

Prosodie

- Das Sprachsignal enthält zusätzlich zur „segmentalen Struktur“, d.h. Information über die Abfolge von Lauten, die Wörter identifizieren, „suprasegmentale“ oder „prosodische“ Information:
 - Sprechgeschwindigkeit, Rhythmus, Pausen
 - Akzent
 - Intonation
- Funktionen prosodischer Information:
 - Gliederung des Satzes
 - Satzmodus (Aussage, Frage, Befehl)
 - Text- und Dialogkohärenz, Informationsstruktur

Sprachsynthese

- Sprachsynthese erscheint einfacher als Spracherkennung. Trotzdem gibt es bisher keine Vollsynthese-Systeme mit einer Qualität, die an menschliche Sprecher heranreicht.
- Einige Probleme der Sprachsynthese:
 - Orthografie (*beinhalten, Staubecken*)
 - Fremdsprachige Wörter, Zahlen, Sonderzeichen (→ Lektüretext)
 - Prosodie: Die richtige Rhythmik und Intonation setzt linguistisches und konzeptuelles Wissen voraus (→ Übung 3.1)

Sprachsynthese: Alternative Techniken

- Wortkonkatenation/ voraufgenommene Sprachausgabe: Wörter bzw. Äußerungen werden mit Sprechern voraufgenommen und geschnitten. Wörter werden bei der Synthese zusammengehängt, bzw. in ein Äußerungsmuster eingehängt und prosodisch modifiziert. – Beispiel für Konkatenation:

Der Flug | Lufthansa | LH | 4 | 7 | 5 | 2 | um | 11 Uhr |
35 | aus | Dresden | ist um | 2 Stunden | und 25 Minuten |
verspätet.

- Synthese: Vollsynthese von Wörtern aus einzelnen Lauten ist unmöglich, da scharfe Begrenzung zwischen Lauten nicht existiert. Diphon-Synthese (vereinfacht): Alle möglichen Kombinationen von zwei Lauten werden voraufgenommen, geschnitten, nachbearbeitet.
 - Web-Demo für eine gute deutsche Synthese (AT&T):
www.research.att.com/~ttsweb/cgi-bin/ttsdemo/

Morphologie

- Morphologie ist der Teilbereich der Linguistik, der sich mit der internen Struktur von Wörtern befasst.
- Die wesentlichen Aufgaben der Morphologie in der Computerlinguistik sind
 - die Reduktion komplexer Wörter bzw. Wortformen auf ihre Bestandteile
 - die Identifikation von grammatischer Information, die in der Wortform kodiert ist (z.B. Kasus, Numerus, Tempus)

Anmerkung: Die folgende Übersicht über morphologische Phänomene ist stichwortartig. Mehr Information zur Morphologie gibt es in der Einführungsvorlesung Sprachwissenschaft.

Teilbereiche der Morphologie

- Flexionsmorphologie:
Deklination, Konjugation von Substantiven, Verben,
Adjektiven, Pronomina
frag+te+st
ge+frag+t
- Derivationsmorphologie (Ableitung)
Er+kenn+ung
- Komposition (Zusammensetzung)
Sprach+erkennung+s+technik

Elemente der morphologischen Struktur

- Stämme, Präfixe, Suffixe (in Flexion und Derivation)
 - Flexionsmorphologie:
frag+te+st
ge+frag+t
 - Derivationsmorphologie:
Er+kenn+ung

Elemente der morphologischen Struktur

- Stämme, Präfixe, Suffixe (in Flexion und Derivation)
 - Flexionsmorphologie:
frag+te+st
ge+frag+t
 - Derivationsmorphologie:
Er+kenn+ung

Elemente der morphologischen Struktur

- Stämme, Präfixe, Suffixe (in Flexion und Derivation)
 - Flexionsmorphologie:
frag+te+st
ge+frag+t
 - Derivationsmorphologie:
Er+kenn+ung

Elemente der morphologischen Struktur

- Grund- , Bestimmungswörter, Fugenelemente (in der Komposition)

Sprach+erkennung+s+technik

Sprach+erkennung+s+technik

- Bestimmungswort: *Sprach* +
Grundwort: *erkennung*
- Bestimmungswort: *Spracherkennung* +
Fugenelement: *s*
Grundwort: *technik*

Elemente der morphologischen Struktur

- Stämme, Präfixe, Suffixe (in Flexion und Derivation)
- Grund- , Bestimmungswörter, Fugenelemente (in der Komposition)
- Modifikation von Stämmen (Umlaut, Ablaut)
 - *Mutter, Mütter*
 - *schwimmen, schwamm, geschwommen*

