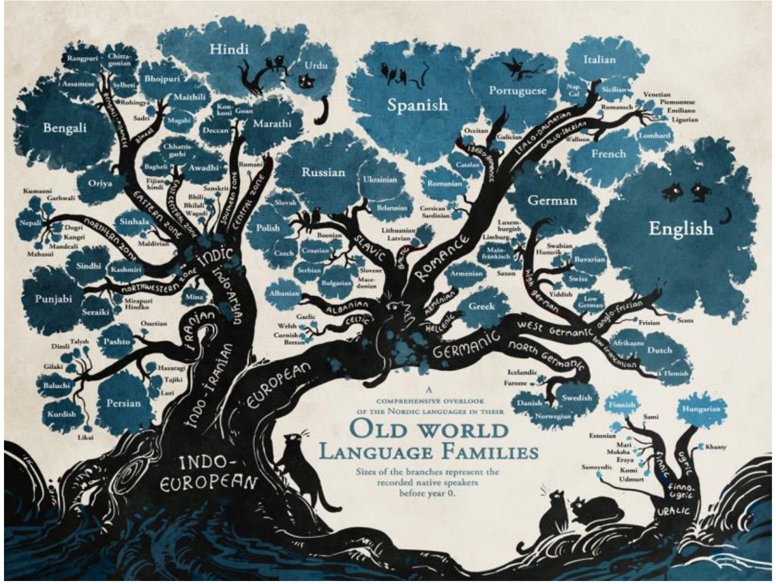
12 Formele talen

Elfde college



verkeerslicht?

신호등을지킵시다

(Automatische) Vertaling van het Koreaans

You should observe the traffic lights

Is Koreaans een formele taal? Nee natuurlijk niet! Alleen, voor iemand die de taal niet kan lezen, doet het ‘formeel’ aan; je ziet vorm, geen betekenis.

Is een formele taal zonder betekenis? Ook al niet! Rekenkundige expressies bijv. vormen een formele taal, maar hebben ook een semantiek.

taal …

letter = symbool

alfabet Σ: eindige (niet-lege) verzameling letters

string = woord: (geordend) rijtje letters (uit het alfabet)
Alle strings: Σ^*
Lege string λ, lengte nul

taal: verzameling woorden
[dus een deelverzameling van Σ^*]
Alle talen: $P(\Sigma^*)$

Op verzamelingen kunnen we de bekende boolese operaties toepassen, maar op talen ook concatenatie, macht en ster (d.w.z. achter elkaar zetten + herhaling). Dit zijn standaardoperaties op woorden en (daarmee ook) op talen.

alfabet

Een *alfabet* is een <u>eindige</u>, niet-lege verzameling *letters*.

Σ = { a,b,c }

B = { 0,1 }

C = { а,б,в,г,д,е,ж,з,и,й,к,л, … э,ю,я }

P = { <u>if</u>, <u>else</u>, <u>while</u>, <u>do</u>, … }

V = { … , appel, koek, ei, … }

Σ alfabet.
Een *string/woord* (over Σ) is een *eindig* *geordend* rijtje letters uit Σ.

ab, abca, abcbabcba
0, 1, 00, 01, 10, 11, 000, 001, …
язык, приключения, Астерикса
"de appel valt niet ver"

Notaties: Σ^*, lege string λ, lengte $|x|$

$|abcbac| = 6$ $|\lambda| = 0$ $a^6b^3 = aaaaaabbb$

$\mathbb{B}^* = \{ \lambda, 0, 1, 00, 01, 10, 11, 100, 101, \dots \}$
= alle woorden over het alfabet $\mathbb{B} = \{ 0,1 \}$

Een *taal* (over Σ) is een verzameling strings over Σ

Σ^* : alle strings

PAL = { λ, aa, bb, abba, baab, abaaba, … }
BIN = { 0, 1, 10, 11, 100, 101, 110, 111, 1000, … }

