

Vorlesung CL

Morphologie

und

**Endliche Automaten
sowie ihre Verwendung
in der morphologischen Verarbeitung**

Hans Uszkoreit

WS 04/05

Wörter und Lexeme

Wort Mehrdeutigkeit des Begriffes Wort

Wortform

Zitierform (headword) Eintragungsform für Wörter in Wörterbüchern z.B. Infinitiv oder erste Person Singular für Verben und Nominativ Singular für Nomen

Wort

In der Sprachtechnologie wird der Terminus Wort oft zur Bezeichnung einzelner Wortformen verwendet werden. Wortformen werden durch Kursivschrift gekennzeichnet: *find, found*.

Definition: Lexem

Ein Lexem ist eine Klasse lexikalisch äquivalenter Wortformen, die es in verschiedenen Umgebungen repräsentieren. Falls erforderlich können Lexeme typographisch durch Fettschrift gekennzeichnet werden: **find, die, tooth**.

L1 = {*sing, sings, singing, sang, sung*}

Das Lexem L1 besteht aus den Wörtern *sing, sings* etc.

Type vs. Token

- Im Umgang mit sprachlichen Objekten unterscheiden wir zwei Arten von “Gleichheit”:
 - Im ersten Fall geht es um wirkliche Identität, es handelt sich um ein- und dasselbe individuelle Objekt. (token identity)
 - Im zweiten Fall basiert die “Gleichheit” auf der Identität bestimmter Eigenschaften oder Merkmale, es handelt sich dabei um denselben Typ. (type identity)
- Diese Unterscheidung kann in der deutschen Sprache durch die Verwendung von *der-, die-, das-selbe* im ersten Fall und *der-, die-, dasgleiche* im zweiten Fall explizit gemacht werden; zur Bezeichnung dieser Unterscheidung dienen im Englischen die Begriffe *type* und *token*. Wir wollen dafür **Typ** (engl. *type*) und **Exemplar** (engl. *token*) sagen:

Morpheme

Das **Morphem** ist die grundlegende Einheit der Morphologie.

Morpheme sind wie die Phoneme oder Lexeme abstrakte Einheiten, die in der Rede durch diskrete, d.h. voneinander deutliche abgrenzbare, Einheiten realisiert werden, und zwar in der mündlichen Sprache als Phonemfolgen, in der schriftlichen als Graphemfolgen. Diese Repräsentations-einheiten werden **Morphe** genannt.

Morphemtypen

Im Deutschen und Englischen können viele Morpheme selbständig als Wörter verwendet werden. Solche Morpheme heißen **frei**.

Definition: *Freie Morpheme*

Morpheme, die selbständig als Wörter vorkommen können heißen freie Morpheme.

Haus, Hund, Wiese, Katze, Baum bzw. *boy, book, sing* etc. sind freie Morpheme.

Definition: *Gebundene Morpheme*

Morpheme, die nicht als selbständige Wörter vorkommen können, heißen **gebundene** Morpheme

Formative

In manchen Sprachen gibt es formbildende Elemente, die nicht den Status von Morphen (bzw. Morphemen) haben, weil sie keine identifizierbare Bedeutung haben. Sie können andererseits aber auch nicht nur als phonologische Erscheinungen behandelt werden.

Dies gilt beispielsweise für die sog. **Fugenelemente** bei der Bildung von zusammengesetzten Wörtern im Deutschen wie *Krankheit-s-zeichen*, *Liebe-s-lied*, *Universität-s-bibliothek*, *Auge-n-braue*, *Schwan-en-hals*.

Diese Fugenelemente sind zwar historisch aus Flexionsmorphemen von Substantiven hervorgegangen, haben jedoch keine identifizierbare Bedeutung mehr.

Pseudomorpheme

Ein ähnliches Problem liegt im Englischen in Wörtern wie *deceive*, *(mis-, pre-)conceive*, *(ap-)perceive*, *receive* vor.

Die Tatsache, daß die “Vorsilben” (Präfixe) *de-*, *con-*, *per-*, *re-* in vielen anderen englischen Wörtern romanischen Ursprungs vorkommen, legt eine Segmentierung in *de-ceive*, *con-ceive*, *per-ceive*, *re-ceive* nahe.

Das Element *-ceive* verhält sich wie ein Morphem, es kann ihm aber keine eindeutige Bedeutung zugeordnet werden. Es handelt sich um “Pseudomorpheme” oder *Formative*.

Ein Beispiel aus dem Deutschen: Erd+beer+e, Erd+beer+en, Erd+beer+marmelade
Johannis+beer+e,
Him+beer+e
Blau+beer+e
Brom+beer+e
Stachel+beer+e

Teilbereiche der Morphologie

Flexionsmorphologie:

Markierung von Tempus, Person, Kasus, Numerus...

geht-ging

der Mann-des Mannes

Derivationsmorphologie:

**Bedeutungsverändernde Bildung von Wörtern aus einem Stamm-Morphem
und einem Derivationsmorphem**

klar - unklar

Sache - sächlich o. sachlich

Komposita:

Zusammensetzung von mehreren Stamm-Morphemen

Bauer u. Hof - Bauernhof

Sonne u. baden - sonnenbaden

Derivationsmorphologie

N+los --> A

mittel+los

heil+los

furcht+los

A+keit --> N

Heiter+keit

Fröhlich+keit

Sauber+keit

V+bar --> A

les+bar

schreib+bar

unverzicht+bar

V+ung --> N

Zeich+nung

Salb+ung

Tarn+ung

Flexionsmorphologie und Affigierung

Affigierung ist die Verbindung einer Wurzel/Stammes mit Morphemen, die bestimmte Merkmale eines Wortes signalisieren.

Wurzel

Eine Wurzel ist die Grundform eines Wortes, die ohne Identitätsverlust nicht weiter analysiert werden kann. Es ist der Teil des Wortes, der verbleibt, wenn alle Affixe entfernt werden.

Das Wort *dishonesty* besteht aus den Morphemen *dis + honest + y*, wobei *dis* und *y* Affixe sind. Wenn man diese wegläßt bleibt als Wurzel *honest* übrig.

Stamm

In der Flexionsmorphologie nennen wir die Wurzel Stamm sprechen wir im Zusammenhang (Konjugation, Deklination, Komparation).

Affix

Affix ist der Sammelbegriff für alle Arten von Formativen, die nur in Verbindung mit einem anderen Morphem (der Wurzel oder dem Stamm) verwendet werden können, d.h. Affixe sind ein Typ gebundener Morpheme.

Affixe werden gewöhnlich in drei Klassen eingeteilt, je nach ihrer Position bezüglich der Wurzel oder des Stammes eines Wortes:

Präfix

Suffix, und

Infix.

Wurzel

Definition: *Wurzel*

Eine **Wurzel** ist die Grundform eines Wortes, die ohne Identitätsverlust nicht weiter analysiert werden kann. Es ist der Teil des Wortes, der verbleibt, wenn alle Affixe entfernt werden.

Das Wort *dishonesty* besteht aus den Morphemen *dis + honest + y*, wobei *dis* und *y* Affixe sind. Wenn man diese wegläßt bleibt als Wurzel *honest* übrig.

Stamm

Von einem **Stamm** sprechen wir also nur im Zusammenhang der Flexionsmorphologie (Konjugation, Deklination, Komparation).

Es gibt Sprachen, die besondere Stammbildungsformative verwenden. Im Lateinischen unterscheidet man bei der Konjugation je nach Stammbildung mehrerer Klassen, z.B. die ā-Stämme wie *laud-ā-re* 'loben' (Wurzel *laud*), die ē-Stämme wie *del-ē-re* 'zerstören' (Wurzel *del*), die ī-Stämme *aud-ī-re* 'hören' (Wurzel *aud*).

Basis

Jede Form, an die ein Affix angefügt werden kann, ist eine **Basis**.

Jede Wurzel und jeder Stamm ist eine Basis. Die Menge der Basen ist jedoch größer als die Vereinigung aller Wurzeln und Stämme, weil der Prozeß der Affigierung mehrfach anwendbar ist. Die Form *touchable*, z.B., fungiert als Basis für die Präfigierung mit *un-* bei der Bildung von *untouchable*. Dabei ist *touchable* jedoch weder eine Wurzel, weil es zusammengesetzt ist (*touch + able*), noch ein Stamm, weil es sich nicht um die Bildung einer Flexionsform handelt.

Undoable kann auf zwei Arten zerlegt werden

un+(do+able)

(un+do)+able

Morphophonologie

Wenn Morpheme miteinander kombiniert werden, können an den Verbindungsstellen eine Reihe von Veränderungen eintreten. Der traditionelle Oberbegriff für diese Art von Veränderungen an Kontaktstellen (Morphem oder Wortgrenzen) ist **Sandhi**.

Definition: *Sandhi*

Mit Sandhi bezeichnet man phonologische Veränderungen an grammatischen Formen (Morphemen, Wörtern), die bei ihrer Verknüpfung auftreten.

Man kann zwischen **internem** (an Morphemgrenzen in Wörtern) und **externem** (an Wortgrenzen) Sandhi unterscheiden.

Morphophonologie

Die phonologische Gestalt des Stammes beispielsweise kann sich ändern, wenn ein bestimmtes Morphem angefügt wird.

Wenn z.B. das englische Pluralmorphem an den Stamm *wife* angefügt wird, wird das stimmlose /f/ durch das stimmhafte /v/ ersetzt: *wife* + s → *wives*. Ähnlich bei {/haus/}+{/iz/} → {/hauziz/}.

Externes Sandhi

Ein bekanntes Beispiel für **externes** Sandhi ist die **liaison** im Französischen.

Ein Wort wie **grand** hat drei Allomorphe, /grɑ̃/, /grɑ̃d/ und /grɑ̃t/, wobei ersteres gewählt wird, wenn kein Vokal folgt. Die zweite Form erscheint im Wortinneren vor einem Vokal, z.B. **grande femme** /grɑ̃dɛ fɑ̃m/, die letzte am Wortende vor vokalischem Anlaut des nächsten Wortes: **grand homme** /grɑ̃tɔ̃m/.

Morphophonologie

Die Untersuchung der Wechselwirkung zwischen Morphologie und Phonologie (z.B. bei Sandhiformen) ist Gegenstand der **Morphophonologie** (engl. *morphophonemics*).

Synkretismus

Wie wir gesehen haben, kommt es bei flektierenden Sprachen häufig vor, daß verschiedene Kategorien durch die gleiche Form repräsentiert werden. Man bezeichnet dies als Synkretismus.

Definition: *Synkretismus*

Die Tatsache, daß innerhalb eines Paradigmas verschiedene grammatische Kategorien durch die gleiche Form repräsentiert werden, nennt man Synkretismus.

In *He came* und *He has come* haben das Präteritum und das Partizip Perfekt von **come** verschiedene Formen. In *He tried* und *He has tried* haben sie die gleiche Form, es handelt sich um einen Fall von Synkretismus.

Morphologische Prozesse

Wir können verschiedene Prozesse unterscheiden, mithilfe derer Wörter aus elementarerer Elementen wie z.B. Morphemen konstruiert werden können.

Segmental

- Affigierung

- Modifikation

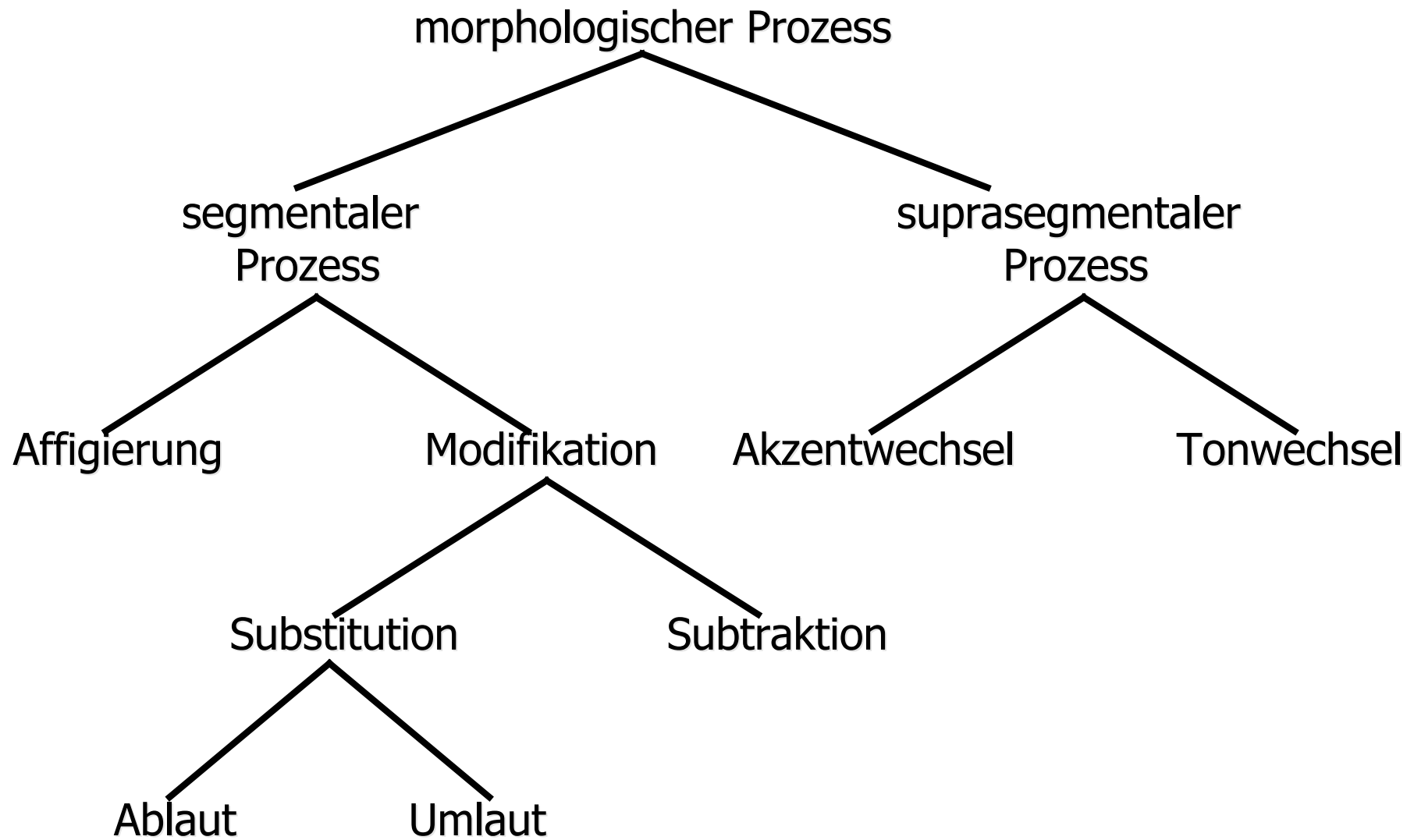
 - Substitution von Segmenten oder Merkmalen

 - Subtraktion (Tilgung) von Segmenten

Suprasegmental

- Akzentwechsel

- Tonwechsel



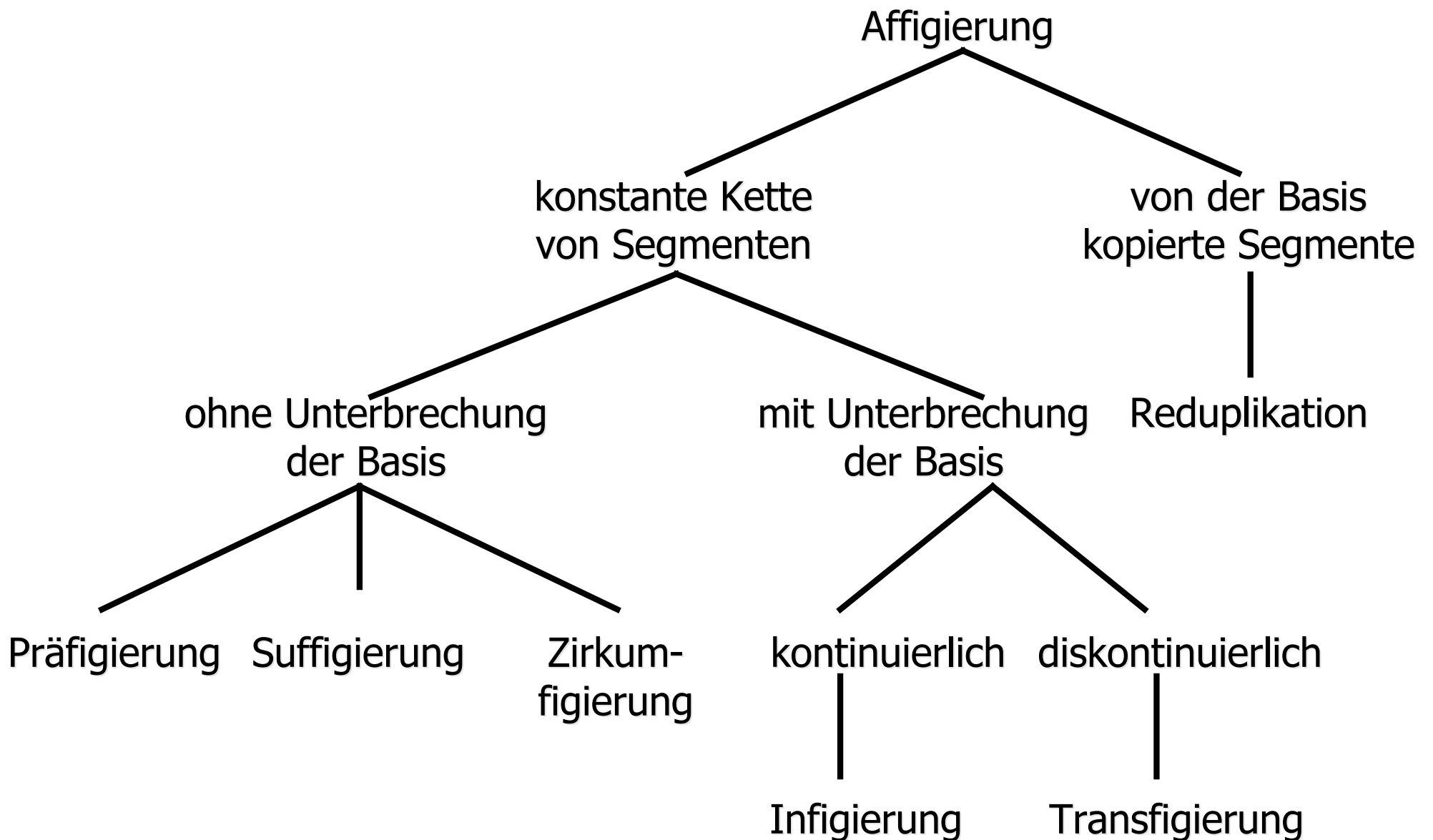
Affigierung

Der gebräuchlichste morphologische Prozeß im Deutschen und Englischen und vielen anderen Sprachen besteht im Anfügen eines Morphems an eine Wurzel oder einen Stamm:

trag + bar → tragbar

sing + ing → singing

Der morphologische Prozeß, durch den grammatische oder Lexikalische Information an einen Stamm angefügt wird, wird **Affigierung** genannt.



