Dualité et théorie équationnelle des langages

Mai Gehrke[1] Serge Grigorieff[2] Jean-Éric Pin[2]

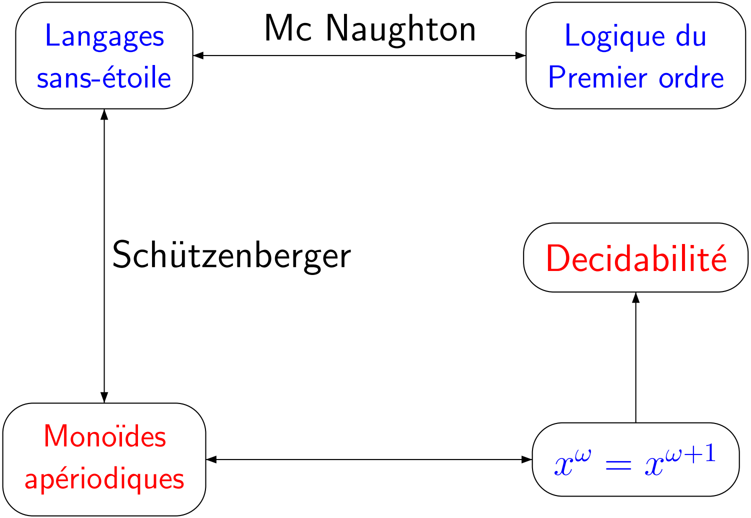
[1]Radboud Universiteit

[2]LIAFA, CNRS et Université Paris Diderot

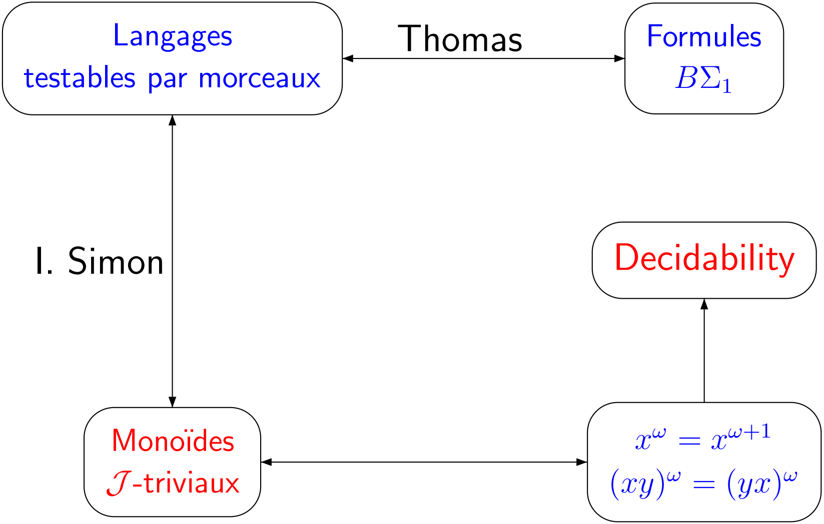
Octobre 2008, PPS

Sommaire

Un cercle vertueux



Un autre cercle vertueux



Première partie I

Treillis de langages

Parties reconnaissables

Un langage L de A^* est reconnu par un monoïde M s'il existe un morphisme surjectif de monoïdes $\varphi : A^* \to M$ qui sature L : si $u \in L$ et $\varphi(u) = \varphi(v)$, alors $v \in L$ [ou encore $L = \varphi^{-1}\varphi(L)$].

Un langage est reconnaissable s'il est reconnu par un monoïde fini. Un langage est reconnaissable ssi il est rationnel (Kleene).

Le monoïde syntactique d'un langage rationnel L de A^* est le plus petit monoïde qui le reconnaît. Sa taille sert de mesure de complexité pour L.

Treillis de langages

On fixe un alphabet fini A. Un treillis de langages est un ensemble de langages reconnaissables de A^* contenant $\emptyset$ et A^* et fermé par union finie et intersection finie.

Définition

Soient u et v des mots de A^*. Un langage L de A^* satisfait l'équation $u \to v$ si la condition $u \in L$ entraîne $v \in L$.

Treillis finis de langages

Soit E un ensemble d'équations de la forme $u \to v$. Alors l'ensemble $\mathcal{L}(E)$ des langages de A^* vérifiant les équations de E est un treillis de langages.

