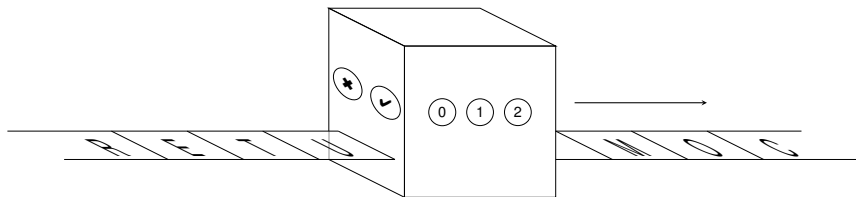


Abschnitt 4

Zwei-Ebenen-Morphologie

Wiederholung: Endlicher Automat



Wiederholung: Endlicher Automat (Definition)

Ein nicht-deterministischer **endlicher Automat** ist ein 5-Tupel $(Q, \Sigma, \delta, q_0, F)$ mit

- einer **endlichen** Menge von Zuständen Q ,
- einem **endlichen** Eingabealphabet Σ ,
- einer Zustandsübergangsfunktion $\delta : Q \times \Sigma \rightarrow \wp(Q)$, die das Steuerungsverhalten des Automaten bestimmt
- **einem** Startzustand $q_0 \in Q$
- einer Menge von Endzuständen $F \subseteq Q$

Endlicher Transduktor (I)

Ein nicht-deterministischer endlicher **Transduktor** ist ein 5-Tupel $(Q, \Sigma, \delta, q_0, F)$ mit

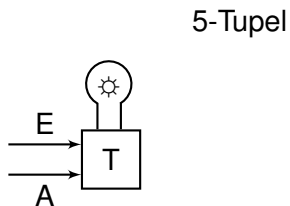
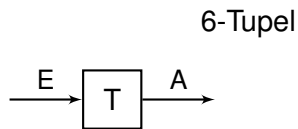
- einer **endlichen** Menge von Zuständen Q ,
- einem **endlichen** Eingabealphabet $\Sigma \subseteq I \times O$ bestehend aus komplexen Symbolen, die aus Eingabe-Ausgabe-Paaren $i : o$ bestehen, mit $i \in I$ (Eingabealphabet) und $o \in O$ (Ausgabealphabet),
- einer Zustandsübergangsfunktion $\delta : Q \times \Sigma \rightarrow \wp(Q)$, die das Steuerungsverhalten des Automaten bestimmt,
- **einem** Startzustand $q_0 \in Q$
- einer Menge von Endzuständen $F \subseteq Q$

Endlicher Transduktor (II)

Ein nicht-deterministischer endlicher **Transduktor** ist ein 6-Tupel $(Q, \Sigma, \Delta, \delta, q_0, F)$ mit

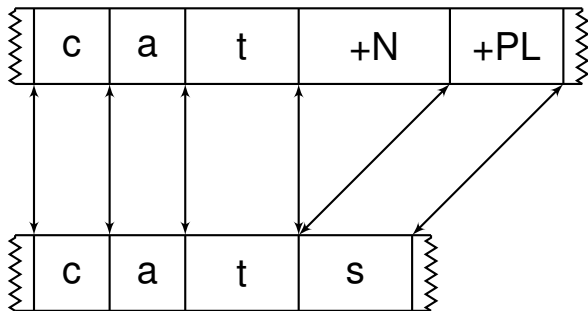
- einer **endlichen** Menge von Zuständen Q ,
- einem **endlichen** Eingabealphabet Σ ,
- einem endlichen Ausgabealphabet Δ ,
- einer Zustandsübergangsfunktion $\delta : Q \times \Sigma \rightarrow \wp(Q \times \Delta^*)$, die das Steuerungsverhalten des Automaten bestimmt
- **einem** Startzustand $q_0 \in Q$
- einer Menge von Endzuständen $F \subseteq Q$

