

Kétszintű morfológia

Rebrus Péter

Számítógépes nyelvészet kurzus
ELTE Elméleti Nyelvészet
2022. 03. 21.

Vázlat

- Eredet
 - Generatív nyelvészet
 - Fonológiai szabályrendszer (SPE)
 - Véges állapotú automata (FSA)
- Architektúra
 - Transzducer (FST)
 - Példák

Generatív grammatika

Cél: a grammatikus/jólformált mondatok megadása formális eszközökkel

A modell feltételezései (egyszerűsítések)

- a jelentésnek nincs szerepe a grammatikalitásban – vitatott
- a mondatok véges számú alapelemből (szó/morféma) állnak és minden mondat véges hosszúságú – általánosan elfogadott
- egy nyelvben véges sok nyelvtani szabály van – vitatott
- a mondatok hossza és az iteratív nyelvtani szabályok alkalmazásának a száma nincs korlátozva – vitatott
- a nyelvi szerkezetek diszkrét, nem-sztocasztikus modellel leírhatók – vitatott

Formális alapfogalmak 1

- **Sztring**: véges hosszúságú szimbólumsorozat
 - pl. a , ab , aaa , $abcca$ stb.;
 - **üres sztring**: λ
- **Konkatenáció**: sztringek összefűzése
 - pl. $a^b = ab$, $a^ba = ab^a = aba$, $a^a = aa$, $a^\lambda = \lambda^a = a$
- **Ábécé** (V): szimbólumok véges halmaza, pl.
 - absztrakt szimbólumok: $\{a, b, c\}$, $\{a\}$, $\{a_1, a_2, \dots, a_n\}$
 - számok: $\{0, 1, 2, \dots, 9\}$, $\{0, 1\}$; számok és műveleti jelek: $\{0, 1, \dots, 9, +, \cdot, -, :\}$
 - logikai nyelv szimbólumai: $\{p, q, \neg, \&, (,)\}$;
 - egyéb szimbólumok: $\{., -, / \}$; $\{A, G, C, T\}$
 - a természetes nyelv alapelemei, pl.: $\{\text{összes magyar szó/morféma}\}$, $\{\text{összes magyar hang}\}$

Formális alapfogalmak 2

- **Sztringhalmazok konkatenációja** (AB): a két halmaz elemeiből előálló összes sztring, pl.
 - $\{a,b\}\{c\} = \{ac, bc\}$, $\{c\}\{a,b\} = \{ca, cb\}$, $\{a,b\}\{a,b\} = \{aa, ab, ba, bb\}$
 - $N\{\text{"ként"}\} = \{\text{összes főnév essivus-formalisban}\}$,
 - $NN = \{\text{összes elm. lehetséges kéttagú főnévi összetétel}\}$
 - $CC = \{\text{összes elm. lehetséges msh-kapcsolat}\}$
- **Kleene-lezárás** (A^*): a halmaz elemeiből készíthető összes sztring akárhány konkatenációval
 - $\{a, b\}^* = \{\lambda, a, b, aa, ab, ba, bb, aaa, aab, \dots\}$,
 - $\{a\}^* = \{\lambda, a, aa, aaa, aaaa, \dots\}$
 - $\{1, 2, \dots, 9\}^* = \{\text{az összes, a 0-t nem tartalmazó pozitív egész szám}\}$
 - $\{\text{összes szó}\}^* = \{\text{összes elméletileg lehetséges mondat (szósaláta)}\}$
- **Nyelv** (L): grammatikus sztringek: az ábécé Kleene-lezárásának egy részhalmaza: $L \subseteq V^*$
 - pl. $V = \{a,b\}$ esetén az 5-nél rövidebb sztringek vagy az ab^n , $(ab)^n$, $a^n b^n$ alakú sztringek
 - $\{\text{grammatikus mondatok}\} \subseteq \{\text{összes szó}\}^*$
 - $\{\text{grammatikus szavak}\} \subseteq \{\text{összes morféma}\}^*, \{\text{összes hang}\}^*$

