# Factorisations pour la théorie des automates

Thomas Colcombet
Cnrs, Irisa

Sminaire 68NQRT, jeudi 3 mai 2007

# QUELQUES MOTIVATIONS

Comment complémenter sans (pouvoir) déterminiser ?

Comment effectuer des preuves sur $A^*$ quand l'induction sur la longueur ne fonctionne pas ?

Comment accélerer un calcul d'automate sur un mot fini?

# PLAN

Théorème de Ramsey

Théorème des forêts de factorisation

Variantes

# Ramsey (version infinie dénombrable)

**Th:** Soit $c : \mathbb{N} \times \mathbb{N} \to C$ ($C$ ensemble fini). Il existe $E \subseteq \mathbb{N}$ infini et $d \in C$ tq

$$\forall (x < y) \in E.\ c(x, y) = d\ .$$

# RAMSEY (VERSION INFINIE DÉNOMBRABLE)

**Th:** Soit $c : \mathbb{N} \times \mathbb{N} \to C$ ($C$ ensemble fini). Il existe $E \subseteq \mathbb{N}$ infini et $d \in C$ tq

$$\forall (x < y) \in E.\ c(x, y) = d\ .$$

**Preuve:**

| | | |
|---|---|---|
| $N_0 = \mathbb{N}$ | | $u_0 = \min N_0$ |
| $N_1 = \{x \in N_0 \setminus \{u_0\}\ :\ c(u_0, x) = c_1\}$ | tq $N_1$ infini | $u_1 = \min N_1$ |
| $N_2 = \{x \in N_1 \setminus \{u_1\}, :\ c(u_1, x) = c_2\}$ | tq $N_2$ infini | $u_1 = \min N_2$ |

$$\vdots$$

# RAMSEY (VERSION INFINIE DÉNOMBRABLE)

**Th:** Soit $c : \mathbb{N} \times \mathbb{N} \to C$ ($C$ ensemble fini). Il existe $E \subseteq \mathbb{N}$ infini et $d \in C$ tq

$$\forall (x < y) \in E.\ c(x, y) = d\ .$$

**Preuve:**

$$\begin{array}{lll} N_0 = \mathbb{N} & & u_0 = \min N_0 \\ N_1 = \{x \in N_0 \setminus \{u_0\}\ :\ c(u_0, x) = c_1\} & \text{tq } N_1 \text{ infini} & u_1 = \min N_1 \\ N_2 = \{x \in N_1 \setminus \{u_1\}, :\ c(u_1, x) = c_2\} & \text{tq } N_2 \text{ infini} & u_1 = \min N_2 \\ \quad \vdots & & \end{array}$$

Les $N_i$ sont tous infinis et forment un chaine décroissante.

Soit $d$ tq $\exists^\infty i.c_i = d$.

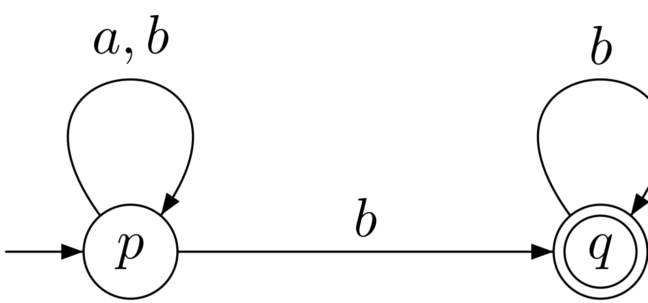
Soit $E = \{u_i\ :\ c_i = d\}$.

# AUTOMATES DE BÜCHI

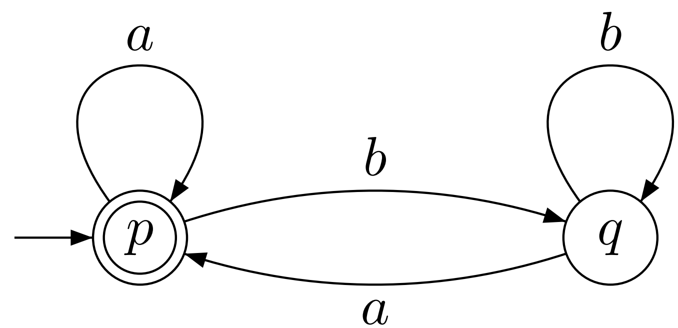
**Automates de Büchi:** Automate non-déterministe, sans état final + ensemble d'*états marqués* $B$.

Un mot infini est *accepté* s'il existe une exécution de l'automate sur le mot visitant infiniement souvent $B$.

