem Englischen übernommen, wie die Belege *Computer*,

Entlehnung

Lehnübersetzung

Internet, Browser, Homepage etc. zeigen. Aber auch Verben wie *downloaden, chatten* und Kurzwörter wie *URL* oder *RAM* werden übernommen. Manchmal tritt an die Stelle eines Fremdworts eine sogenannte LEHNÜBERSETZUNG; aus der *Mouse* wird dann die *Maus*, aus *downloaden* wird *herunterladen*. Solche Übersetzungsprozesse kommen besonders dann zur Anwendung, wenn die Metaphorik eines Begriffs auch im Deutschen ersichtlich bleiben soll, wie es bei der *Datenautobahn* der Fall ist, die allerdings nicht aus *data highway*, sondern aus *information highway* entlehnt ist.

Vielleicht stellen Sie sich jetzt die Frage, inwieweit man bei Entlehnungen überhaupt von Wortbildungsprodukten sprechen kann, wenn doch eigentlich gar keine Wortbildung stattfindet. Das ist aus synchroner Sicht ein sehr berechtigter Einwand. Gleichzeitig ist aber zu berücksichtigen, dass jedes Wort seine Etymologie (griech: *etymologia* = Ableitung), also seine Wortgeschichte hat. Dabei entpuppen sich viele vermeintlich indigene (oder auch native (lat.: *sermo natus* = Muttersprache)) Wörter diachron betrachtet als Entlehnungen. Aus dem Lateinischen kommen *Mauer* von *murus*, *Fenster* von *fenestra*, *krass* von *crassus* (= dick, grob), *opfern* von *operari*; aus dem Alt-Griechischen stammen z. B. *logisch* (*lógos* = u. a. Vernunft) oder *Meter* (*metréo* = messen); aus dem (Alt-) Französischen kommen *Jacke* (afz.: *jacque* = Waffenrock) oder *Puder* (*poudre* = Staub, *Pulver*; vgl. lat.: *pulvis* = Staub, Sand, Asche); englischer Herkunft sind neben vielen offensichtlichen Anglizismen u. a. auch die sehr einheimisch wirkenden Lexeme *stoppen*, *Dogge*, *dribbeln* (= tröpfeln, d. h. den Ball durch kurze Stöße vorwärts treiben), *Elf(e)*, *Tank* usw.

Fremdwort

Anhand dieser Beispiele lässt sich festhalten, dass die Identifizierung von sogenannten Fremdwörtern im Deutschen nicht immer ganz einfach ist. Auf den Terminus „Lehnwort“ kann man aus synchroner Perspektive gänzlich verzichten, da sich ein Lehnwort nach dem Verständnis vieler traditioneller Grammatiken ja gerade dadurch auszeichnet, dass es sich den strukturellen Eigenheiten der deutschen Sprache so weit angepasst hat, dass man ihm seine Herkunft nicht mehr ansieht. Unter einem FREMDWORT wird in dieser Einführung ein Wort verstanden, das Produkt einer Entlehnung ist und sich phonologisch und/oder graphematisch für das Deutsche untypisch verhält.

Assimilation

Im Rahmen einer Entlehnung kommt es in der Regel zu einer syntaktischen, morphologischen, phonetischen und/oder orthographischen ASSIMILATION (lat.: *assimulare* = ähnlich machen, angleichen). Im Deutschen sagt man nicht **ich chat* oder **ich habe chatted mit ihm*, sondern versieht das Verb mit indigenen Flexionssuffixen; außerdem verdoppelt man im Präsens orthographisch den Konsonanten bzw. realisiert im Partizip Perfekt sowohl phonetisch als auch orthographisch die Auslautverhärtung (von der eingedeutschten Aussprache des Vokals ganz zu schweigen) und verwendet es gemäß der im Deutschen üblichen Wortstellung: *ich chat-te* bzw. *ich habe mit ihm ge-chatt-et*.

Viele entlehnte Lexeme werden ihrerseits im Deutschen produktiv, d. h. sie finden sich in sogenannten HYBRIDEN (lat.: *hybrida* = Mischling) KOMPOSITA (*Internetanbieter, Computerfachmann* sowie die bereits erwähnte, semantisch redundante *URL-Adresse*) oder manchmal sogar HYBRIDEN DERIVATIONEN (*to log in* → *einloggen*) wieder. Einige Fremdwörter kommen nicht frei, sondern ausschließlich in Komposita vor. Dazu gehört *Bio*, dessen Beteiligung in *Bioladen* in jedem Fall zu einem hybriden Kompositum führt. Bei *Biotechnologie* hingegen sieht man dem Wort wie so häufig bei hybriden Bildungen nicht an, ob es sich um ein im Deutschen gebildetes Wortbildungsprodukt aus den Morphemen {bio}, {techn(o)}, {log} und {ie} handelt oder ob das Wort komplett aus dem Englischen entlehnt wurde.

Die hier aufgezählten Beispiele dürfen nicht darüber hinwegtäuschen, dass es im Deutschen zahlreiche Fremdwörter aus anderen Sprachen gibt. In vielen Fällen kann man an der Übernahme bestimmter Lexeme aus anderen Sprachen etwas über gesellschaftliche Phänomene lernen – die Entlehnung von Wörtern ist oft mit der Einführung von technischen Innovationen, mit allgemeinen kulturellen Entwicklungen o. Ä. verknüpft. Während in der Computersprache eine relativ hohe Anzahl von Anglizismen mit dem Status des Englischen als *Lingua franca* einhergeht, sieht es im Bereich des Kulinarischen deutlich vielfältiger aus. Auch Sie benutzen in Ihrem Alltag vermutlich Entlehnungen aus dem Englischen (*Burger*), Italienischen (*Spaghetti, Frikadelle*), Spanischen (*Tapas*), Französischen (*Pommes frites, Frikassee*), Russischen (*Wodka*), Sorbischen (*Quark*), Türkischen (*Kaviar, Döner*), Griechischen (*Gyros*), Chinesischen (*Tee*), Japanischen (*Sushi*) sowie nicht zuletzt der mittelamerikanischen Indianersprache Nahuatl (*Kakao, Schokolade*). Dabei lässt sich am Zeitpunkt, zu dem bestimmte Wörter auftauchen, und an der Häufigkeit von Entlehnungen aus einer bestimmten Sprache durchaus etwas über kulturelle Prozesse aussagen; das Wort *Döner* z. B. ist noch relativ jung und eines der wenigen türkischen Fremdwörter im Deutschen. Gleichzeitig lässt sich an manchen Wörtern förmlich die Reise nachvollziehen, die die entsprechenden Inhalte historisch erlebt haben. So wurde *Schokolade* über das Spanische und Niederländische ins Deutsche entlehnt.

Ein interessantes Phänomen, das eng mit der Entlehnung zusammenhängt, ist die sogenannte VOLKSETYMOLOGIE. Wenn Sie bisher dachten, eine *Hängematte* sei gebildet aus dem verbalen Basismorphem {häng} und dem substantivischen Basismorphem {Matte}, unterliegen Sie einem verbreiteten Irrtum; das Wort stammt von dem karibischen *hamáka* ab, mit dem die Eingeborenen in Haiti ihre Schlafnetze bezeichneten, und wurde nachträglich dem plausibler erscheinenden *Hängematte* phonetisch angeglichen. Auch der *Maulwurf* heißt nicht so, weil er sein Maul zum Graben benutzt, sondern stammt vom ahd. *mülwerfo* ab, das *Haufenwerfer* bedeutet. Bei der Volksetymologie weist die Sprachgemeinschaft also aus synchroner Perspektive einem Lexem eine Etymologie zu, die mit der historisch korrekten Etymologie nichts zu tun hat.

Hybride Komposita

Hybride Derivationen

Volksetymologie

Bedeutungswandel Im Gegensatz zu den bisher beschriebenen sonstigen Verfahren der Wortschatzerweiterung spielt sich die **BEDEUTUNGSVERÄNDERUNG** bzw. der **BEDEUTUNGSWANDEL** ausschließlich auf der Inhaltsseite von Lexemen ab.

Bedeutungsverengung Dabei kann die **BEDEUTUNG VERENGT** werden; typische Beispiele hierfür sind *Hochzeit*, womit im Mhd. noch ganz allgemein ein Fest bezeichnet wurde, oder *Schirm*, dessen ursprünglich weitere Bedeutung „Schutz“ heute noch im Ausdruck *Schirmherrschaft* anklingt. Umgekehrt kann man eine **BEDEUTUNGSERWEITERUNG** an den Wörtern *Horn* und *packen* beschreiben; während Ersteres ursprünglich nur einen Auswuchs an der Stirn eines Tieres bezeichnete, war Letzteres im 16. Jahrhundert nur im Sinne von *einpacken* und nicht wie heute auch als *fassen*, *er-* bzw. *begreifen* gebräuchlich. Eine **BEDEUTUNGSVERSCHIEBUNG** oder **-ÜBERTRAGUNG** kann man in *Zweck* erkennen, das heute nur noch in *Reißzwecke* seine frühere Bedeutung „Nagel“ konserviert hat. Oftmals steht dabei eine übertragene Bedeutung neben der eigentlichen, wie z. B. an der metaphorisch (griech.: *metaphorá* = Übertragung) verwendeten Konstituente *Rücken* in *Buchrücken* oder der metonymischen (griech.: *metonymía* = Umbenennung) Verwendung von *Leder* für *Fußball*, obwohl dieser heutzutage in der Regel aus Kunststoff besteht. Erwähnenswert sind schließlich noch **BEDEUTUNGS-AUF- und -ABWERTUNGEN**, wie sie in *toll* (ursprünglich etwa *verwirrt*, *verrückt*) und *gemein* (eigentlich *allgemein*) auftreten. Der Bedeutungswandel eines Wortes gibt oftmals Aufschluss über gesellschaftliche Einstellungen zu bestimmten Sachverhalten, wie man z. B. daran sehen kann, dass das Adjektiv *schwul* in vielen Bereichen als Schimpfwort verwendet wird. Insgesamt gilt für alle Phänomene des Bedeutungswandels, dass sie sich ausschließlich auf die Inhaltsseite eines Lexems beziehen. Sie sind also nicht im engeren Sinne der Wortbildung zuzurechnen.

Wortschöpfung Schließlich stellt sich die Frage, ob der Wortschatz eigentlich auch noch mit völlig neuen Morphemen erweitert wird, mit anderen Worten: Gibt es das Phänomen der **WORTSCHÖPFUNG**? In diesem Zusammenhang wird oft der vermeintliche Anglizismus *Handy* erwähnt, der gar nicht aus dem Englischen stamme, weil es dort *mobile* oder *cellphone* heiße. Tatsächlich scheint es so zu sein, dass der Begriff aus einem bereits in den 1940er Jahren in den USA gebräuchlichen Begriff abgeleitet wurde. Offenbar handelt es sich beim sogenannten *Handie talkie* um ein „handliches“, also tragbares Funkgerät, das im Zweiten Weltkrieg eingesetzt wurde. Es liegt die Vermutung nahe, dass die Werbefachleute irgendeines Mobilfunkunternehmens sich an dieses Lexem erinnern haben, als sie das Mobiltelefon *Handy* taufen. Aus der Perspektive des Deutschen könnte man also zwar von einer Wortschöpfung sprechen, die jedoch an ein englisches Lexem angelehnt ist.

Grundsätzlich kann man nach Wortschöpfungen in der Werbesprache suchen, wo viele Produktnamen reine Phantasiekreationen sind, was allerdings nicht heißt, dass bei der Benennung von Produkten alles dem Zufall

Arten der Wortbildung

Arten der Wortbildung	
Ausdruckserweiterung	
Komposition	Determinativkompositum: <i>Haustür, Rotwein, hellbraun</i> ; Spezialfall: Rektionskompositum: <i>Deutsch-Lehrer</i> Kopulativkompositum: <i>Radlowecker, süß-sauer</i> exozentrisches Kompositum: <i>Zwölfzylinder, Nachmittag</i>
Derivation	explizite Derivation: Präfixderivation: <i>verschreiben, zerreden, unschön, Unmenge</i> Suffixderivation: <i>Sportler, Frechheit, Umleitung, absichtlich, radeln</i> kombinatorische Derivation/Zirkumfixderivation: <i>beabsichtigen, Gerede</i> Zusammenbildung: <i>Geldgeber, blauäugig</i> implizite Derivation: <i>Wurf</i> Rückbildung: <i>Sanftmut; Häme, Klinisi</i>
Ausdruckskürzung	
unisegmentales Kurzwort	Anfangswort: <i>Uni, Auto, Lok</i> Endwort: <i>Rad, Bahn</i> Rumpfwort: <i>Lisa</i> Klammerform: <i>Bierdeckel, Kirschblüte</i>
multisegmentales Kurzwort/ Akronym	Initialabkürzungswort: <i>PKW, URL; UNO, TÜV</i> Silbenkurzwort: <i>Kripo, Hanuta</i> Mischformen: <i>Azubi, BAföG</i> Partielles Kurzwort: <i>U-Bahn</i>
Kontamination	Wortkreuzung: <i>Demokratur</i> Wortverschmelzung: <i>verschlimmbessern</i> Kontraktion: <i>am, beim, zur</i>
Konversion	
syntaktische Konversion	<i>das Treffen, das Fenster ist gestrichen, das Gegenüber, im Heute, ohne Wenn und Aber, das So-tun-als-ob, Du Guter!</i>
morphologische Konversion	<i>der Treff, grünen, computern</i>
Sonstige Arten der Wortschatzerweiterung	
Entlehnung	<i>Computer, downloaden, Schokolade</i>
hybrides Kompositum/ Derivationen	<i>Computerfachmann, Bioladen, URL-Adresse, einloggen</i>
Lehnübersetzung	<i>Maus, herunterladen, Datenautobahn</i>
Bedeutungsveränderung/ Bedeutungswandel	Bedeutungsverengung: <i>Hochzeit, Schirm</i> Bedeutungserweiterung: <i>Horn, packen</i> Bedeutungsverschiebung/ -übertragung: <i>Zweck, Rücken</i> Bedeutungsaufwertung: <i>toll</i> Bedeutungsabwertung: <i>gemein</i>
Wortschöpfung	<i>sitt</i> , Kinder- und Werbesprache

Tab. 6.1
Die Wortbildungs-
typen im Überblick

überlassen wird. Beispielsweise sind Autofirmen sehr darauf bedacht, dass ihre Markennamen nicht in irgendeiner Sprache etwas Unanständiges bedeuten wie im berühmten Beispiel des Mitsubishi Pajero, der in spanischsprachigen Ländern Montero heißt – Pajero bedeutet dort in der Vulgärsprache „Wichser“. Ein anderes Beispiel für den Versuch einer Wortschöpfung sind die Bemühungen, ein deutsches Wort für den Zustand zu kreieren, dass man genug getrunken hat, mit anderen Worten: Wie heißt das Gegenteil von *durstig*? Einer der Vorschläge lautete *sitt*, ein anderer *schmöll* (vgl. Kapitel 12.2.2). Tatsächliche Wortschöpfungen finden Sie heutzutage in gehäufte Form aber nur in der Kindersprache. So hat einer der Autoren dieses Buches im Alter von etwa vier Jahren, als er zum ersten Mal Kaninchen sah, diese auf Nachfrage seiner Eltern zu deren großen Erstaunen als *Jüjois* bezeichnet – offensichtlich war er zu stolz zuzugeben, dass er weder diese Tiere noch ihren Namen kannte. Bei anderen, oft onomatopoetischen Bildungen wie *Wauwau* ist fraglich, ob man diese wirklich als Wortschöpfungen von Kindern betrachten kann. Abgesehen von solchen kindersprachlichen Ad-hoc-Bildungen und der Werbung spielen Wortschöpfungen in der Gegenwartssprache aber praktisch keine Rolle.