K = { a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, … }
= { $x \in \{a,b\}^*$ | x eindigt op een a }
L = { λ, aa, ab, ba, bb, aaaa, aaab, aaba, aabb, … }
= { $x \in \{a,b\}^*$ | x heeft even lengte }

∅ : lege taal

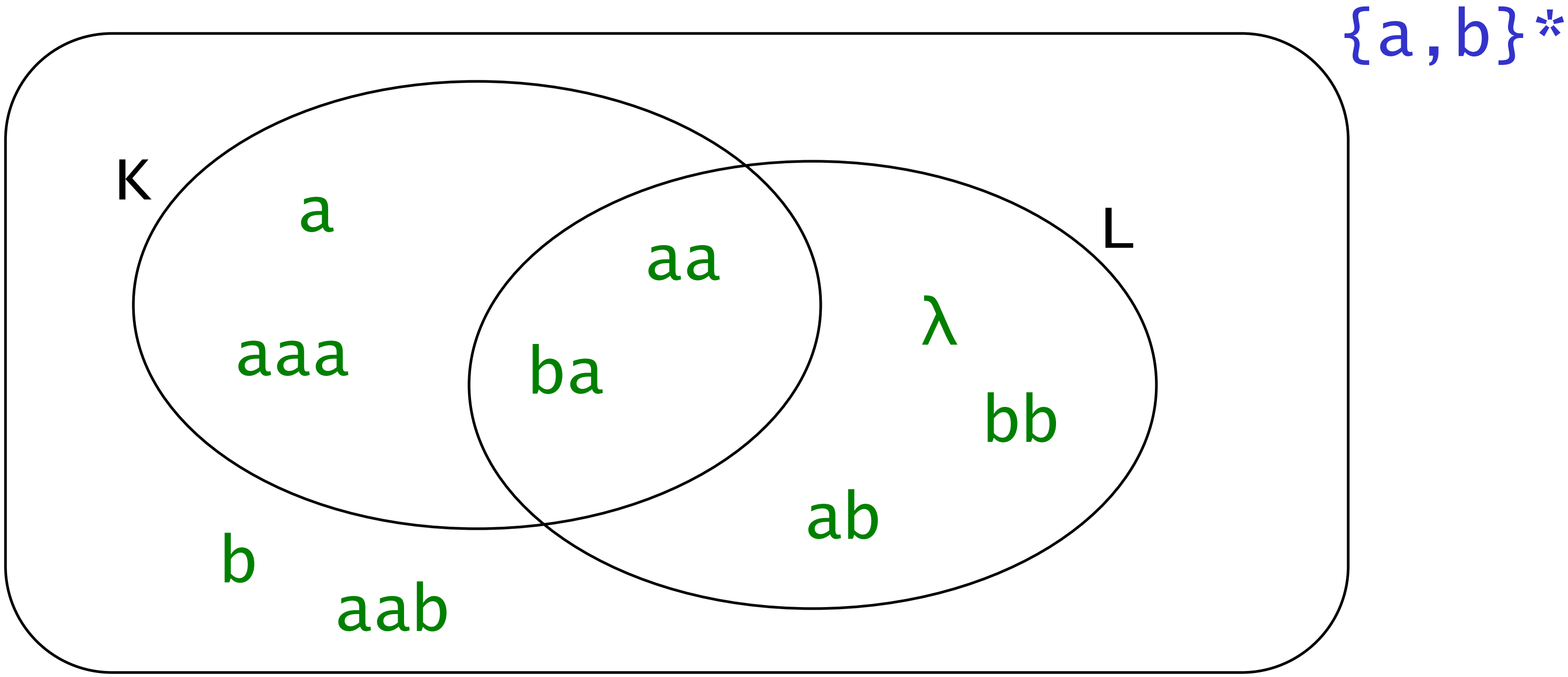
$\mathcal{P}(\Sigma^*)$: collectie van alle talen over Σ
collectie van alle deelverzamelingen van Σ^*

Een *taal* (over Σ) is een verzameling strings over Σ

en dus een deelverzameling van Σ*

K = { a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, … }
= { x∈{a,b}* | x eindigt op een a }