Affigierung

Affigierung ist ein rekursiver Prozeß insofern er wiederholt auf seinen eigenen Output angewandt werden kann, zum Beispiel:

de + scribe → describe

describe + able → describable

in + describeable → indescribable

Die Reihenfolge, in welcher die Affixe an den Stamm angefügt werden, ist signifikant, d.h. Wörter können eine interne Struktur haben, die über die bloße sequentielle Anordnung hinaus geht:

[in[[de scribe] able]]

Affix

Affix ist der Sammelbegriff für alle Arten von Formativen, die nur in Verbindung mit einem anderen Morphem (der Wurzel oder dem Stamm) verwendet werden können, d.h. Affixe sind ein Typ gebundener Morpheme.

Affixe werden gewöhnlich in Klassen eingeteilt, je nach ihrer Position bezüglich der Wurzel oder des Stammes eines Wortes.

Ohne Unterbrechung der Basis

Präfix

Suffix

Zirkumfix

Superfix

Mit Unterbrechung der Basis

Infix

Transfix

Präfix - Präfigierung

Ein **Präfix** ist ein Affix, das am Anfang einer Wurzel oder Stammes angefügt wird.

Der Prozeß des Anfügens eines Präfixes wird **Präfigierung** genannt.

Der Prozeß der Präfigierung wird im Deutschen und Englischen häufig zur Bildung neuer Lexeme verwendet nicht jedoch zur Bildung von Flexionsformen:

un- + glücklich → unglücklich

un- + happy → unhappy

mini- + computer → minicomputer

Suffix - Suffigierung

Ein **Suffix** ist ein Affix, das am Ende einer Wurzel oder eines Stammes angefügt wird.

Der Prozeß des Anfügens eines Suffixes wird **Suffigierung** genannt.

Suffigierung wird im Deutschen und Englischen sehr häufig verwendet, sowohl zur Derivation neuer Lexeme als auch zum Ausdruck grammatischer Beziehungen:

heiter + keit → Heiterkeit

Kopf + los → kopflos

national + -ise → nationalise

generate + -ion → generation

Flexionsendungen wie dt. *-t, -st, -en* bzw. engl. *-s, -ed, -ing*

Zirkumfix - Zirkumfigierung

Ein **Zirkumfix** ist ein diskontinuierliches Affix, das um eine Wurzel oder einen Stamm gelegt wird, also aus einem präfigierenden und einem suffigierendem Teil besteht.

Der Prozeß des Anfügens eines Zirkumfixes wird **Zirkumfigierung** genannt.

Als Beispiel für eine Zirkumfigierung wird häufig die Bildung des Partizips im Dt.

herangezogen: ge + worf + en, ge + mach + t. Da das ge- aber auch fehlen kann (bestell+t, be + komm + en) scheint das Suffix aber enger zum Stamm zu gehören.

Vgl. aber die Pluralbildung im Toiko (Nord-Togo): bara 'Frau' vs. m+bara+m 'Frauen'

Infix - Infigierung

Ein **Infix** ist ein Affix das im Inneren einer Wurzel oder eines Stammes eingefügt wird.

Der Prozeß des Einfügens eines Infixes wird **Infigierung** genannt.

Infigierung ist in den europäischen Sprachen eine sehr seltene Erscheinung, findet sich jedoch häufig in asiatischen, amerikanischen und afrikanischen Sprachen. Historisch gesehen ist das *-n-* im deutschen *stand* (im Gegensatz zu *stehen*) ein **Infix**.

Tagalog:

sulat 'schreiben' – s-um-ulat 'schrieb' – s-in-ulat 'wurde geschrieben'

Transfix - Transfigierung

Ein **Transfix** ist ein mehrteiliges diskontinuierliches Affix, das mit einer Wurzel oder einem Stamm verzahnt wird.

Der Prozeß des Einfügens eines Transfixes wird **Transfigierung** genannt.

Die Transfigierung ist ein häufig verwendetes Mittel zur Flexion in den semitischen Sprachen. Wortformen entstehen durch die Kombination von diskontinuierlichen konsonan-tischen Wurzeln, die die lexikalische Bedeutung kodieren, mit vokalischen Transfixen.

Transfix - Transfigierung

Die arabische Wurzel ktb bedeutet 'schreiben'. Daraus lassen sich z.B. im ägyptischen Arabisch Formen ableiten wie:

kátab 'er schrieb'

jíktib 'er wird schreiben'

maktúub 'geschrieben'

maktába 'Buchhandlung'

makáatib 'Buchhandlungen'

kitáab 'Buch'

káatib 'Schreiber'

kutub 'Bücher'

k t b

u u

k u t u b

Reduplikation

Unter Reduplikation versteht man die Verdoppelung von an- oder auslautenden Teilen einer Wurzel oder eines Stammes zum Ausdruck morphosyntaktischer Kategorien

Gotisch:

haldan 'halten' – haihald

haitan 'heissen' – haihait

Latein:

tundo 'stoße' – tutudi 'stieß'

pello 'treibe' – pepuli 'trieb'

Maori:

tau 'Mann' – ta-tau 'Männer'

mero 'Junge' – me-mero 'Jungen'

Modifikation

Ein weiterer wichtiger morphologischer Prozeß ist die **Modifikation**, eine Veränderung in der Wurzel oder im Stamm eines Wortes.

Ein Beispiel dafür ist der Vokalwechsel zwischen den Singular- und Pluralformen vieler deutscher sowie einiger englischer Substantive:

dt. *Sohn* ~ *Söhne*, *Hut* ~ *Hüte*, *Lamm* ~ *Lämmer*

eng. *man* ~ *men*, *mouse* ~ *mice*, *goose* ~ *geese*).

Ein verbreiteter Vorgang dieser Art ist der **Ablaut**.

Ablaut

Ablaut nennt man den regelhaften Vokalwechsel in Wörtern des gleichen Lexems, der nicht phonologisch konditioniert ist.

Einschlägige Beispiele finden wir bei vielen sog. starken Verben

dt. *singen* ~ *sang* ~ *gesungen*, *finden* ~ *fand* ~ *gefunden*, *werden* ~ *ward* ~ *geworden*,

engl. *sing* ~ *sang* ~ *sung*, *find* ~ *found* ~ *found*, *give* ~ *gave* ~ *given* etc.

Der Ablaut ist vom **Umlaut** zu unterscheiden.

Umlaut

Umlaut ist eine Vokalalternation zwischen verwandten Vorderzungen- und Hinterzungenvokalen, die — zumindest historisch betrachtet — phonologisch konditioniert ist (regressive Assimilation unter dem Einfluß von /i, j/ in der Folgesilbe).

Wo jedoch die Bedingungsfaktoren verlorengegangen sind, muß Umlaut als ein morphologischer Prozeß aufgefaßt werden.

Beispiele:

*Mutter ~ Mütter, Vater ~ Vater, Vogel ~ Vögel,
man ~ men, mouse ~ mice, fox ~ vixen (dt. Fuchs ~ Füchsin).*

Subtraktion

Genus im französischen Adjektiv

maskulinum		femininum	
Schrift	Laut	Schrift	Laut
<i>grand</i>	[gʁɑ̃]	<i>grande</i>	[gʁɑ̃d]
<i>petit</i>	[pti]	<i>petite</i>	[ptit]
<i>gris</i>	[gʁi]	<i>grise</i>	[gʁiz]
<i>gentil</i>	[ʒɑ̃ti]	<i>gentille</i>	[ʒɑ̃tij]

Orthographisch betrachtet werden die femininen Formen durch Anhängen von –e an den Stamm gebildet.

Phonologisch jedoch sind die maskulinen Formen durch Tilgung des Auslautkonsonanten abgeleitet.

Superfix (Suprafix) - Superfigierung

Ein **Superfix** (oder **Suprafix**) suprasegmentales Affix, das eine Wurzel oder einen Stamm überlagert.

Der Prozeß der Modifikation durch eine Superfix wird **Transfigierung** genannt. Manifestationen sind Akzentwech-sel und Tonwechsel zum Ausdruck grammatischer Bedeutungen:

Akzentwechsel (engl.):

prodúce (v.) vs. próduce (n), permít (v.) vs. pérmit (n.)

impórt (v.) vs. ímport (n.), insúlt (v.) vs. ínsult (n.), discóunt vs. díscout

Tonwechsel (Kanuri, Nigerien; ´ = hoher Ton, ` = fallender Ton):

lezè (Konj.) vs. lezé (Opt.) 'gehen'

tussè (Konj.) vs. tussé (Opt) 'ruhen'

Konversion

Konversion ist ein besonderer Ableitungsprozeß, wobei ein Lexem in eine neue Lexemklasse überführt wird, ohne daß ein Affix angefügt wird.

Beispiele: Verb → Nomen:

schau-en → *Schau*

bau-en → *Bau*

fall-en → *Fall*

Da das Englische keine sehr ausgeprägte Flexion hat, ist die Konversion ein sehr verbreitetes Wortbildungsmittel;

Verb → Nomen: *smell, taste, hit, walk*

Adjektiv → Verb: *dirty, empty, lower*

Komposition

Komposition ist der morphologische Prozeß, durch den neue zusammengesetzte Lexeme durch die Kombination zweier oder mehrerer freier Formen gebildet werden.

Ein durch Komposition gebildetes Wort heißt **Kompositum** (engl. *compound*).

Beispiele:

dt. *Haus* + *Tür* → *Haustür*, *groß* + *Stadt* → *Großstadt*,
engl. *bed* + *room* → *bedroom*, *black* + *bird* → *blackbird*,
washing + *machine* → *washing machine*

Morphophonologische Prozesse (Beispiele)

❑ Epenthese

z.B. bad + st ⇔ badest

❑ Elision

z.B. ras + st ⇔ rast

❑ Merkmalsveränderungen an Phonemen

z.B. Auslautverhärtung

(stimmhafte Konsonanten werden am
Silbenende stimmlos)

/bad/ ⇔ [bat]

Automaten

Automaten in der weiteren Bedeutung des Wortes sind ein zentrales Konzept aber nicht formal definiertes Konzept in der Informationsverarbeitung.

Wenn wir durch eine Sequenz von Handlungen bestimmte Effekte auslösen, dann ist das nicht unbedingt Informationsverarbeitung. Normalerweise wird mit einer jeden Handlung kausal eine Wirkung erzeugt.

Beispiel: ich öffne eine Tür mit zwei Schlössern, indem ich jedes Schloß mit einem Schlüssel aufschließe.

Wenn wir aber ein System haben, das so gebaut ist, daß es in Abhängigkeit von meinen Handlungen „Entscheidungen“ über seine Handlungen trifft, dann liegt Informationsverarbeitung vor. In diesem Fall bewirken meine Handlungen die Handlungen der Maschine nicht direkt kausal.

Automaten sind Systeme, die in Abhängigkeit von meinen Handlungen, bestimmte Aktionen ausführt bzw auslöst.

Damit wir von Automaten sprechen, muß es mindestens einen Entscheidungspunkt geben.

Automaten

Automaten werden verwendet, um Symbol- oder Handlungsabfolgen zu überprüfen, zu analysieren oder abzuarbeiten.

Sie werden z.B. bei der Verarbeitung von Benutzereingaben verwendet.

Dabei wird geprüft,

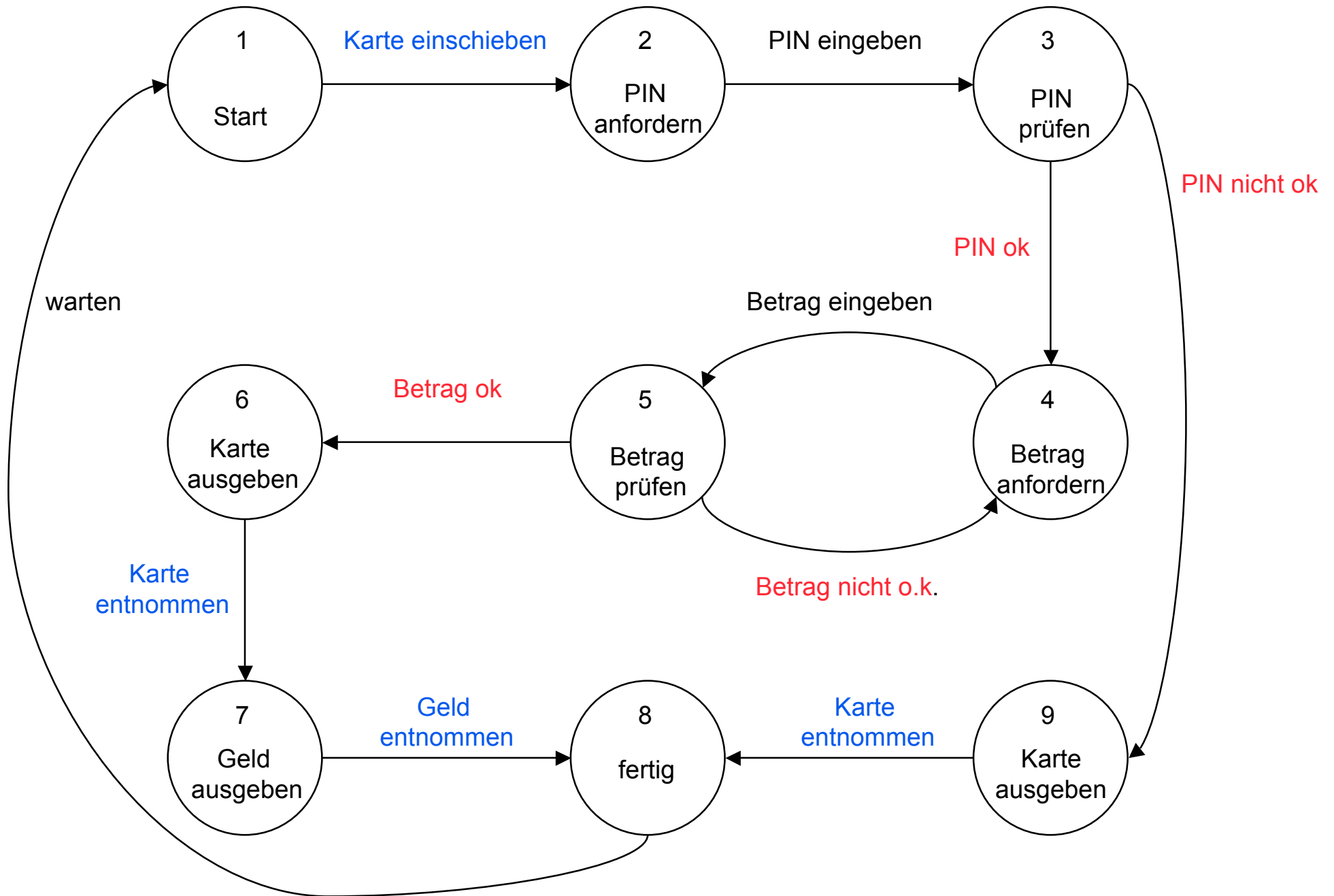
ob Eingaben korrekt sind und
in der richtigen Reihenfolge bzw. im richtigen Kontext erfolgen.

Automaten spielen auch eine Rolle bei der Überprüfung bzw. Übersetzung von Programmier- oder Spezifikationssprachen in die internen Sprachen des Computers.

In der Sprachverarbeitung werden Automaten an vielen Stellen eingesetzt, so z.B. bei der morphologischen Analyse von Wörtern.

„Endlicher Geld-Automat“

Das Beispiel ist leicht abgewandelt von Prof. Dr. Reinhard Völler (Hamburg) übernommen



Bestandteile eines Automaten:

endliche Menge von Zuständen

Startzustand, Endzustand (auch mehrere)

Eingaben: Worte einer Sprache

Ausgaben: Worte einer Sprache

Regeln für die Auswirkungen einer Eingabe

Wichtig: diese Regeln müssen vollständig sein, d.h. für jede Eingabe muß eine Reaktion des Automaten spezifiziert sein.

Notfalls wird ein Fehlerzustand definiert.

Endliche Automaten

Ein endlicher Automat ist ein mathematisches Modell eines Systems mit Ein- und Ausgaben.

Ein solches System befindet sich immer in einer aus einer endlichen Anzahl möglicher interner Konfigurationen.