Proposition

Soient $L, L_1, \ldots, L_n$ des langages. Si L satisfait toutes les équations satisfaites par $L_1, \ldots, L_n$, alors L appartient au treillis de langages engendré par $L_1, \ldots, L_n$.

Caractérisation des treillis finis de langages

Corollaire

Un ensemble fini de langages de A^* est un treillis de langages ssi il peut être défini par un ensemble d'équations de la forme $u \to v$ avec $u, v \in A^*$.

Il y a donc une théorie équationnelle pour les treillis finis de langages. Que peut-on dire des treillis infinis ?

Caractérisation des treillis finis de langages

Corollaire

Un ensemble fini de langages de A^* est un treillis de langages ssi il peut être défini par un ensemble d'équations de la forme $u \to v$ avec $u, v \in A^*$.

Il y a donc une théorie équationnelle pour les treillis finis de langages. Que peut-on dire des treillis infinis ?

Il faut se plonger dans le monde profini...

Deuxième partie II

Le monde profini

Citation (M. Stone)

A cardinal principle of modern mathematical research may be stated as a maxim : One must always topologize.

Séparer deux mots

Un monoïde M sépare deux mots u et v de A^* s'il existe un morphisme de monoïde $\varphi : A^* \to M$ tel que $\varphi(u) \neq \varphi(v)$.

Par exemple, le morphisme qui envoie chaque mot sur sa longueur modulo 2 est un morphisme de $\{a,b\}^*$ sur $\mathbb{Z}/2\mathbb{Z}$ qui sépare $abaaba$ et $abaabab$.

Soit $M = \{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right), \left(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}\right), \left(\begin{smallmatrix}0&1\\0&1\end{smallmatrix}\right)\}$ et soit $\varphi : \{a,b\}^* \to M$ defini par $\varphi(a) = \left(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}\right)$ and $\varphi(b) = \left(\begin{smallmatrix}0&1\\0&1\end{smallmatrix}\right)$. Alors, pour chaque mot u, φ sépare ua et ub car $\varphi(ua) = \varphi(a)$ et $\varphi(ub) = \varphi(b)$.

La distance profinie

Soient u et v deux mots. Posons

$$r(u,v) = \min\{|M| \mid M \text{ est un monoïde fini qui sépare } u \text{ et } v\}$$

$$d(u,v) = 2^{-r(u,v)}$$

Dans ce cas, d est une distance ultramétrique. Pour tout $x, y, z \in A^*$,

(1) $d(x,x) = 0$,
(2) $d(x,y) = d(y,x)$,
(3) $d(x,z) \leqslant \max\{d(x,y), d(y,z)\}$

Propriétés de d

Intuitivement, deux mots sont proches pour d s'il faut un monoïde de grande taille pour les séparer.

Une suite de mots u_n est une suite de Cauchy ssi, pour tout morphisme φ de A^* vers un monoïde fini, la suite $\varphi(u_n)$ est ultimement constante.

Une suite de mots u_n converge vers un mot u ssi, pour tout morphisme φ de A^* sur un monoïde fini, la suite $\varphi(u_n)$ est ultimement égale à $\varphi(u)$.

Le monoïde profini libre

La complétion de l'espace métrique (A^*, d) est le monoïde profini libre sur A, noté $\widehat{A^*}$. Ses éléments sont appelés des mots profinis.

Le produit est uniformément continu sur A^* et s'étend donc par continuité à $\widehat{A^*}$. Si A est fini, $\widehat{A^*}$ est compact.

Tout morphisme $\varphi : A^* \to M$, où M est fini (discret) est uniformément continu. Puisque A^* est dense dans $\widehat{A^*}$, φ s'étend de façon unique en un morphisme uniformément continu $\hat{\varphi} : \widehat{A^*} \to M$.

Le monoïde profini libre comme limite projective

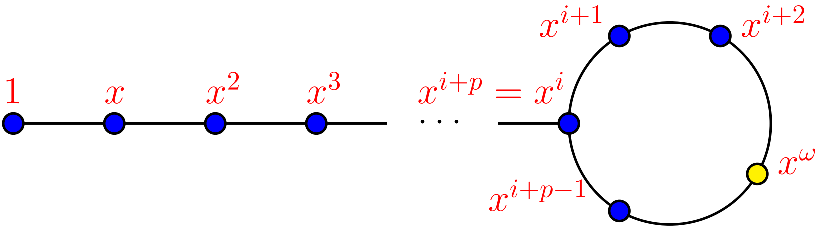
Le monoïde profini peut être défini comme la limite projective du système dirigé formé par les morphismes surjectifs entre monoïdes finis.