L = { λ, aa, ab, ba, bb, aaaa, aaab, aaba, aabb, … }
= { x∈{a,b}* | x heeft even lengte }



standaardoperaties op woorden en/of talen:

concatenatie = achter elkaar plakken

macht = idem, vast aantal herhalingen

ster = idem, willekeurig herhaald

operaties op woorden

standaardoperaties:

concatenatie = achter elkaar plakken

macht = idem, vast aantal herhalingen

maar ook:

prefix

spiegelbeeld

concatenatie

$\cdot : \Sigma^* \times \Sigma^* \to \Sigma^*$

$a_1 \ldots a_m \cdot b_1 \ldots b_n = a_1 \ldots a_m b_1 \ldots b_n$

- λ: één(element) $\lambda \cdot x = x = x \cdot \lambda$
- associatief $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
- niet commutatief bei·aard ≠ aard·bei
- $|x \cdot y| = |x| + |y|$
- $\cdot$ wordt vaak weggelaten

prefix

$x, y \in \Sigma^*$
x is een *deelwoord* van y als $y = u \cdot x \cdot v$ voor $u, v \in \Sigma^*$
… *prefix* … als $y = x \cdot v$ voor $v \in \Sigma^*$
… *suffix* … als $y = u \cdot x$ voor $u \in \Sigma^*$

óók: *sub*woord

lekkerkerker: deelwoord kerk (twee *voorkomens*)

aarzelaar: aar prefix én suffix (net als λ)

eigenschappen van prefix

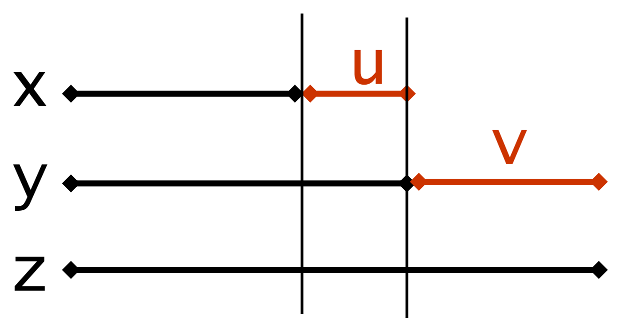
$x, y \in \Sigma^*$
x is een *prefix* van y als $y = x \cdot v$ voor $v \in \Sigma^*$

De relatie $\preceq$ definiëren we als $x \preceq y$ als x een prefix is van y

reflexief	$x \preceq x$ voor alle x
anti-symmetrisch	als $x \preceq y$ en $y \preceq x$ dan $x = y$
transitief	als $x \preceq y$ en $y \preceq z$ dan $x \preceq z$

$\Rightarrow$ partiële ordening

$y = x \cdot u$
$z = y \cdot v$
$\quad = x \cdot u \cdot v$



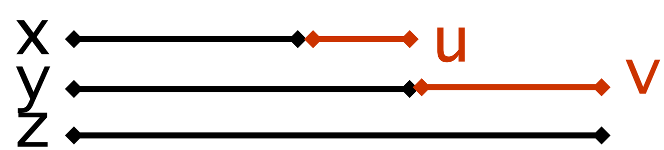
eigenschappen van prefix

$x, y \in \Sigma^*$
x is een *prefix* van y als $y = x{\cdot}v$ voor $v \in \Sigma^*$

De relatie $\preceq$ definiëren we als $x \preceq y$ als x een prefix is van y

reflexief $x \preceq x$ voor alle x
anti-symmetrisch als $x \preceq y$ en $y \preceq x$ dan $x = y$
transitief als $x \preceq y$ en $y \preceq z$ dan $x \preceq z$

$\Rightarrow$ partiële ordening



$x = x{\cdot}\lambda$, dus x prefix van x
$y = x{\cdot}u$ en $x = y{\cdot}v$ dan $|y| \le |x|$ en $|y| \le |x|$.
Daarom $|x| = |y|$ en dus $u = v = \lambda$ (lengte 0), $x = y$
$y = x{\cdot}u$ en $z = y{\cdot}v$ dan $z = (x{\cdot}u){\cdot}v = x{\cdot}(u{\cdot}v)$