Formális alapfogalmak 3

- **Újraírási szabály** (RWR): a bemeneti sztringet kicseréli a kimenetire, kétféle ábécé:
 - **nemterminális** (újraírható):, pl. $S \rightarrow NP VP$, $NP \rightarrow Det N$, $VP \rightarrow V NP$
 - **terminális** (kimeneti): $N \rightarrow kutyám, macskád$, $Det \rightarrow a, egy$, $V \rightarrow kergeti, szereti$
- **Nyelvtan** (generatív szabályrendszer): véges sok szabály
 - rekurzivitás: egy szabály kimenete ugyanazon/másik szabály bemenete lehet
- **Nyelvtan típusa** (3, 2, 1, 0): komplexitási sorrend
 - legelőnyösebb: **reguláris** nyelvtan (3), ld. véges állapotú automata (FSA)
 - szintaxis: ált. **környezetfüggetlen** (2), de vannak **környezetfüggő** (1) jelenségek
 - fonológia: a szabályok környezetfüggők (1), de!

Generatív fonológia (SPE)

- *Chomsky–Halle: Sound Pattern of English* (1968)
- Fonológiai szabály formátuma: $A \rightarrow B / C_D$
 - tipikusan környezetfüggő
 - A, B, C, D: hangok természetes osztálya jegyértékekkel kifejezve
 - pl. $\{p, t, k\} \rightarrow \{b, d, g\} / _ \{b, d, g\}$
 $[-zönge] \rightarrow [+zönge] / _ [+zönge]$
- Fonológiai szabályrendszer: rendezett szabályok sorozata, pl.
 - mögöttes alak:** /alma+t/ /elme+t/
 - szabály(ok): $\{a, e\} \rightarrow \{á, é\} / _+$
 - felszíni alak:** [almát] [elmét]

A fonológiai szabályrendszer komplexitása

- a fonológiai szabályok környezetfüggők (1-es típus)
- de ***nincs valódi rekurzivitás*** \Rightarrow nem környezetfüggő (Johnson 1972, Kaplan & Kay 1981)
- modellezhető ***input–output*** közötti ***reguláris relációval***
- automata-modellje: ***véges állapotú transzducer*** (FST)

mindkét irányban effektív:

- ***generálás***: input \rightarrow output
- ***elemzés***: output \rightarrow input