Konzeptioneller Vergleich von 5- und 6-Tupel-Definition

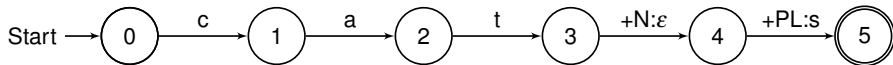


Anwendung

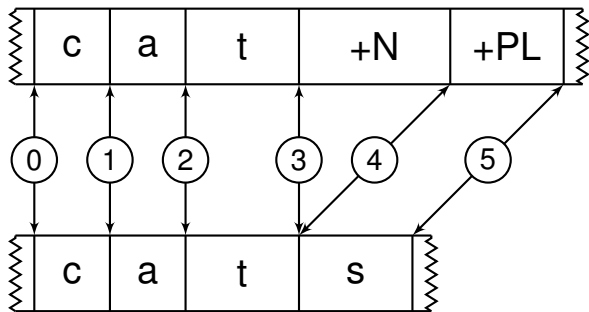
- Morphologisches Parsing (\uparrow)
- Flexionsformgenerierung (\downarrow)



Transduktor zur Flexionsformgenerierung



Beispiellauf mit Konfigurationen des Transduktors



1. (0, cat +N:ε +PL:s)
2. (1, at +N:ε +PL:s)
3. (2, t +N:ε +PL:s)
4. (3, +N:ε +PL:s)
5. (4, +PL:s)
6. (5, ε)

Trie

- Wird “am Anfang” eines Transduktors verwendet, damit dieser beliebige Wörter erkennen/umwandeln kann (Lexikon)
- Ein Trie (sprich wie engl. *try*) ist (im Kontext dieses Moduls) ein Automat zum Abspeichern von Wörtern natürlicher Sprachen
- Wesentliche Merkmale:
 - Bei jedem Übergang wird nur ein Zeichen gelesen
 - Mehrfach vorkommende Prä- und Suffixe teilen sich die jeweiligen Zustände

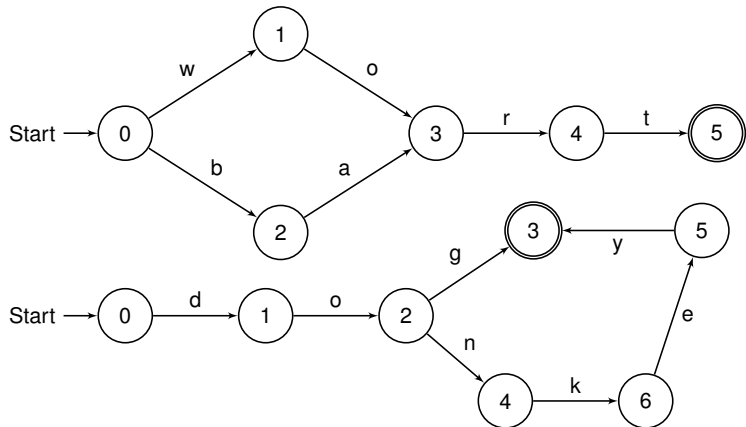
Trie

- Wird “am Anfang” eines Transduktors verwendet, damit dieser beliebige Wörter erkennen/umwandeln kann (Lexikon)
- Ein Trie (sprich wie engl. *try*) ist (im Kontext dieses Moduls) ein Automat zum Abspeichern von Wörtern natürlicher Sprachen
- Wesentliche Merkmale:
 - Bei jedem Übergang wird nur ein Zeichen gelesen
 - Mehrfach vorkommende Prä- und Suffixe teilen sich die jeweiligen Zustände

Trie

- Wird “am Anfang” eines Transduktors verwendet, damit dieser beliebige Wörter erkennen/umwandeln kann (Lexikon)
- Ein Trie (sprich wie engl. *try*) ist (im Kontext dieses Moduls) ein Automat zum Abspeichern von Wörtern natürlicher Sprachen
- Wesentliche Merkmale:
 - Bei jedem Übergang wird nur ein Zeichen gelesen
 - Mehrfach vorkommende Prä- und Suffixe teilen sich die jeweiligen Zustände