$x^n = x \cdot \ldots \cdot x$ (n maal)

inductief:

$x^0 = \lambda$

$x^{n+1} = x^n \cdot x$

dus $x^1 = x^0 \cdot x = \lambda \cdot x = x$

$x^{m+n} = x^m \cdot x^n$

$|x^n| = n \cdot |x|$

Bewijs m.b.v. inductie naar n (m vast);
zie dictaatje H2, Lemma 2.8

$(\mathrm{ker})^5 = \mathrm{kerker} \cdot \mathrm{kerk} \cdot \mathrm{erker}$

Voor een voorbeeld:

$x^4 \cdot x^3 \stackrel{?}{=} x^7$

$x^4 \cdot x^3 = x^4 \cdot (x^2 \cdot x) = (x^4 \cdot x^2) \cdot x$
$= (x^4 \cdot (x \cdot x)) \cdot x = ((x^4 \cdot x) \cdot x) \cdot x$
$= (x^5 \cdot x) \cdot x = x^6 \cdot x = x^7$

associativiteit !!

spiegelbeeld

inductief:

$\text{mir}(\lambda) = \lambda$

$\text{mir}(xa) = a \cdot \text{mir}(x)$ voor elke $a \in \Sigma$

Voorbeeld:

$\text{mir}(aab)=b\cdot\text{mir}(aa)=b\cdot a\cdot\text{mir}(a)=b\cdot a\cdot a\cdot\text{mir}(\lambda)=b\cdot a\cdot a\cdot\lambda=baa$

Notatie ook wel: x^R

Palindroom: $x = \text{mir}(x)$

snorfrons netebeten

Eigenschappen:

$\text{mir}(\text{mir}(x)) = x$

$\text{mir}(xy) = \text{mir}(y) \cdot \text{mir}(x)$

$\text{mir}(x^n) = \text{mir}(x)^n$

bewijs gaat met inductie naar de lengte van x resp. y en gebruik van de inductieve definitie

operaties op talen

standaardoperaties:

concatenatie

macht

ster, plus

maar ook:

$\cap$, $\cup$, $-$ want een taal is een verzameling

spiegelbeeld

taal: Boolese operaties

Een *taal* (over Σ) is een verzameling strings over Σ

$K = \{$ a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, … $\}$
$= \{ x \in \{a,b\}^* \mid x$ eindigt op een a $\}$
$L = \{ \lambda,$ aa, ab, ba, bb, aaaa, aaab, aaba, aabb, … $\}$
$= \{ x \in \{a,b\}^* \mid x$ heeft even lengte $\}$

$\mathcal{P}(\Sigma^*)$ talen over Σ
vereniging, doorsnede, complement (t.o.v. Σ^*)

$K \cap L = \{$ aa, ba, aaaa, aaba, abaa, abba, baaa, … $\}$
$K - L = \{$ a, aaa, aba, baa, bba, aaaaa, aaaba, … $\}$
$L - K = \{ \lambda,$ ab, bb, aaab, aabb, abab, abbb, baab, … $\}$
$\{a,b\}^* - (K \cup L) = \{$ b, aab, abb, bab, bbb, aaaab, … $\}$
$= (K \cup L)^C$

verzamelingsrekenregels

commutatief	$A \cap B = B \cap A$	$A \cup B = B \cup A$
associatief	$(A \cap B) \cap C = A \cap (B \cap C)$	$(A \cup B) \cup C = A \cup (B \cup C)$
distributief	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
wetten van De Morgan	$(A \cap B)^C = A^C \cup B^C$	$(A \cup B)^C = A^C \cap B^C$
absorptie	$A \cap (A \cup B) = A$	$A \cup (A \cap B) = A$
idempotentie	$A \cap A = A$	$A \cup A = A$
nulelement	$A \cap \varnothing = \varnothing$	$A \cup \varnothing = A$
éénelement	$A \cap U = A$	$A \cup U = U$
dubbel complement	$(A^C)^C = A$	
complementregels	$A \cap A^C = \varnothing$	$A \cup A^C = U$