6.6 | Flexion

Flexion Während die Wortbildung der Bildung von Lexemen dient, die einer Sprache zur Verfügung stehen, werden mit Hilfe der FLEXION die Lexeme in Wortformen umgewandelt. Flexion ist also dazu da, aus abstrakten lexikalischen Wörtern Wörter in einer konkreten, syntaktischen Verwendung zu machen, d. h. sie mit zusätzlichen grammatischen Merkmalen zu versehen, welche es ermöglichen, dass sich die einzelnen Wörter aufeinander und auf außersprachliche Sachverhalte beziehen. So sagt das Lexem *Baum* an sich noch nichts darüber aus, ob es sich um einen oder mehrere Bäume handelt, ob ein Geschehen aus der Perspektive des Baums beschrieben wird (Im Satz *Der Baum wächst* ist *Baum* Nominativ und damit Subjekt) oder sich das Geschehen auf den Baum richtet (Im Satz *Er fällt den Baum* ist *Baum* Akkusativ und damit Objekt). Die Schwierigkeit, Wortbildung und Flexion voneinander abzugrenzen, wird deutlich am Beispiel der Komparation (*groß* → *größer* → *größt-*) und der Bildung des Partizips I (*schreib(en)* → *schreibend*). Obwohl beide Prozesse viele Gemeinsamkeiten aufweisen – z. B. verlaufen sie beide sehr regelhaft und in beiden Fällen kann an das Komparations- bzw. Partizipsuffix noch ein weiteres (Flexions-)Suffix angehängt werden (*größer-e*, *schreibend-e*) – wird die Komparation eher der Flexion, die Partizipbildung hingegen der Wortbildung, genauer gesagt: der Adjektivderivation, zugeordnet. Zu erklären, warum das so ist, würde an dieser Stelle zu weit führen; die im Anhang zu dieser Einheit angeführte Literatur hilft hier weiter. Stattdessen seien nur zwei grundlegende Unterschiede zwischen Flexion und Wortbildung genannt: Bei der Flexion

eines Lexems ändert sich erstens niemals dessen Wortart. Und zweitens wird ein Lexem mit Hilfe von Flexion in eine auch morphologisch sichtbare Beziehung zu anderen Lexemen gesetzt. Wenn Sie also den Altersunterschied zu Ihren Eltern verbalisieren wollen, können Sie das entweder lexikalisch machen, indem Sie sagen: *Das Alter meiner Eltern liegt um soundso viele Jahre über meinem Alter*. Oder Sie verwenden einfach ein Komparationssuffix (inkl. Umlaut): *Meine Eltern sind soundso viele Jahre älter als ich*.

Flexion: Prozess, bei dem aus abstrakten lexikalischen Wörtern konkrete, syntaktische Wörter entstehen, welche sich aufeinander und auf außersprachliche Sachverhalte beziehen.

Definition

Rechnet man also die Komparation mit zur Flexion, kommt man zu folgenden Flexionsarten: die NOMINALFLEXION oder DEKLINATION (Substantive werden nach Numerus und Kasus, Adjektive, Pronomen und Artikel zusätzlich in verschiedenen Genera flektiert), die VERBFLEXION oder KONJUGATION (Verben werden nach Person, Numerus, Tempus und Modus flektiert) und eben die KOMPARATION der Adjektive. Wenn Sie an dieser Stelle die Passivbildung vermissen, liegt das daran, dass im Deutschen die Passivformen (*behandelt werden*), ähnlich wie z. B. das Perfekt (*behandelt haben*), das Futur (*behandeln werden*) und bestimmte Formen des Konjunktivs (*behandeln würden*, *behandelt hätten*), nicht allein mit Hilfe von Flexionsmorphemen gebildet werden, sondern durch sogenannte ANALYTISCHE (griech.: *analyō* = auflösen) oder PERIPHRASTISCHE (griech.: *periphrássō* = ringsum einschließen) FORMEN. Mit anderen Worten: das Verb *behandeln* wird im Passiv aufgelöst in das Hilfsverb *werden* und das Partizip *behandelt*, wobei die beiden Teile der Verbform in der Lage sind, Teile eines Satzes zu umschließen: *Ich werde im Krankenhaus vom Chefarzt behandelt*. Diese analytischen bzw. periphrastischen Formen gehören nicht zur Flexionsmorphologie im engeren Sinne; dazu zählen nur die SYNTHETISCHEN (griech.: *synthesis* = Zusammensetzung) FORMEN wie z. B. *er behandel-t*.

Nominalflexion/
Deklination

Verbflexion/
Konjugation

Komparation

Analytische/
periphrastische
Formen

Synthetische Formen

Deutsche Flexion –
Schwere Flexion

Synkretismus

Das deutsche Flexionssystem ist – jeder, der Deutsch als Fremdsprache lernt, wird Ihnen das gern bestätigen – zu kompliziert, als dass es in diesem Rahmen erschöpfend dargestellt werden könnte. Deswegen soll hier nur auf einige Besonderheiten hingewiesen werden. Zunächst einmal fällt bei der Betrachtung der Verbformen im Indikativ Präteritum auf, dass man mit vier Formen auskommen muss, obwohl es sechs Positionen gibt (vgl. Tab. 6.2).

Das Phänomen, dass sich die 1. und 3. Person hier weder im Singular noch im Plural unterscheiden, nennt man SYNKRETISMUS (griech.: *syn* = zusammen, *krētismós* = Lug und Trug (nach Art der Kreter)). Es taucht auch in der Nominalflexion auf; z. B. lauten die Pluralformen des Lexems *Umleitung* in allen Kasus gleich, nämlich *Umleitungen*. Mit der Erklärung dieses Phänomens

beschäftigt sich die Linguistik gerade in jüngerer Zeit, ohne bisher in allen Fällen zu einem voll überzeugenden Ergebnis zu kommen.

Tab. 6.2 | Das Verb rufen im Indikativ Präteritum

Verbstamm	Person	Numerus	
		Singular	Plural
rief	1.	{ø}	{-en}
	2.	{-st}	{-t}
	3.	{ø}	{-en}

Starke und schwache Flexion

Ein weiteres erwähnenswertes Phänomen ist die Unterscheidung zwischen STARKER und SCHWACHER FLEXION. Diese findet man im Bereich der Nominalflexion. Während z. B. Substantive entweder stark, gemischt oder schwach dekliniert werden, je nachdem, zu welchem Typ sie gehören (vgl. Tabelle 6.3), hängt bei Adjektiven die Flexion davon ab, ob die Nominalphrase (also eine Wortgruppe, die als Kern ein Substantiv oder Pronomen enthält), in der sie stehen, mit einem bestimmten oder unbestimmten Artikel gebildet wird (vgl. Tabelle 6.4).

Tab. 6.3 | Starke und schwache Deklination am Beispiel verschiedener maskuliner Substantivtypen

Substantivstamm Typ 1 (starke Maskulina/Neutra)	Kasus	Numerus	
		Singular	Plural
Berg	Nom.	{ø}	{-e}
	Gen.	{-(e)s}	{-e}
	Dat.	{{-e}}	{-en}
	Akk.	{ø}	{-e}
Substantivstamm Typ 2 (schwache Maskulina)	Kasus	Numerus	
		Singular	Plural
Mensch	Nom.	{ø}	{-en}
	Gen.	{-en}	{-en}
	Dat.	{{-en}}	{-en}
	Akk.	{{-en}}	{-en}
Substantivstamm Typ 3 (gemischte Maskulina/Neutra)	Kasus	Numerus	
		Singular	Plural
Staat	Nom.	{ø}	{-en}
	Gen.	{-(e)s}	{-en}
	Dat.	{{-e}}	{-en}
	Akk.	{ø}	{-en}

Adjektivflexion (stark)	Kasus	Numerus	
		Singular	Plural
grober Unfug/ große Menschen	Nom.	<i>grob-er Unfug</i>	<i>groß-e Menschen</i>
	Gen.	<i>grob-en Unfugs</i>	<i>groß-er Menschen</i>
	Dat.	<i>grob-em Unfug</i>	<i>groß-en Menschen</i>
	Akk.	<i>grob-en Unfug</i>	<i>groß-e Menschen</i>
Adjektivflexion (gemischt)	Kasus	Numerus	
		Singular	Plural
ein grober Unfug/ keine großen Menschen	Nom.	<i>ein grob-er Unfug</i>	<i>keine groß-en Menschen</i>
	Gen.	<i>eines grob-en Unfugs</i>	<i>keiner groß-en Menschen</i>
	Dat.	<i>einem grob-en Unfug</i>	<i>keinen groß-en Menschen</i>
	Akk.	<i>einen grob-en Unfug</i>	<i>keine groß-en Menschen</i>
Adjektivflexion (schwach)	Kasus	Numerus	
		Singular	Plural
der grobe Unfug/ die großen Menschen	Nom.	<i>der grob-e Unfug</i>	<i>die groß-en Menschen</i>
	Gen.	<i>des grob-en Unfugs</i>	<i>der groß-en Menschen</i>
	Dat.	<i>dem grob-en Unfug</i>	<i>den groß-en Menschen</i>
	Akk.	<i>den grob-en Unfug</i>	<i>die groß-en Menschen</i>

Tab. 6.4

Starke und schwache Deklination des Adjektivs im Maskulinum in Abhängigkeit vom Artikel

22

Noch weitaus komplizierter ist die Situation bei den Pronomina und Artikeln; hier empfiehlt sich ein Blick in eine einschlägige Grammatik.

Abgesehen von ihrer Funktion, aus lexikalischen Wörtern syntaktische zu machen, werden die unterschiedlichen Flexionseigenschaften von Wörtern in der Linguistik als eines von mehreren Kriterien zur Unterscheidung von Wortarten herangezogen. Die Klassifikation von Wortarten wird häufig im Rahmen der Morphologie behandelt; da hierbei jedoch auch syntaktische Kriterien eine Rolle spielen, wird darauf in der folgenden Einheit eingegangen.

Übungen

16.7

1 Um welche Wortbildungsarten handelt es sich bei den folgenden Beispielen?

- Geschirrspüler, Geschirrspülmaschine, Leasing, verfassungswidrig, Privatisierung, Brunch, rahmen, HSV, (das) Schreiben, (das) Hoch, filtern, unhöflich, Gerede, ver-rechnen.

- 2 Beschreiben Sie ausgehend von der zugrunde liegenden Wortbildung den Bedeutungswandel bei den Adjektiven *dämlich* und *herrlich*.
- 3 Mit welchen Allomorphen wird der Gen. Sg. bei deutschen Substantiven im Maskulinum realisiert?
- 4 Diskutieren Sie, ob es sich bei der Bildung des Partizips II im Deutschen um Wortbildung oder um Flexion handelt.

6.8 | **Verwendete und weiterführende Literatur**

Best, Karl-Heinz (2005): Linguistik in Kürze. Mit einem Ausblick auf die Quantitative Linguistik. 3., überarb. Aufl. Göttingen: Skript.

Donalies, Elke (2005): Die Wortbildung des Deutschen. 2., überarb. Aufl. Tübingen: Narr.

Duden (2005) – Die Grammatik. 7., völlig neu bearb. u. erw. Aufl. Hrsg. v. d. Dudenredaktionen. Mannheim et al.: Dudenverlag.

Duden (2000) – Das große Wörterbuch der deutschen Sprache. Hrsg. v. Wissenschaftlichen Rat der Dudenredaktion. CD-ROM-Ausgabe auf Basis der 3., völlig neu bearb. u. erw. Aufl. der Buchausgabe in 10 Bänden von 1999. Mannheim: Bibliographisches Institut.

Elsenberg, Peter (2004): Grundriß der deutschen Grammatik. Bd. 1. Das Wort. 2., überarb. u. aktual. Aufl. Stuttgart, Weimar: Metzler.

Fleischer, Wolfgang; Barz, Irmhild (1995): Wortbildung der deutschen Gegenwartssprache. 2., durchges. u. erg. Aufl. Tübingen: Niemeyer.

Habermann, Mechthild; Müller, Peter O.; Munske, Horst Haider (Hrsg.) (2002): Historische Wortbildung des Deutschen. Tübingen: Niemeyer.

Kluge (2002) – Etymologisches Wörterbuch der deutschen Sprache. Bearb. v. Elmar Seebold. 24., durchges. und erw. Aufl. Berlin, New York: de Gruyter.

Lexikon der Sprachwissenschaft (2002). Hrsg. v. Hadumod Bußmann. 3., aktual. u. erw. Aufl. Stuttgart: Kröner.

Lohde, Michael (2006): Wortbildung des modernen Deutschen. Ein Lehr- und Übungsbuch. Tübingen: Narr.

Metzler-Lexikon Sprache (2004). Hrsg. v. Helmut Glück. CD-ROM-Ausgabe. Berlin: Directmedia Publishing.

Motsch, Wolfgang (1999): Deutsche Wortbildung in Grundzügen. Berlin, New York: de Gruyter.

Muthmann, Gustav (2001): Rückläufiges deutsches Wörterbuch: Handbuch der Wortausgänge im Deutschen mit Beachtung der Wort- und Lautstruktur. 3., überarb. u. erw. Aufl. Tübingen: Niemeyer.

Olschansky, Heike (1999): Täuschende Wörter. Kleines Lexikon der Volksetymologien. Stuttgart: Reclam.

Traditionelle Syntaxanalyse

Inhalt

7.1	Traditionelle Syntaxanalyse – Was ist ein Satz?	116
7.2	Wie analysiert man einen Satz?	117
7.3	Syntaktische Tests	119
7.4	Syntaktische Kategorien und Funktionen	120
7.4.1	Wortarten als syntaktische Kategorien	120
7.4.2	Phrasen als syntaktische Kategorien	124
7.4.3	Syntaktische Funktionen	124
7.5	Ein Modell zur operationalen Satzanalyse	127
7.6	Übungen	131
7.7	Verwendete und weiterführende Literatur	133

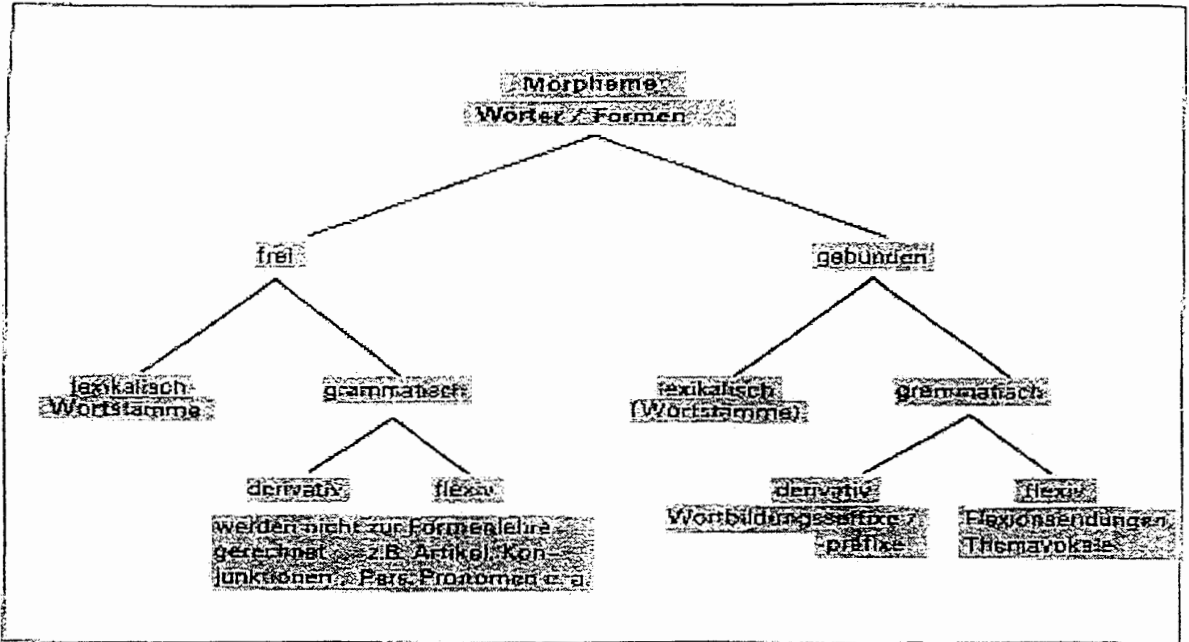
Morphologie I

Lehrmaterial

- Bussmann, H. (1990): *Lexicon der Sprachwissenschaft*, Stuttgart: Alfred Kroener.
- König, W. (1985): "Zur allgemeinen Einführung: Das Wort (Morphologie)" in: *dtv Atlas zur deutschen Sprache*, 6te Auflage, München: Deutscher Taschenbuch Verlag.
- Trommer, J. (2004): "Morphologie", in: *Computerlinguistik und Sprachtechnologie, Eine Einführung*, Carstensen, K.U., Ebert, C., Endriss, C., Jekat, S., Klabunde, R., Langer, H. (Hrsg.), 2te überarbeitete und erweiterte Auflage, München: Spektrum Akademischer Verlag.