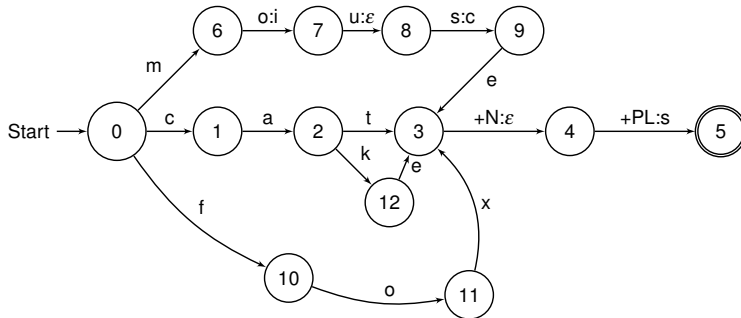
Trie (Beispiele)



Aufgabe zu Tries

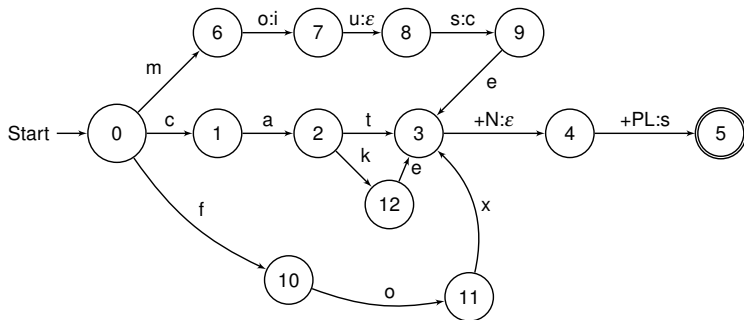
Geben Sie einen Trie an, der die Wörter “Pappe”, “Papier” und “Papa” kodiert.

Flexionsformgenerierung mit Transduktor und Trie



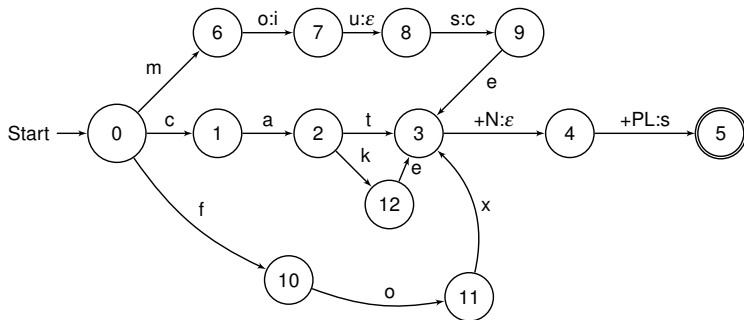
- Transduktor zur Pluralbildung mit eingebautem Trie
- Aber: Problematische Phonologie-/Orthographie-Regeln
 - fox +N +PL ist *foxes*, nicht **foxs*
 - city +N +PL ist *cities*, nicht **citys*
- Lösung: Zwei-Ebenen-Morphologie

Flexionsformgenerierung mit Transduktor und Trie



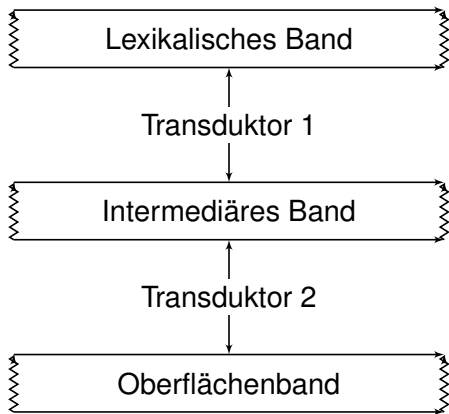
- Transduktor zur Pluralbildung mit eingebautem Trie
- Aber: Problematische Phonologie-/Orthographie-Regeln
 - fox +N +PL ist *foxes*, nicht **foxs*
 - city +N +PL ist *cities*, nicht **citys*
- Lösung: Zwei-Ebenen-Morphologie