En particulier, un mot profini u est complètement déterminé par ses images $\hat{\varphi}(u)$, où φ décrit la classe des morphismes de A^* sur un monoïde fini.

$$\text{Mot profini } = \{\hat{\varphi}(u)\}_{\varphi : A^* \to M}$$

Un mot profini non fini

Pour chaque $u \in \widehat{A^*}$, la suite $u^{n!}$ est une suite de Cauchy et donc converge dans $\widehat{A^*}$ vers une limite, notée u^ω. Si φ est un morphisme de A^* sur a monoïde fini M, $\hat{\varphi}(u^\omega)$ est l'unique idempotent du sous-semigroupe de M engendré par $x = \varphi(u)$.



Troisième partie III

Théorie équationnelle des langages

Équations

Soit L un langage reconnaissable de A^* et soit $\eta : A^* \to M$ son morphisme syntactique. On sait que η se prolonge en un morphisme uniformément continu de $\hat{A}^*$ vers M. On note $\overline{L}$ la fermeture topologique de L pour la distance profinie.

Définition

Soient u et v des mots profinis de $\hat{A}^*$. Alors L satisfait l'équation $u \to v$ si la condition $u \in \overline{L}$ entraîne $v \in \overline{L}$.

Une définition équivalente

Soit L un langage reconnaissable de A^* et soit $\eta : A^* \to M$ son monoïde syntactique.

Définition

Soient u et v des mots profinis de $\hat{A}^*$. Alors L satisfait l'équation $u \to v$ si la condition $\hat{\eta}(u) \in \eta(L)$ entraîne $\hat{\eta}(v) \in \eta(L)$.

Caractérisation équationnelle des treillis

Soit E un ensemble d'équations de la forme $u \to v$. Le sous-ensemble de $\mathrm{Rec}(A^*)$ défini par E est l'ensemble de tous les langages reconnaissables de A^* satisfaisant toutes les équations de E.

Théorème

Un ensemble de langages reconnaissables de A^* est un treillis de langages ssi il peut être défini par un ensemble d'équations de la forme $u \to v$, où $u, v \in \hat{A}^*$.

Quatrième partie IV

Dualité

Petit résumé sur la dualité

L'espace dual d'un treillis distributif est l'ensemble de ses filtres premiers.

Élements	$\leftrightarrow$	Filtres premiers
Algèbres de Boole	$\leftrightarrow$	Espaces topologiques
Treillis distributifs	$\leftrightarrow$	Espaces topologiques ordonnés
Sous-treillis	$\leftrightarrow$	Espaces quotients
Opérations n-aires	$\leftrightarrow$	Relations $(n+1)$-aires

Filtres premiers et valuation

Soit D un treillis distributif borné (avec 0 et 1).

Un filtre premier est une partie F de D telle que

(1) $1 \in F$, $0 \notin F$,

(2) si $x \in F$ et $x \leqslant y$, alors $y \in F$ (axiome d'extension),

(3) si $x, y \in F$, alors $x \wedge y \in F$,

(4) si $x \vee y \in F$, alors $x \in F$ ou $y \in F$.

Une valuation est un morphisme de treillis v de D dans le treillis booléen $\{0, 1\}$.

(1) $v(0) = 0 \quad v(1) = 1$

(2) $v(x \vee y) = v(x) + v(y) \quad v(x \wedge y) = v(x)v(y)$

Espace dual

Si $v : D \to \{0, 1\}$ est une valuation, l'ensemble $v^{-1}(1)$ est un filtre premier.
Si p est un filtre premier, l'application définie par

$$v(x) = \begin{cases} 1 & \text{si } x \in p \\ 0 & \text{sinon} \end{cases}$$

est une valuation.

L'espace dual d'un treillis distributif est l'ensemble de ses filtres premiers, ou, de façon équivalente, l'ensemble de ses valuations.

Dualité de Stone-Priestley

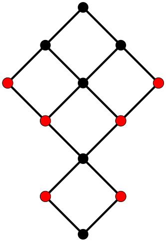
Soit X_D l'ensemble des filtres premiers de D (dual de D). L'application $e : D \to \mathcal{P}(X_D)$ donnée par

$$e(d) = \{ \text{ filtres premiers contenant } d \ \}$$

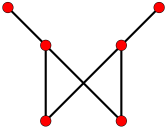
définit un morphisme de treillis injectif (mais pas surjectif en général).