**Einführung in die Morphologie:
Definitionen und Erklärungen aus
König, W. (1985): "Zur allgemeinen Einführung: Das
Wort (Morphologie)" in: *dtv Atlas zur deutschen
Sprache*, 6te Auflage, München: Deutscher Taschenbuch
Verlag.**

14 Zur allgemeinen Einführung



Herkömmliche und strukturell-taxonomische Terminologie in der Morphologie

	...	darf	...		arbeiten		↑	
		soll			laufen		↑	
Waltraud	muss	früh			aufstehen		↑	
Horst	kann	schnell			essen		↑	
Liss		spät					↑	
...							↑	
← syntagmatische Beziehungen →							↑	paradigmatische Beziehungen

Zwischen sprachlichen Einheiten gibt es grundsätzlich zwei Arten von Beziehungen:

Syntagmatisch nennt man alle, die linear innerhalb einer sprachlichen Äußerung zwischen den Elementen feststellbar sind, z. B. Beziehungen der Satzteile untereinander, der Morphemgruppen in Satzteilen, der Phoneme in Morphemen (Distribution).

Paradigmatisch nennt man alle, die diese Elemente zu Elementen anderer Äußerungen haben. Paradigmatische Beziehungen bestimmen das Verhältnis der Elemente eines Sprachsystems untereinander. Sie machen eine Klassifizierung von Satztypen, Wortarten, Phonemen möglich.

Syntagmatische und paradigmatische Beziehungen

Sprache verwirklicht sich in Texten, Texte konstituieren sich aus Sätzen, Sätze aus Wörtern. Mit den Wörtern beschäftigt sich die Morphologie.

So verschiedene Auffassungen es über die Definition des Satzes gibt, so umstritten ist »Wort« als linguistischer Begriff. Wort ist eine Einheit, die nur in ihrem jeweiligen Gebrauchszusammenhang definiert werden kann, z. B. als phonologische oder graphematische Einheit. Das phonologische Wort *dt. le:ren* kann zwei graphematische Wörter darstellen: *le:ren* und *leh:ren*. Ein phonolog. oder graphemat. Wort kann aber auch verschiedene grammat. Wörter repräsentieren: *band* z. B. repräsentiert die Vergangenheitsform von *binden* sowie das Nomen *Band*. Ein lexikalisches Wort (*Lexem*) wiederum schließt (nach den Bedürfnissen des Wörterbuches) alle grammat. Formen und die gängigen Ableitungen mit ein. So erscheint in einem Wörterbuchstichwort zwar *Haus*, aber kaum der Genitiv (*des Hauses*) oder das Diminutiv (*Häuschen*).

Ähnlich problematisch ist es mit dem Begriff der Silbe. Primär ist die Silbe eine Sprechheit, bei der Worttrennung am Zeilenende ist sie aber auch eine graphische Einheit, die nicht mit dem grammat. Bau eines Wortes übereinstimmen muß. *Loben* z. B. wird nach grammat. Kriterien in *lob-en* getrennt, von der Silbenstruktur her in *lo-ben*.

Wörter erhält man bei der Zerlegung von Sätzen nach den oben (S. 12f.) beschriebenen Segmentierungsverfahren. Diese dienen auch dazu, Wörter in kleinere Bestandteile aufzuspalten. Die Sätze *Der Arbeiter arbeitet* und *Der Besitzer besitzt* lassen sich damit in folgende Segmente zerlegen: *der, arbeit, be, sitz, er, er, t. Besitzer* z. B. wird auf Grund folgender Ersatzproben in drei Einheiten zerlegt:

be: *be-nachrichtigung*
be-schlag
sitz: *vor-sitz*
sitz-möbel
er: *arbeit-er*
fisch-er

Auf diese Weise kommt man zu einzelnen kleineren Einheiten, den Morphemen, die, in Klassen zusammengefaßt, die Morpheme einer Sprache bilden. Man definiert Morphem als die kleinste bedeutungstragende Einheit eines Sprachsystems. Das Morphem *-er* drückt bei den obigen Beispielen jeweils aus, daß es hier um »handelnde« Personen geht. Bei *heb-t* wird durch das *-t* ausgedrückt, daß nur ein »er«, »sie« oder »es« hier heben kann. Der erste Teil des Wortes, *heb-*, trägt eine konkrete Bedeutung, die in der herkömmlichen Sprachbeschreibung im Wörterbuch aufzufinden ist. Morpheme dieser Art nennt man deshalb *lexikalische* (wortfähige) Morpheme (Grundmorpheme). Sie heißen in der traditionellen Grammatik: »Wurzeln« oder »Stämme«.

Ein Morphem, das selbständig, ohne Bindung an ein anderes, als Wort im Satz vorkommen

kann, nennt man *freies Morphem*: *Buch, Art, schön*. Im Gegensatz dazu stehen *gebundene Morpheme*. Sie sind meist *grammat. Morpheme* und treten als *Flexions- und Derivationsmorpheme* auf.

Durch die Kombination des Grundmorphems *geh-* mit Flexionsmorphemen werden die grammat. Formen dieses Verbs gebildet:

ich geh-e, du geh-st, er geh-t, wir geh-en usw.

Entsteht durch Hinzufügung eines gebundenen Morphems an ein freies Morphem ein anderes neues Wort, so handelt es sich um ein *Wortbildungsmorphem* (Derivations- oder Formationsmorphem): *geh-en* (mit Infinitiv anzeigende Endung *-en*), *Geh-er* (die Endung *-er* macht das Verb bzw. den Verbalstamm zum Nomen agentis).

Unfale Morpheme sind Morpheme, die in der Regel nur in einer einzigen Kombination vorkommen. Die in *Brombeere* und *Himbeere* vorkommenden Morpheme *Brom* und *Him* haben selbständig keine Bedeutung, sie sind weder wortfähig noch grammat. doch der Vergleich mit *Stachelbeere, Vogelbeere* zeigt, daß diese Segmentierung durchaus sinnvoll ist. Es handelt sich hier um den Typ des gebundenen lexikalischen Morphems.

Flexionsprozesse verlaufen meist regelmäßig. Man kann von (fast) jedem Verbum bei Beachtung gewisser Regeln die 3. Pers. Sg. bilden: *er geht, er steht, er liebt*, oder von Substantiven die Nom. Pl.: *der Tisch/die Tische, das Kind/die Kinder*. Derivationsprozesse erfordern sehr viele mehr Regeln: es gibt zu den Verben *gehen, stehen* und *lieben* zwar einen *Geker* und *Stehe* (beim Sport), aber keinen *Lieber*, dafür aber einen *Liebenden* oder einen *Liebhaber* (mit verschiedener Bedeutung). Es gibt Wortbildungsmorpheme, die noch produktiv sind, d. h. unter Beachtung gewisser Regeln verwendet werden können, wie *-bar* (*machbar, essbar, brauchbar, definierbar*); andere sind zwar noch als ehemalige produktive Affixe zu erkennen; sie sind aber nur noch in gewissen Wörtern vorhanden und unproduktiv, z. B. *-de* in *Zierde, Gemälde, B. schwerde* zu den Verben *zieren, malen, b. schweren*.

Bei Flexions- und Derivationsprozessen handelt es sich immer um die Verbindung von freien (oder wortfähigen) Morphemen mit gebundenen Morphemen. Treten nun aber zwei wortfähige Morpheme aneinander, so handelt es sich um eine *Zusammensetzung* (Kompositum): *Werbefunk, Frühlingsfrische, Blödsinn*. Genus und Wortart richten sich im Deutschen nach dem zweiten Glied: *haushoch* ist Adjektiv, *Hochhaus* Substantiv. Bei *Determinativkomposita* wird ein Teil der Zusammensetzung durch den anderen näher bestimmt (*Kinokasse, Fünfkampf, Automotor*). Bei *Kopulativkomposita* sind die beiden Konstituenten einander gleich geordnet, sie sind – im Gegensatz zu den *Determinativkomposita* – im Prinzip unstellbar (*Jackenkleid, Strichpunkt, Pringregen*).

**Einführung in die Morphologie:
Definitionen und Erklärungen aus
Bußmann (1990), *Lexikon der Sprachwissenschaft***

**Einführung in die Morphologie:
Definitionen und Erklärungen aus**

**Trommer, J. (2004): "Morphologie", in:
*Computerlinguistik und Sprachtechnologie, Eine
Einführung*, Carstensen, K.U., Ebert, C., Endriss, C.,
Jekat, S., Klabunde, R., Langer, H. (Hrsg.), 2te
überarbeitete und erweiterte Auflage, München:
Spektrum Akademischer Verlag.**

3.2 Morphologie

Jochen Trommer

Wörter erscheinen im Deutschen wie in vielen anderen Sprachen in verschiedenen Formen (z.B. *Zwerg: Zwerge, Zwergs, Zwergen*) und dienen gleichzeitig als Basis zur Bildung neuer Wörter (etwa ausgehend von *Zwerg: Zwerglein, zwerghaft, Gartenzwerg*). Die Morphologie (Formenlehre) untersucht die systematischen Beziehungen zwischen Wörtern und Wortformen oder – prozedural ausgedrückt – die Regeln, nach denen Wörter/Wortformen gebildet werden. Die formale Umsetzung solcher Regeln in der Computerlinguistik dient dazu, Vollformenlexika zu ergänzen oder zu ersetzen.

3.2.1 Überblick

Im Zentrum dieses Unterkapitels stehen zwei Formalismen (DATR – eine Wissensrepräsentationssprache für lexikalisches Wissen – und endliche Automaten), anhand derer gezeigt werden soll, welche Probleme bei der Formalisierung morphologischer Phänomene auftreten. Dabei ergibt sich, dass einfache Finite-State-Ansätze auf der Basis endlicher Automaten für viele Phänomene inadäquat sind (Abschnitt 3.2.4). DATR erlaubt, die aufgeworfenen Probleme zu lösen, ist aber für praktische Anwendungen problematisch (Abschnitt 3.2.5). In Abschnitt 3.2.6 werden erweiterte Finite-State-Ansätze vorgestellt, die bestimmte Vorzüge von DATR integrieren. Die einleitenden Abschnitte führen die relevanten Begriffe und Probleme ein (Abschnitt 3.2.2) und stellen grundlegende Modelle aus der generativen Morphologie vor (Abschnitt 3.2.3). Das Unterkapitel schließt mit einigen Bemerkungen zur mathematischen Komplexität von natürlichsprachiger Morphologie (Abschnitt 3.2.7), einer Zusammenfassung (Abschnitt 3.2.8) und weiterführenden Literaturhinweisen (Abschnitt 3.2.9).

3.2.2 Grundbegriffe und -probleme

Die genaue Definition vieler morphologischer Begriffe ist – angefangen mit dem Begriff *Wort* – immer noch Gegenstand lebhafter Diskussionen. Im Folgenden soll mit *Wort* eine abstrakte Einheit bezeichnet werden, die verschiedenen Formen zugrundeliegt und dem Eintrag eines Wörterbuchs (*Lexikon*) entspricht. Ein Synonym ist der technische Begriff *Lexem*. Im Gegensatz dazu sind *Wortformen* die verschiedenen Formen eines *Lexems*. Wortformen in diesem Sinn sind z.B. die *Nominativ-Singular-Form* (*der*) *Mensch* und die *Akkusativ-Singular-Form* (*den*) *Menschen*. Verschiedene Wortformen können phonologisch oder orthographisch zusammenfallen, z.B. die entsprechenden Formen des *Lexems* *Haken*: (*der*) *Haken* und (*den*) *Haken*. Dieses Zusammenfallen nennt man auch *Synkretismus*. Die Menge der Wortformen eines *Lexems* (oder eine Teilmenge davon) heißt *Paradigma*.

Flexion und Wortbildung

In engem Zusammenhang mit der Unterscheidung zwischen Wörtern und Wortformen steht die Aufteilung der Morphologie in Flexion (lat. *Beugung*) und Wortbildung. Letztere lässt sich weiter unterteilen in Komposition (Zusammensetzung, Bildung von neuen Wörtern auf der Basis mehrerer Ausgangswörter, etwa *Garten-Zwerg* oder *Zwergen-Garten* aus *Garten* und *Zwerg*) und Derivation, die Bildung von Wörtern auf der Basis einzelner Ausgangswörter (zu *Zwerg*: *zwergenhaft*, *Zwerglein*). Abbildung 3.4 illustriert die Beziehungen zwischen den bisher eingeführten Begriffen anhand einiger Beispiele.

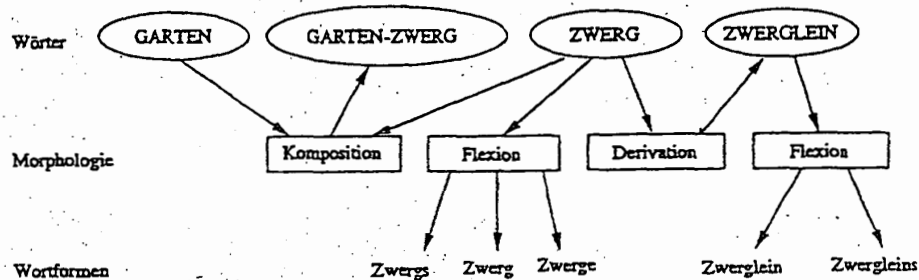


Abbildung 3.4: Flexion und Wortbildung

Morpheme

Die Graphemkette *Gartenzwerg* lässt sich in *Garten* und *Zwerg* zerlegen. Genau wie *Gartenzwerg* haben *Garten* und *Zwerg* eine bestimmte Bedeutung, die auch in der Bedeutung des Kompositums enthalten ist: Ein *Gartenzwerg* ist ein (künstlicher) *Zwerg*, der typischerweise im *Garten* aufgestellt wird. Natürlich kann man auch *Garten* und *Zwerg* weiter zerlegen, z.B. *Zwerg* in *zw* und *erg*. Das sind zwar auch Graphemketten, aber ohne identifizierbare Bedeutung. *Garten* und *Zwerg* sind sogenannte Morpheme: Minimale Phonem/Graphem-Ketten mit einer festgelegten Bedeutung. *Gartenzwerg* fällt nicht unter diese Definition, weil es nicht minimal ist. *zw* und *erg* sind keine Morpheme, weil sie keine festgelegte Bedeutung haben. Eine abstraktere Definition des Morphembegriffs, die auf dieser Grundidee beruht, findet sich im Folgenden.

Klassifikation von Morphemen

Garten und *Zwerg* sind sehr spezielle, sogenannte freie Morpheme. Freie Morpheme können auch ohne den Kontext anderer Morpheme geäußert werden. Morpheme, die diese Eigenschaft nicht haben, heißen gebunden. Z.B. ist das Morphem *-s* in *(des) Zwergs* mit der Bedeutung *Genitiv Singular* gebunden, da es nur nach einem Nomen auftauchen kann.

Eine weitere Einteilung für Morpheme ist die in Grund- und periphere Morpheme. Grundmorpheme stellen den Ausgangspunkt für Derivation und Flexion dar (*Zwerg* in den angeführten Beispielen). Statt von Grundmorphem spricht

man auch von Wurzel. Periphere gebundene Morpheme heißen *Affixe*. *Affixe*, die vor Wurzeln auftreten, heißen *Präfixe*, solche, die nach Wurzeln stehen, *Suffixe*.

Morphemkombinationen ohne Flexionsaffixe werden als *Stämme* bezeichnet. Z.B. ist *kauf* sowohl eine Wurzel als auch ein *Stamm*, *ver-kauf* ist ein *Stamm*, aber keine Wurzel; *ver-kauf-st* und *kauf-st* sind weder *Stämme* noch *Wurzeln*, da sie beide das Flexionsaffix *-st* enthalten.