regels ook geldig voor oneindige verzamelingen

Bij talen en hun operaties gebruiken we soms oneindige verenigingen. Als we twee verzamelingen verenigen geldt de regel van De Morgan.
Die geldt ook voor meerdere (eindig veel) verzamelingen; dat kunnen we laten zien door de regels herhaald toe te passen (en wat associativiteit).

Ook oneindige vereniging kent de regel van De Morgan, maar daar moeten we een beetje logica bij gebruiken, met de begrippen ‘er is’ en ‘voor alle’.
Bv. een woord x zit in de vereniging van A_i als er een index i is zodat x in A_i.

concatenatie

$\cdot : \mathcal{P}(\Sigma^*) \times \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^*)$

$K \cdot L = \{\ x{\cdot}y \mid x \in K \text{ en } y \in L\ \}$

eindig·eindig

$\{\ a,ab\ \}\cdot\{\ a,ba\ \} = \{\ a{\cdot}a, a{\cdot}ba, ab{\cdot}a, ab{\cdot}ba\ \}$
$= \{\ aa, aba, abba\ \}$

$\{\ a,ba\ \}\cdot\{\ a,ab\ \} = \{\ a{\cdot}a, a{\cdot}ab, ba{\cdot}a, ba{\cdot}ab\ \}$
$= \{\ aa, aab, baa, baab\ \}$

oneindig·eindig

$\{\ \lambda, a, b, aa, ab, ba, bb, aaa, aab, \ldots\ \}\cdot\{\ a\ \} =$
$\{\ a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, \ldots\ \}$

Voor eindige talen hebben we:
aantal woorden K·L ≤ aantal woorden K * aantal woorden L

concatenatie

$\cdot : \mathcal{P}(\Sigma^*) \times \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^*)$

$K \cdot L = \{ x{\cdot}y \mid x \in K \text{ en } y \in L \}$

oneindig·oneindig

{ a,ab,abb,abbb, … }·{ a,ba,aba,baba,ababa, … }

= { aa,aba,aaba,abba,ababa,abbba,aababa,abbaba, … }

	a,	ba,	aba,	baba,	ababa,	…
a	aa,	aba,	aaba,	ababa,	aababa,	…
ab	aba,	abba,	ababa,	abbaba,	abababa,	…
abb	abba,	abbba,	abbaba,	abbbaba,	abbababa,	…
abbb	abbba,	abbbba,	abbbaba,	abbbbaba,	…	
…						

concatenatie

$\cdot : \mathcal{P}(\Sigma^*) \times \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^*)$

$K \cdot L = \{ x \cdot y \mid x \in K \text{ en } y \in L \}$

$\{\lambda\}$: één(element) $\quad \{\lambda\} \cdot L = L \cdot \{\lambda\} = L$

$\varnothing$: nul(element) $\quad \varnothing \cdot L = L \cdot \varnothing = \varnothing$

associatief $\quad (K \cdot L) \cdot M = K \cdot (L \cdot M)$

niet commutatief

macht

$K^n = K \cdot \ldots \cdot K$ (n maal)

inductief:

$K^0 = \{ \lambda \}$

$K^{n+1} = K^n \cdot K$

dus $K^1 = K^0 \cdot K = \{\lambda\} \cdot K = K$

$\{\lambda, a, ab\}^1 = \{\lambda, a, ab\}$

$\{\lambda, a, ab\}^2 =$

$\{\lambda, a, aa, ab, a \cdot ab, aba, abab\}$

$\{\lambda, a, ab\}^3 =$

$\{\lambda, a, aa, ab, aaa, a \cdot ab, aba, aaab, a \cdot ab \cdot a, abaa,$
$abab, a \cdot ab \cdot ab, abaab, ababa, ababab\}$

ster & plus

$K^n = K \cdot \ldots \cdot K$ (n maal)

inductief:

$K^0 = \{ \lambda \}$

$K^{n+1} = K^n \cdot K$

$K^* = \bigcup_{n \geq 0} K^n$, $K^+ = \bigcup_{n > 0} K^n$

ab* = {a}·{b}* = { a,ab,abb,abbb,abbbb … }

(ab*)* = { a,ab,abb,abbb,abbbb … }* =

= { λ,a,aa,ab,aaa,aab,aba,abb,aaaa,aaab,… }

= {λ} ∪ {a}·{a,b}*

{λ,a,ab}* = {a,ab}* =

{ x | voor elke b staat een a }

ster & plus

$K^n = K \cdot \ldots \cdot K$ (n maal)

inductief:

$K^0 = \{ \lambda \}$

$K^{n+1} = K^n \cdot K$

$K^* = \bigcup_{n\geq 0} K^n$, $K^+ = \bigcup_{n>0} K^n$

alternatief:

K^* is de kleinste verzameling die λ bevat en gesloten is onder concatenatie met K

oftewel (inductief):

- $\lambda \in K^*$
- als $x \in K^*$ en $w \in K$ dan $x \cdot w \in K^*$

Wanneer is

$$\boxed{L = L^2}$$

- voorbeelden
- wat weten we van L
- opgave 59, 60

oefening

$\boxed{L = L^2}$

als $\lambda \in L$ dan $L \subseteq L^2$ (opgave 60a)

dus nog testen of $L^2 \subseteq L$?

$\{\lambda\}$

$\varnothing$

Σ^*

- heeft subwoord/prefix/suffix x
- heeft even lengte/ veelvoud n
- subwoord x én/of subwoord y

- ? heeft géén subwoord x
- is een palindroom
- K^* van *willekeurige* K (!)
- $\{ ab^2, ab^3, ab^5, ab^7, \dots \}$
- $\{ ab, aabb, aaabbb, \dots \}$

  $= \{ a^n b^n \mid n \geq 1 \}$

$\boxed{L = L^*}$

taal specificeren

- opsommen:

$K = \{ a, aa, ba, aaa, aba, baa, bba, aaaa, aaba, \ldots \}$

- beschrijven:

$K = \{ x \in \{a,b\}^* \mid x$ eindigt op een $a \}$

- operaties $*$ $\cup$ $\cdot$

$K = \{a,b\}^*a = (\{a\} \cup \{b\})^* \cdot \{a\}$

- reguliere expressie: $(a+b)^*a$ of $(a \vee b)^*a$

We hebben ook voorbeelden gezien van inductief gedefinieerde talen. Eigenschappen kunnen dan vaak met behulp van (structurele) inductie bewezen worden.

Voorbeeld.

Inductieve definitie van de taal $L \subseteq \{a,b\}^*$:

(i) $a \in L$

(ii) als $x \in L$ en $y \in L$, dan ook $bxy \in L$

(iii) L bevat geen andere woorden

Bewijs met inductie dat voor elke $z \in L$ geldt dat $\#a(z) = \#b(z) + 1$.

Bij recursief gedefinieerde talen is vaak niet goed te zien of ze regulier zijn of niet. Hierna focussen we ons op de *reguliere talen*.

reguliere talen

te maken met * ∪ ·

Alfabet: {a,b}

L_1 = { x∈{a,b}* | x bestaat uit 1 of meer a's, gevolgd door 1 of meer b's }

{a}·{a}*·{b}·{b}*   aa*bb*

L_2 = { x∈{a,b}* | x bevat precies twee a's }

{b}*·{a}·{b}*·{a}·{b}*   b*ab*ab*

L_3 = { x∈{a,b}* | voor elke b staat een a }

{a,ab}* = ({a} ∪ {a}·{b})*   (a+ab)*   (a∪ab)*

sommetje

alfabet $\{a,b\}^*$

$\{ x \in \{a,b\}^* \mid$ voor en na elke a staat een b $\}$

Fout: $\{b,bab\}^*$ $(b + bab)^*$

$\{b,bab\}^* = \{ \lambda,$ b, bb, bbb, bab, bbab, babb, bbbab, bbabb, babbb, … $\}$

Echter babab zit er bijvoorbeeld niet in!