Man sagt auch: das System befindet sich in einem Zustand .

Beispiele:

Ein Schaltkreis mit n Gattern befindet sich in einem von 2^n möglichen Zuständen.

Texteditoren oder lexikalische Analysatoren von Compilern kann man als endliche Automaten modellieren.

Auch ein Computer ist ein endlicher Automat. Allerdings ist dieses Modell wegen der großen Anzahl möglicher Zustände nicht besonders hilfreich.

Deterministischer endlicher Automat (DFSA)

Ein deterministischer endlicher Automat ist ein Fünftupel $A = (Z, E, \delta, z_0, F)$

Z Menge der Zustände

E Menge der Eingabesymbole

$\delta: Z \times E \rightarrow Z$ Zustandsübergangsfunktion

$z_0 \in Z$ Anfangszustand

$F \subseteq Z$ Menge der Endzustände

DFSA als Akzeptor

Für eine Eingabekette $w = e_1, e_2, \dots, e_n$ soll überprüft werden, ob sie durch einen Automaten A akzeptiert wird

Wir definieren uns zwei Variablen, q für den gegenwärtigen Zustand und e für das gegenwärtige Eingabesymbol

Wir setzen $q = z_0$ und wiederholen dann für jedes Symbol der Eingabekette e_i

$q := d(q, e_i)$

$e := e_{i+1}$

Wenn $e = e_n$ und $q \in F$ dann ist die Eingabekette durch A akzeptiert, ansonsten gilt sie als zurückgewiesen.

Beispiel 1

Namensliste:

Peter Müller

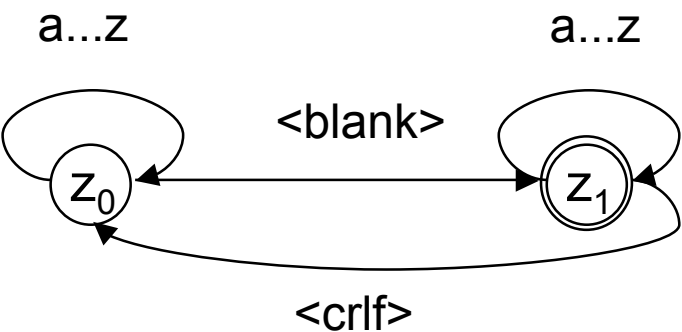
.

.

.

Doris Steckler

	z₀	z₁
a	z ₀	z ₁
.	z ₀	z ₁
.	z ₀	z ₁
.	z ₀	z ₁
z	z ₀	z ₁
<blank>	z ₁	
<crlf>		z ₀



	z₀	z₁
z₀	a...z	<blank>
z₁	<crlf>	a...z

Nichtdeterministischer endlicher Automat

Ein nichtdeterministischer endlicher Automat ist ein Fünftupel $A = (Z, E, \delta, z_0, F)$

Z Menge der Zustände

E Menge der Eingabesymbole

$\delta: Z \times E \rightarrow 2^Z$ Zustandsübergangsfunktion

$z_0 \in Z$ Anfangszustand

$F \subseteq Z$ Menge der Endzustände

Automaten mit Ausgabe

Ein Mealey-Automat ist ein Sechstupel $A = (Z, E, A, \delta, z_0, \lambda)$

Z Menge der Zustände

E Menge der Eingabesymbole

A Menge der Ausgabesymbole

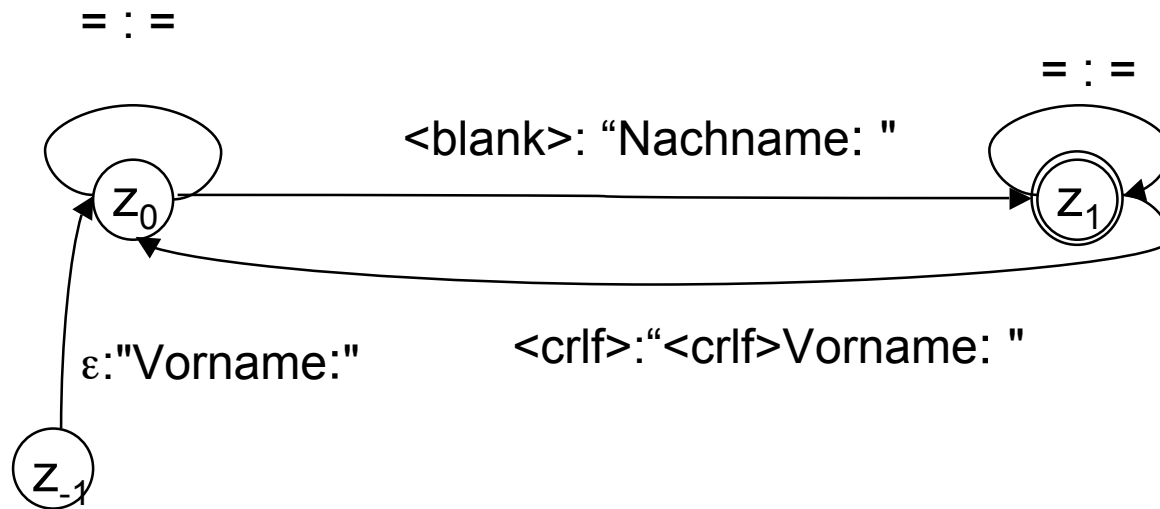
$\delta: Z \times E \rightarrow 2^Z$ Zustandsübergangsfunktion

$\lambda: Z \times E \rightarrow A$ Ausgabefunktion

$z_0 \in Z$ Anfangszustand

Beispiel 2

	z_{-1}	z_0	z_1
ϵ : "Vorname:"		z_0	
= : =		z_0	z_1
<blank>: "Nachname:"		z_1	
<crLf> : "Vorname:"			z_0