Allomorphe

Oft scheinen verschiedene Morpheme zusammenzugehören: Sie realisieren dieselbe Bedeutung in verschiedenen morphologischen Kontexten. Z.B. wird Plural in *Kind-er* durch *-er* in *Wind-e* aber durch *-e* ausgedrückt. Diese Beobachtung ist die Grundlage für einen etwas abstrakteren Morphembegriff, bei dem man Ketten mit festgelegter Bedeutung nicht als Morpheme, sondern als *Morphe* bezeichnet und *Morphem* in Abhängigkeit von *Morph* definiert.

Man spricht davon, dass eine Menge von Morphem in komplementärer Verteilung zu einander steht, wenn es keinen Kontext gibt, in dem wahlweise das eine oder das andere Morph auftauchen kann: Das trifft auf *-e* und *-er* zu, da *-er* nicht rechts von *Wind* (**Wind-er*) und *-e* mit der Bedeutung Plural nicht rechts von *Kind* (**Kind-e*) vorkommen kann.

Ein Morphem ist dann eine maximale Menge von bedeutungsgleichen Morphem in komplementärer Verteilung. Die Elemente dieser Menge sind die *Allomorphe* oder *Morphemalternanten* des Morphems. Ein Beispiel ist die Menge aller Pluralmorpheme des Deutschen: $\{-e, \text{er}, -s, \dots\}$, die zusammen das Pluralmorphem bilden, dessen *Allomorphe* wiederum *-e, er, -s, \dots* sind.

Morphophonologie

Oft unterscheiden sich verschiedene *Allomorphe* eines Morphems nur minimal voneinander. Man spricht dann von *partieller Allomorphie*. Ein Beispiel ist das Nomen das in *hand-lich* als *hand*, aber in *Händ-chen* als *Händ* realisiert wird.

Derartige *Allomorphie* lässt sich in vielen Fällen auf reguläre phonologische Prozesse zurückführen, das heißt auf regelmäßige Lautveränderungen, die in bestimmten lautlichen Kontexten einheitlich angewendet werden.

Beispiel 3.2.1

Ein typischer phonologischer Prozess ist die *Auslautverhärtung*, die bestimmte stimmhafte Konsonanten (*b, d, g*) am Ende einer Silbe stimmlos werden lässt (*p, t, k*) (siehe Unterkapitel 3.1). Die *Auslautverhärtung* sorgt dafür, dass z.B. *Hund, Trieb* und *Weg* am Ende mit stimmlosem Konsonanten ausgesprochen werden:

[hundə]	→	[hʌnt]	(Hunde/Hund)
[tri:bə]	→	[tri:p]	(Triebe/Trieb)
[we:gə]	→	[we:k]	(Wege/Weg)

□

Allomorphie dieser Art wird auch als phonologisch konditionierte oder uneigentliche Allomorphie bezeichnet.

Auch der Prozess der Umlautung, welcher aus *Hand* vor *-chen* *Händ* macht, sieht auf den ersten Blick wie ein phonologischer Prozess aus. Hier wird der Stammvokal des Ausgangsmorphems einheitlich nach vorne ($a \rightarrow \bar{a}$, $u \rightarrow \bar{u}$, $o \rightarrow \bar{o}$) verschoben.

<i>Hand</i>	<i>Schloss</i>	<i>Mut</i>
<i>Händchen</i>	<i>Schlösschen</i>	<i>Mütchen</i>

Der Umlaut ist aber zumindest teilweise morphologisch determiniert oder morphophonemisch, da sich kein einheitlicher phonologischer Kontext angeben lässt, in dem Umlaut eintritt. Ein extremes Beispiel für diese morphologische Konditionierung ist die Umlautung bei der Pluralbildung von Wörtern wie *Mutter* (pl.: *Mütter*) und *Vater* (pl.: *Väter*). Die (nicht umgelautete) Ausgangsform für *Väter* ist offensichtlich *Vater*. Wenn der Umlaut eine rein phonologische Regel wäre, müsste diese phonologische Form also immer die Umlautung führen. Das ist aber für die Singular-Form offensichtlich falsch, die ebenfalls *Vater* heißt, aber keine Umlautung auslöst.

Nichtkonkatenative Morphologie

Eine alternative Möglichkeit ist es, den Umlaut als eine Form von nichtkonkatenativer Morphologie zu betrachten. *Nichtkonkatenativ* bedeutet in diesem Zusammenhang, dass eine Kategorie wie Plural nicht durch ein Morphem, sondern durch andere Mittel wie den Umlautungsprozess ausgedrückt wird.

Weitere Formen von nichtkonkatenativer Morphologie sind Prozesse, bei denen Morpheme verkürzt (subtraktive Morphologie) oder teilweise wiederholt werden (Reduplikation). Ein Fall von subtraktiver Morphologie ist die Bildung von *-*Wörtern (z.B. *Katharina* \Rightarrow *Kathi*, siehe Beispiel 3.2.3). Reduplikation taucht etwa bei der Bildung des lateinischen Perfekt auf: *pend-o-ich hänge*; *pe-pend-i-ich hing*.

Eine Form von Lautmodifikation, die noch deutlicher morphologisch bedingt ist als der Umlaut, ist der sogenannte Ablaut, der im Deutschen bei der Flexion vieler unregelmäßiger (starker) Verben relevant ist. Während der Umlaut zwar nicht in phonologisch einheitlichen Kontexten auftritt, aber einen phonologisch einheitlichen Prozess darstellt (er frontiert Vokale), hat der Ablaut u.U. entgegengesetzte Effekte. Z.B. wird im Imperfekt von *schieben* *i* zu *o* (*schob*) während bei *stoßen* *o* zu *i* wird (*stieß*).

Grundprobleme der morphologischen Beschreibung

Von den spezifischen Problemen, die sich bei der Beschreibung und Formalisierung von Morphologie ergeben, sollen im weiteren Verlauf dieses Unterkapitels vor allem die folgenden diskutiert werden:

Neutralisierung Wörter/Wortformen spiegeln ihre Bedeutung oder Funktion oft nur teilweise wieder. Z.B. kann *Haken* sg. oder pl. sein. Man spricht in

Einführung in die Morphologie: Definitionen und Erklärungen aus Bußmann (1990), *Lexikon der Sprachwissenschaft*¹

Morph: [griech. morphé] >Form<. "Kleinstes bedeutungstragendes lautliches Segment einer Äußerung auf der Ebene der Parole, das noch nicht als Repräsentant eines bestimmten Morphems (auf der Ebene der Langue) klassifiziert ist".

"Haben zwei oder mehrere Morphe gleiche Bedeutung bei verschiedener Verteilung, so gehören sie zu demselben Morphem bzw. werden als Allomorphe bezeichnet.

So sind, zum Beispiel /-e/ in "Hunde", /-er/ in "Kinder" und /-n/ in "Opern" drei verschiedene Repräsentanten der deutschen Pluralmorphems".

"Vergleicht man die Morphe -er in "Kinder", "härter" und "er", so handelt es um ein homonymes Morph, d.h., -er ist Allomorph verschiedener Morpheme, nämlich "Plural", "Komparativ", "Personalpronomen".

(Bußmann, 1990)

Morpheme: "kleinste bedeutungstragende Elemente der Sprache, die als phonologisch-semantische Basiselemente nicht mehr in kleinere Elemente zerlegt werden:

zum Beispiel: Buch, drei, es, lang "

"Morpheme sind abstrakte Einheiten, sie werden phonetisch-phonologisch repräsentiert durch Morphe, als kleinste bedeutungstragende aber noch nicht klassifizierte Lautsegmente.

Haben solche Morphe gleiche Bedeutung und komplementäre Verteilung oder stehen sie in freier Variation, so handelt es um Allomorphe desselben Morphems.

Zum Beispiel: repräsentieren im Deutschen die Flexionselemente -e,-en, -n, -s in Hunde, Betten, Katzen, Autos Allomorphe des Morphems "Plural" wobei dies (wie im Himmel) auch durch eine Leerstelle, einen sogenannten Null(allo)morph repräsentiert sein kann".

"Morpheme entsprechen also nur in Sonderfällen der grammatischen Kategorie Wort (zum Beispiel bei "Wort", "wir", "bald") und sind prinzipiell von der phonetischen Einheit Silbe zu unterscheiden:

Silben sind konkrete Lauteinheiten des Wortes auf Parole-Ebene, Morpheme aber Abstrahierungen des Sprachsystems, formale Identität zwischen beiden ist zufällig".

"Je nach dem unterschiedlichen Untersuchungsaspekt ergeben sich verschiedene Typologien der Klassifizierung und Differenzierung von Morphemen:

- (a) nach Form und Bedeutung (diskontinuierlichen Morphemen (viele Morpheme-eine Bedeutung) vs Port-manteau Morphemen (ein Morphem-viele Bedeutungen)

[1] Bussmann, H. (1990): *Lexikon der Sprachwissenschaft*, Stuttgart: Alfred Kroener.

diskontinuierliche Morphemen: ge+lieb+t : Morphem "Partizip" wird durch die getrennten Morphemen ge- und -t gebildet.

Port-manteau Morphem: sprach: ist in den folgenden Morphemen zu analysieren: vom "sprech" als lexikalische Bedeutung / "Tempus", "Person", "Numerus"

(b) nach Bedeutungsfunktion (Lexikalisches Morphem vs Grammatisches Morphem)

(c) nach Selbständigkeit: (freie vs gebundene Morpheme)

Freie Morpheme: zum Beispiel: "Buch", "rot", "schnell" (auch lexikalische Morpheme), "aus", "und" "es" (grammatische Morpheme).

Gebundene Morpheme: Stamm-Morpheme (z.B. Sprach in "Sprachanalyse", Flexionsmorpheme, Ableitungsmorpheme (Affixe), z.B: -zer, -bar, -nis)"

(Bußmann, 1990)

Allomorph: [griech. morphé] >Form<. Auch: Polymorph (ie).

"Konkret realisierte Variante eines Morphems.

Die Klassifizierung von Morphem als Allomorph bzw. als Repräsentanten eines bestimmten Morphems beruht auf

(a) Bedeutungsgleichheit und

(b) Komplementärer Verteilung.

Allomorphe von "geben": sind z.B. sind [ga:p], [ga:b], [gi:p], [ge:p], [geb:] "

"Wird die lautliche Form des Allomorphs durch die phonetische Umgebung bestimmt, so handelt es sich um phonologisch bedingte Allomorphe.

Zum Beispiel im Deutschen die durch Auslautverhärtung bedingten Allomorphe: [ba:t] vs [ba:d] für "Bad", "bad+en". "

"Liegen keine phonetischen Bedingungen für unterschiedliche Allomorphe vor, so spricht man von morphologisch bedingten Allomorphe.

Zum Beispiel: [ʃwim] vs [ʃwam] für "schwimm+en", "schwamm". "

(Bußmann, 1990)

***Auslaut:** "Position von phonologischen oder phonetischen Segmenten im Wurzel-, Stamm- oder Silbenauslaut".

(Bußmann, 1990)

***Auslautverhärtung:** [Auch: (Auslaut)-verschärfung]. "Vorgang und Ergebnis des Stimmtonverlusts von Obstruenten in der Silbenkoda (= Schlußsegment einer Silbe zwischen den Silbenkern und den Anlaut (Kopf) der folgenden Silbe, z.B [kt] in "Jagd".):

Zum Beispiel (Auslautverhärtung):

"lieblich" ['li:pliç] vs. "lieben" ['li:bɛn]

"jagen" ['ja:gɛn] vs. "Jagd" ['ja:kt] "

(Bußmann, 1990)

Morphologie I

Übungen und Aufgaben

1. Morpheme

1. Bitte erklären Sie kurz, anhand eines Beispiels, was ein **Morphem** ist.
2. Was sind **Silben**? Auf welcher Weise unterscheiden sich die Silben von den Morphemen?
3. *Diskussion:* Auf welcher Weise unterscheiden sich die **Morpheme** von den **Phonemen**?
4. Wieviele **Morpheme** können Sie in der folgenden Liste erkennen?

Buch, drei, es, lang
 -e, -en, -n, -s
 Hunde, Betten, Katzen, Autos

2. Klassifizierung und Differenzierung von Morphemen

5. Erwähnen Sie bitte die drei **(3) Typologien der Klassifizierung und Differenzierung von Morphemen**.
6. Bitte erklären Sie kurz, anhand eines Beispiels, was die **diskontinuierliche Morphemen** sind.
7. Bitte erklären Sie kurz, anhand eines Beispiels, was ein **Port-manteau Morphem** ist.
8. Bitte analysieren Sie die folgende Morphemen:

Bär-en
 ge+lieb+t:
 sprach:
 Kind-er

9. Erklären Sie bitte, anhand eines Beispiels, was ein **Lexikalisches Morphem** ist.
10. Was ist ein **Grammatisches Morphem**? Bitte geben Sie dabei drei (3) typische Beispiele.
11. Bitte erklären Sie kurz, anhand eines Beispiels, was ein **freies Morphem** ist (nach Bußmann, 1990, nach König, 1985).
12. Was ist ein **gebundenes Morphem**? Ihre Antwort soll eine Beschreibung mindestens vier (4) Beispiele gebundener Morpheme enthalten.

"Wort" "wir", "bald", Autos

"Buch", "rot", "schnell", "aus", "und" "es"

Sprach in "Sprachanalysen"

"zerlegbar", "Verhältnissen", "Geheimnis"

13. Welche der Morpheme in der obigen Liste sind zugleich auch **lexikalische Morpheme**?
14. Welche dieser Morpheme sind zugleich auch **grammatische Morpheme**?
15. Bitte bestimmen Sie welche dieser Morpheme zugleich auch **Flexionsmorpheme** sind.
16. Welche der Morpheme in der obigen Liste sind zugleich auch **Ableitungsmorpheme (Affixe)**?
17. Bitte bestimmen Sie welche dieser Morpheme zugleich auch **Suffixe** sind.
18. Welche dieser Morpheme sind zugleich auch **Präfixe**?
19. Bitte erklären Sie kurz, anhand eines Beispiels, was ein **Wortbildungsmorphem** ist.
20. Was ist ein **Unikales Morphem**? Bitte geben Sie dabei zumindest einen (1) typischen Beispiel.

3. Wortbildungsprozesse

21. Erwähnen Sie bitte die Unterschiede zwischen die drei **Wortbildungsprozesse**.

22. Welcher Wortbildungsprozess hat in den folgenden Beispielen stattgefunden?

essbar
Reichtum
Geldsegen
verkauftst
Hundes
Busbahnhofhaltestelle

23. Bitte erklären Sie kurz, anhand eines Beispiels, was ein **Flexionsprozess** ist.

24. Bitte erklären Sie kurz, anhand eines Beispiels, was ein **Derivationsprozess** ist.

25. Bitte betrachten Sie die folgende Beispiele: Muss Derivation immer Wortart verändernd sein?

grün-lich
Teller-chen
er-finden
be-halten

26. Bitte füllen Sie die Lücken der folgenden Tabelle mit angemessenen Beispielen:

	ADJEKTIV	NOMEN	VERB
ADJEKTIV	<i>bunt-kariert</i>		<i>fest-machen</i>
NOMEN			<i>not-landen</i>
VERB	<i>schreib-fähig</i>		<i>Mäh-dreschen</i>

27. Bitte bestimmen Sie den Unterschied zwischen die **Determinativkomposita** und die **Kopulativkomposita**. Bitte geben Sie dabei zumindest zwei typische Beispiele.
28. Erwähnen Sie bitte zwei Beispiele eines **Determinativkompositums**.
29. Erwähnen Sie bitte zwei Beispiele eines **Kopulativkompositums**.