Elemente der morphologischen Struktur

- Stämme, Präfixe, Suffixe (in Flexion und Derivation)
- Grund- , Bestimmungswörter, Fugenelemente (in der Komposition)
- Modifikation von Stämmen (Umlaut, Ablaut)
 - *Mutter, Mütter*
 - *schwimmen, schwamm, geschwommen*
- Morpho-phonologische Prozesse

Morpho-phonologische Prozesse

- Systematische Modifikation, Einfügung und Tilgung von Lauten/Phonemen aus phonetisch-phonologischen Gründen

Einfügung: $ba\underline{d} + \underline{s}t \rightarrow ba\underline{de}st$

Tilgung: $ra\underline{s} + \underline{s}t \rightarrow ra\underline{st}$

Modifikation (phonetisch): $[ba\underline{t}] + [\underline{e}s] \rightarrow [ba\underline{de}s]$

Modifikation (orthografisch, alt): $na\underline{\beta} + \underline{e} \rightarrow na\underline{sse}$

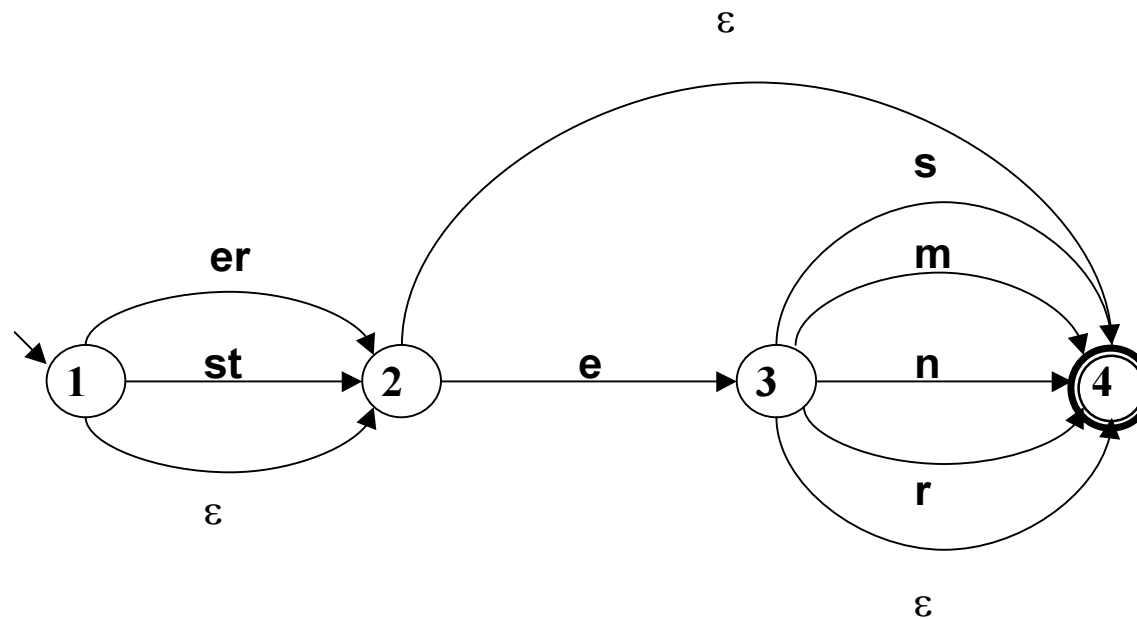
Morphologische Verarbeitung in der Computerlinguistik

- Systeme zur morphologischen Analyse in der Computerlinguistik arbeiten meist auf der orthografischen Struktur (z.B. für Rechtschreib- und Grammatikkorrektur)
- Teilaufgaben der morphologischen Verarbeitung:
 - Lemmatisierung: Flexionsmorphologische Analyse: Ermittlung des Stammes/Lemmas („stemming“) und ggf. der in den Flexionsformen enthaltenen grammatischen Information
 - Derivativ- und Komposita-Zerlegung: Reduktion komplexer Wörter (Ableitungen und Zusammensetzungen) auf ihre Bestandteile
- Methodisches Werkzeug für alle Aufgaben morphologischer Analyse sind „Endliche Automaten“.
- Wir betrachten die Verwendung endlicher Automaten an der vergleichsweise einfachen Teilaufgabe der Lemma-Ermittlung.