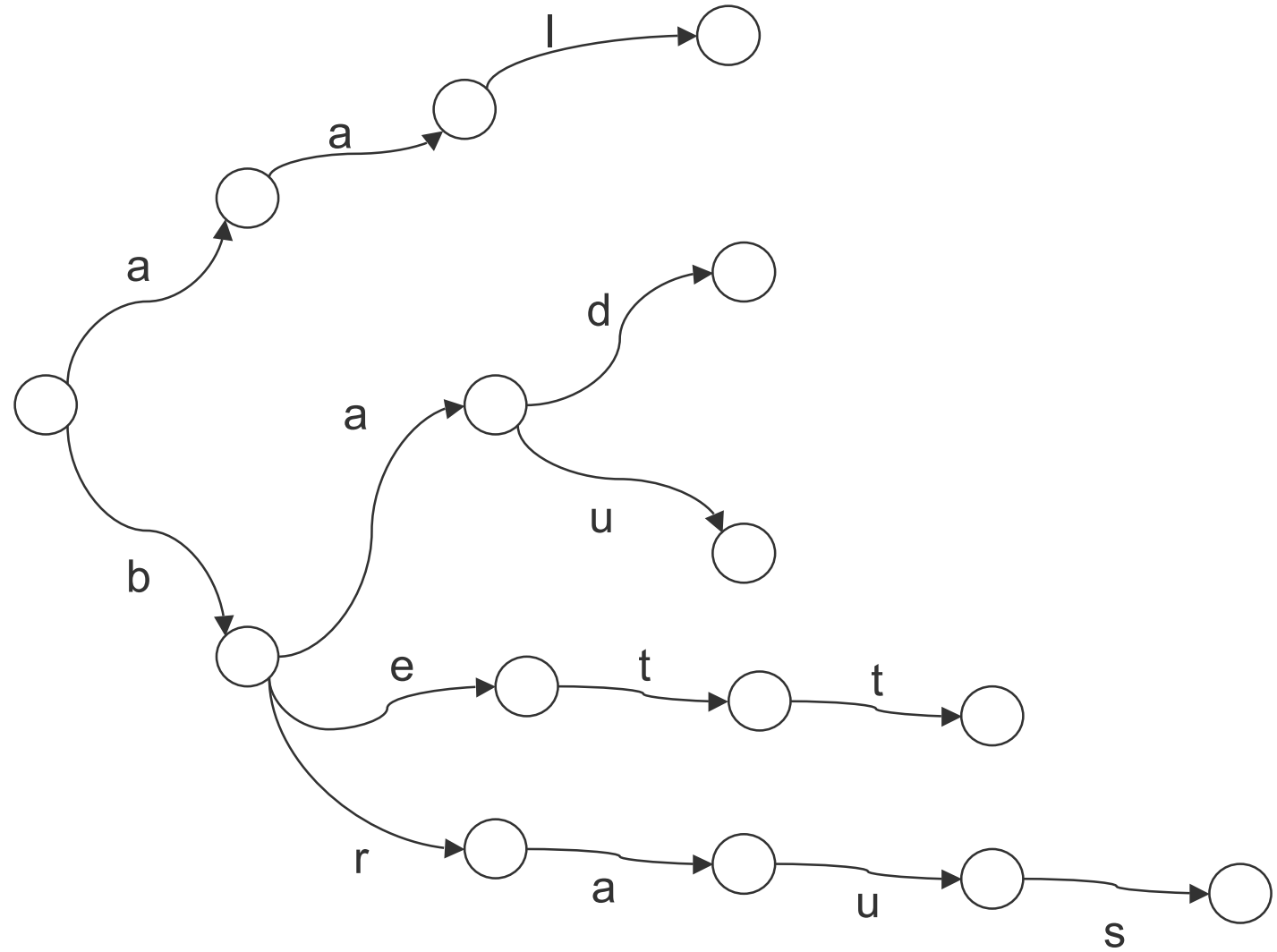
Beispiel 2

Vorname: Marion Nachname: Abbecker

Vorname: Klaus Nachname:Becker

Vorname: Günter Nachname:Bruck

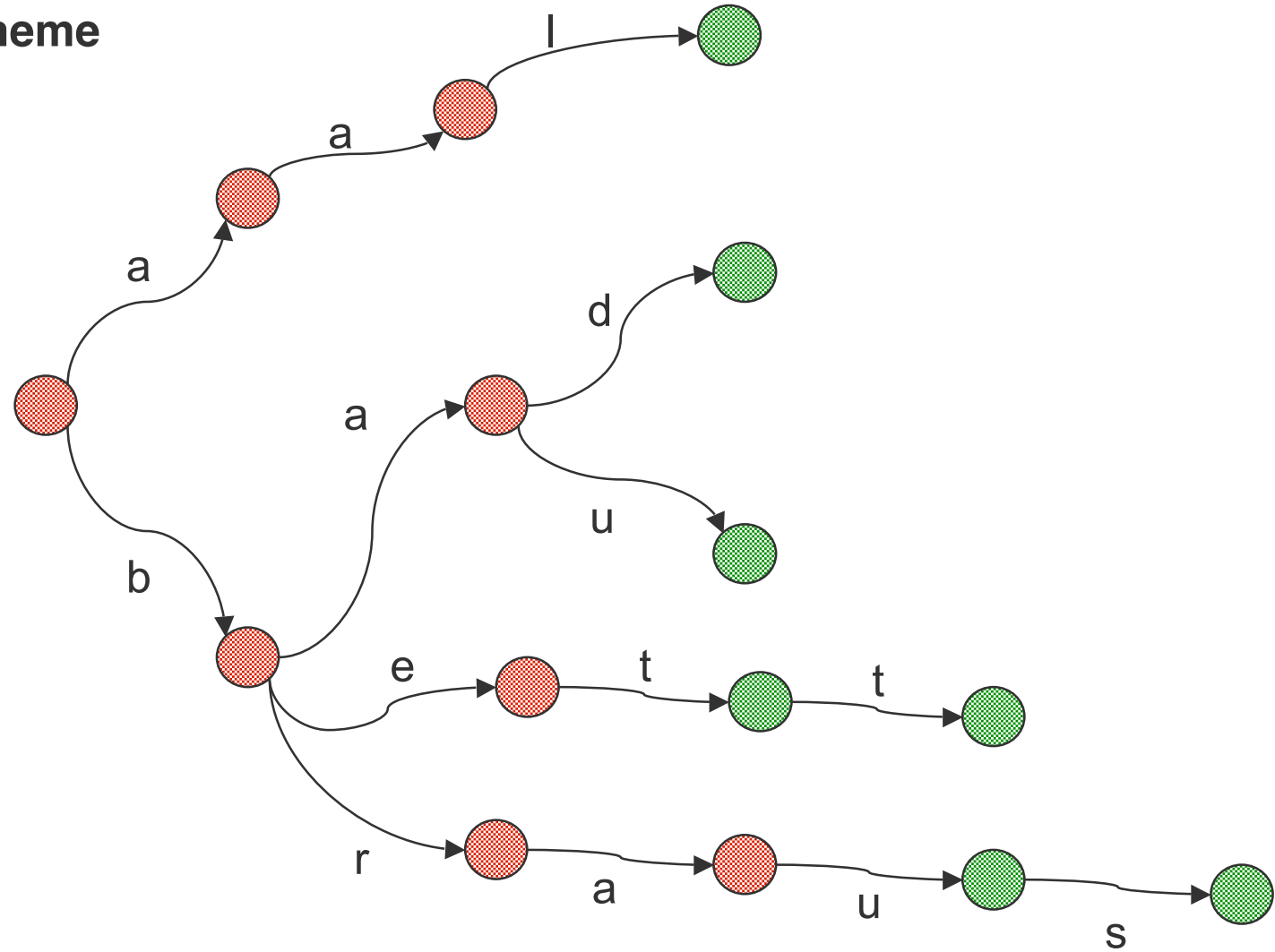
Buchstabenbaum



Buchstabenbaum

Verbstamm-Morpheme

aal
bad
bau
bet
bett
brau
braus

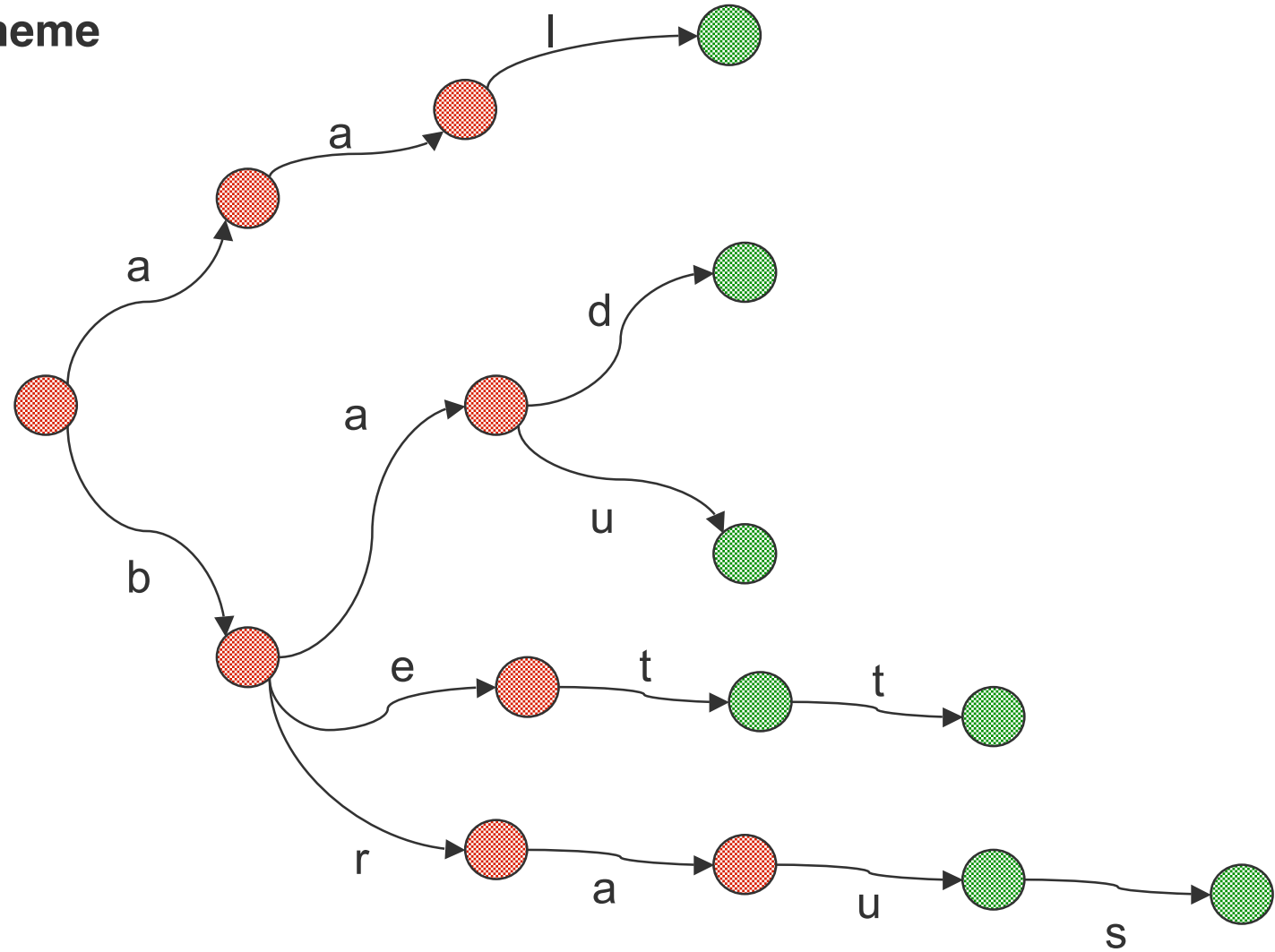


Buchstabenbaum

VSTEM: aal

Verbstamm-Morpheme

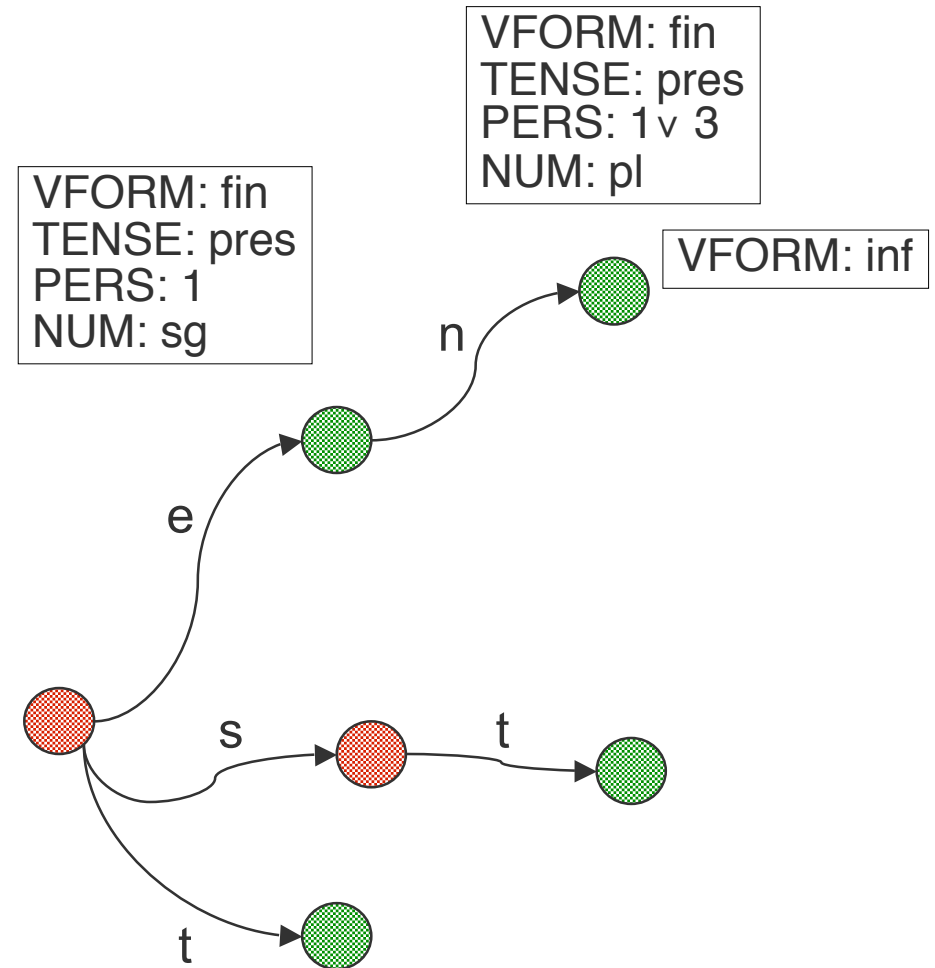
aal
bad
bau
bet
bett
brau
braus



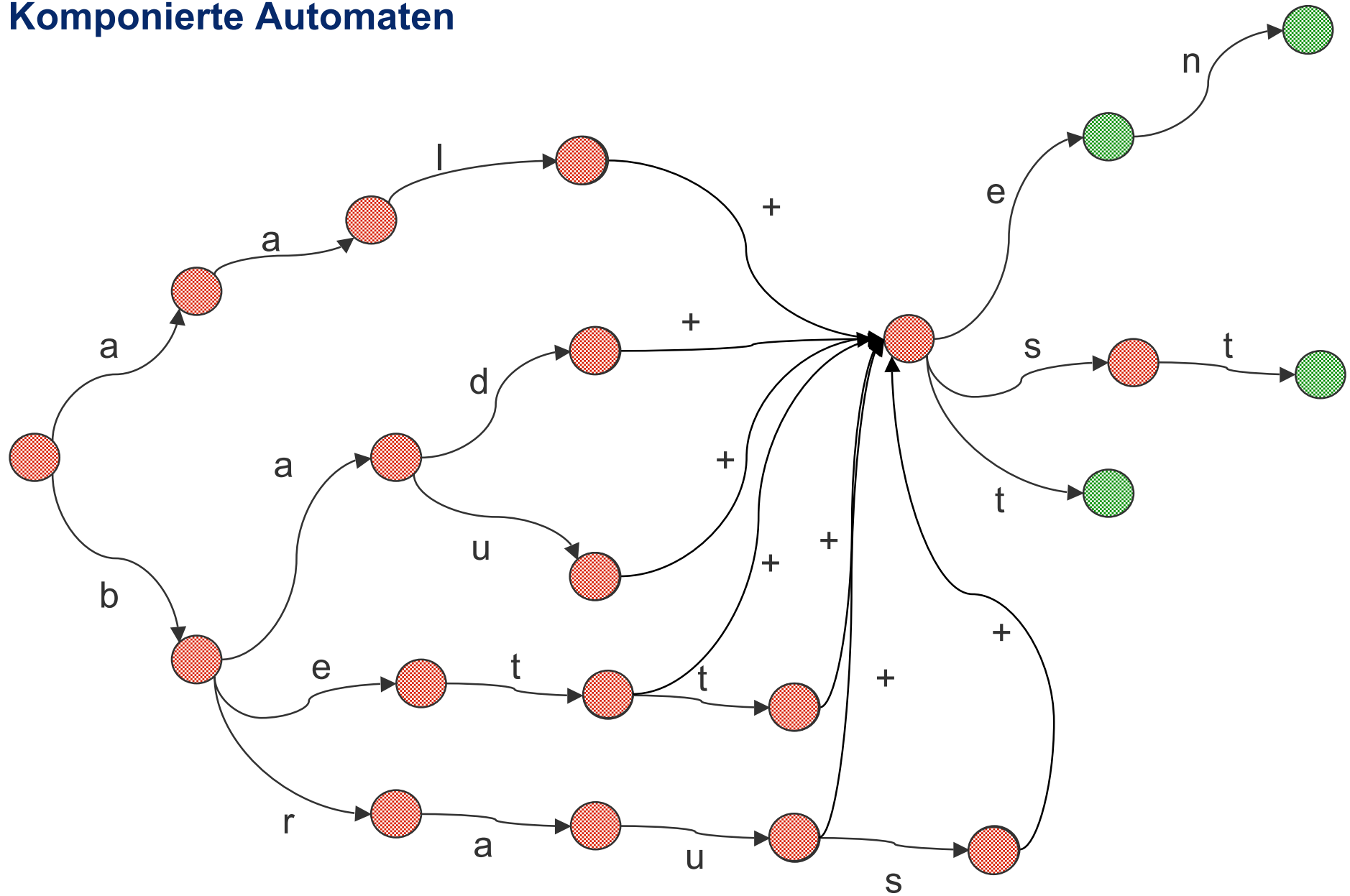
Buchstabenbaum

Verbsuffix-Morpheme

-e
-en
-st
-t



Komponierte Automaten



Beispiele

a	a	l	+	s	t
a	a	l		s	t

a	a	l	+	s	t
a	a	l	e	s	t

b	a	d	+	s	t
b	a	d		s	t

b	a	d	+	s	t
b	a	d	e	s	t

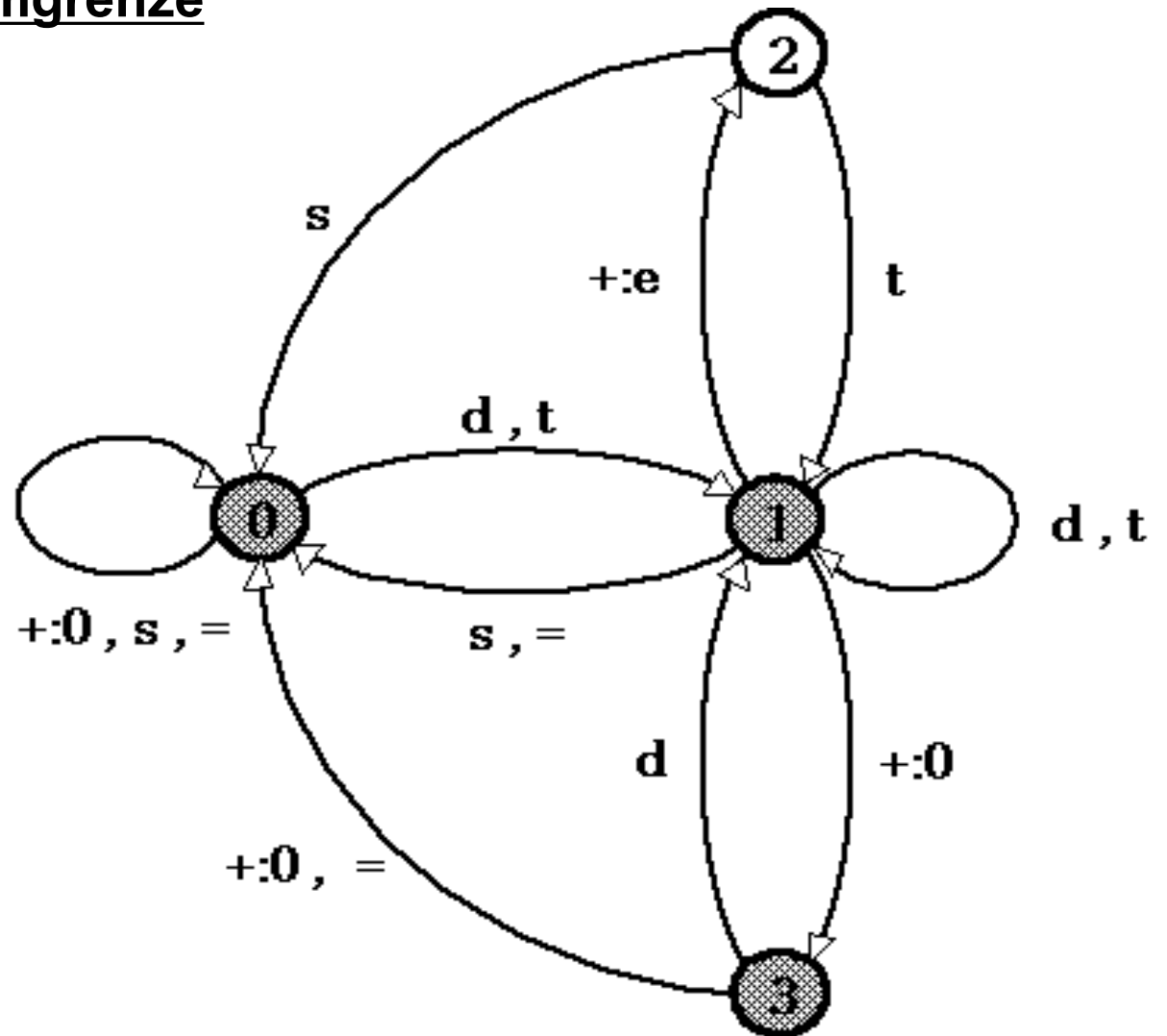
Zwei-Ebenen Morphologie

e-Epenthese an der Morphemgrenze (vereinfacht)

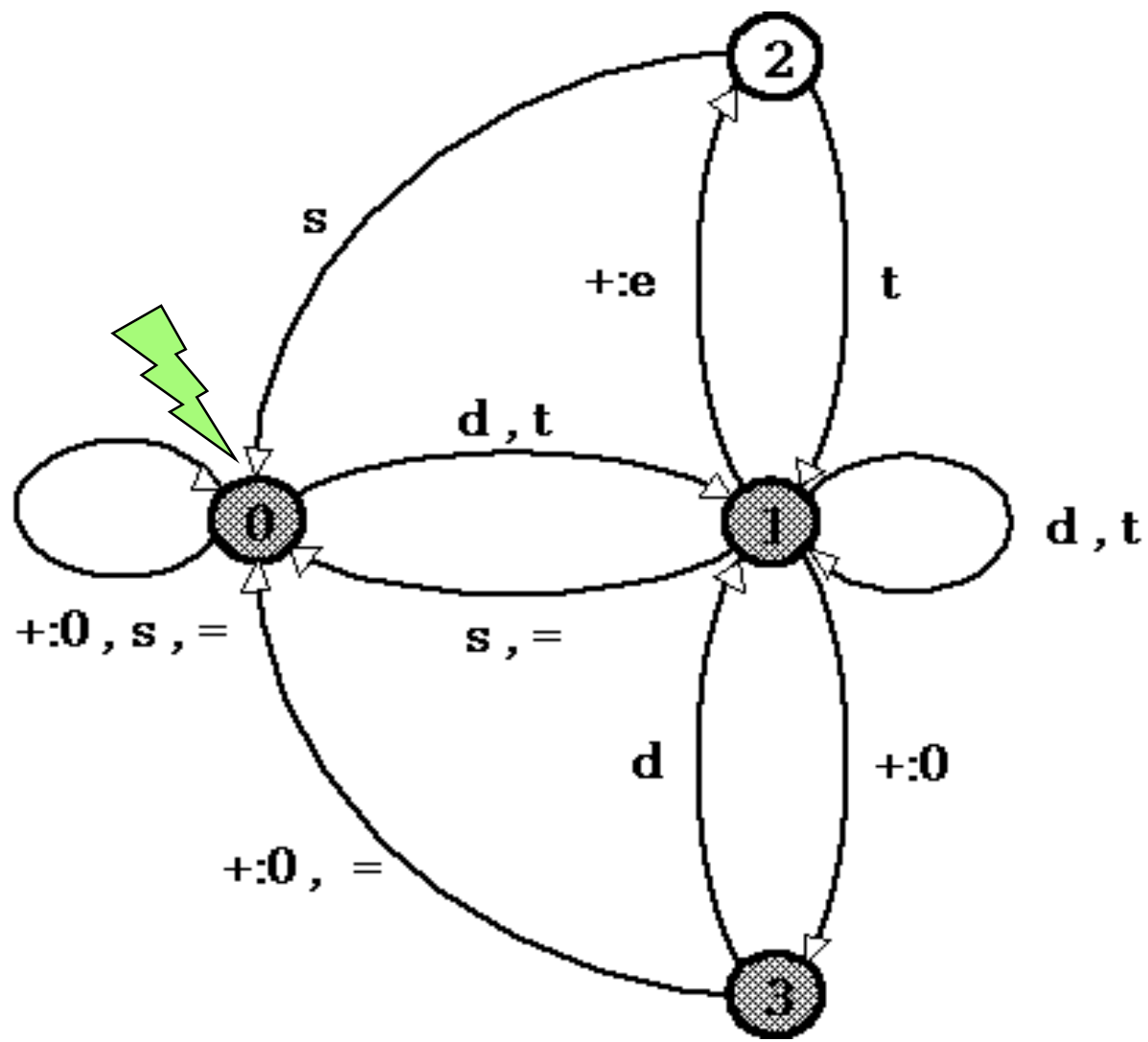
Regel: $+:e \Leftrightarrow \{d, t\} _ \{s, t\}$

Übergangstabelle:

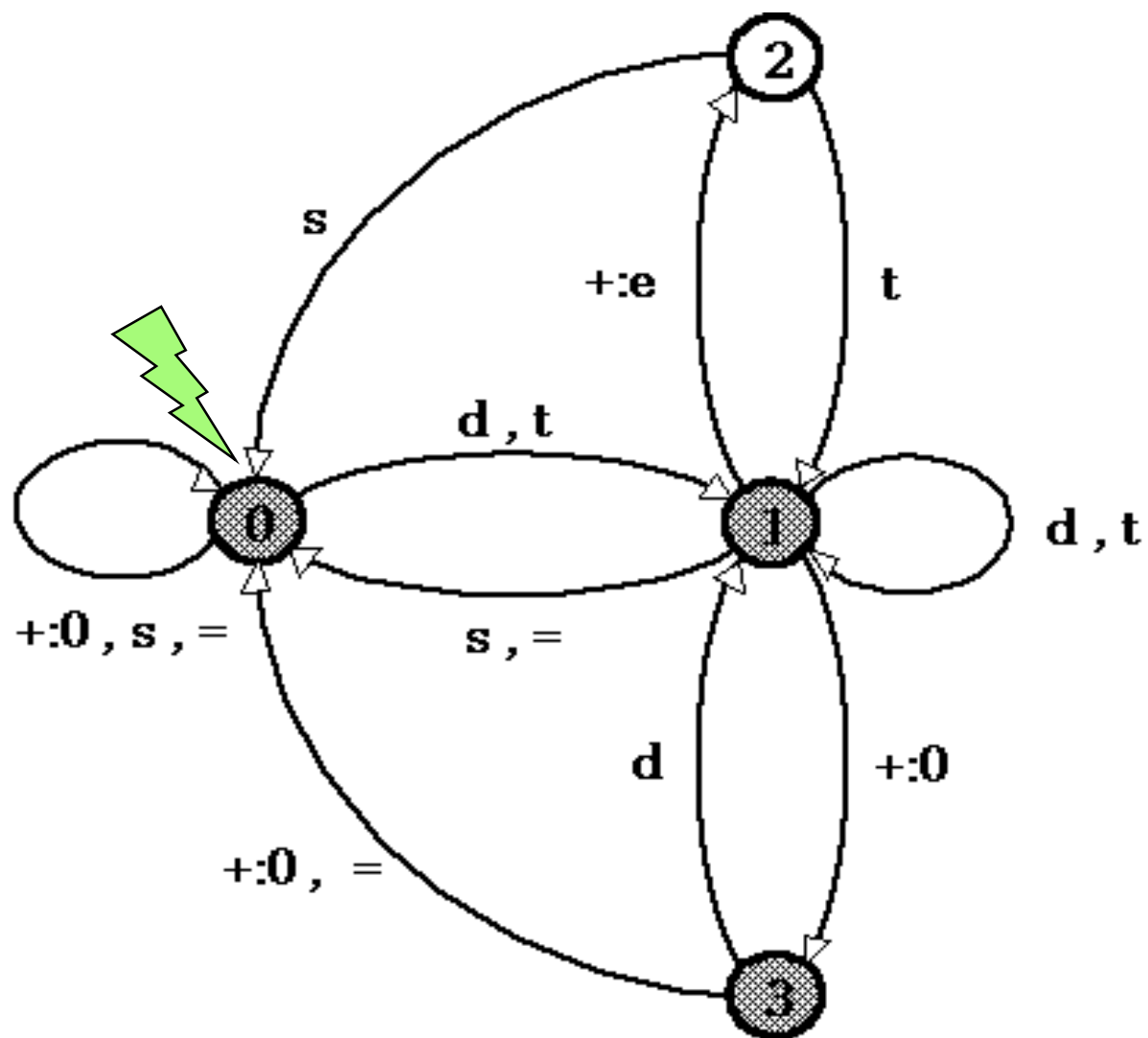
	0	1	2	3
$+:e$	-	2	-	-
$+:0$	0	3	-	0
$d:d$	1	1	-	1
$s:s$	0	0	0	-
$t:t$	1	1	1	-
$:=$	0	0	-	0