**Exemple:**




Ultimement $b$.



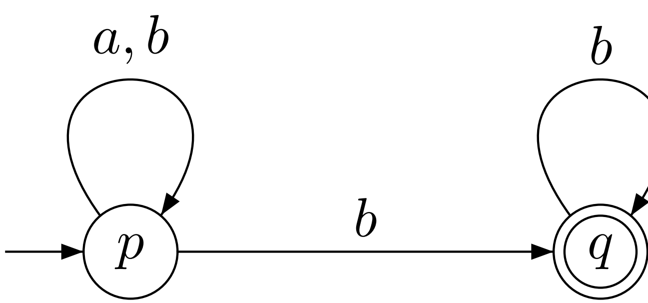

Infiniement souvent $a$

# AUTOMATES DE BÜCHI

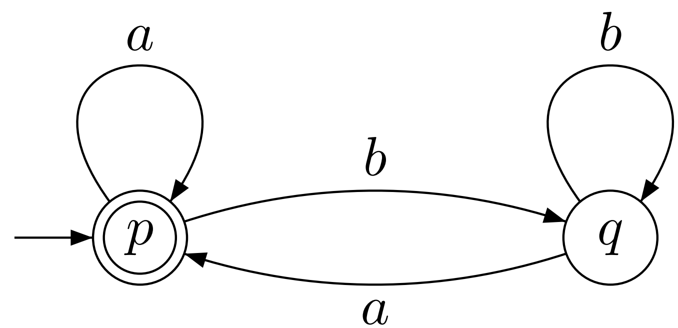
**Automates de Büchi:** Automate non-déterministe, sans état final + ensemble d'états marqués $B$.

Un mot infini est accepté s'il existe une exécution de l'automate sur le mot visitant infiniement souvent $B$.

**Exemple:**




Ultimement $b$.




Infiniement souvent $a$

**Th(Büchi 60):** Les langages acceptés par les automates de Büchi sont clos sous union, intersection, projection et **complémentation**.

# SEMIGROUPE $\Leftrightarrow$ AUTOMATE

**Structure de semigroupe:** $(S, .)$, avec

$$S = \mathcal{P}(Q \times \{0,1\} \times Q)$$
$$a.b = \{(p, \max(i,j), r) \ : \ (p,i,q) \in a,\ (q,j,r) \in b\}$$

**Morphisme de semigroupe:** Étant donné un mot $u$,

$$\varphi(u) \ni \begin{cases} (p,0,q) & \text{s'il existe une exécution de } p \text{ à } q \text{ sur } u \text{ ne visitant pas } B \\ (p,1,q) & \text{s'il existe une exécution de } p \text{ à } q \text{ sur } u \text{ visitant } B \end{cases}$$

$$\varphi(u.v) = \varphi(u).\varphi(v)$$

**Rq:** Marche pour tout type d'automate (sur mots finis/infinis, déterministes, non-déterministes, alternants, bidirectionnels. . . )

**Application de Ramsey:** Tout mot infini $u$ s'écrit:

$$u = vw_1w_2\ldots \qquad \text{avec} \quad \varphi(w_1) = \varphi(w_2) = \ldots$$

# Complémentation automates de Büchi

**Application de Ramsey:** Tout mot infini $u$ s'écrit:

$$u = vw_1w_2\ldots \qquad \text{avec} \quad \varphi(w_1) = \varphi(w_2) = \ldots$$

**Equivalence sémantique:** et dans ce cas

$$\begin{aligned} vw_1w_2\cdots \in L \qquad & \Leftrightarrow \qquad vw_1^\omega \in L \\ & \Leftrightarrow \qquad (\varphi(v), \varphi(w_1)) \in A = \{(\varphi(u), \varphi(w)) \; : \; uw^\omega \in L\} \end{aligned}$$

# Complémentation automates de Büchi

**Application de Ramsey:** Tout mot infini $u$ s'écrit:

$$u = vw_1w_2\dots \qquad \text{avec} \quad \varphi(w_1) = \varphi(w_2) = \dots$$

**Equivalence sémantique:** et dans ce cas

$$\begin{aligned} vw_1w_2\dots \in L \qquad & \Leftrightarrow \qquad vw_1^\omega \in L \\ & \Leftrightarrow \qquad (\varphi(v), \varphi(w_1)) \in A = \{(\varphi(u), \varphi(w)) \ : \ uw^\omega \in L\} \end{aligned}$$