Adjektivflexion: Paradigma (nur sog. „starke Flexion“)

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

Adjektivendungen: Darstellung durch Zustandsdiagramm



Zustandsdiagramm

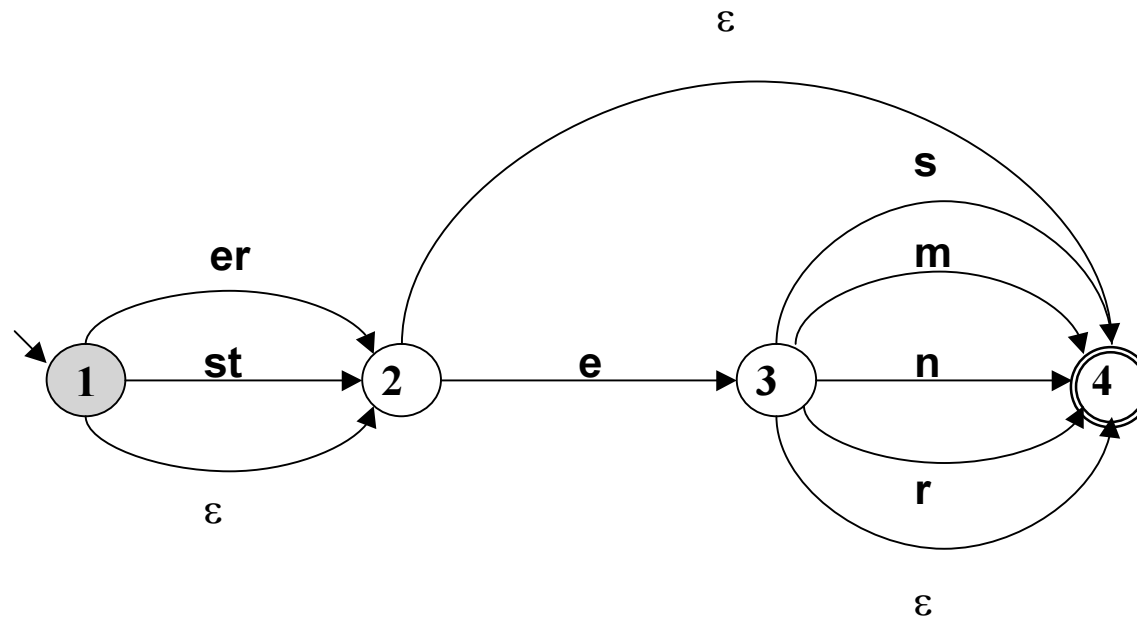
Ein Zustandsdiagramm besteht aus

- Knoten (Zuständen) (im Beispiel: 1,2,3,4)
- davon ein Startknoten (1)
- einem oder mehreren Zielknoten (4)
- Kanten zwischen den Knoten, die
 - gerichtet und
 - beschriftet sind
- Die Kanteninschriften bestehen aus Ketten von Symbolen über einem Alphabet (im Beispiel: e,r,m,n,s,t).
- Auch die leere Kette (ε) ist als Kantenbeschriftung zugelassen; ε ist kein Symbol des Alphabets, sondern bezeichnet den Grenzfall der „aus 0 Symbolen bestehenden“ leeren Kette.

Die Interpretation des Zustandsdiagramms

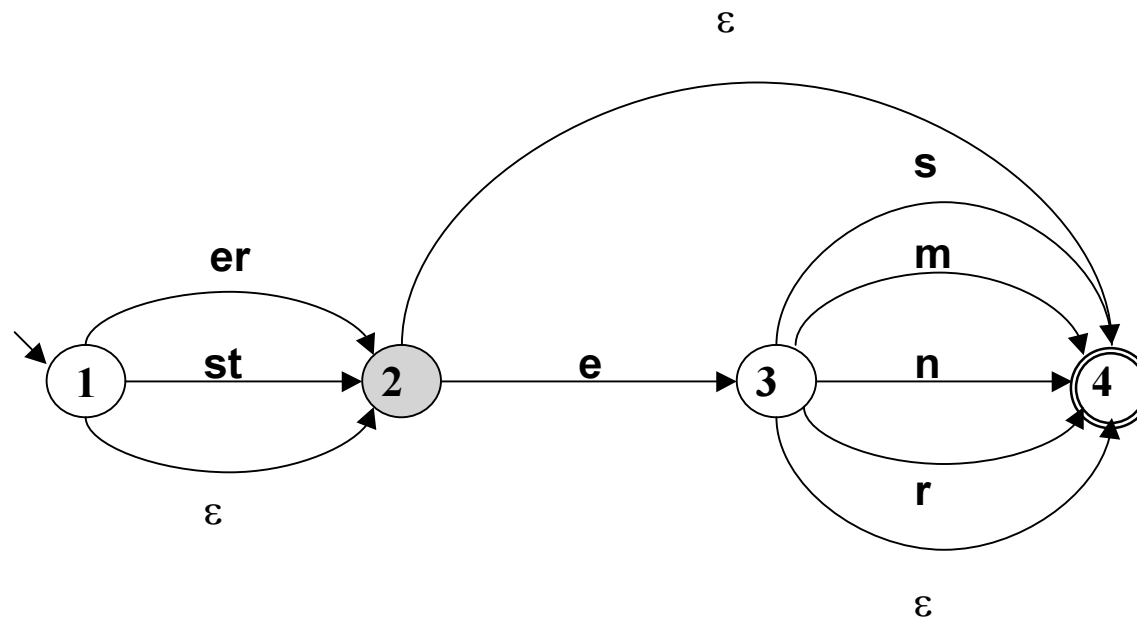
- Das Zustandsdiagramm beschreibt alle Symbolkombinationen oder „Worte“, die sich dadurch ergeben, dass man das Diagramm vom Startknoten zu einem Zielknoten durchläuft und die Inschriften der Kanten, die man dabei beschreitet, aufliest und aneinanderhängt. Man nennt die Menge der Worte, die sich so erzeugen lassen, die vom Diagramm beschriebene „Sprache“.
- Üblicherweise werden Diagramme verwendet, um Eingabeketten zu testen. Man spricht von endlichen Automaten, und sagt, dass ein Automat ein Wort akzeptiert.