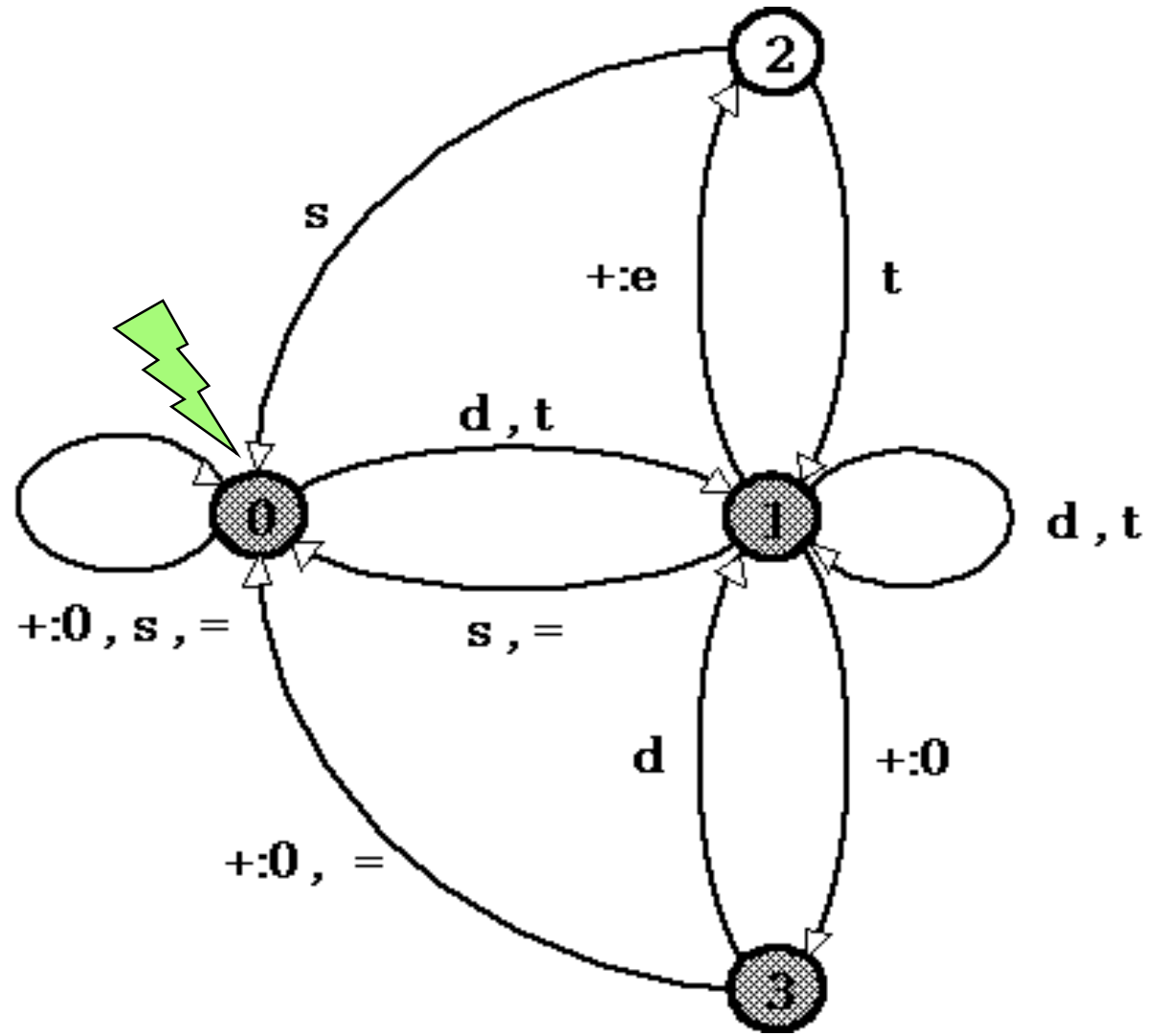
b a d + s t
b a d s t



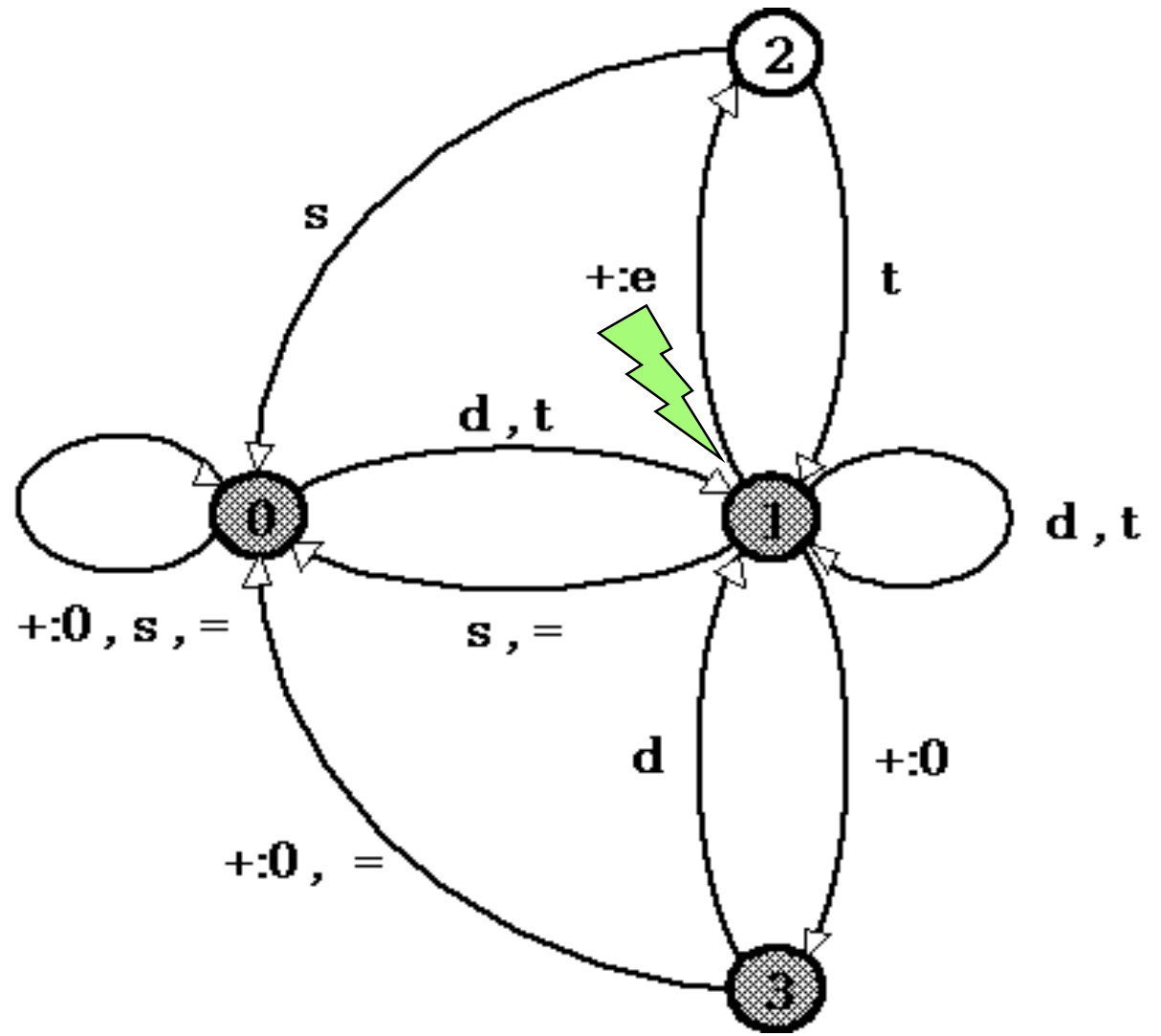
b a d + s t
b a d s t



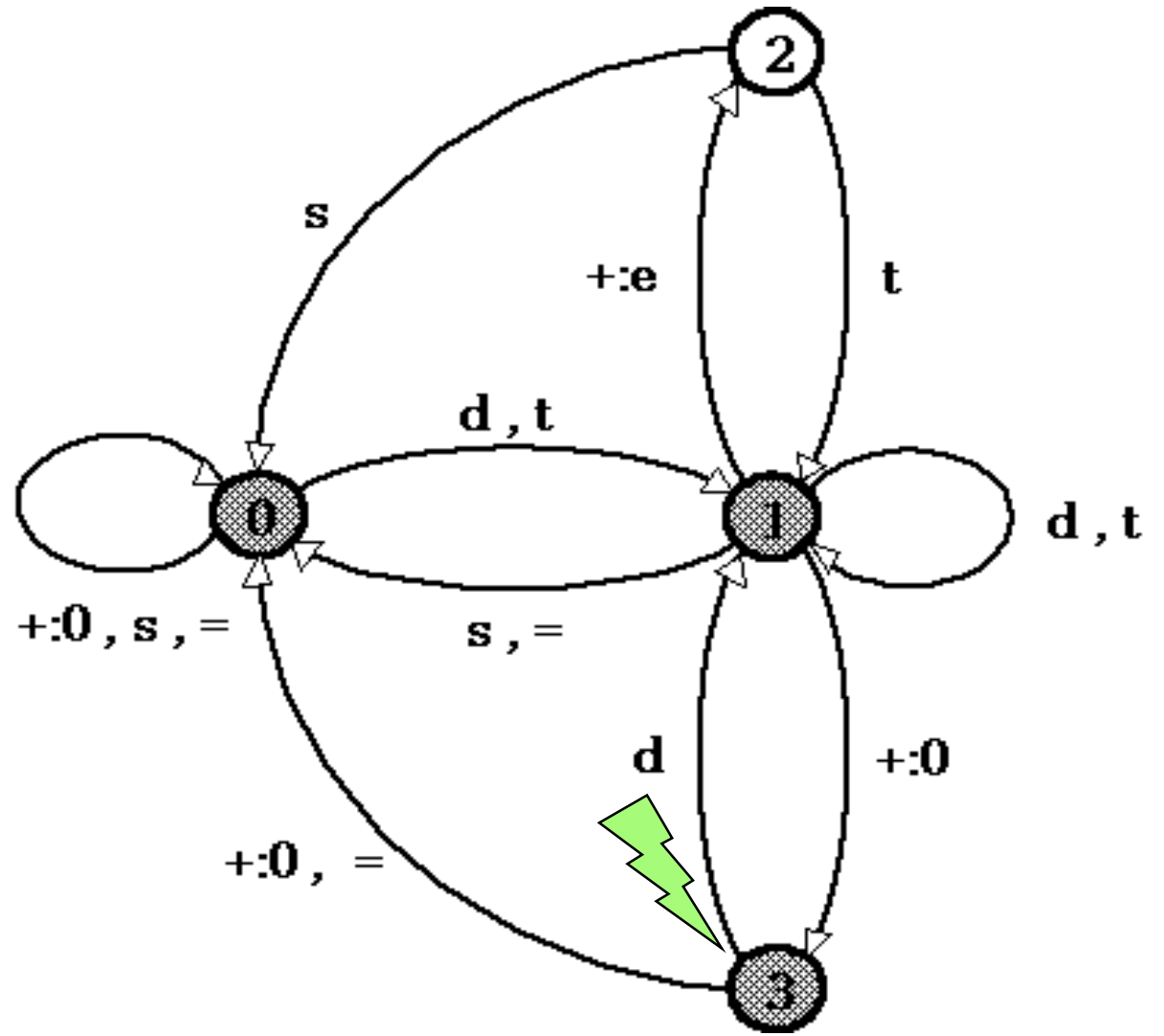
b a d + s t
b a d s t



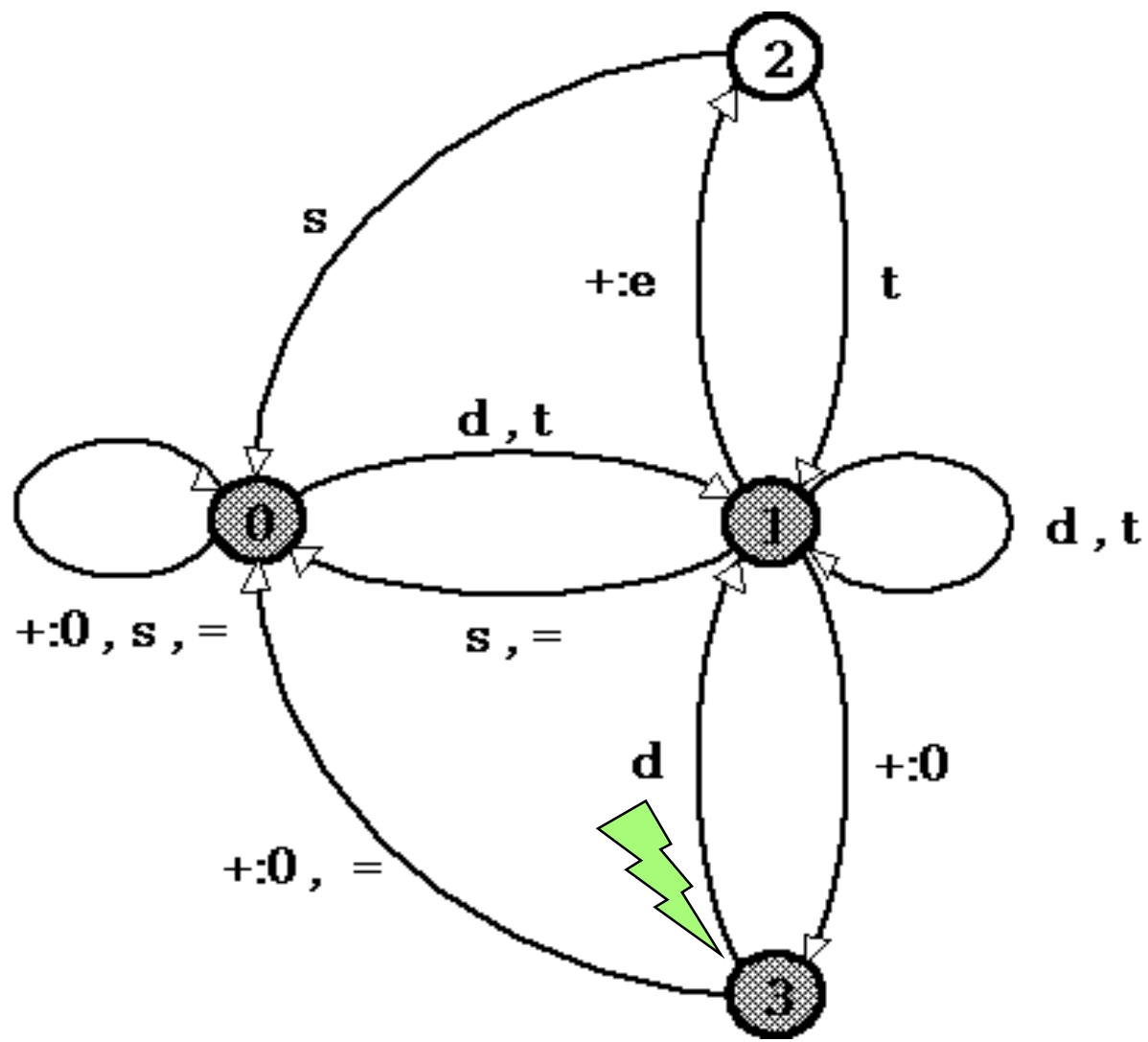
b a d + s t
b a d s t



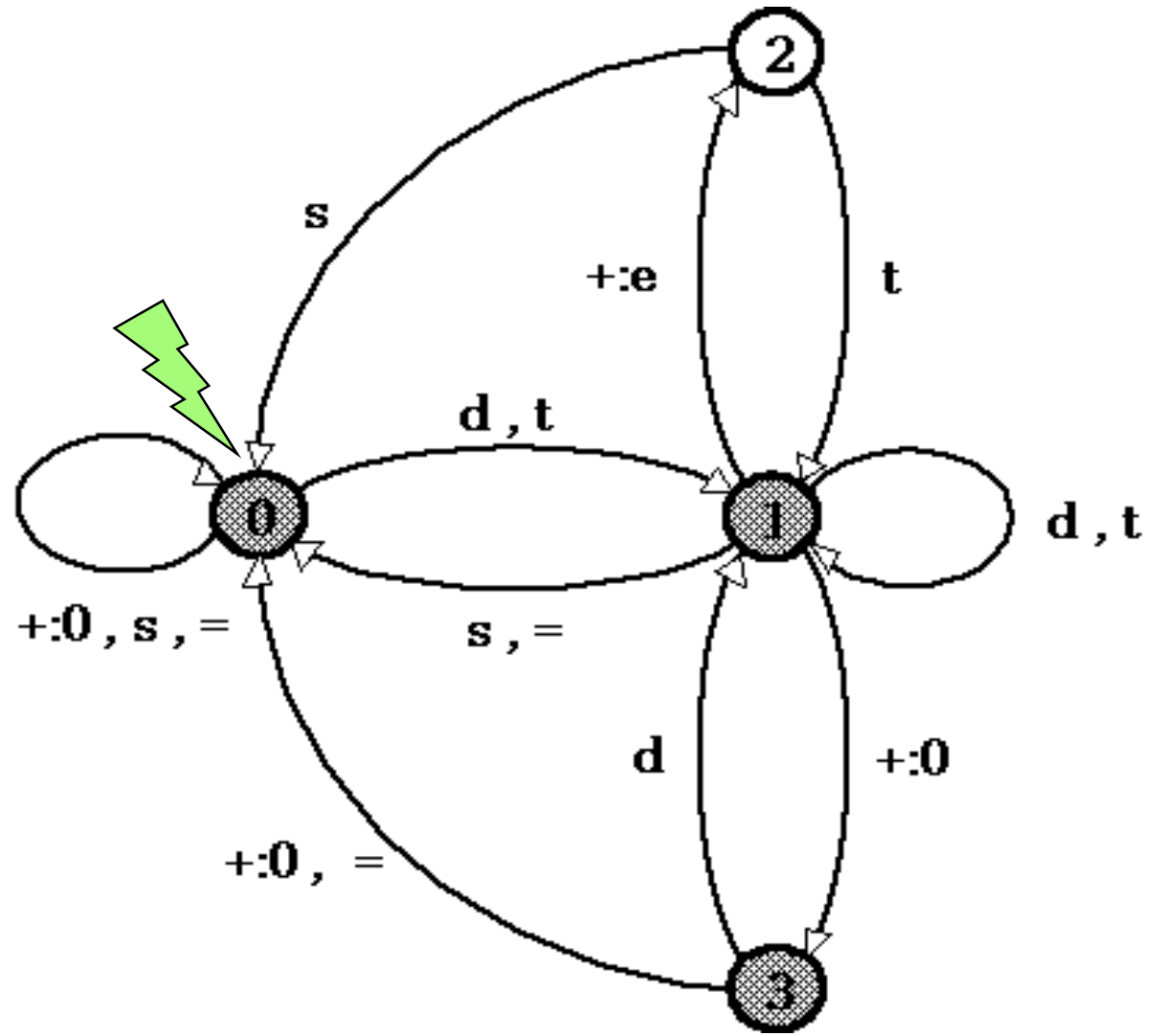
b a d + s t
b a d s t



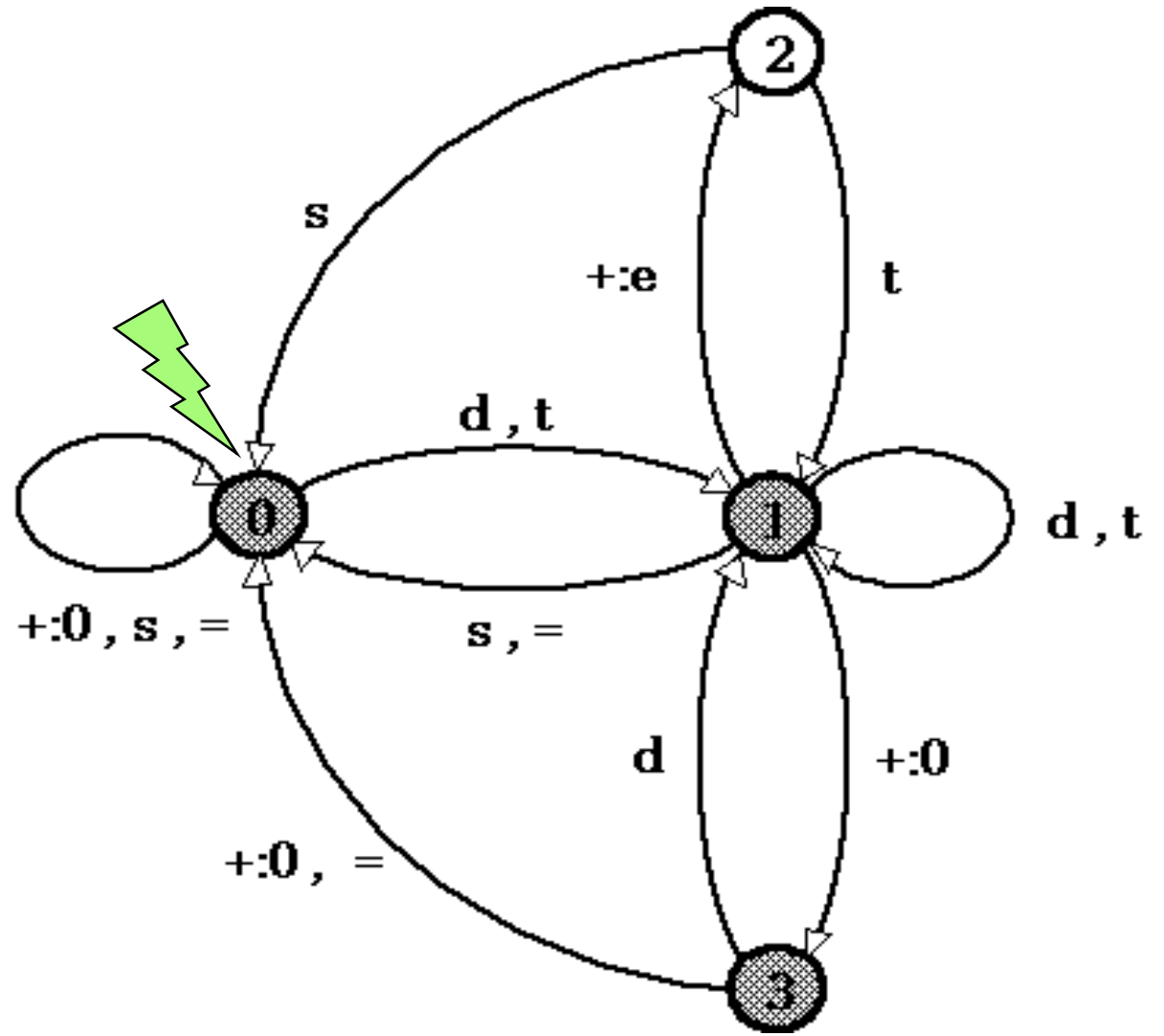
b a d + s t
b a d s t



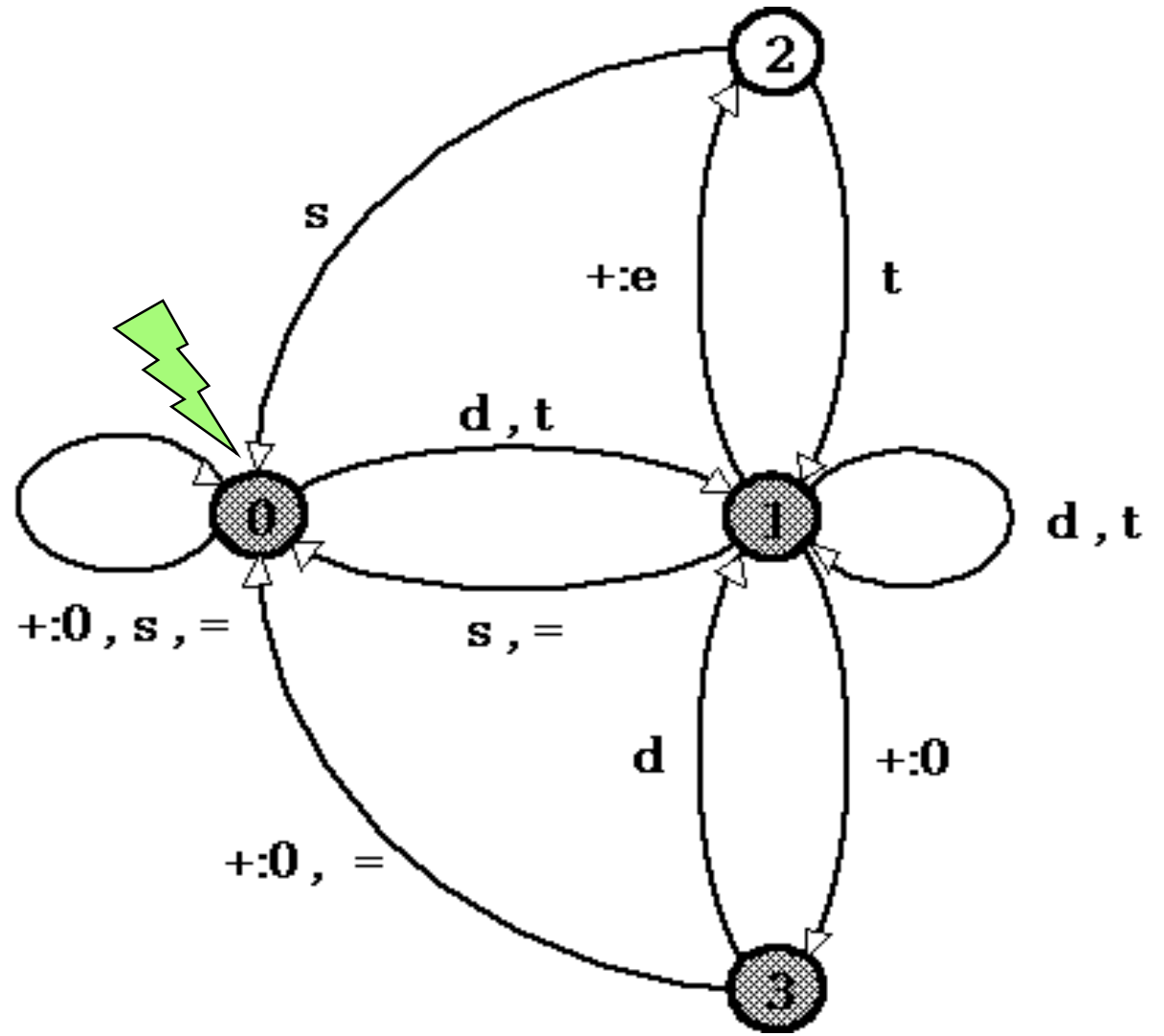
b a d + s t
b a d e s t



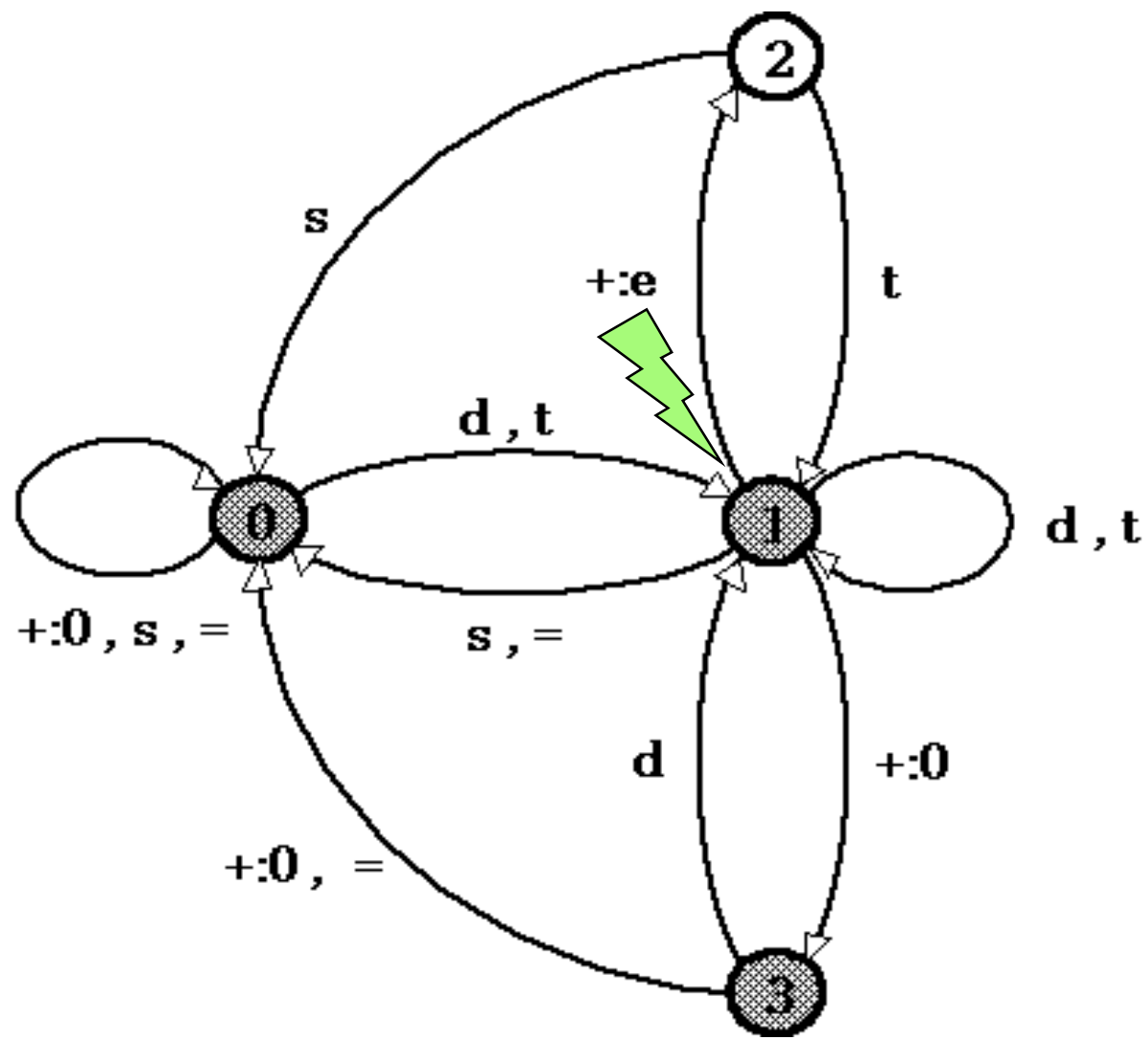