**Donc:**

$$\complement L = \sum_{(a,b)\notin A} \varphi^{-1}(a).\varphi^{-1}(b)^\omega$$

# Complémentation automates de Büchi

**Application de Ramsey:** Tout mot infini $u$ s'écrit:

$$u = vw_1w_2\dots \qquad \text{avec} \quad \varphi(w_1) = \varphi(w_2) = \dots$$

**Equivalence sémantique:** et dans ce cas

$$\begin{aligned} vw_1w_2\cdots \in L \qquad &\Leftrightarrow \qquad vw_1^\omega \in L \\ &\Leftrightarrow \qquad (\varphi(v), \varphi(w_1)) \in A = \{(\varphi(u), \varphi(w)) \ : \ uw^\omega \in L\} \end{aligned}$$

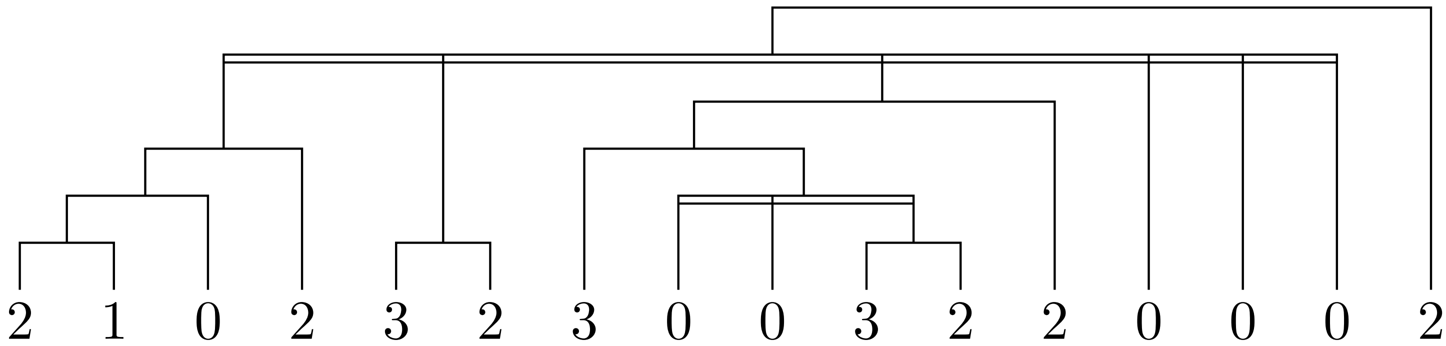
**Donc:**

$$\complement L = \sum_{(a,b) \notin A} \varphi^{-1}(a).\varphi^{-1}(b)^\omega$$

**Conlusion:** Et cette expression se traduit aisément en automate car $\varphi^{-1}(a)$ et $\varphi^{-1}(b)$ sont réguliers.

# Forêts de factorisation

On considère $S = (\mathbb{Z}/5\mathbb{Z}, +)$ et le mot '210232300322002'.

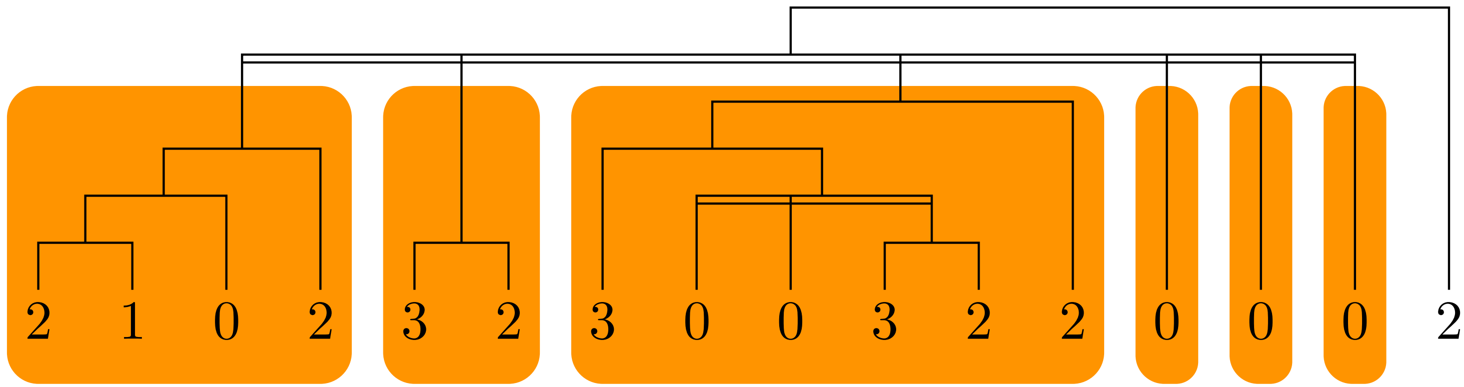



Un arbre est ramseyen si tout nœud

- est une feuille étiquetée par une lettre, ou
- a exactement deux fils,
- tous ses fils ont pour valeur le même idempotent de $S$.