Funktion des Zustandsdiagramms [1]



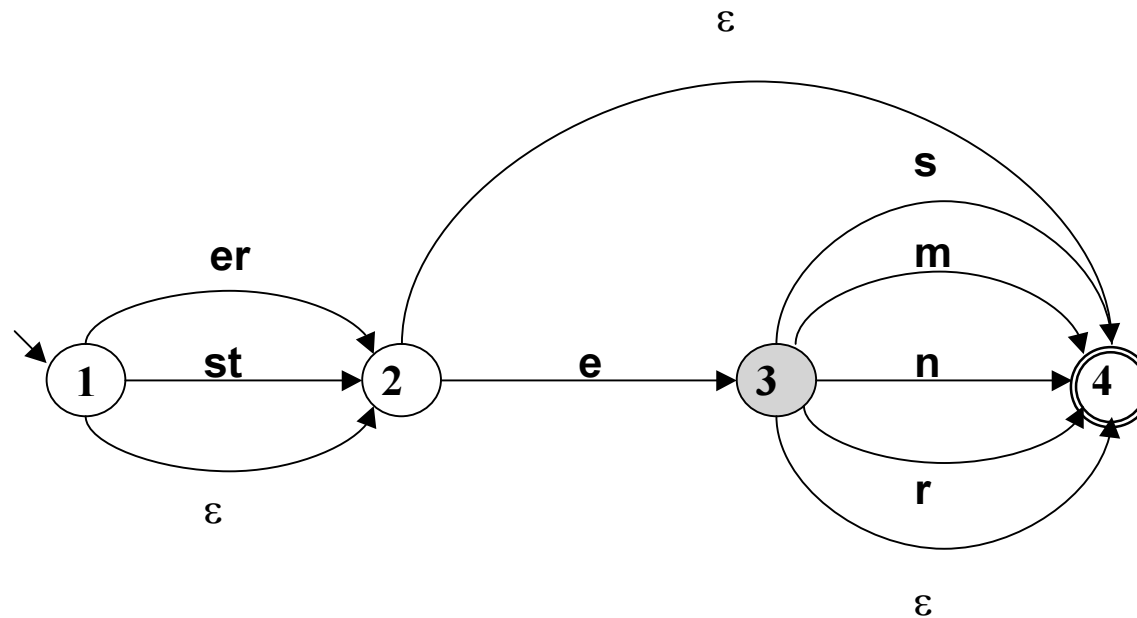
klein eres

Funktion des Zustandsdiagramms [2]