4. Allomorphe:

30. *Diskussion:* Bitte erklären Sie kurz, anhand eines Beispiels, was eine **Morphe** ist.
31. Bitte erklären Sie kurz, anhand eines Beispiels, was ein **Allomorph** ist.
32. Auf welchen **Kriterien** beruht die Klassifizierung von **Morphem als Allomorph**?
33. Erklären Sie bitte, anhand eines Beispiels, was **morphologisch bedingten Allomorphe** sind.
34. Bitte bestimmen Sie die Beziehung folgender Morphemen: [ʃwim] vs [ʃwam] für "schwimm+en", "schwamm"?
35. Bitte bestimmen Sie die Beziehung folgender Morphemen: [ga:p], [ga:b], [gi:p], [ge:p], [geb:] ?
36. Bitte erklären Sie kurz, anhand eines Beispiels was **phonologisch bedingte Allomorphe** sind.
37. Bitte bestimmen Sie die Beziehung folgender Morphemen: [ba:t] vs [ba:d] für "Bad", "bad+en"?
38. Was ist ein **Null(allo)morph**? ? Ihre Antwort soll eine Beschreibung des Beispiels "Himmel" enthalten.
39. Was ist ein **homonymes Morph**? Bitte geben Sie dabei zumindest einen (1) typischen Beispiel.
40. Bitte beschreiben Sie die Beziehung folgender Morphemen: "Kinder", "härter" und "er"?

Lehrmaterial

(EXTRA)

Speech and Language Processing

An Introduction to Natural Language Processing,
Computational Linguistics, and Speech Recognition

Second Edition

Daniel Jurafsky
Stanford University

James H. Martin
University of Colorado at Boulder



Upper Saddle River, New Jersey 07458

Jurafsky and Martin, 2000

(25.3) English: *The bottle floated out.*

Spanish: *La botella salió flotando.*
The bottle exited floating.

Verb-framed

Satellite-framed

Verb-framed languages mark the direction of motion on the verb (leaving the satellites to mark the manner of motion), like Spanish *acercarse* 'approach', *alcanzar* 'reach', *entrar* 'enter', *salir* 'exit'. **Satellite-framed** languages mark the direction of motion on the satellite (leaving the verb to mark the manner of motion), like English *crawl out*, *float off*, *jump down*, *run after*. Languages like Japanese, Tamil, and the many languages in the Romance, Semitic, and Mayan languages families, are verb-framed; Chinese as well as non-Romance Indo-European languages like English, Swedish, Russian, Hindi, and Farsi are satellite framed (Talmy, 1991; Slobin, 1996).

Finally, languages vary along a typological dimension related to the things they can omit. Many languages require that we use an explicit pronoun when talking about a referent that is given in the discourse. In other languages, however, we can sometimes omit pronouns altogether, as the following example from Spanish shows, using the \emptyset -notation introduced in Chapter 21:

(25.4) [El jefe]_i dio con un libro. \emptyset _i Mostró a un descifrador ambulante.

[The boss] came upon a book. [He] showed it to a wandering decoder.

Pro-drop

Referential density

Cold language

Hot language

Languages that can omit pronouns are called **pro-drop** languages. Even among the pro-drop languages, there are marked differences in frequencies of omission. Japanese and Chinese, for example, tend to omit far more than does Spanish. We refer to this dimension as **referential density**; languages that tend to use more pronouns are more referentially dense than those that use more zeros. Referentially sparse languages, like Chinese or Japanese, that require the hearer to do more inferential work to recover antecedents are called **cold** languages. Languages that are more explicit and make it easier for the hearer are called **hot** languages. The terms *hot* and *cold* are borrowed from Marshall McLuhan's (1964) distinction between hot media like movies, which fill in many details for the viewer, versus cold media like comics, which require the reader to do more inferential work to fill out the representation (Bickel, 2003).

Typological differences between languages can cause problems for translation. Obviously, translating from SVO languages like English to SOV languages like Japanese requires huge structural reorderings, since all the constituents are at different places in the sentence. Translating from a satellite-framed to a verb-framed language or from a head-marking to a dependent-marking language requires changes to sentence structure and constraints on word choice. Languages with extensive pro-drop, like Chinese or Japanese, cause huge problems for translation into non-pro-drop languages like English since each zero has to be identified and the anaphor recovered.

25.1.2 Other Structural Divergences

Many structural divergences between languages are based on typological differences. Others are simply idiosyncratic differences characteristic of particular languages or language pairs. For example, in English the unmarked order in a noun phrase has ad-

Isolating

Polysynthetic

Agglutinative

Fusion

Morphologically, languages are often characterized along two dimensions of variation. The first is the number of morphemes per word, ranging from **isolating** languages like Vietnamese and Cantonese, in which each word generally has one morpheme, to **polysynthetic** languages like Siberian Yupik (“Eskimo”), in which a single word may have very many morphemes, corresponding to a whole sentence in English. The second dimension is the degree to which morphemes are segmentable, ranging from **agglutinative** languages like Turkish (discussed in Chapter 3), in which morphemes have relatively clean boundaries, to **fusion** languages like Russian, in which a single affix may conflate multiple morphemes, like *-om* in the word *stolom* (table-SG-INSTR-DECL1), which fuses the distinct morphological categories instrumental, singular, and first declension.

SVO

SOV

VSO

Syntactically, languages are saliently different in the basic word order of verbs, subjects, and objects in simple declarative clauses. German, French, English, and Mandarin, for example, are all **SVO (subject-verb-object)** languages, meaning that the verb tends to come between the subject and object. Hindi and Japanese, by contrast, are **SOV** languages, meaning that the verb tends to come at the end of basic clauses, and Irish, Arabic, and Biblical Hebrew are **VSO** languages. Two languages that share their basic word-order type often have other similarities. For example, **SVO** languages generally have **prepositions**, whereas **SOV** languages generally have **postpositions**.

For example, in the following SVO English sentence, the verb *adores* is followed by its argument VP *listening to music*, the verb *listening* is followed by its argument PP *to music*, and the preposition *to* is followed by its argument *music*. By contrast, in the Japanese example that follows, each of these orderings is reversed; both verbs are *preceded* by their arguments, and the postposition follows its argument.

(25.1) English: *He adores listening to music*

Japanese: *kare ha ongaku wo kiku no ga daisuki desu*
 he music to listening adores

Head-marking

Another important dimension of typological variation has to do with **argument structure** and **linking** of predicates with their arguments, such as the difference between **head-marking** and **dependent-marking** languages (Nichols, 1986). Head-marking languages tend to mark the relation between the head and its dependents on the head. Dependent-marking languages tend to mark the relation on the non-head. Head-marking Hungarian, for example, marks the possessive relation with an affix (A) on the head noun (H), and English marks it on the (non-head) possessor:

(25.2) English: *the man-A's^H house*

Hungarian: *az ember^H ház-Aa*
 the man house-his

Typological variation in linking can also relate to how the conceptual properties of an event are mapped onto specific words. Talmy (1985, 1991) noted that languages can be characterized by whether direction of motion and manner of motion are marked on the verb or on the “satellites”: particles, prepositional phrases, or adverbial phrases. For example, a bottle floating out of a cave would be described in English with the direction marked on the particle *out*, while in Spanish the direction would be marked on the verb:

ELEMENTS OF GERMAN PHONOLOGY AND MORPHOLOGY

Elmer H. Antonsen

Elements of German is obviously the result of the long experience of the author, one of the world's greatest experts in the area. Both college teachers of German and students will benefit from this book. It requires no previous knowledge of linguistic theory, introduces the basic concepts of phonology and morphology, and applies them consistently and lucidly to the facts of Modern German.

—Anatoly Liberman,

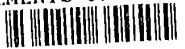
author of *Germanic Accentology*

Exercises are included throughout to help better absorb the rules for real world language use. This volume provides an in-depth look of the German language from the ground up. Its detailed approach makes this book an excellent complement to the world's specific grammar textbooks and reviews.

Elmer H. Antonsen is Professor Emeritus of Linguistics and of Germanic Languages of the University of Illinois at Urbana-Champaign. Former editor of the journal *Studies in the Linguistic Sciences*, he is author or editor of six books, including *Runes and Germanic Linguistics*, *Grimm, Brothers and the Germanic Past*, and *A Concise Grammar of the Old Norse Runic Inscriptions*.

COVER DESIGN: PAUL MOXON

The University of Alabama Press
Tuscaloosa, Alabama 35487-0380
www.uapress.ua.edu/

ISBN-13: 978-0-8173-5450-7
ANTONSEN, ELMER H.
9780817354503 0807
ELEMENTS OF GERMAN

9780817354505
1050--LING

THE
CORNELL STORE

ANTONSEN ELEMENTS OF GERMAN

ELEMENTS OF GERMAN

PHONOLOGY AND MORPHOLOGY

ELMER H. ANTONSEN

The University of Alabama Press
Tuscaloosa, Alabama 35487-0380

Copyright © 2007
Elmer H. Antonsen
All rights reserved
Manufactured in the United States of America

∞
The paper on which this book is printed meets the minimum requirements of
American National Standard for Information Science—Permanence of Paper for
Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Antonsen, Elmer H.

Elements of German: phonology and morphology / Elmer H. Antonsen.
p. cm.

Includes bibliographical references.

ISBN-13: 978-0-8173-1583-2 (alk. paper)

ISBN-10: 0-8173-1583-7 (alk. paper)

ISBN-13: 978-0-8173-5450-3 (pbk. : alk. paper)

ISBN-10: 0-8173-5450-6 (pbk. : alk. paper)

1. German language—Phonology. 2. German language—Morphology.
3. German language—Textbooks for foreign speakers—English. I. title

PF3131.A58 2007
431'5—dc 22

2006051779

Elements of German

Phonology and Morphology

Elmer H. Antonsen

The University of Alabama Press
Tuscaloosa

VERB PHRASE, but do not fit the morphological definition of a verb (i.e., inflection for person + number) are called VERBALS, and in German these are deverbal nouns (infinitives), as in *Rechts fahren!*, and deverbal adjective/adverbs (participles), as in *Stillgestanden!* We therefore have two categories of terms which need to be kept apart, those that refer to SYNTACTIC FUNCTIONS and those that designate MORPHOLOGICAL FORM CLASSES:

Syntactic functions	Morphological form classes:
NOMINALS	NOUNS
ADJECTIVALS	PRONOUNS
ADVERBIALS	DETERMINERS
VERBALS	ADJECTIVE/ADVERBS
	ADVERBS
	VERBS
	PREPOSITIONS
	CONJUNCTIONS

The morphological form classes will be discussed in detail, but it should be kept in mind that members of these classes may fulfill various syntactic functions.

3.1 Morphological analysis

MORPHOLOGY is the scientific study of the formation of words. A **MORPHEME** can be defined as the smallest **DISTINCTIVE UNIT OF MEANING**, in which 'meaning' must be taken broadly to include both lexical meaning (as found in a dictionary) and grammatical meaning. A **NONDISTINCTIVE VARIANT** of a morpheme is an **allomorph**. Morphemes are written in phonemic transcription within braces { }; allomorphs are also given in phonemic transcription, but within slant lines / /.

The term 'word' presents problems of definition, and any attempt to define it must be language-specific (i.e. restricted to a particular language). Compare, for example, the different evaluations of similar items in such closely related languages as German and English: *Straßenkehrer*, but *street sweeper*. A German would undoubtedly call *Straßenkehrer* one word, since it is written as such, whereas an English-speaker would call *street sweeper* two words, since they are written separately; or compare *life insurance company* and *Lebensversicherungsgesellschaft*. The native analysis of these items in each of the languages can be justified from different points of view. It is clear that the English items consist of separate words and therefore are written with intervening spaces. It is equally clear that the German items can be considered to be single words, since the gender of the entire word is determined by the last element and can differ from that of the other individual parts (*die Straße*, but *der Straßenkehrer*). On the other hand, it is also evident that English *street sweeper* is a single concept and in a way a single 'word' since the plural is not **streets sweepers*, but *street sweepers*, and it would be completely justifiable to write *streetsweeper*, *streetsweepers*. Similarly, it is evident that the German items also consist of separate 'words', and we have such anomalies as *auskommen* 'to get along' (considered one 'word'), but *wir kommen gut aus*. In both languages, the unity of such expressions is also indicated by the compound accent assigned to them, i.e., although each root has a root accent, normally only one receives the syntactic accent, e.g., Ger. /°štrásən + kërər↓/, Engl. /°strít + swípər↓/. We therefore have a problem with the term 'word', because it is ambiguous. To avoid this ambiguity, we will not use the term 'word' in our discussion, nor will we refer to arbitrary spelling conventions. Instead of 'word', we will use a linguistic term that we can define with some precision: *stem* (Ger. *der Stamm*). A **STEM** is A

MORPHOLOGICAL ITEM THAT CAN STAND ALONE AND CAN BE ASSIGNED TO A PARTICULAR MORPHOLOGICAL CLASS (e.g. noun, pronoun, determiner, adjective/adverb, adverb, preposition, verb, etc.).

STEM-FORMATION, or DERIVATION. Every stem must contain a **ROOT** (or if foreign, a **BASE**). In German, ALL NATIVE ROOTS ARE MONOSYLLABIC, BEAR THE ROOT ACCENT, AND CARRY THE BASIC LEXICAL MEANING, e.g., {áug-} 'eye' in such stems as *Auge*, *Augapfel*, *Augenblick*, *liebäugeln*, etc. Note that in *das Auge*, the root {áug-} is identifiable by the root accent, but it is not assignable to any morphological class. Only the addition of the stem-forming (derivational) element *-e* makes it possible to assign the stem to the class of nouns, thus: {áug-} + {-ə(n)}ⁿ. Similarly, {áug-} + {-láiŋ}ⁿ = *das Äuglein*, is a noun (indicated by the designation of the fixed gender of the stem), whereas the stem {áug-} + {-iŋ}^{adj/adv} = *äugig* is an adjective/adverb (traditionally called a 'descriptive adjective'), while {áug-} + {-əl}^{wv} = *äugel-* is a weak verb. Since the formation of stems is called derivation, ALL STEM-FORMANTS are DERIVATIONAL AFFIXES. It follows that any time a new stem is formed, whether from a root/base or from another existing stem, we are dealing with a derivation, which is the same thing as saying 'We are forming a new stem' (i.e., a new 'word').

Like the *-e* of *Auge*, that of *Rose*, *Sonne*, *Löwe*, *Name*; the *-en* of *Wagen*, *Garten*, *Kaufen*; the *-er* of *Vater*, *Mutter*, *Maler*, the *-el* of *Nagel*, *Schnitzel*, the *-ung* of *Warnung*, etc., are clearly not a part of the 'root', since they are easily separable from the root, as can be seen in:

Rose, *Rosenknospe*, but *Rös-lein*, *ros-ig*;
Sonne, *Sonnenschein* but *Sonn-tag*, *sonn-ig*, (*er*) *sonn-t*
 (*sich*);

<i>Löwe</i> , <i>Löwenzahn</i>	but <i>Löw-in</i> ;
<i>Name</i> , <i>Namenbuch</i>	but <i>nam-haft</i> , <i>näm-lich</i> ,
<i>Wagen</i>	but <i>Wäg-lein</i> , <i>Wäg-el-chen</i> ;
<i>Garten</i>	but <i>Gärt-lein</i> , <i>Gärt-chen</i> ;
<i>Kaufen</i>	but <i>kauf-t</i> , <i>käuf-lich</i> , <i>Ver-käuf-er</i> ;
<i>Vater</i>	but <i>Vat-i</i> ;
<i>Mutter</i>	but <i>Mutt-i</i> , <i>Mutt-chen</i> ;
<i>Maler</i>	but <i>mal-t</i> , <i>Mal-en</i> ;
<i>Nagel</i>	but <i>Näg-lein</i> ;
<i>Schnitzel</i>	but <i>schnitz-t</i> , <i>Schnitz-en</i> ;
<i>Warnung</i>	but <i>warn-t</i> , <i>Warn-en</i> .

Although it might be suggested that the stem-formant /-əl/ is still present in the guise of a zero-variant in a form like *Näglein*, this is in fact not the case, as may be determined from the presence of open juncture, /näg+láin/, which is clearly detectable through the voiceless pronunciation of /g/ and the absence of any devoicing of /l/. The open juncture is required before the combining root *-lein* (see §3.2), but would not be present after /g/ if /-əl/ were present.

While it is possible to find forms in which the apparent derivational element is never dropped (as in the case of our *Hobel*, *Hummer*, *Muster*, *Kloster*, and many others), nevertheless, an economical analysis of the morphological patterns of German requires that these lexical items also be considered complex forms consisting of a monosyllabic root and a stem-formant. There is a clear, fundamental difference in the formation of these native (or naturalized) complex stems, on the one hand, and compound or foreign stems, on the other.