b a d + s t
b a d e s t



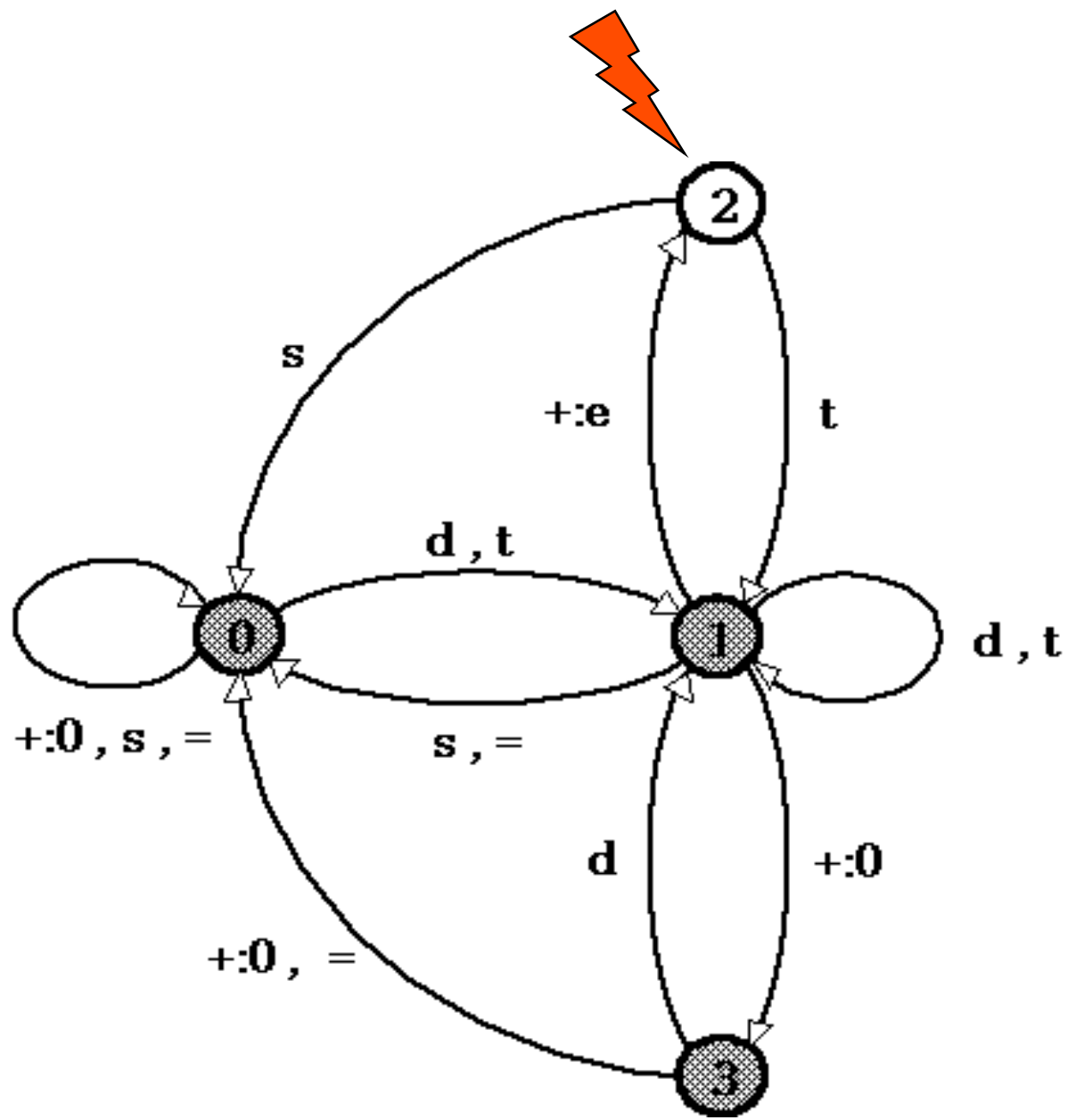
b a d + s t
b a d e s t



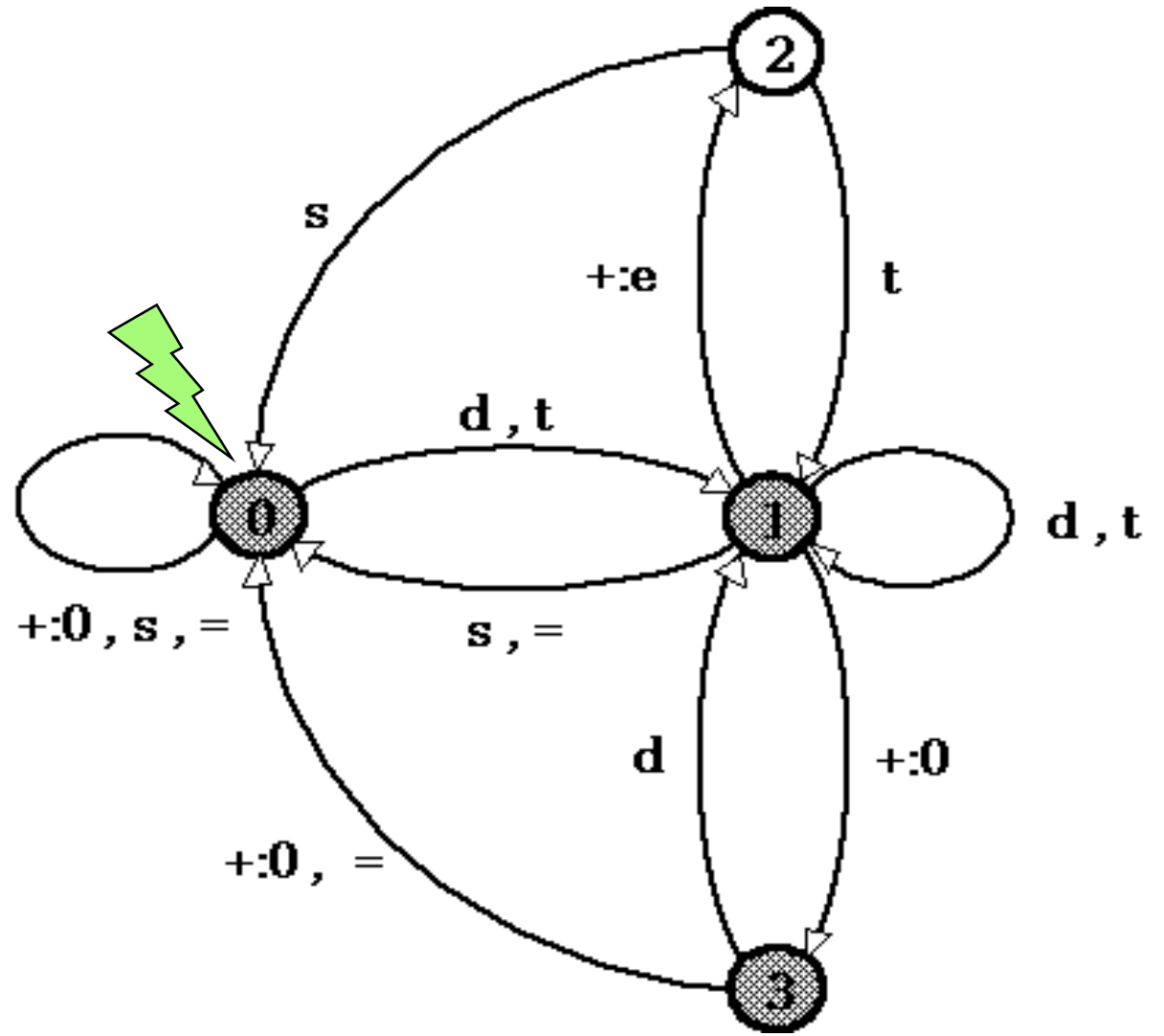
b a d + s t
b a d e s t



b a d + s t
b a d e s t

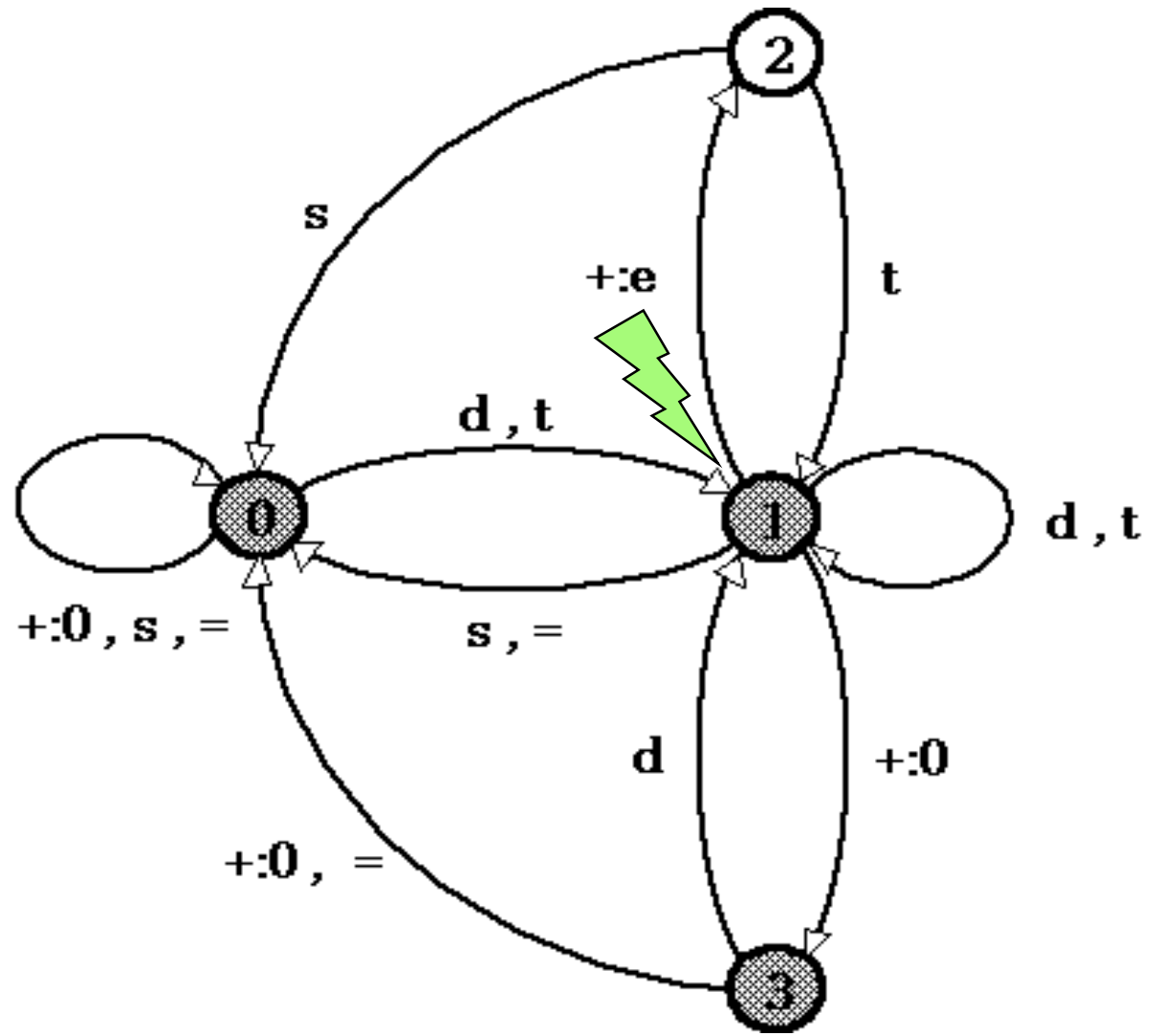


b a d + s t
b a d e s t



b a d + s t

b a d e s t



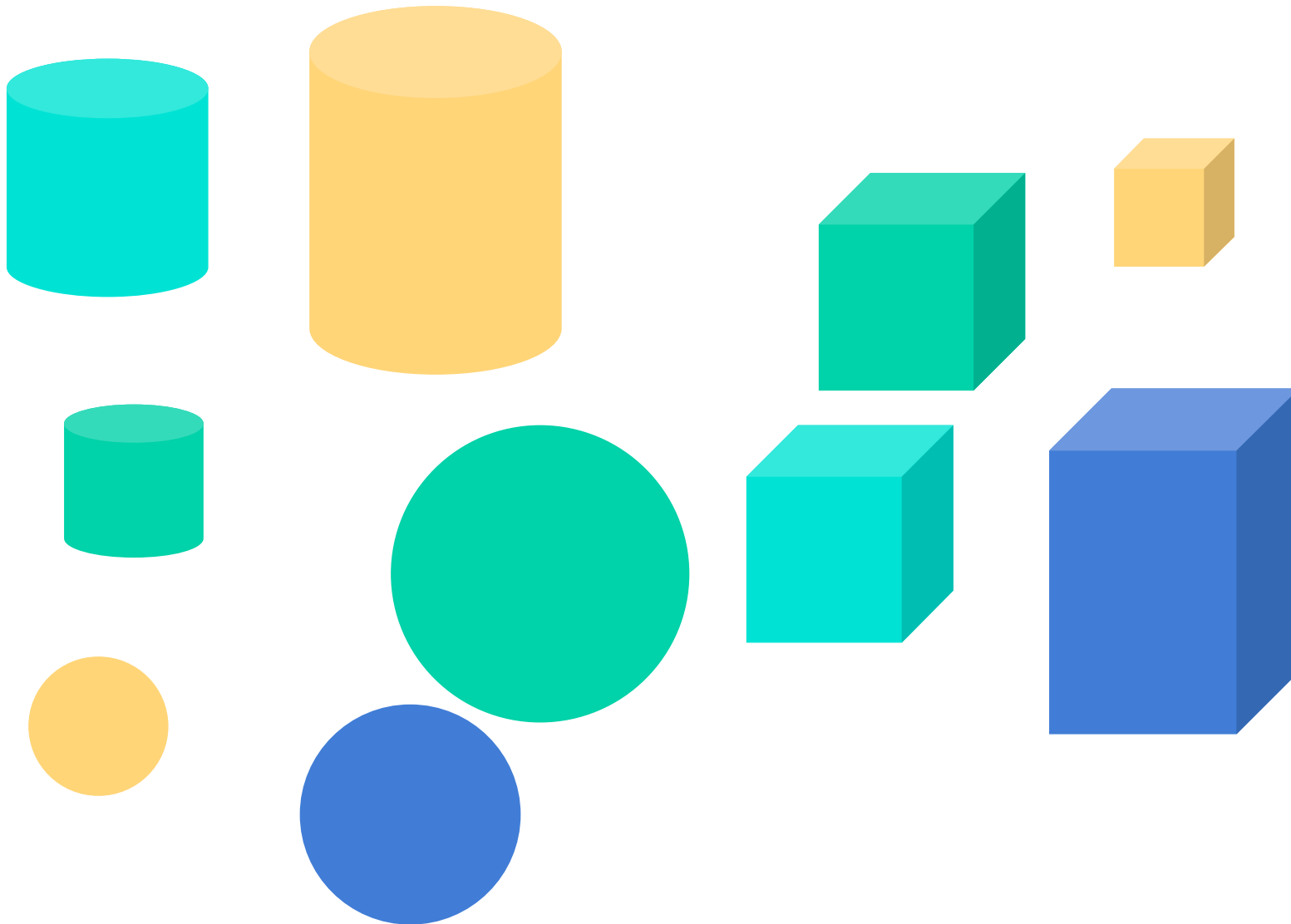
Merkmale

Das Wort '**Merkmal**' bedeutet im Prinzip soviel wie 'Eigenschaft' und bezieht sich auf die individuellen **Attribute**, die ein bestimmtes Objekt