Hoe kunnen we deze taal dan wel uitdrukken m.b.v. $*$, $\cup$ en $\cdot$ of een reguliere expressie?

kleinste familie van talen die
- $\varnothing$ $\{\lambda\}$ $\{a\}$ bevat
- gesloten is onder $*$ $\cup$ $\cdot$

- $\varnothing$ $\{\lambda\}$ $\{a\}$ $\in$ REG
- als $K,L \in$ REG dan
  - $K \cup L \in$ REG
  - $K \cdot L \in$ REG
  - $K^* \in$ REG
- verder geen talen in REG

Uit de definitie volgt onmiddellijk dat elke eindige taal regulier is

Alle talen die je kan maken met de operaties $*$ $\cup$ $\cdot$

rekenregels 'algebra'

$\varnothing \cdot L = L \cdot \varnothing = \varnothing$
$\{\lambda\} \cdot L = L \cdot \{\lambda\} = L$
$(K \cdot L) \cdot M = K \cdot (L \cdot M)$
$K \cdot (L \cup M) = K \cdot L \cup K \cdot M$
$(K \cup L) \cdot M = K \cdot M \cup L \cdot M$
$(K^*)^n = K^*$ voor alle $n \geq 1$†
$(K^*)^* = K^*$
$(K^* L^*)^* = (K \cup L)^* = (K^* \cup L^*)^*$

Lemma 4.5, 4.7
dictaatje H4

$K \subseteq L$
dan $K \cup M \subseteq L \cup M$
$K \cdot M \subseteq L \cdot M$, $M \cdot K \subseteq M \cdot L$
$K^n \subseteq L^n$ voor alle $n \geq 1$
$K^* \subseteq L^*$

Volgt rechtstreeks
uit de definities

† In het bijzonder: $(K^*)^2 = K^*$

spiegelbeeld

$mir(K) = \{ mir(x) \mid x \in K \}$ ook K^R

1. $mir(mir(K)) = K$
2. $mir(K \cup L) = mir(K) \cup mir(L)$
3. $mir(K \cdot L) = mir(L) \cdot mir(K)$
4. $mir(K^n) = mir(K)^n$
5. $mir(K^*) = mir(K)^*$

Stelling 4.9, 4.10
dictaatje H4

- $\varnothing$ $\{\lambda\}$ $\{a\}$ $\in$ REG
- als K, $L \in$ REG dan
  - $K \cup L \in$ REG
  - $K \cdot L \in$ REG
  - $K^* \in$ REG

REG is gesloten onder mir (volgt uit 2,3,5)
d.w.z.: K regulier dan ook mir(K) regulier

- beschrijven, genereren

  K: w_0, w_1, w_2, w_3, …

  - recursieve definitie (gezien)
  - operaties * ∪ ·
  - reguliere expressies

  } **REG**
  - grammatica (productieregels)
  - …

- herkennen, algoritme

  w ∈ K? ja/nee

  - automaat, Turing machine
  - parser

grammatica

- $MI \in L$
- als $xI \in L$ dan $xIU \in L$ ♣
  als $Mx \in L$ dan $Mxx \in L$ ♦
  als $xIIIy \in L$ dan $xUy \in L$ ♥
  als $xUUy \in L$ dan $xy \in L$ ♠
- L bevat geen andere elementen

axioma MI

'herschrijf'regels

I	→	IU	mits uiteinde
Mx	→	Mxx	mits hele woord (?!)
III	→	U	in elke context
UU	→	λ	