In the case of FOREIGN STEMS, THE BASE CONTAINS MORE THAN ONE SYLLABLE, each of which bears a 'root' accent and is preceded by open juncture, but none of them is identifiable as a

'root', since none of the individual syllables of the base morpheme can be identified as the carrier of the basic meaning. All of them together function in this way, e.g., {gé+ó+lóg-} in *Geologe*, *Geologie*, *geologisch*; {lá+ó+kó+ón}ᵐ *Laokoon*, {fá+kúl+tá+tív}adj/adv *fakultativ*, {zól+dát-}+{-(-ən)}ᵐ *Soldat*, etc. This absence of a single, identifiable root is a prime indicator of a foreign stem. Each syllable in the base thus enjoys the status of a quasiroot, which explains the diffidence with which German-speakers handle these foreign words, i.e., the care with which they pronounce them (compared to English-speakers, for instance). The syllable of the base that receives the syntactic accent is determined by essentially non-German rules, whereas in native German words and loanwords with their clearly identifiable roots, it is always the root that receives the syntactic accent, as long as contrastive accent is not involved. (A few exceptions to the general rule can be found in words like *lebendig* and *Holunder*.)

Some roots can stand alone and function as stems (see below, s.v. 'simplex'), whereas others cannot, but always occur in conjunction with other morphological elements, either stem-formants or other roots/bases or stems. In German, every root and every stem is preceded by open juncture.

SIMPLEX, COMPLEX, AND COMPOUND STEMS. A stem that consists of a ROOT/BASE ALONE, without any additional morphological elements, is called a **SIMPLEX**, e.g., {tág} is a root (indicated by the root accent), and ALSO a stem {tág}ᵐ, since it can stand alone and can be assigned to the morphological class of nouns on the basis of its fixed gender, all of which is indicated by the superscript *m* for 'masculine'. A foreign BASE that is simultaneously a stem is *das Militär* {mí+lí+tár}ⁿ. The presence of a superscript always indicates that the preceding morpheme, or sequence of morphemes, is a stem, e.g.,

{grún}adj/adv, {káum}adv, which are simultaneously roots and stems (i.e. = 'words').

COMPLEX STEMS contain no more than ONE ROOT/BASE AND AT LEAST ONE STEM-FORMANT, e.g., the root {vág-} of *der Wagen* cannot stand alone and does not belong to any morphological category. By adding the syllable /-ən/ we DERIVE a complex stem consisting of a root + a stem-formant: {vág-}+{-ən}ᵐ = (*der*) *Wagen*, a noun. Similarly, *geologisch* represents a complex stem consisting of a base and a stem-formant: {gé+ó+lóg-}+{-ɪʃ}adj/adv. There may be more than one stem-formant in a complex stem, as in {bə-}+{gláit}wv+{-ər}ᵐ = (*der*) *Begleiter*. (Note that in German, true PREFIXES, traditionally called 'inseparable prefixes', are ALWAYS STEM-FORMANTS.)

COMPOUND STEMS contain MORE THAN ONE ROOT/BASE MORPHEME, e.g., (*der*) *Augapfel* = the root {áug-} + the complex stem {áp-}+{-əl}ᵐ. The compound (*der*) *Augenblick*, contains two roots: {áug-} and {blík}, and is analyzed morphologically as containing two stems, i.e. {áug-}+{-ə(n-)}ⁿ and {blík}ᵐ. The stem-formant {-ə(n-)}ⁿ has two allomorphs (non-distinctive variants): /-ə/ and /-ən/. The former occurs only in the complex stem (*das*) *Auge*; the latter occurs only when this stem is the first element of a compound stem, as in (*der*) *Augenblick*, (*das*) *Augenlid*, *augenscheinlich*. It is therefore understood that the element within parentheses in the morpheme {-ə(n-)}ⁿ occurs only when another root follows in a compound, or before {pl}. This variant is called the CONNECTIVE ALLOMORPH. (See §3.4 for discussion of these stem-formants and their allomorphs.) *Militärübung* is a compound stem consisting of a base + root and a stem-formant, which has a combining allomorph with /-s/: {mí+lí+tár}ⁿ+{úb}wv+{-ung(s-)}ᶠ.

Stems containing the so-called 'heavy suffixes' or 'heavy prefixes' are always compound stems, since these assumed suffixes or prefixes are in fact BOUND ROOTS that always carry a root accent and are always preceded by open juncture: *Gleichheit* = {gláix}adj/adv+{-háit(s-)}^f, *Gärtchen* = {gárt-}+{-xín}ⁿ, *Untier* = {ún-}+{tír}ⁿ, *biegsam* = {bíg}^{sv-2}+{-zám}adj/adv. (See Bound roots, §3.2.)

INFLECTION. In addition to ROOTS/BASES and DERIVATIONAL AFFIXES, there is a third group of morphemes called **inflectional affixes**. These are always ADDED TO PRE-EXISTING STEMS. There can only be ONE SUCH MORPHEME attached to any given stem at a time. They are always SYNTACTICALLY DETERMINED, i.e. their presence is dictated by the particular use of the stem within an utterance. (See below in the sections on the various morphological classes.)

CLITICS. Certain stems can occur in colloquial speech with especially weak stress in an utterance and thereby undergo phonological changes, including that of /e/ to /ə/, as in /der/ > /dər/, and even the loss of syllabicity, in which case the phonological remainder of the morpheme attaches itself to a more strongly-stressed neighboring morpheme and the open juncture between them is removed, as when /ér+hát+és+gə+zúgt↓/ > (becomes) /ér+háts+g^ozúgt↓/, /mft+áinəm+^omés-ər↓/ > /mftnəm+^omésər↓/, or /fn+dém/ > /fm/. Note that /fm/ must be analyzed as {fn}prep+{d-}det+{masc. dat.} (or {neut. dat.}) with the allomorphs: /í/ + Ø + /m/. (Cf. English /dʒ^oífdʒet↑/ = *Did you eat yet?* and contractions like *I'm*, *you're*, *could've*, *shouldn't*, *'tis*, etc.)

3.2 Bound roots

BOUND ROOTS can never occur in a simplex, i.e. they are found only in stems containing an additional morphological element, either a stem-formant, or another root/base, or a stem, e.g., {zón-} of *Sonne*, {hám-} of *Hammer*, {áug-} of *Auge*, and {gé+ó+lóg-} of *Geologie*. We indicate the bound nature of a given morpheme by putting a hyphen at the appropriate place within the braces.

There are a number of bound roots in German that never occur without another root (cf. the *cran-* of English *cranberry*, which functions like the first elements of *blueberry*, *blackberry*, *strawberry*, etc., but does not occur alone). While such elements in German are sometimes called 'heavy suffixes' or 'heavy prefixes', it is clear that these morphemes are actually roots, since they are always preceded by open juncture and always have root accent. Therefore, the stems in which they appear always have compound accent, as in:

Bildnis = [³bíltnìs¹ ↓] = /^obíld+ní:s↓/ = {bíld}ⁿ + {-ní:s}ⁿ

Freiheit = [³frájhàjt^h¹ ↓] =
/^ofrái+háit↓/ = {frái}adj-adv + {-háit(s-)}^f

unehrlich = [³ʔún[?]èxlìç¹ ↓] =
/^oún+ér+lí:x↓/ = {ún-}+{ér-}+{-lí:x}adj-adv

We can therefore refer to such bound roots as **compounding roots**. (The (s-) in feminine nouns is a combining allomorph.)

The most common compounding roots:

-bar	{-bár}adj-adv	-haft	{-háft}adj-adv
-chen	{-xín} ⁿ	-heit	{-háit(s-)} ^f

-keit	{-káit(s-)} ^f	-sal	{-zál} ⁿ or {-zál} ^f
-lein	{-láin} ⁿ	-sam	{-zúm} ^{adj-adv}
-lich	{-líx} ^{adj-adv}	-schaft	{-šáft(s-)} ^f
-ling	{-líng} ^m	-tum	{-túm} ⁿ or {túm} ^m
-nis	{-nfs} ⁿ or {-nfs} ^f	-wärts	{-vérts} ^{ad}
-los	{-lóz} ^{adj-adv}		

Erz-, erz-	{érts-}	Un-, un-	{ún-}
Miß-, miß-	{mfs-}	Ur-, ur-	{úr-}

3.3 Noun morphology

In the noun paradigms given in phonemic transcription in Table 5 under **A. Strong Nouns**, mark off the root/base morpheme from all other elements of each stem. Separate the remainder (if any) of each stem into its constituent parts and determine whether each of these parts is a derivational or an inflectional affix. Remember that an inflectional affix is one that is determined by the syntax of the utterance in which it occurs and will vary within the given paradigm for each stem, whereas a derivational affix is not determined by the syntax.

In these paradigms, we find the roots

{tóg}^m, {zón}^m, {váld}^m, {párk}^m; {néts}ⁿ,
{flós}ⁿ, {kráut}ⁿ, {hémd}ⁿ, {vánd}^f, {fráu}^f,

which are simultaneously simplex noun stems, as indicated here by the superscript gender-marker. They require no further stem-forming elements. All of the other roots require a stem-formant to become nouns:

{hób-} + {-əl}^m
{vág-} + {-ən}^m
{húm-} + {-ər}^m

{šnfts}^{wv} + {-əl}ⁿ
{káuf}^{wv} + {-ən}ⁿ
{klóst-} + {-ər}ⁿ
{áug-} + {-ə(n-)}ⁿ

{róz-} + {-ə(n-)}^f
{várn}^{wv} + {-ung(s-)}^f
{kránk}^{adj/adv} + {-háit(s)}^f

Note that the roots {šnfts}^{wv}, {káuf}^{wv}, and {várn}^{wv} are actually simplex stems also, but they belong to the class of weak verbs, as indicated by the superscript. They can, however, be converted into noun stems by adding the proper stem-formants: {šnfts}^{wv} + {-əl}ⁿ, {káuf}^{wv} + {-ən}ⁿ (the latter stem is called the infinitive), and {várn}^{wv} + {-ung(s-)}^f. Similarly, the simplex stem {kránk}^{adj/adv} is converted into a noun stem through addition of the compounding root {-háit(s)}^f.

For *das Auge* and *die Rose*, the stem-formants are represented as {-ə(n-)}ⁿ and {-ə(n-)}^f, respectively, since the formant *-e* given in the singular paradigm has a variant *-en* (= the CONNECTIVE ALLOMORPH) when the stem occurs before another root/base or the plural morpheme (i.e., in a compound), as:

Augenlid 'eyelid' =
{áug-} + {-ə(n-)}ⁿ + {líd}ⁿ

augenscheinlich = 'apparent(ly)'
{áug-} + {-ə(n-)}ⁿ + {šáin}^m + {-líx}^{adj/adv};

will always also add the stem formants /-tə/ and /(gə-) + -t/ (e.g., *rennen, rannte, gerannt*).

c. **Preterite-present verb:** a verb stem whose present-tense forms take the secondary endings normally reserved for past and subjunctive stems. The present-tense forms look like old-fashioned strong past-tense forms with different vowels in the singular and plural (except for *sollen*), e.g., *ich darf, wir dürfen*, and the actual past-tense forms and past participles are formed like weak verbs with /-tə/ and /gə- + -t/ with a vowel change, e.g., *durfte, gedurft*. These verbs are commonly referred to as 'the modals + *wissen*'.

3.11. Morphological analysis of a text. Give a complete morphological analysis of the first paragraph of 'Der Wolf und die sieben Geislein' in §2.8. Arrange your material in the following columns. Under 'Word' give each item in normal spelling.

WORD	ROOT/BASE	DERIVATION	INFLECTION
------	-----------	------------	------------

If you come across a form IDENTICAL IN ALL RESPECTS to one you have already analyzed, simply write '(see above)'. Be sure to enter the allomorphs in which the derivational or inflectional morphemes occur in the text.

Suggested further reading

- Benware, Wilbur A. 1986. *Phonetics and Phonology of Modern German: An Introduction*. Washington: Georgetown University Press.
- Bithell, J. 1952. *German Pronunciation and Phonology*. London: Methuen.
- Curme, George O. 1952. *A Grammar of the German Language* [1904]. 2d 1952 rev. ed. New York: F. Ungar.
- Dinneen, Francis P. 1967. *An Introduction to General Linguistics*. New York: Holt, Rinehart & Winston.
- Eisenberg, Peter, Karl Heinz Ramers, Heinz Vater (eds.). 1992. *Silbenphonologie des Deutschen*. (Studien zur deutschen Grammatik, 42.) Tübingen: G. Narr.
- Fox, Anthony. 2005. *The Structure of German*. 2d ed. (Oxford Linguistics.) Oxford / New York: Oxford University Press.
- Gleason, H. H. 1961. *Introduction to Descriptive Linguistics*. Rev. ed. 1961 New York: Holt, Rinehart & Winston.
- Heidolph, Karl Erich, Walter Flämig, and Wolfgang Motsch (eds.). 1981. *Grundzüge einer deutschen Grammatik*. Berlin: Akademie-Verlag.
- Kloeke, Wus van Lessen. 1982. *Deutsche Phonologie und Morphologie: Merkmale und Markiertheit*. (Linguistische Arbeiten, 117.) Tübingen: Niemeyer.

Text, Translation, Computational Processing

8

Editors

Annely Rothkegel
John Laffling

Mouton de Gruyter
Berlin · New York

Text Resources and Lexical Knowledge

Selected Papers from the 9th Conference
on Natural Language Processing
KONVENS 2008

Edited by

Angelika Storrer
Alexander Geyken
Alexander Siebert
Kay-Michael Würzner

Mouton de Gruyter
Berlin · New York

4 Conclusions

We have presented the results of the lexical and textual resource we have built aligning the WN glosses in English-Spanish. The lexical resource contains the translation of the English variants. Sometimes the equivalence is not a one-to-one since a language can have more synonyms for a concept than the other; thus, synsets have not necessarily the same number of variants in both languages. As for the glosses, they are annotated with POS and semantic information. They parallel WordNet 3.0 entries by keeping the annotation from this source whenever possible. We have tried to make translations as literal as possible, since, even though the morphosyntactic annotation is easily done, the semantic annotation is an arduous task and it is worthwhile to take advantage of work already completed.

Both resources will be very useful for NLP researchers working with Spanish since currently there is not any completely public resource for this language linked to any version of WordNet and, on the other hand, there are very few corpora for Spanish with annotation at semantic level. Also, it presents the added value of it being aligned to the English corpus and therefore it can contribute information in both languages from a contrastive perspective.

References

- Bentivogli, Luisa, Emanuele Pianta, and Marcello Ranieri (2005). Multisemcor: an english-italian aligned corpus with a shared inventory of senses. In *Proceedings of the Meaning Workshop*, 90, Trento, Italy.
- Castellón, Irene, Ana Fernández, and Gloria Vázquez (2005). Creación de un recurso textual para el aprendizaje del inglés. In *NOVATICA. Revista de la Asociación de técnicos de informática*, 51-54.
- Fellbaum, Christiane (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
- Gelbukh, Alexander, Grigori Sidorov, and José Angel Vera-Félix (2006). Paragraph-level alignment of an english-spanish parallel corpus of fiction texts using bilingual dictionaries. In *Text, Speech and Dialogue*, 61-67, Berlin: Springer.
- Izquierdo, Marlen (2006). Contrastive Analysis and Translation English-Spanish: functions of the English -ing form and its equivalents in Spanish. In *Multilingua*, <http://multilingua.uib.no/marlen.page>.
- McEnery, Tony, Andrew Wilson, Fernando Sánchez-León, and Amalio Nieto-Serano (1997). Multilingual resources for European languages: Contributions of the Crater Project. In *Literary and Linguistic Computing*, 219-226.
- Miller, George, Richard Beckwith, Christiane Fellbaum, David Gross, and Katherine Miller (1990). Introduction to WordNet: An On-line Lexical Database. In *International Journal of Lexicography*, 235-244.
- Ramon, Noelia (2004). Building an English-Spanish Parallel Corpus for Teaching and Research: The ACTRES Project. In *Proceedings of the Sixth Teaching and Language Corpora Conference*.
- Ranieri, Marcello, Emanuele Pianta, and Luisa Bentivogli (2004). Browsing Multilingual Information with the MultiSemCor Web Interface. In *Proceedings of the LREC 2004 Workshop The Amazing Utility of Parallel and Comparable Corpora*, 38-41, Lisbon, Portugal.