Flexionsformgenerierung mit Transduktor und Trie



- Transduktor zur Pluralbildung mit eingebautem Trie
- Aber: Problematische Phonologie-/Orthographie-Regeln
 - fox +N +PL ist *foxes*, nicht **foxs*
 - city +N +PL ist *cities*, nicht **citys*
- Lösung: Zwei-Ebenen-Morphologie

Schema der Zwei-Ebenen-Morphologie

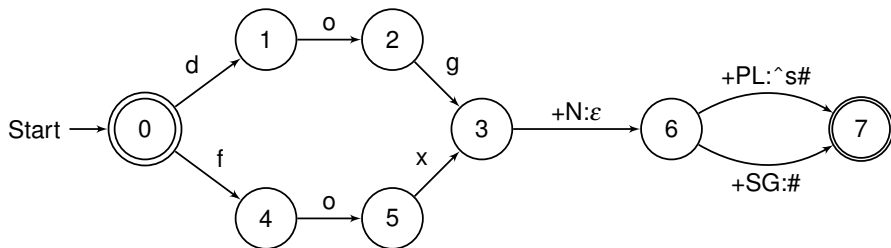


- Intermediäres Band: Zwischenrepräsentation, die Morphem- und Wortgrenzen explizit markiert

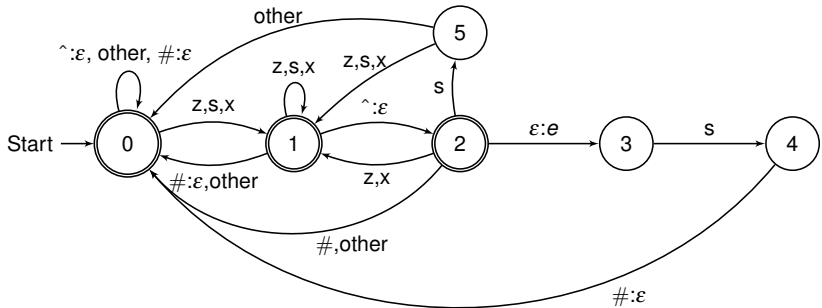
Schreibkonventionen für Zwei-Ebenen-Morphologie

$\alpha : \gamma$	Paare, sodass α von Eingabe und γ von Ausgabealphabet ist
α	Identische (default) Paare der Form $\alpha : \alpha$
\wedge	Morphemgrenze
$\#$	Wortgrenze
other	Ein beliebiges, sonst nicht benutztes Zeichen
@	beliebiges Symbol

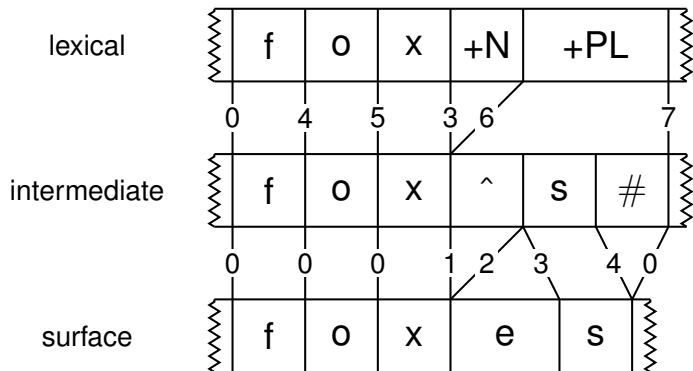
Zwei-Ebenen-Morphologie für e-Insertion: Lexikon-Intermediär-Transduktor



Zwei-Ebenen-Morphologie für e-Insertion: Intermediär-Oberflächen-Transduktor



Zwei-Ebenen-Morphologie für e-Insertion: Beispiellauf



Übung: ZEM für k-Insertion

panic +V +G \mapsto panicking

panic +V +P \mapsto panicked

traffic +V +P \mapsto trafficked

Aber auch: flicker +V +P \mapsto flickered