On munit X_D d'une topologie dont les ensembles de la forme $e(d)$ ($d \in D$) et leur complément forment une base d'ouverts (et donc de clopen). L'espace X_D est un espace topologique ordonné compact (pas nécessairement séparé).

Dual d'un treillis distributif fini



Treillis



Filtres premiers

L'espace dual de $\mathrm{Rec}(A^*)$

Almeida (1989, implicitement) and Pippenger (1997, explicitement) ont prouvé :

Théorème

L'espace dual du treillis des langages rationnels est l'espace des mots profinis. Le plongement canonique est donné par la clôture topologique : $e(L) = \overline{L}$.

En effet, filtres premiers de $\mathrm{Rec}(A^*)$, valuations et mot profinis peuvent être identifiés !

L'espace dual de $\mathrm{Rec}(A^*)$ (suite)

- Si p est un filtre premier, il existe un unique mot profini u tel que, pour tout $\varphi : A^* \to M$, $\hat{\varphi}(u)$ soit l'unique élément $m \in M$ tel que $\varphi^{-1}(m) \in p$.

- Si u est un mot profini, l'ensemble

$$p_u = \{\, L \in \mathrm{Rec}(A^*) \mid \hat{\varphi}(u) \in \varphi(L) \text{ pour un morphisme } \varphi \text{ de } A^* \text{ sur un monoïde fini} \,\}$$

est un filtre premier.

- Si v est une valuation, on peut définir un mot profini u par la condition $\hat{\varphi}(u) = m$ ssi, pour chaque $\varphi : A^* \to M$, la valuation du langage $\varphi^{-1}(m)$ vaut 1.

Dualité de Priestley

Donc X_A est l'ensemble des mot profinis. D'après le théorème de dualité de Priestley, on a un morphisme injectif de treillis distributif de $\mathrm{Rec}(A^*)$ dans $\mathcal{P}(X_A)$:

$$\begin{aligned} L &\to \{\text{filtres premiers contenant } L\} \\ &\to \{\text{valuations telles que } v(L) = 1\} \\ &\to \{\text{mot profinis } u \text{ tels que } \hat{\eta}(u) \in \eta(L)\} \end{aligned}$$

Ces ensembles sont des ouverts et fermés de X_A. De plus, comme chaque singleton $\{u\}$ est un langage reconnaissable, A^* se plonge dans $\mathcal{P}(X_A)$.

Quelques formules

Les applications $L \mapsto \overline{L}$ et $K \mapsto K \cap A^*$ définissent des isomorphismes réciproques entre les algèbres de Boole $\mathrm{Rec}(A^*)$ et $\mathrm{Clopen}(\widehat{A^*})$. Pour tout $L, L_1, L_2 \in \mathrm{Rec}(A^*)$:

(1) $\overline{L^c} = (\overline{L})^c$,

(2) $\overline{L_1 \cup L_2} = \overline{L_1} \cup \overline{L_2}$, $\overline{L_1 \cap L_2} = \overline{L_1} \cap \overline{L_2}$,

(3) Si $x, y \in A^*$, on a $\overline{x^{-1}Ly^{-1}} = x^{-1}\overline{L}y^{-1}$.

(4) Si $\varphi : A^* \to B^*$ est un morphisme et $L \in \mathrm{Rec}(B^*)$, alors $\hat{\varphi}^{-1}(\overline{L}) = \overline{\varphi^{-1}(L)}$.

Un résultat de dualité

Les résiduels gauche et droit de L par K sont :

$$K\backslash L = \{u \in A^* \mid Ku \subseteq L\}$$
$$L/K = \{u \in A^* \mid uK \subseteq L\}$$

Théorème

Le produit des mots profinis est le dual de l'opération de résiduation sur les langages rationnels.

L'identité $(H\backslash L)/K = H\backslash(L/K)$ dans $\mathrm{Rec}(A^*)$ équivaut au fait que le produit est associatif.

Une preuve du résultat principal

Théorème

Un ensemble de langages reconnaissables de A^* *est un treillis de langages ssi il peut être défini par un ensemble d'équations de la forme* $u \to v$, *où* $u, v \in \hat{A}^*$.

La preuve résulte de la dualité entre sous-treillis de $\mathrm{Rec}(A^*)$ et préordres de l'espace dual $\widehat{A^*}$.