MI ⇒ MII ⇒ MIIII ⇒ MUI ⇒ MUIUI ⇒ MUIUIU ⇒…

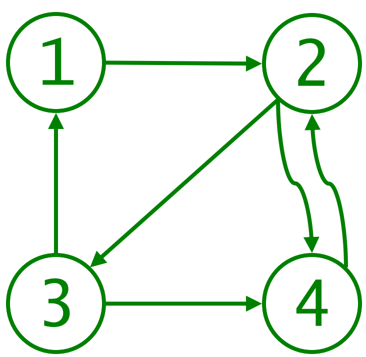
Chomsky: type regels *families* FI2
rechts-lineair, context-vrij, monotoon, type-0

probleem = taal

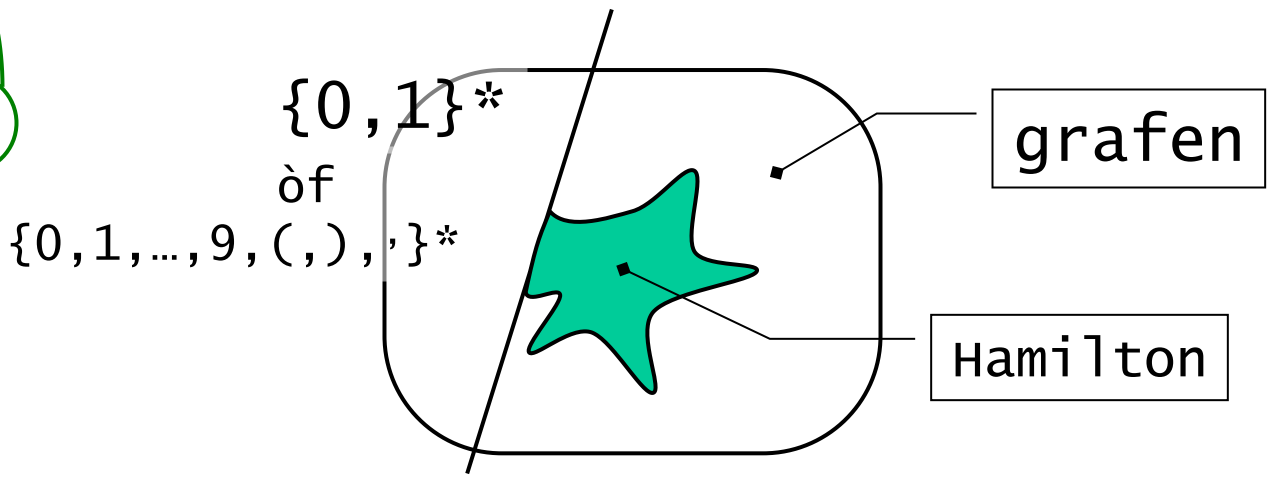
HP: Hamilton Path Problem
gegeven: gerichte graaf G, begin- & eindpunt
gevraagd: is er een Hamilton pad (elke knoop éénmaal)

aantal punten	pijlen						begin & eind
4	(1,2)	(2,3)	(2,4)	(3,1)	(3,4)	(4,2)	1,2

000 100 001010 010011 010100 011001 011100 100010 000 001010



G graaf -> woord w(G)
Hamilton grafen -> taal Ham



probleem = taal

Bij het vak Complexiteit komt de formele definitie van het begrip NP-compleet aan de orde. Dit begrip kunnen we bijvoorbeeld formuleren met behulp van formele talen. Hiervoor moeten we dan het begrip ‘probleem’ omzetten naar een ‘taal’.

Hoeveel tijd kost het om instantie X van probleem P op te lossen wordt dan: hoeveel tijd kost het om te beslissen of X tot de taal P behoort.

antwoord

$\{ x \in \{a,b\}^* \mid$ voor en na elke a staat een b $\}$

$(\ \{b\} \cup \{b\} \cdot \{ab\}^*\ )^*$ $\quad$ $(\ b + b(ab)^*\ )^*$

College volgende week:
dinsdag 27 november,
13.30 – 15.15 in zaal E004 (Steenis)

Werkcollege volgende week:
vrijdag 30 november,
9.00 – 10.45 in Snelliuszalen 402, 405