aufweist, z.B. **Farbe**, **Form**, **Größe**, **Aggregatzustand** (fest, flüssig, gasförmig) etc...

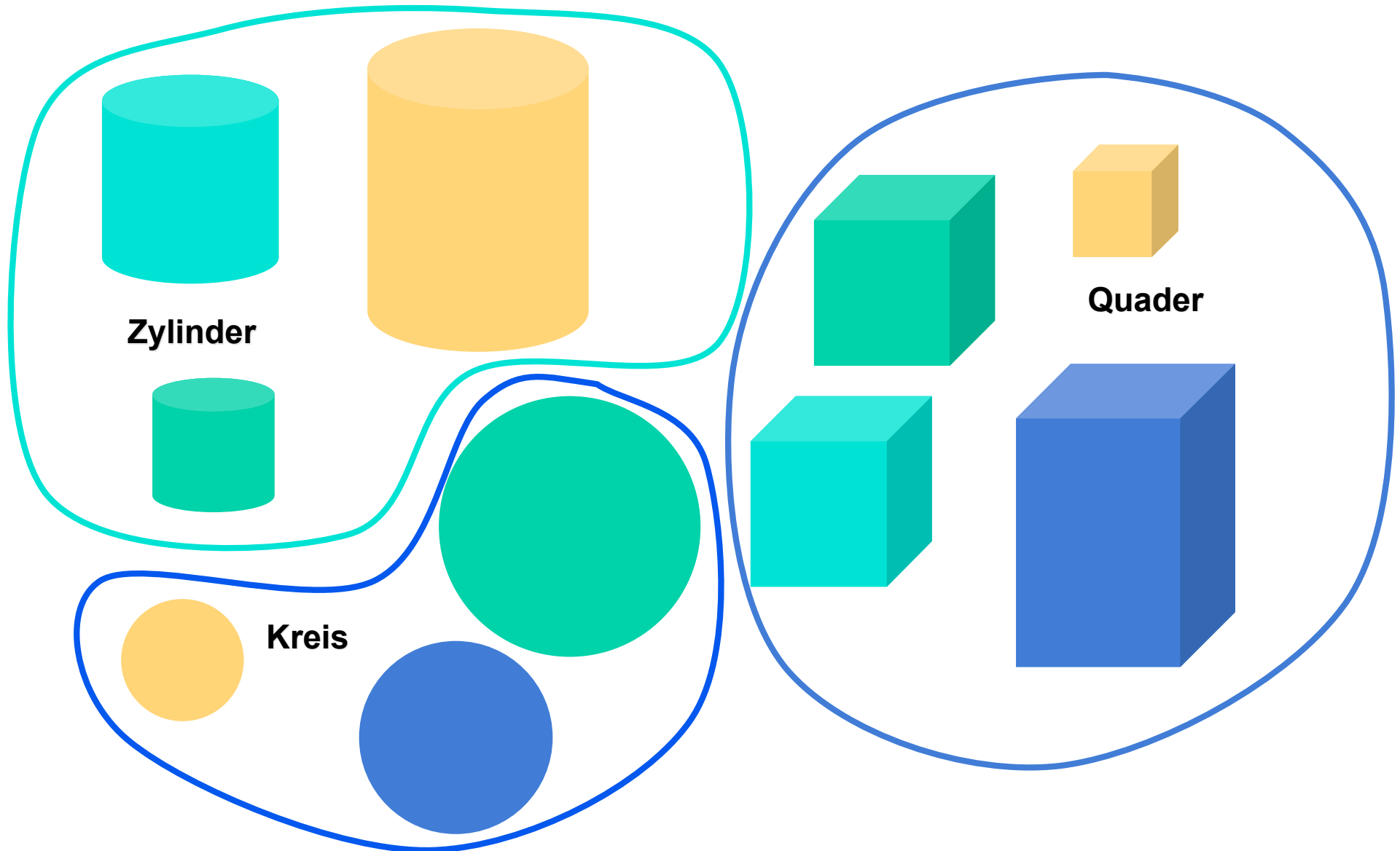
Objekte können über Mengen von Merkmalen beschrieben werden, und wenn bestimmte Objekte dieselben Merkmale aufweisen, kann man sie möglicherweise in einer Klasse zusammenfassen.

Die Verwendung von Merkmalen erlaubt es, Objekte miteinander zu vergleichen und sie dabei entweder voneinander zu differenzieren oder Klassen von Objekten mit denselben Merkmalen zu bilden.