An OLIF-based open inflectional resource and yet another morphological system for German

Simon Clematide

58

Abstract. This paper describes the implementation of finite-state based, high precision morphological tools for the generation and analysis of open word classes based on the inflection classes for German of the Open Lexicon Interchange Format (OLIF). Productive compounding and derivations are treated by simple word formation rules. The latter is constrained by selective frequency checks over the web and corpora. Minimal lexicographic requirements (only stem and a numeric inflectional code) allow simple expandability and define a morphological abstraction layer which existing finite state morphological systems do not exhibit. Although a lot of lexical information is freely available for end users over the web, the same is not true for resources which will be used in NLP applications. Therefore, we initiate an open and shared morphological OLIF-based resource where we integrate material from sources which allows for such a term of use.

1 Introduction

The acquisition of morphological resources is commonly viewed as expensive in terms of "expert knowledge and labour" (Demberg 2007). Well, in fact it is expensive if it is done again and again by different academic researchers without sharing their resulting resources, and even more important, without a well-thought and well-agreed standard classification system which covers the needs of common text technology systems. For a highly inflected language as German, lemmatization and generation of inflected word forms is crucial for almost any text technological application. Simple and clear-cut interfaces for the coupling and extension of morphological and lexical resources are vital and should be based on standardized linguistic data categories. The EAGLES specification for German morpho-syntax (cf. EAGLES 1996) provides such a resource. For inflectional classes (in a very broad sense), the OLIF (Open Lexicon Interchange Format)¹ consortium has provided a list of "Recommended Values for OLIF Data Categories" for several languages including German (McCormick et al. 2004). In section 2, we described some more recent systems for German morphology. In section 3, we present our work in implementing a finite-state based morphological framework based on a minimal, but standard-oriented lexicographic interface.

1. See <http://www.olif.net>.

2 Other works

Perera and Witte (2005) have built a self-learning system called *DurmLemmatizer* that induces a German full-form lexicon for nouns by processing raw text corpora². Their linguistic processing is embedded in the GATE framework (cf. Cunningham et al. 2002) and restricted to a standard part-of-speech tagger (*TreeTagger*), a base NP chunker (*JAPE*), and their own case and grammatical number tagger based on Hidden Markov Model. Lemmatization is done by stripping off native German inflection suffixes, therefore plural forms involving umlaut as in "Ärzte (pl); Arzt (sg)" (*doctor*) can't be treated correctly. In these and some other difficult cases, their algorithm inserts alternative possible lemma forms to gain recall (e.g. the possible lemma "*Öfen", "*Öfe", "*Öf" (*oven*)). These alternative forms may be reduced, if a further analysis appears with only one of the previously possible lemma forms. An assessment of the quality of the lemmatization based on this resource is more difficult than it may seem. Firstly, the evaluation results in their paper is based on a rather small lexicon with about 13'000 entries whereas the currently distributed resource contains about 84'000. Secondly, their own evaluation numbers need careful interpretation. They are gained against a subset of 88% of all noun occurrences where the *TreeTagger* was also able to produce a lemma. About 75% of the noun occurrences thereof are lemmatized by their system with a precision of around 95%. However, it's unclear how they treat cases where the lexicon contains alternative lemmas – the current distribution of their lexicon has about 14'000 ambiguous lemmatizations.

The SOAP services from <http://wortschatz.uni-leipzig.de> allow the request for the generation of other word forms from a given one. This service is described as "For a given word form returns all other word forms of the same lemma". The word form "geben" (*to give*) produces the output "gibt gab geben gegeben gebe gaben gäbe gibt's Gibt gab's" which makes obvious that only forms which are covered in the corpus are returned. The word form "lieben" (verb *to love* or adjective *dear*) seems to return adjective forms only: "lieber liebsten lieben liebe lieb liebste liebstes liebes liebster liebstem". Although a verb and an adjective reading is returned by their base form service.

Geyken and Hanneforth (2006) present their German morphological analyzer based purely on finite state methods with weighted transitions³. The architecture of this system basically allows free combinations of the items from their stem (80'000 entries) and affix lexicon. About 1'000 morphotactic constraints (word grammar) restrict the possible combinations according to the language specific rules and limit morphological overanalyses. However, there are still lots of unwanted and irrelevant though possible morphological segmentations which one would like to get rid off.

2. <http://www.ipd.uka.de/~durm/tm/Lemma/>

3. An online demo is available from <http://www.tagh.de>.

With the use of penalty weights associated with morphological boundaries and rare morphemes, an optimality ranking between competing analyses emerges from the analyses itself. Volk (1999) showed in the context of GERTWOL (cf. Koskeniemmi and Haapalainen 1996) that the heuristic "prefer simple analyses" is very effective in determining the intended lemma. Without weighted automata, one has to do this in a separate postprocessing filter.⁴ Still, the weighted automata do not suppress unwanted analyses. The TAGH stem lexicon consists of complex entries because every stem alternation gives rise to a separate entry: E.G. the German verb "werfen" (*to throw*) needs the following lemma-stem pairs "werf:warf", "werf:werf", "werf:wirf", "werf:worf", "werf:wüf" with their corresponding morphological features which determine the distribution of the stems in the inflectional paradigm. But there is also a lot of redundancy in this entries for the information which belongs to the lemma itself. The following two entries for past and past participle illustrate this point.

```
(werf:warf) [VIRREG VType=main PrefVerb=no Latinate=no StDef=yes St23SgInd=no
             StPret=yes StSubjI=no StSubjII=no StPartII=no StImpSg=no
             St23SgIndVowelChange=yes]
(werf:worf) [VIRREG VType=main PrefVerb=no Latinate=no StDef=yes St23SgInd=no
             StPret=no StSubjI=no StSubjII=no StPartII=yes StImpSg=no
             St23SgIndVowelChange=yes]
```

The TAGH system is optimized towards coverage.⁵ For the 100 million word corpus "DWDS-Kerncorpus"⁶ the authors give a coverage of 98.2%. Although no published quantitative evaluation on the correctness of the analyses is available, its effective use in two large scale and public lemmatization applications grants high quality.

Schmid et al. (2004) present a morphological analyser that recognizes derivation and composition. Stems may therefore be basic, derived or compounds. Affixes have the origin classes native, foreign, classical. They select their stems by word class features. An illustrating extract from the SMOR lexicon included in the SFST software distribution is shown below:

```
<Base_Stems>haus<PREF>:<><ge>ha:i<>:elt<V><base><nativ><VVPastStr>
<Base_Stems>haus<PREF>:<><ge>ha:ält<V><base><nativ><VVPres2t>
<Base_Stems>haus<PREF>:<><ge>halt<V><base><nativ><VPPP-en>
<Base_Stems>haus<PREF>:<><ge>halt<V><base><nativ><VVPres1>
<Base_Stems>g:bu:et:<><ADJ><base><nativ><AdjSup>
```

4. Such a post-processing filter has an extreme low memory and processing cycle footprint if it's done using a standard UNIX flex tool as our own reimplementations of the original PERL code shows.

5. However, on the demo web site they mention that rare word form (a threshold of 10 over a corpus of 500 million tokens) are omitted for efficiency reasons.

6. <http://www.dwds.de>

```
<Base_Stems>g:bu:et:s<>:s<ADJ><base><nativ><AdjComp>
<Base_Stems>Roß:s<>:s<>:e<NN><base><nativ><NNeut/Pl>
```

The entries include structural (<PREFIX> "prefix"), morphotactic (<nativ>) and inflectional (<VVPres2t>) information. As in the case of TAGH, each stem alternation (e.g. a: i) is encoded by a separate lexicon entry. This is also true for suppletive gradation as "gut" (good), "besser" (better).

3 Architecture of mOLIFde

Other than the discussed SMOR or TAGH systems, our morphological system has minimal requirements for the lexicographic interface: An atomic stem⁷ and an OLIF inflection code: E.G.

```
haus|halt 387
obig 531
Reichtum 111
```

For our internal lexical grammar, we strictly follow the EAGLES specification for German morpho-syntax (EAGLES 1996) which grants us compliance with STTS (Schiller et al. 1999) and documentation. We use the morpho-syntactic features and values verbatim (e.g. "&pos" "=noun"⁸) and serialize them top down according to the hierarchy presented in the standard. The raw EAGLES format and its corresponding shorter STTS representation look like

```
Reichtütern &pos=noun&type=com&declin=no&num=pl&case=dat&gend=masc&infl=--
Reichtütern NN:Masc.Dat.Pl.*
```

3.1 The struggle with OLIF inflection categories

The recommended OLIF data categories for inflection codes contain more than 700 quite fine-graded word classes. For the open inflectional word classes, we find the following numbers: verbs (388), nouns (216), adjectives (34). These classes are more or less directly taken from the LOGOS machine translation system (cf. Scott 2004). To our knowledge, other lexical standardization initiatives (e.g. ISLE/MILE (Ide et al. 2003)) have not produced data category sets comparable to this list. Fig. 1

7. The only exception is a boundary marker after separable verb prefixes that marks also the place for the insertion of "ge" in past participles.

8. For a concise documentation on the syntax of the Xerox regular expression calculus see <http://www.xrce.xerox.com/competencies/content-analysis/fsCompiler/fssyntax.html>.

displays an extract of the noun inflection codes. Roughly said, they define a morphological abstraction layer which also covers some lexical and distributional informations needed for common text technological applications. Although the number of classes may be seen as high, coverage is not perfect.⁹

OLIF systematically shows separate classes for root verbs ("handeln" *to trade*), verbs with inseparable prefix ("behandeln" *to treat*), verbs with separable prefix ("herunterhandeln" *to beat down*), and verbs with a separable and an inseparable prefix ("wiederbehandeln" *to treat again*). The latter are quite uncommon as finite forms, however, adjectival use of past participles built out of them or nominalizations are more frequent. The German dictionary WAHRIG (Wahrig and Wahrig-Burfeind 2006) contains a list of 188 inflection paradigms for strong verbs, which would lead to an upper limit of 752 verb classes.

The high number of noun classes is mostly due to foreign words with foreign or alternate inflection paradigms ("Klima" *climate*, with 3 plural forms in nominative plural as "Klimata", "Klimate", "Klimas") and the fact that every OLIF class has its determined gender even with identical inflection (e.g. "Vater" (*father*) masculine 51, "Kloster" (*convent*) neuter 141). There is also suppletive plural formation (e.g. the plural "Streuzuckersorten" for the uncountable German "Streuzucker" (*castor sugar*)) which may be practical for machine translation systems, but may seem idiosyncratic otherwise. Additional classes evolve from nouns with singular or plural forms only. Nouns with alternate paradigms get their own OLIF class which may lead to many additional classes when done consequently. Another more lexicographic question arises with nouns with alternate gender (often attributed to regional preferences, e.g. the masculine form "Gehalt" used in Austria in the sense of salary in contrast to the standard German neuter gender). And last but not least, spelling reforms of German have produced additional classes.

The linguistic characterisations of the different OLIF inflection classes are often sparse, as can be seen in Fig. 1. The use of arbitrary numbers as class identifiers may seem odd at first. The use of prototype lemmata in the style of "inflects like" should give a intuitive access to the classes. Still, an explicit explanation about the intended sense of a class would have made our work a lot easier. The example lemma itself may also be a source of confusion. For example, OLIF has an inflection class 105 -s/-"e exemplified by the lemma "Sonnenbrand" (*sunburn*) which therefore disallows "*Sonnenbrandes"¹⁰. Neither the Canoo language tools¹¹ nor WAHRIG support this limitation, and an exact Google search gives about 2'000 hits for "Son-

9. Unfortunately, the integer IDs for the classes are not even unique across different part-of-speech. On the other hand, there are quite a few classes which are redundant, i.e. they cover the same phenomena.

10. There exists a noun class 55 "Wunsch" (wish) -es/-"e that seems to enforce schwa in genitive singular.

11. <http://www.canoo.net>

POS	Gender	Example	Inflects Like	Code
noun	feminine	Mutter	-/- like Mutter/Mütter	53
noun.	feminine	Hand	-/-e like Hand/Hände	57
noun	feminine	Frau	-/-en like Frau/Frauen	64
adjective		arm	With umlaut and st in superlative like arm, ärmer, ärmst	96
verb		herausschinden	Irregular with separable prefix, like herausschinden - herausschund - herausschunden	645

Figure 1. Information contained in the OLIF inflection classes for German

nenbrands”, but 8'000 for “Sonnenbrandes”. There exist quite a few classes with overlapping or identical extension. The decision whether there is real redundancy has to be done painstakingly. In short, OLIF inflection codes were not as perfect as initially imagined. Along our development, we detected various problems and omissions which the OLIF consortium used to correct things according to our feedback.

3.2 Our word-and-paradigm finite state morphology

Our system is implemented using the Xerox finite state tool *xfst* (cf. Beesley and Karttunen 2003). The benefits of transducers for morphology systems are common place now: Bidirectionality (generation and analysis), non-determinism (regular relations encode many-to-many mappings, i.e. a word form allows more than one analysis and the same morphological features may produce more than one word form), efficiency in processing time and memory.

One special feature of our system is the ability to generate word forms in a class based fashion. Our demo web service¹² generates any desired inflectional paradigm for a given lemma by specifying the corresponding OLIF inflection class. Though monolithic morphologic systems as SMOR or TAGH can generate, they are limited to their lexical content which can't be extended simply by a pair of stem and inflection class.

Finite state morphology engineering is either based on a two-level rule component as GERTWOL (Koskeniemmi and Haapalainen 1996), or on composition of replacements and restrictions since the invention of the replacement operator (Karttunen 1995). We decided to use the latter serial approach because our lemma lexicon does not contain stem alternation, and therefore a lot has to be done by rules to ensure the correct word forms.

12. See <http://www.cl.uzh.ch/kitt/molif> for morphological generation and analyses.

IA B C D E F G H I J K L OLIFC example	IA B D E F G OLIFC, STERNULE, example
A0,B0,CO,DO,EO,FO,GO,HO,IO,JO,KO,LO, 90 "klein"	A4,BX,CX,DX,EX,FX,CO, 11, SEIN, "sein"
A0,B0,CO,DO,EO,F2,GO,HO,IO,JO,KO,LO, 96 "reicher"	A3,B1,C9,DO,E1,F2,G2, 831, WACHSEN, "aufwachsen"
A0,B0,CO,DO,EO,FO,GI,HO,IO,JO,KO,LO, 132 "arm"	A0,B0,CO,DO,EO,FO,GO, 41, MACHEN, "machen"
A0,B0,CO,DO,E1,FO,GO,HO,IO,JO,KO,LO, 136 "dunkel"	A3,B0,C1,D4,EO,F1,GO,281, SCHNEIDEN, "schneiden"
A0,B0,CO,DO,EO,FO,G2,HO,IO,JO,KO,LO, 422 "schmal"	A3,B0,C1,D4,EO,F1,G1,281, SCHNEIDEN, "anschneiden"
A0,B0,CO,D3,EO,FO,GO,HO,IO,JO,KO,LO, 446 "gut"	A3,B0,C1,D4,EO,F1,G2, 391, SCHNEIDEN, "anlechnen"
A1,B0,C1,DO,EO,FO,GO,HO,IO,JO,KO,LO, 622 "rosa"	A3,B0,C1,D4,EO,F1,G3, 491, SCHNEIDEN, "mitbeschneiden"
A0,B1,C1,DO,EO,FO,GO,HO,IO,JO,KO,LO, 631 "obig"	

Figure 2. Extract of the matrix of linguistic features for adjectives and verbs

The huge number of inflection classes which had to be managed required a systematic specification approach with as much as possible automated reuse thereof. In the first place, every OLIF inflection class had to be reconstructed as a matrix of linguistic features.

Figure 2 shows some sample feature vectors. For adjectives, we have e.g. A0 = inflectional, A1 = non-inflectional; B0 = attributive and/or predicative use, B1 = attributive use only; C0 = unlimited gradation, C1=positive only; D3=irregular gradation stem; F2=optional elision of e in comparative forms; G0=no umlaut, G1=umlaut, G2=optional umlaut.