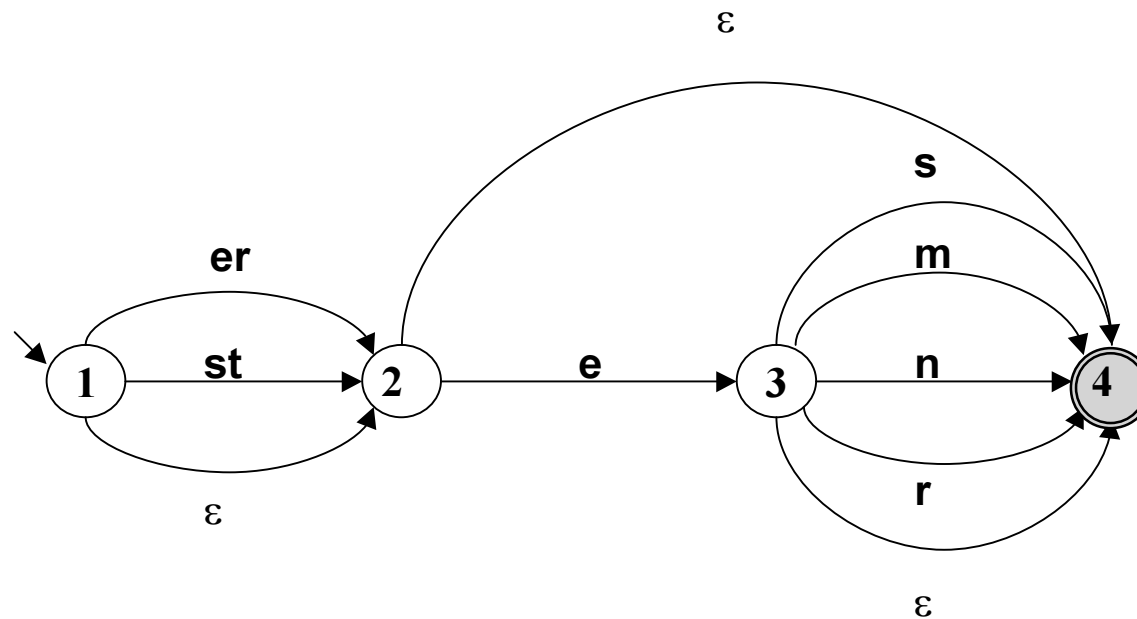
klein eres

Funktion des Zustandsdiagramms [3]



klein eres

Funktion des Zustandsdiagramms [4]



klein eres_

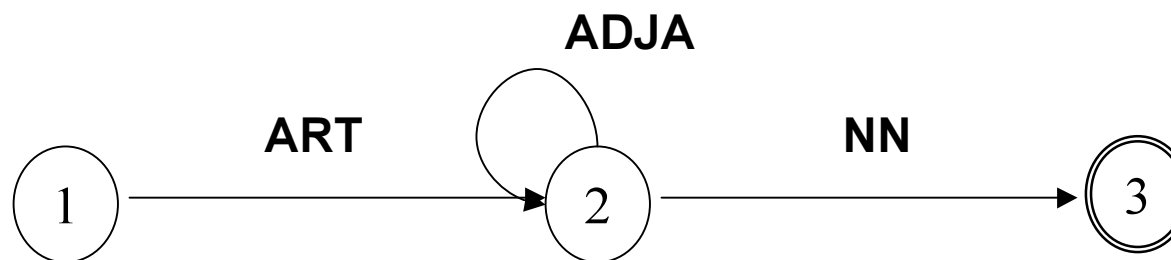
Zustandsdiagramme: Ein zweites Beispiel [1]

- Wortarten kombinieren sich in bestimmter Weise zu Satzteilen
- Um zu testen, ob eine Wortfolge in einem Dokument eine erlaubte Abfolge von Wortarten darstellt, können Zustandsdiagramme benutzt werden.
- Das folgende Zustandsdiagramm akzeptiert bestimmte einfache Nominalausdrücke, wie

der Wagen

eine interessante Vorlesung

das neue schöne rote Dach



Zustandsdiagramme: Ein zweites Beispiel [2]

- Das „Alphabet“ des Zustandsdiagramms sind Wortart-Bezeichnungen („Wortart-Tags“ oder auch „POS-Tags“, POS für „part of speech“, engl. für „Wortart“), in unserem Beispiel Art, ADJA, NN
 - ART (Artikel)
 - ADJA (adjektivisches Attribut)
 - NN (Gattungssubstantiv, Gattungsnomen)
- Erkannte „Worte“ sind erlaubte Abfolgen von Wortartsymbolen, z.B. „ART NN“, „ART ADJA NN“, „ART ADJA ADJA ADJA NN“
- Im Gegensatz zum Adjektivendungsdiagramm akzeptiert das Nominalausdrucksdiagramm beliebig lange Worte und beschreibt eine unendliche Sprache. Grund: Es enthält eine Schleife, es ist zyklisch.

Zustandsdiagramme: Ein zweites Beispiel [3]

Das abgebildete Diagramm akzeptiert auch Adjektive, die mit (einer oder mehreren) Gradpartikeln (GPRT) versehen sind, wie z.B.

eine ziemlich interessante Vorlesung

das recht neue sehr sehr schöne rote Dach