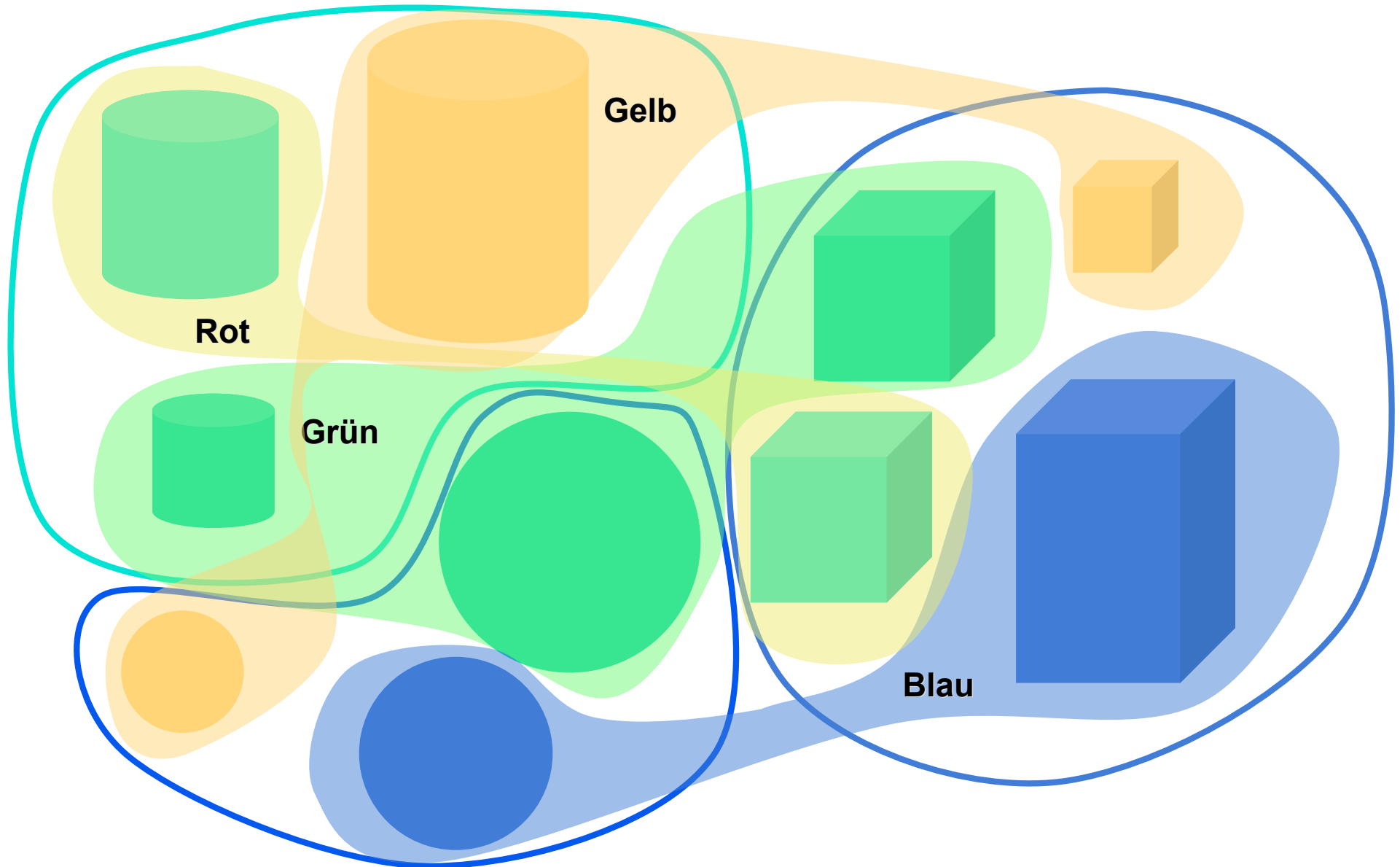
Merkmale von Objekten



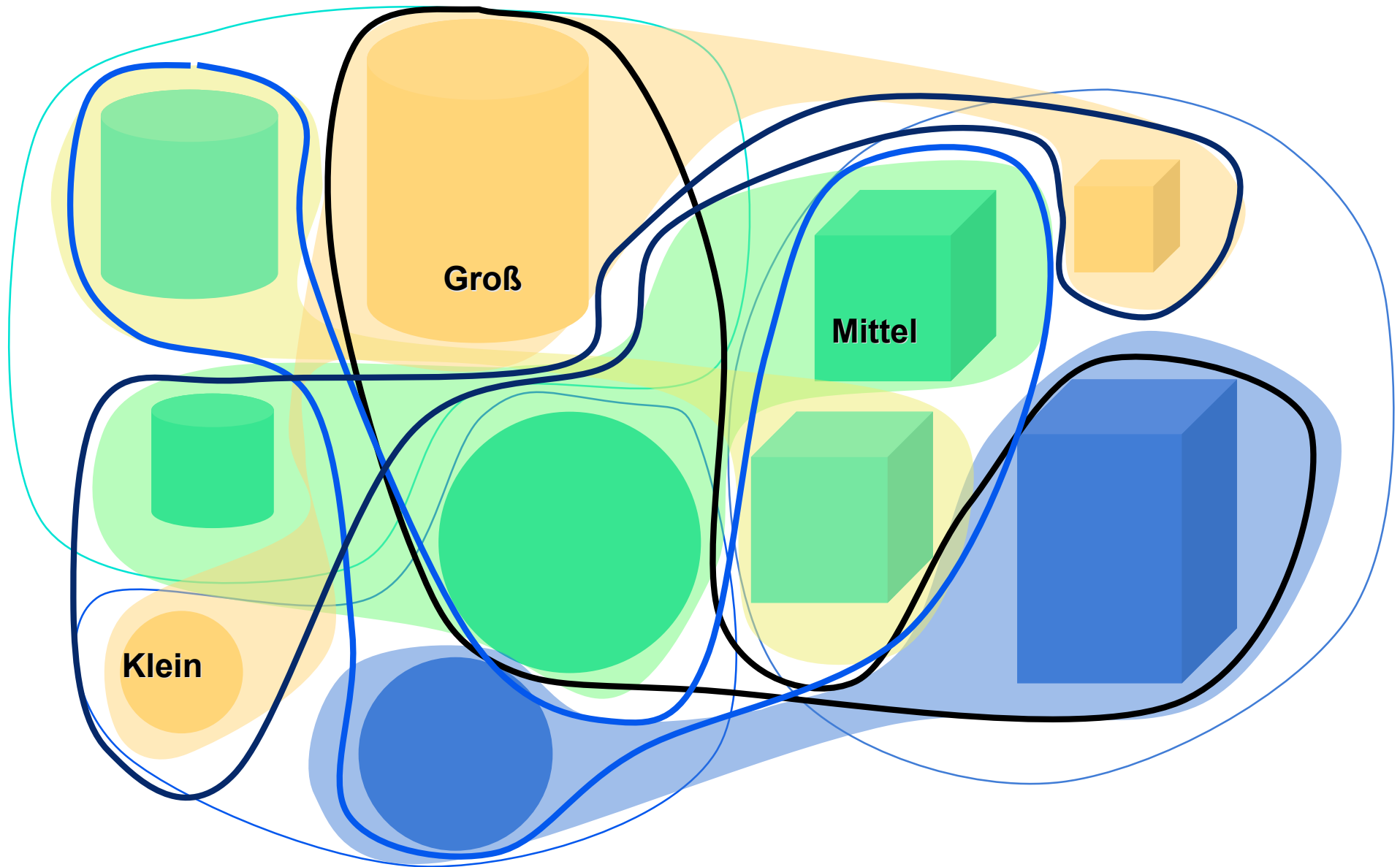
Merkmale von Objekten: Form



Merkmale von Objekten: Farbe



Merkmale von Objekten: Größe



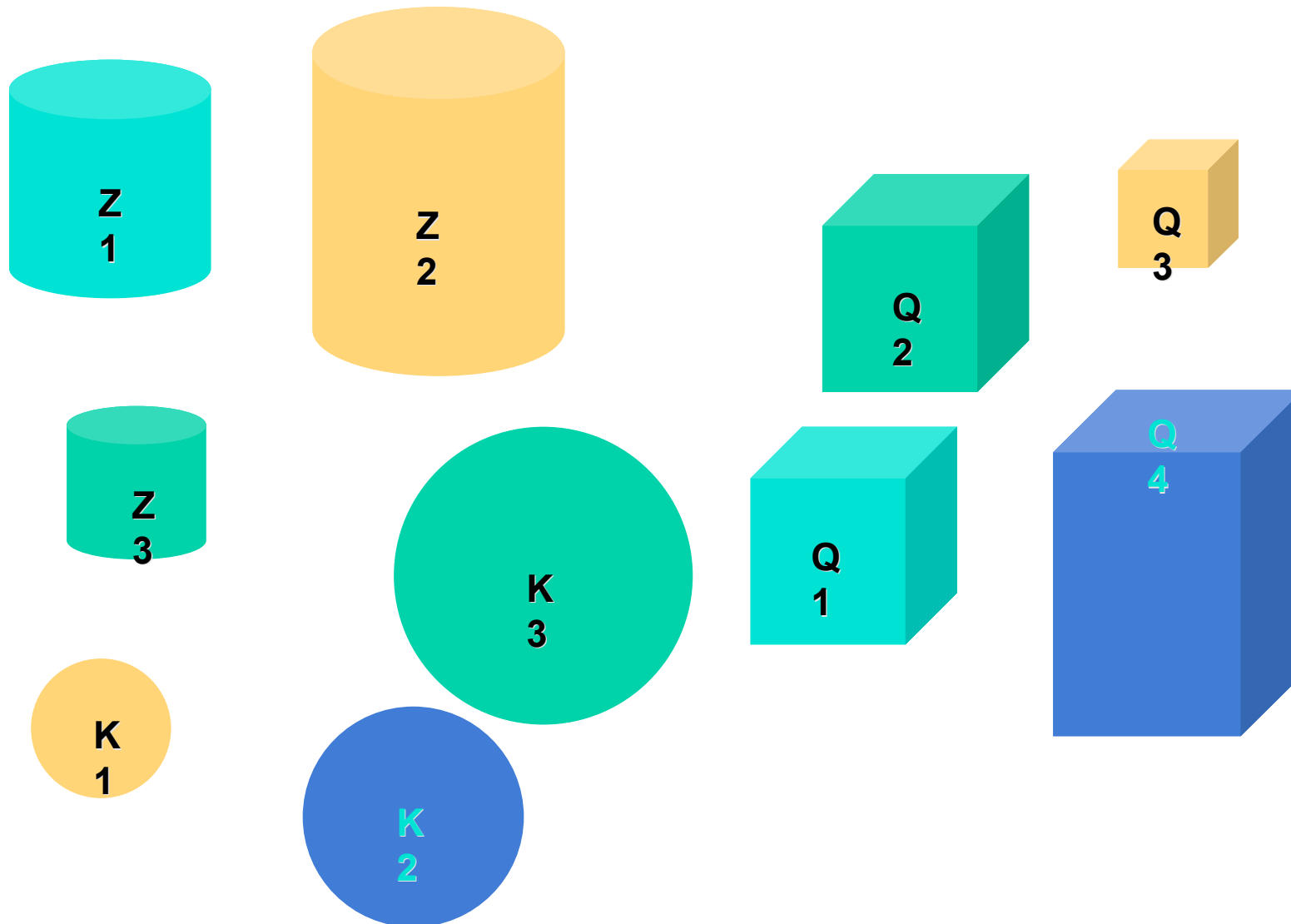
Objekte, Attribute, Werte

Objekte sind entweder physische Entitäten oder begriffliche Einheiten.

Attribute sind allgemeine Charakteristika oder Eigenschaften, die mit Objekten assoziiert werden. **Größe**, **Form** und **Farbe** sind typische Attribute von physischen Objekten.

Der **Wert** eines Attributs kennzeichnet die spezifische Beschaffenheit (Ausprägung) eines Attributs in einer bestimmten Situation.

Merkmale von Objekten



Objekt-Attribut-Wert-Tripel: Beispiele

Objekt

Attribut

Wert

Z1

Farbe

rot

Z1

Form

Zylinder

Z1

Größe

mittel

Q3

Farbe

gelb

Q3

Form

Quader

Q3

Größe

klein

K3

Farbe

grün

K3

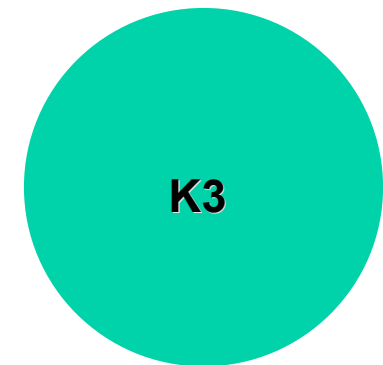
Form

Kreis

K3

Größe

groß



Merkmale in der Linguistik

Auch in der Linguistik dienen Merkmale dazu, Objekte zu charakterisieren und Klassen von Objekten zu bilden. Dabei ist klar, daß die Merkmale und die Objekte, um die es geht, im weitesten Sinne linguistischer Natur sind.

In den folgenden Abschnitten soll dargestellt werden, wie solche Merkmale aufgebaut sind, und welche Funktionen sie für die Repräsentation von linguistischer Information auf den verschiedenen sprachlichen Ebenen ([Phonetik](#), [Morphologie](#), [Syntax](#), [Semantik](#)) haben können.

Objekt-Attribut-Wert-Tripel: Beispiele

Objekt	Attribut	Wert
<i>Kindes</i>	Kategorie	Nomen
<i>Kindes</i>	Genus	Neutrum
<i>Kindes</i>	Numerus	Singular
<i>Kindes</i>	Kasus	Genitiv
<i>Kindes</i>	Person	3

Objekt-Attribut-Wert-Tripel → Attribut-Wert-Paar

Objekt

Attribut

Wert

Kindes

Kategorie

Nomen

Genus

Neutrum

Numerus

Singular

Kasus

Genitiv

Person

3

Attribut-Wert-Paare

Objekt

Kindes

Attribut

Kategorie

Genus

Numerus

Kasus

Person

Wert

Nomen

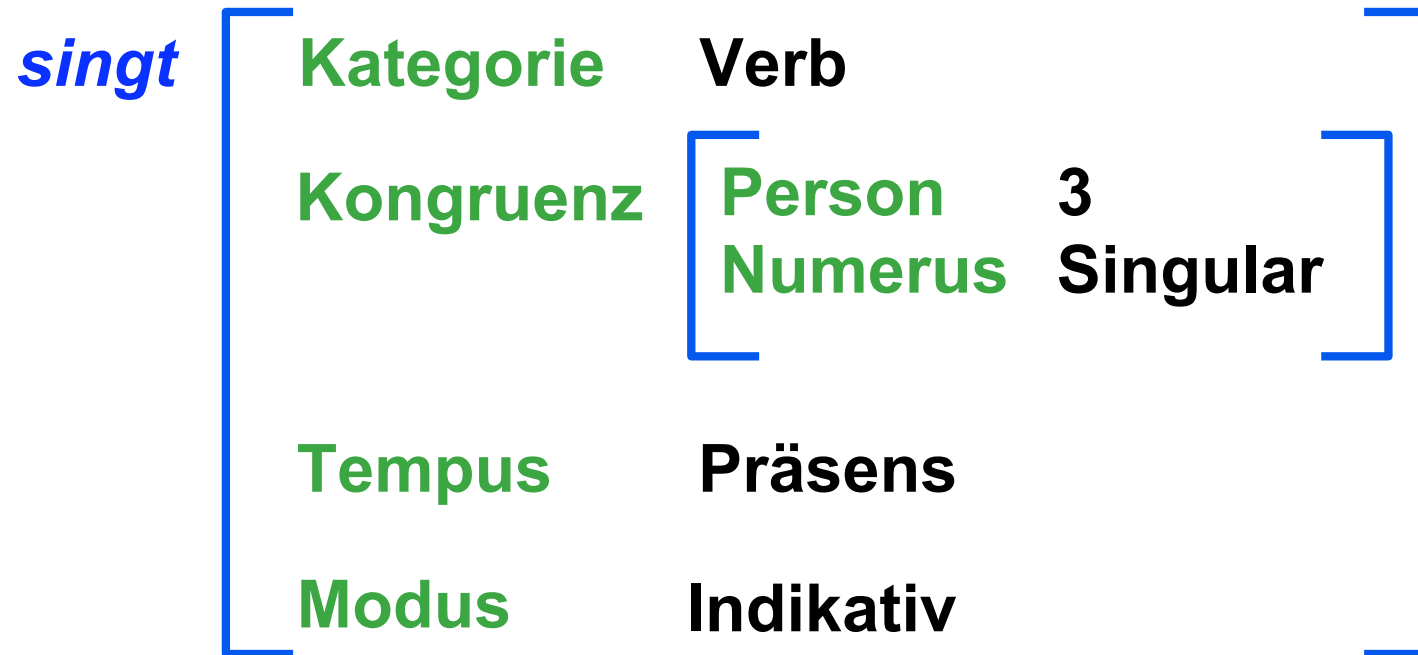
Neutrum

Singular

Genitiv

3

Attribut-Wert-Paare: Merkmalstrukturen



Merkmale in der Morphologie

Dieses der Wortform *singt* zugeordnete Merkmalsbündel besteht aus Angaben über die lexikalische **Kategorie** (also die Wortart, hier: Verb) der entsprechenden Wortform, und über deren Markierung hinsichtlich **Person**, **Numerus**, **Tempus** und **Modus**.

Merkmale in der Morphologie

Ein Merkmal wie beispielsweise [Kasus: Genitiv], welches man sich gut für ein Substantiv wie z.B. *Wetters* in *wegen des schlechten Wetters* vorstellen kann, besteht aus der Zuordnung eines Attributes und eines bestimmten Wertes für dieses Attribut. Dieser Wert entstammt einem festgelegten Wertebereich oder Wertevorrat.

Für das Attribut **Kasus** umfaßt der Wertebereich in der deutschen Sprache die Elemente {Nominativ, Akkusativ, Genitiv, Dativ}, für das Attribut **Numerus** umfaßt der Wertevorrat die Elemente {Singular, Plural} usw.

Attribute – Wertebereiche

Attribut	Wertebereich
Kategorie	{Nomen (N), Verb (V), Adjektiv (A), Präposition (P), ...}
Person	{1, 2, 3}
Numerus	{Singular, Plural, ...}
Kasus	{Nominativ, Akkusativ, Genitiv, Dativ}
Genus	{Maskulin, Feminin, Neutrum}
Tempus	{Präsens, Präteritum, Futur}
Modus	{Indikativ, Konjunktiv, Optativ}
Aspekt	{perfektiv, imperfektiv}
Diathese	{aktiv, passiv}