For verbs, we have e.g.: A=main verb class: A0=regular A4=special inflection A3=strong verb; B=special present forms: B0=no umlaut B1=umlaut; C=ablaut in past and past participle: C0=no change C1=ei-i-i C9=a-u-a; D: additional stem changes (consonant): D0=no change D4=d-tt; E=umlaut in past subjunctive; E1=normal umlaut; F=final sound classes: F1= dental (-d,-t) F2=sibilant (-s,-z) without -sch; G=verb prefix: G0=no prefix G1=inseparable prefix G2=separable prefix G3=both prefixes.

The inflection component for each adjective class has an architecture as depicted in Fig. 3. Similar architectures are used for verbs and nouns. In order to keep the manual writing of class-specific replacement rules consistent and short, two mappings are automatically built by processing the feature matrix.

- Feature macros (e.g. AdjectiveMacroG2) contain the union of every OLIF class tag exhibiting the corresponding feature. The restriction concerning attributive use can be written as:

```
define AdjectiveUseRestr [
    "&use" => .#. [$. AdjectiveMacroB1] _ "=attr"
    , .#. [$. AdjectiveMacroB2] _ "=natr" ];
```

- Class rules (e.g. AdjectiveRule135) contain the composition of general restrictions together with all the class specific feature rules (AdjectiveRuleE1) which have to be coded manually. The rule for feature E1 (deletion of “e” in attributive positive and comparative forms in lemmata as “dunkel” (dark)) looks

like¹³:

```
define AdjectiveRuleE1 [
  [ $ [{e1} "<DEGR/>" ] # precondition: ensure lemma is ending on -el
  .o. [ e -> 0 || _ 1 "<DEGR/>" ["<COMP/>"|"<POS/>"][$. [{"kuse" "=attr"}]]];
```

The precondition ensuring “-el” is only necessary for keeping generation of paradigms specific and discriminating, because it excludes any stem with feature E1 not ending on “-el” from producing word forms. This is essential if we try to induce the OLIF inflection class from full form lexica.

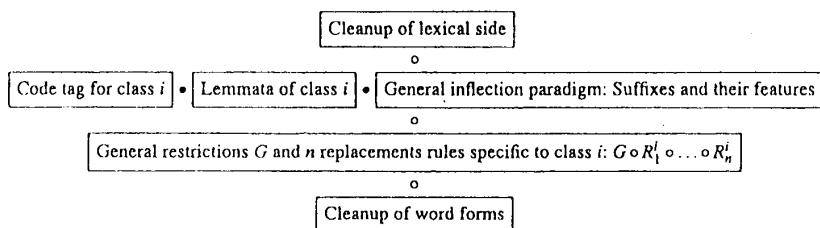


Figure 3. Main architecture of the MOLIFde inflection component for a single inflection class: o means composition, • concatenation.

For class based generation, we need to keep the composed replacement rules separated from the lexicon. The composition of replacement rules (which are typically cyclic and reentrant) can quickly lead to huge transducers and long compilation times. A careful explicit definition of the lexical language and its composition to the rules has been proven critical to reach our goals. The compiler needs some hints where morphological values may appear and where they won't. An extreme example is the purely rule-based treatment for the stem “sein” (be) where we additionally to stem changes specify the real inflection paradigm.

```

...
.o. [{sei} "<VINFL/>" {en} -> {vär} "<VINFL/>" e || _ ?* "kfin" "-fin"
  "kvf-m" "=konj" "knumb" "=ag" "kpers" ["=1"|"=3"] "ktense" "=past"]
.o. [{sei} "<VINFL/>" {en} -> {vär} "<VINFL/>" (e) {st} || _ ?* "kfin" "-fin"
  "kvf-m" "=konj" "knumb" "=ag" "kpers" "=2" "ktense" "=past"]
.o. [{sei} "<VINFL/>" {en} -> {vär} "<VINFL/>" {en} || _ ?* "kfin" "-fin"
  "kvf-m" "=konj" "knumb" "=pl" "kpers" ["=1"|"=3"] "ktense" "=past"]
.o. [{sei} "<VINFL/>" {en} -> {vär} "<VINFL/>" (e) t || _ ?* "kfin" "-fin"
  "kvf-m" "=konj" "knumb" "=pl" "kpers" "=2" "ktense" "=past"]
...

```

13. The gradation suffixes “er” and “st” are represented internally by abstract morphemes “<COMP/>” and “<POS/>” and realized in the cleanup step of the word form. This keeps the size of the composed transducers reasonable.

Without the lexical language the resulting transducer which generates all inflectional and non-inflectional forms, *xfst* gives the following properties: 110.1 Mb. 214706 states, 8886612 arcs, Circular. Composing the lexical language drastically reduces compilation time and size: 151.5 Kb. 812 states, 11507 arcs, Circular. Although this may still seem big, further composition with the lexicon entry “sein” results in a normal lexical transducer: 8.7 Kb. 280 states, 307 arcs, 34 paths.

3.3 Derivation, conversion, and compounding

Our lexicon doesn't provide origin information as SMOR. In contrast to compounding, derivation is a bounded process. Therefore, we can easily produce all derived lemmas¹⁴ and validate them afterwards by frequency checks over web-based search engines and corpora¹⁵. Applying a threshold to the frequency counts gives us quite reliable results, although no systematic evaluation has yet been done. In the current state, we derive all verb forms with separable prefixes from a list of around 100 prefixes. For the frequency checks of this verbs, the past participle is a good choice. Productive and regular derivations which we would like to treat properly appear often in iterated suffixation (adjectives ending on “-ig” derive nouns on “-igkeit”). The corresponding OLIF inflection classes of the source stem and the derived stem can be predicted with high precision.

Productive noun compounding is done as in SMOR with inflected forms (nominative singular and plural, genitive singular) for the first element using the inflection suffix as the linking morpheme. This has to be enriched by feminine noun classes with linking elements “-s-” that are not part of their inflectional paradigm, as well as some nouns as “Schule” (*school*) where final “e” is deleted as in “Schulhaus” (*school building*).

The problem of overanalyses introduced by compounding is also present in our system. Within the finite state calculus we implemented optionally a method called “lexicon prioritizing” to effectively remove overanalyses in the lexical transducer which are already covered by the lexicon. First, we determine a transducer that has all word forms of analyses from simple lexicon entries which can be reanalyzed by compounds on one side, and on the other side the corresponding compound analyses we want to suppress. Second, we use the side with the compound analyses to remove them from the lexical side of the original transducer. The calculation for this operations takes some minutes for the current lexicon size (see Fig. 4) and it's the most expensive compilation step regarding memory consumption and processing time.

14. Of course, conversion has also to be done. We have implemented a fix point computation that stops when conversion and derivation do not produce further new forms.

15. The SOAP services from <http://wortschatz.uni-leipzig.de> are very useful for this.

3.4 An open OLIF-based German lexicon

The lack of open and shared high-quality morphological resources adapted for the use in text technological applications is a dissatisfying situation for a language as German. Although, there is currently an interest in the automatic learning of morphological segmentation Demberg (2007)¹⁶, the results in the DurmLemmatizer lexicon show the difficulties of purely data-oriented boot-strapping approaches.

When we decided to adhere to the OLIF inflection classes, we had the aim to find preclassified entries which could be easily integrated. One hope was the lexicon of the OpenLogos¹⁷ translation system which contains a huge relational database and which was the original source of the OLIF inflection classes. Unfortunately, we had some problems to access it and to take it apart. Currently we are in the process of integrating and mass validating its 165'000 lemmas into our resources we converted in the meantime. The number of lemmas is high, because conversion results as nominalized infinitives and deverbal adjectives are separately listed.

In the first time, we used the full form lexicon which can be exported from the public, but closed source Windows-based system Morphy (Lezius 2000) to induce the inflection classes. Our morphology produced the possible paradigms for each stem, then we compared the results with Morphy's paradigm, and tried to identify a single class. In the course of this work, we found several omissions and errors on our side as well as some peculiarities how Morphy treats the rare past subjunctive forms of strong verbs. For about 21'000 noun lemmas, 5'500 adjective lemmas, 4'000 verbs lemmas (without separable verb prefixes) a single class was identified. One interesting point of this resource in terms of analyses coverage is the tendency of Morphy to postulate a lot of singulare tantum nouns and non-gradable adjectives – although in many cases, it's morphologically sound to produce plural or comparative forms. The restrictions stem from semanto-lexicographic determinations of the words which normally takes place when word forms are coupled with specific meanings. The same kind of frequency checks we use for the validation of derived word forms, can be used to check and quantify the tendency for restricted use of such words.

Third, we used open bilingual resources¹⁸ and extracted adjectives and nouns with frequent and regular suffixes. Classification validation can be done more quickly this way.

16. <http://www.cis.hut.fi/morphochallenge2008>

17. <http://logos-os.dfki.de>

18. <http://www.dict.cc>

Category	Lexicon	Conversion	Derived	All
Verb	4745	0	17393	22138
Noun	20474	21987	15526	57987
Adjective	12173	43865	2997	59035
All:	37392	65852	35916	139160

Figure 4. The current distribution of lexical entries with derivative forms filtered by a threshold of 5 occurrences.

4 Conclusion

We think that a shared, simply extendable, and standard-based morphological resource for German fills a gap for text technology and lexicography. High precision lemmatization and generation of word forms should be standard techniques, self-learning systems may help to extend or optimize further. Huge and well supported corpora with application interfaces are an invaluable service therefore. The use of closed-source software for our morphological tools may seem inconsistent. However, our approach needed powerful and developer-friendly finite state tools already two years ago when the development started. For the finite part of the lexicon we have created a textual export into the open-source SFST tools. A current project will use our morphology in a web-service for generation of inflected forms for the automatic recognition of glossary entries in the OLAT¹⁹ learning management system.²⁰

References

- Beesley, Kenneth R. and Lauri Karttunen (2003). *Finite-State Morphology: Xerox Tools and Techniques*. CSLI Publications.
- Cunningham, Hamish, Diana Maynard, Kalina Bontcheva, and Valentin Tablan (2002). A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 168–175, University of Pennsylvania, URL <http://www.aclweb.org/anthology/P02-1022.pdf>.
- Demberg, Vera (2007). A Language-Independent Unsupervised Model for Morphological Segmentation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 920–927, Prague, Czech Republic: Association for Computational Linguistics, URL <http://www.aclweb.org/anthology/P/P07/P07-1116>.
- EAGLES (1996). ELM-DE: EAGLES Specifications for German morphosyntax: Lexicon Specification and Classification Guidelines. electronic, URL http://www.ilc.cnr.it/EAGLES96/pub/eagles/lexicons/elm_de.ps.gz.

19. <http://www.olat.org>

20. Thanks to Thomas Kappeler and Luzius Thöny for implementing the verb and noun part of the system.

- Geyken, Alexander and Thomas Hanneforth (2006). *Finite-State Methods and Natural Language Processing, 5th International Workshop, FSMNLP 2005, Helsinki, Finland, September 1-2, 2005. Revised Papers*, chapter TAGH: A Complete Morphology for German Based on Weighted Finite State Automata, 55–66. Springer, URL http://dx.doi.org/10.1007/11780885_7.
- Ide, Nancy, Alessandro Lenci, and Nicoletta Calzolari (2003). RDF Instantiation of ISLE/MILE Lexical Entries. In *Proceedings of the ACL 2003 workshop on Linguistic annotation*, 30–37, Morristown, NJ, USA: Association for Computational Linguistics, doi:<http://dx.doi.org/10.3115/1119296.1119301>.
- Karttunen, Lauri (1995). The Replace Operator. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 16–23, Cambridge, Mass, URL <http://www.aclweb.org/anthology/P95-1003.pdf>.
- Koskeniemmi, Kimmo and Mariikka Haapalainen (1996). GERTWOL – Lingsoft Oy. In Roland Hauser (ed.), *Linguistische Verifikation : Dokumentation zur Ersten Morpholympics 1994*, number Band 34 in *Sprache und Information*, 121–140, Tübingen: Niemeyer.
- Lezius, Wolfgang (2000). Morphy - German Morphology, Part-of-Speech Tagging and Applications. In Ulrich Heid, Stefan Evert, Egbert Lehmann, and Christian Rohrer (eds.), *Proceedings of the 9th EURALEX International Congress*, 619–623, Stuttgart.
- McCormick, Susan M, Christian Lieske, and Alexander Culum (2004). OLIF v.2: A Flexible Language Data Standard. URL http://www.olif.net/documents/OLIF_Term_Journal.pdf.
- Perera, Praharshana and Rene Witte (2005). A Self-Learning Context-Aware Lemmatizer for German. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 636–643.
- Schiller, Anne, Simone Teufel, and Christine Stöckert (1999). Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset). URL <http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-1999.pdf>.
- Schmid, Helmut, Arne Fitschen, and Ulrich Heid (2004). SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. In *Proceedings of the 14th International Conference on Language Resources and Evaluation (LREC 2004)*, 1263–1266.
- Scott, Bernard (Bud) (2004). The Logos Model: An Historical Perspective. *Machine Translation* 18:1–72, URL <http://dx.doi.org/10.1023/B:COAT.0000021745.20402.59>.
- Volk, Martin (1999). Choosing the Right Lemma when Analysing German Nouns. In *Multilinguale Corpora: Codierung, Strukturierung, Analyse. 11. Jahrestagung der GLDV*, 304–310, Frankfurt.
- Wahrig, Gerhard and Renate Wahrig-Burfeind (eds.) (2006). *Wahrig Deutsches Wörterbuch: mit einem Lexikon der Sprachlehre*. Gütersloh: Wissen Media Verlag, 8. edition.

Tools for exploring GermaNet in the context of cl-teaching

Irene Cramer and Marc Finthammer

Abstract. Word nets, such as Princeton WordNet or GermaNet, are resources organizing a (more or less extensive) fraction of the vocabulary of a language according to lexical semantic relations. Such resources are widely used in natural language processing (NLP) and computational linguistics (CL) both for research and teaching purposes. While several graphical user interfaces (GUI) exist for Princeton WordNet – some of which are also available online – GermaNet still lacks such utilities. In this paper we describe two GUI-based tools meant to facilitate the navigation through and exploration of GermaNet. Both are freely available for download from our project web page (www.hytex.info). We additionally discuss ways of deploying these tools in teaching. We argue that the GUI-based access to word nets, which can be regarded as a fundamental resource in CL and NLP, enhances the students' understanding of basic lexical semantic concepts, computational semantics and lexicography.

1 Motivation

Word nets are lexical semantic resources modeled according to the principles introduced in Princeton WordNet (e.g. Fellbaum 1998). The central idea of word nets is to group synonymous lexical units, also including compounds and multi-word-units, into so-called synsets (synonym sets) and link them according to lexical semantic relations, such as hyponymy, meronymy, antonymy etc. Currently, Princeton WordNet (Version 3.0) contains approximately 150,000 synsets¹ and approximately 200,000 lexical units. The conceptual design and the resource itself are upgraded continuously – e.g. over the past years proper names have been added and tagged accordingly (Miller and Hristea 2006) and non-classical, i.e. psycho-linguistically motivated, link types have been included as an additional layer of relations (Boyd-Graber et al. 2006). Many NLP-applications, such as information retrieval and information extraction (e.g. Mandala et al. 1998) or word sense disambiguation (e.g. Banerjee and Pedersen 2002), highly rely on word nets as a (lexical) semantic resource². Therefore, in recent years, word nets have been developed for many languages, e.g. in the context of EuroWordNet (Vossen 1998) for seven European languages, and

1. Please refer to <http://wordnet.princeton.edu/man/wnstats.7WN> for more information.

2. Cp. Fellbaum (1998), Kunze (2001), and the Proceedings of the Global WordNet Conferences e.g. Tanács et al. (2008)



ΤΥΠΩΘΗΚΕ ΣΤΟ Ε.Κ.Π.Α.

Σταδίου 5, Τ.Κ. 105 62 - Αθήνα,

Τηλ. 210 36.89.374-210 36.89.375-210 36.89.388,

Fax: 210 36.89.433