Si $\mathcal{L}$ est un treillis de langages, le préordre qui définit le quotient dans l'espace dual est exactement la théorie équationnelle de $\mathcal{L}$.

Structures uniformes associées à un treillis

Si $\mathcal{L}$ est un treillis de langages, la structure (quasi)-uniforme pro-$\mathcal{L}$ est engendrée par les ensembles $U_L = (L \times A^*) \cup (A^* \times L^c)$ où $L \in \mathcal{L}$.

Théorème

Une fonction $f : A^* \to A^*$ est uniformément continue pour la structure (quasi)-uniforme pro-$\mathcal{L}$ ssi f^{-1} préserve les langages de $\mathcal{L}$.

La topologie associée définit un espace spectral.

Cinquième partie V

Retour sur la théorie équationelle

Treillis de langages fermés par quotient

On dit que L satisfait l'équation $u \leqslant v$ si, pour tout $x, y \in \hat{A}^*$, il satisfait l'équation $xvy \rightarrow xuy$.

Proposition

Soit $L \in \mathrm{Rec}(A^*)$, soit $(M, \leqslant_L)$ son monoïde syntactique ordonné et soit $\eta : A^* \rightarrow M$ son morphisme syntactique. Alors L satisfait l'équation $u \leqslant v$ ssi $\hat{\eta}(u) \leqslant_L \hat{\eta}(v)$.

Treillis de langages fermés par quotient (2)

On dit que L satisfait l'équation $u \leqslant v$ si, pour tout $x, y \in \hat{A}^*$, il satisfait l'équation $xvy \to xuy$.

Théorème

Un ensemble de langages reconnaissables de A^* est un treillis de langages fermé par quotient ssi il peut être défini par un ensemble d'équations de la forme $u \leqslant v$, où $u, v \in \hat{A}^*$.

Algèbres de Boole fermées par quotient

On note $u = v$ si $u \leqslant v$ et $v \leqslant u$.

Théorème

Un ensemble de langages reconnaissables de A^* *est une algèbre de Boole fermée par quotient ssi il peut être défini par un ensemble d'équations de la forme* $u = v$, *où* $u, v \in \hat{A}^*$.

On retrouve ainsi une conséquence des théorèmes d'Eilenberg et de Reiterman.

Interprétation des équations (Notation : $P = \eta(L)$)

Fermé par		Interprétation
$\cup, \cap$	$u \to v$	$\hat{\eta}(u) \in P \Rightarrow \eta(v) \in P$
quotient	$u \leqslant v$	$xvy \to xuy$
complément (L^c)	$u \leftrightarrow v$	$u \to v$ et $v \to u$
quotient et L^c	$u = v$	$xuy \leftrightarrow xvy$
φ^{-1}		Variables $=$ mots
φ^{-1} strict. alpha.		Variables $=$ lettres
φ^{-1} length multiplying		Variables $=$ mots de même longueur

Langages avec zéro

Un langage avec zéro est un langage dont le monoïde syntactique possède un zéro.

Proposition

La classe des langages reconnaissables avec zéro est fermée pour les opérations booléennes et par quotients.

Par conséquent, il peut être défini par un ensemble d'équations de la forme $u = v\ldots$

Un mot profini

Fixons un ordre total sur l'alphabet A. Soit $u_0, u_1, \ldots$ la suite ordonnée de tous les mots de A^+ dans l'ordre des mots croisés (longueur d'abord).

$$1, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, \ldots$$

Reilly et Zhang (voir aussi Almeida-Volkov) ont prouvé que la suite $(v_n)_{n \geqslant 0}$ définie par

$$v_0 = u_0, \; v_{n+1} = (v_n u_{n+1} v_n)^{(n+1)!}$$

converge vers un idempotent ρ_A de l'idéal minimal du monoïde profini libre sur A.

Équations définissant les langages avec zéro

Proposition

Un langage reconnaissable possède un zéro ssi il satisfait l'équation $x\rho_A = \rho_A = \rho_A x$ *pour tout* $x \in A^*$.

Langages non denses

Un langage L de A^* est dense si, pour tout mot $u \in A^*$, $L \cap A^*uA^* \neq \emptyset$.

Les langages non-denses ou pleins sont fermés par union finie, intersection finie et quotients. Ils ne sont pas fermés par inverse de morphismes, même pour les morphismes strictement alphabétiques.

Théorème

Un langage de A^* est non-dense ou plein ssi il satisfait l'équation $x \leqslant 0$ pour tout $x \in A^*$.