Szabályrendszer és transzducer

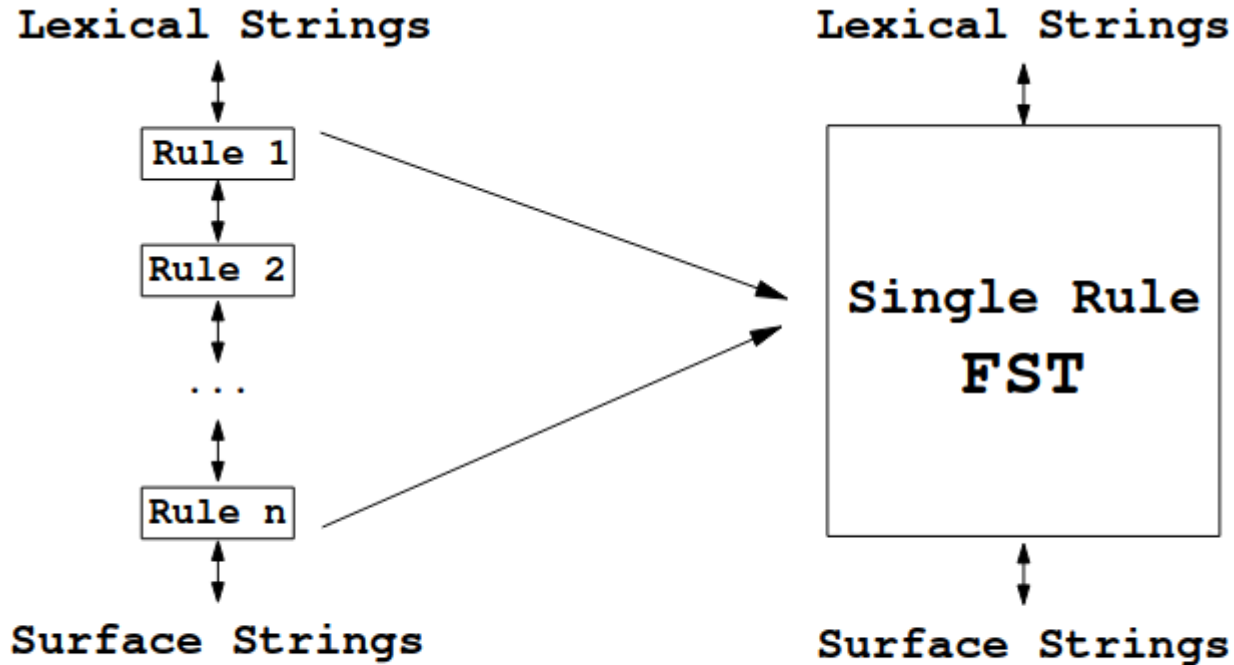


Figure 1: A Cascade of Rewrite Rules Composed into a Single **FST**

Problémák

- Hogyan lesz a fonológiai szabályrendszerből transzducer?
- opcionális szabályok \Rightarrow több output
- **neutralizáló** szabályok \Rightarrow több input \Rightarrow túlelemzés

mögöttes alakok:

take+t také+t

szabály:

$\{a, e\} \rightarrow \{á, é\} / _+$

\ /

felszíni alak:

takét

Neutralizáció és túlelemzés

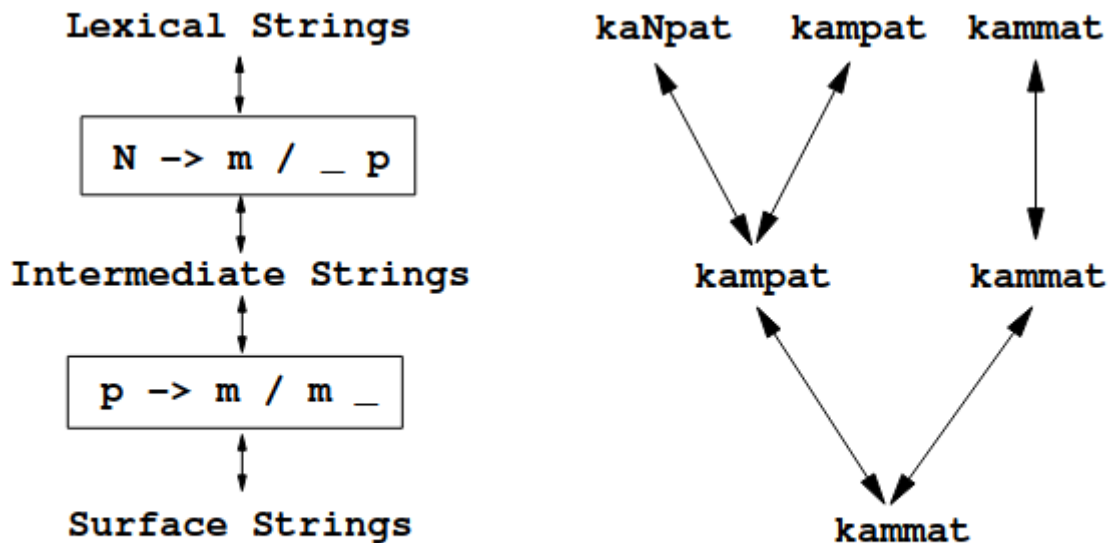


Figure 2: Rules Mapping *kammat* to *kaNpat*, *kammat*, *kammat*

KIMMO-rendszer (Koskenniemi 1983)

- megszorítások szabályok helyett
 - szimbólumpárok: Input : Output
 - párhuzamosan érvényesülnek (nem szekvenciálisan)
 - a környezet megadásánál lehet hivatkozni b melyikre
 - környezet lehet: megengedett, kötelező v. tiltott
- példa: palatalizáció

t:c <=> ___ @:i

KIMMO-formalizmus



Figure 3: Example of Two-Level Constraints

szabályok:

$N:m \Leftrightarrow _ p:@$

$p:m \Leftrightarrow _ @:m$

A kétszintű szabályok szemantikája

<code>a:b <=> l _ r ;</code>	<code>! lar</code>	<code>lar</code>	<code>lbr</code>	<code>xay</code>
	<code>! lbr</code>	<code>lar</code>	<code>lbr</code>	<code>xby</code>
<code>a:b <= l _ r ;</code>	<code>! lar</code>	<code>lar</code>	<code>lbr</code>	<code>xay</code>
	<code>! lbr</code>	<code>lar</code>	<code>lbr</code>	<code>xby</code>
<code>a:b => l _ r ;</code>	<code>! lar</code>	<code>lar</code>	<code>lbr</code>	<code>xay</code>
	<code>! lbr</code>	<code>lar</code>	<code>lbr</code>	<code>xby</code>
<code>a:b /<= l _ r ;</code>	<code>! lar</code>	<code>lar</code>	<code>lbr</code>	<code>xay</code>
	<code>! lbr</code>	<code>lar</code>	<code>lbr</code>	<code>xby</code>

Figure 17: Rule Constraints Allow Certain String Pairs and Disallow Others. Review the semantics of two-level rules until you understand why each rule blocks and allows what it does.

Törlés, beillesztés



Figure 4: A Two-Level View of $y \sim ie$ Alternation in English

Szabályok:

- zéró környezet $y:i \Leftrightarrow 0:e$ _

- tudni kell, hol van a zéró:

0:e

+:0

Egy FST több helyett

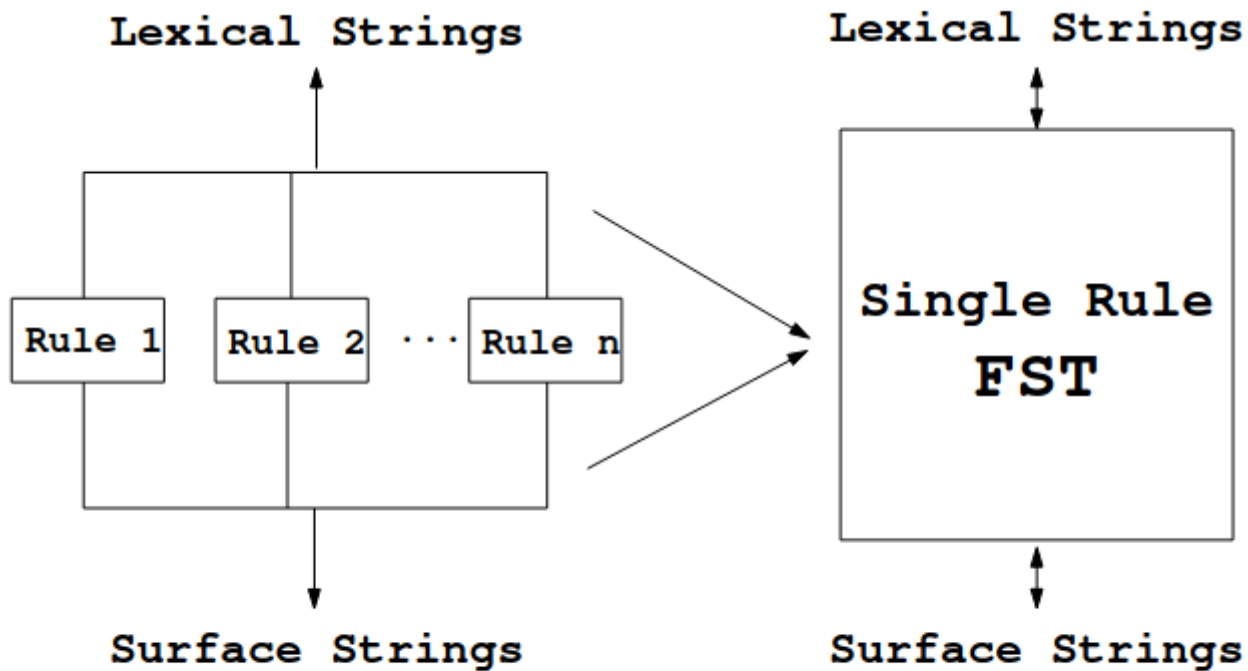
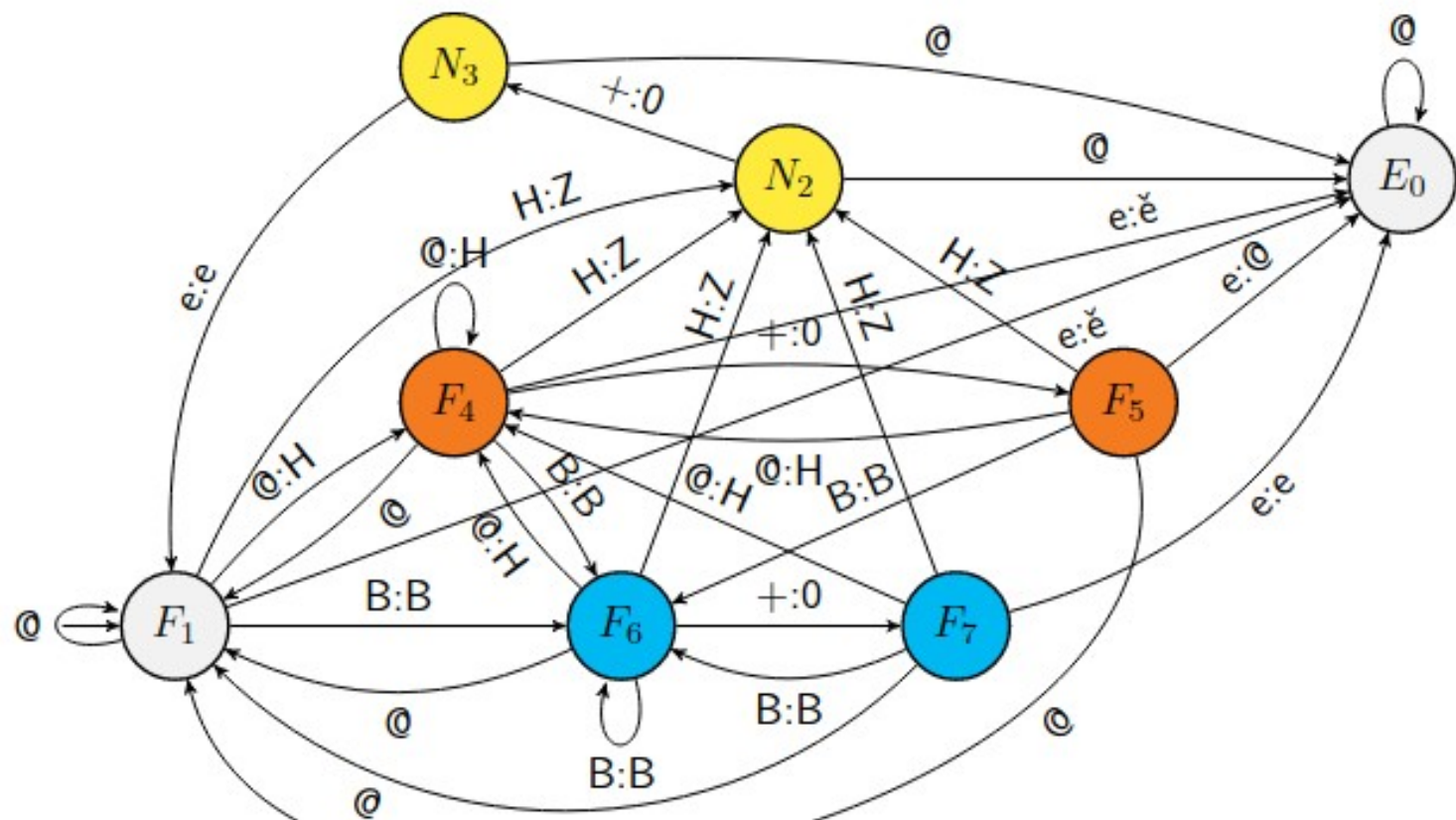


Figure 8: A Set of Two-Level Rules Intersected into a Single **FST**



Czech Feminine Noun Consonant Changes



H:Z =
 g:z | h:z
 | ch:š |
 k:c | r:ř

B:B =
 b:b | f:f |
 m:m |
 p:p | v:v
 | w:w |
 q:q | d:d
 | t:t | n:n
 | ḍ:ḍ | ṭ:ṭ
 | ň:n