Langages minces (slender)

Un langage L est mince si sa densité $d_L(n) = |L \cap A^n|$ est $O(1)$.

Théorème (Folklore)

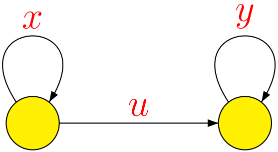
Un langage reconnaissable est mince ssi il est union finie de langages de la forme xu^*y.

Ces langages sont fermés par union finie, intersection finie, quotients et morphismes mais pas par inverse de morphismes, même pour les morphismes strictement alphabétiques.

Automates reconnaissant les langages minces

Théorème

Un langage reconnaissable est mince ssi son automate déterministe minimal ne contient aucune paire connectée de cycles simples.



Deux cycles connectés, où $x, y \in A^+$ et $u \in A^*$.

Équations définissant les langages minces

On note $i(u)$ l'initiale du mot u.

Théorème

Supposons que $|A| \geqslant 2$. Un langage reconnaissable de A^* est mince ou plein ssi il satisfait les équations $x \leqslant 0$ et $0 \leqslant x^\omega u y^\omega$ pour tout $x, y \in A^+$, $u \in A^*$ tels que $i(uy) \neq i(x)$.

Langages clairsemés (sparse)

Un langage L est clairsemé si sa densité est polynomiale.

Théorème (Folklore)

Un langage reconnaissable est clairsemé ssi il est union finie de langages de la forme $u_0 v_1^* u_1 \cdots v_n^* u_n$.

Ces langages sont fermés par union finie, intersection finie, produit de concatenation, quotients et morphismes mais pas par inverse de morphismes, même pour les morphismes strictement alphabétiques.

Équations définissant les langages clairsemés

Théorème

Supposons que $|A| \geqslant 2$*. Un langage reconnaissable de* A^* *est clairsemé ou plein ssi il satisfait les équations* $0 \leqslant (x^\omega y^\omega)^\omega$ *pour tout* $x, y \in A^+$*, tels que* $i(x) \neq i(y)$*.*

Applications

En résumé, tout treillis de langages rationnels a une caractérisation équationelle.

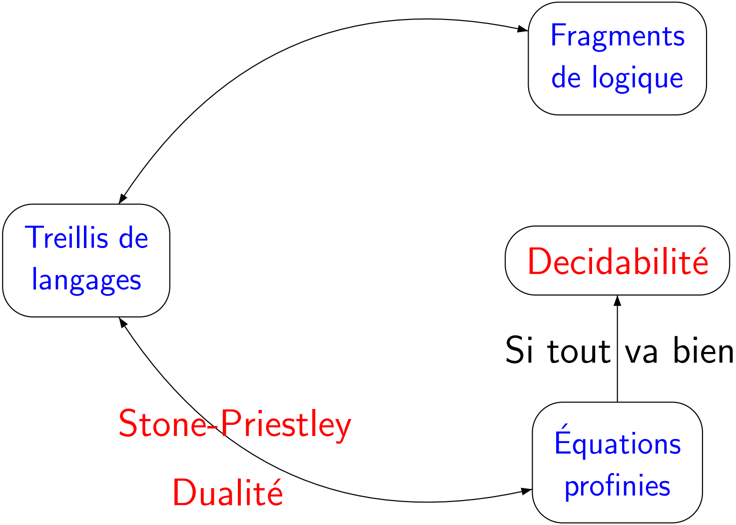
En particulier, tout fragment de logique (du premier ordre, du second ordre monadique, temporelle ou autre) fermé par conjonction et disjonction définit une classe de langages rationnels qui peut être décrite par des équations.

Résumé

On a montré que tout treillis de langages admet une description équationnelle, un résultat qui étend le théorème des variétés d'Eilenberg et ses extensions.

En particulier, toute classe de langages rationnels définie par un fragment de logique fermé par conjonctions et disjonctions (premier ordre, second ordre monadique, logique temporelle, etc.) admet une description équationnelle.

Retour sur les cercles vertueux



Extensions potentielles

Notre résultat s'étend sans difficulté aux mots infinis (en utilisant des ω-semigroupes) et très certainement aux mots sur des ordinaux ou des ordres linéaires en utilisant les travaux de Bruyère-Carton.

D'après les spécialistes, il s'étend également aux langages d'arbres (finis, infinis).

Le point de vue de la dualité permet d'étudier beaucoup d'autres problèmes…