# Forêts de factorisation

On considère $S = (\mathbb{Z}/5\mathbb{Z}, +)$ et le mot '210232300322002'.

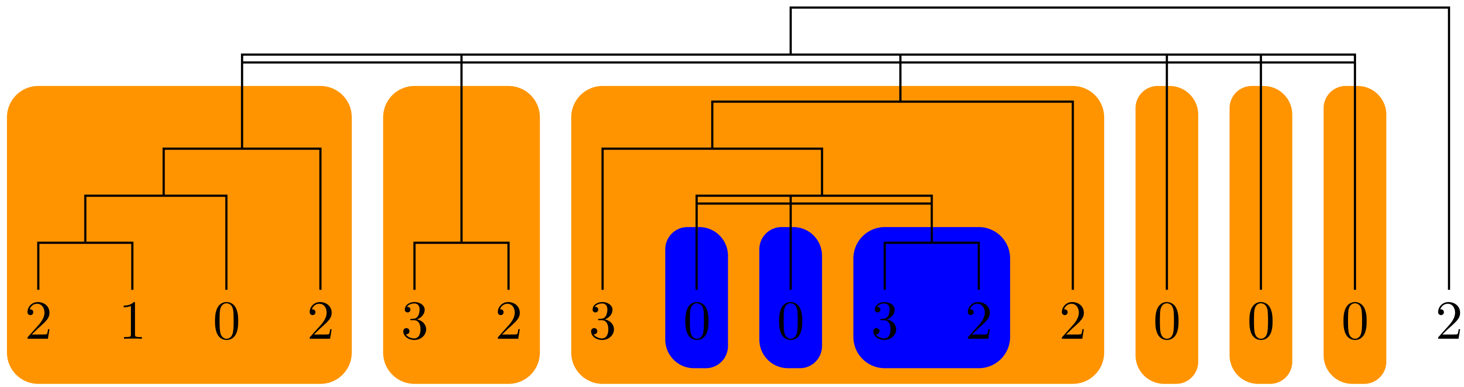



Un arbre est ramseyen si tout nœud

- est une feuille étiquetée par une lettre, ou
- a exactement deux fils,
- tous ses fils ont pour valeur le même idempotent de $S$.

# Forêts de factorisation

On considère $S = (\mathbb{Z}/5\mathbb{Z}, +)$ et le mot '210232300322002'.

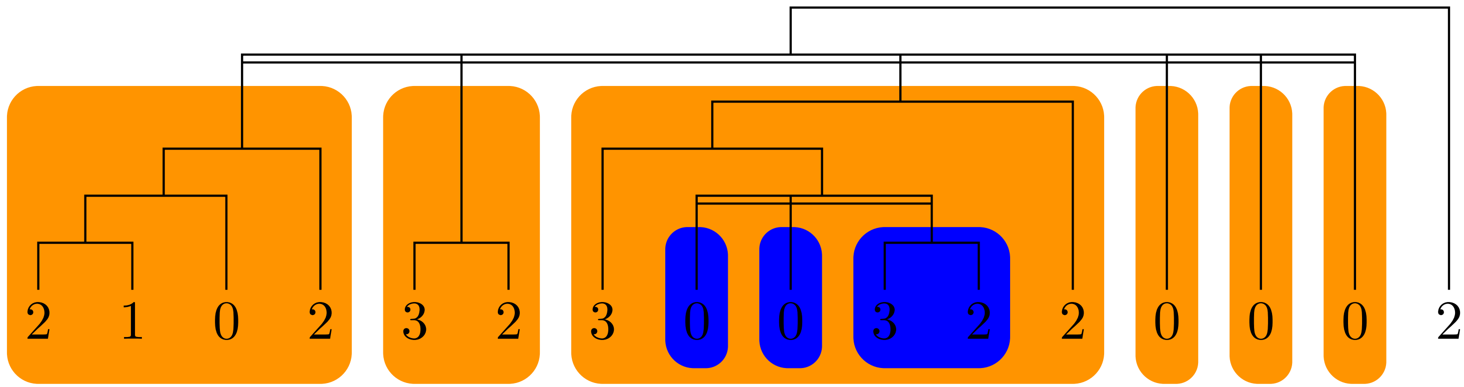



Un arbre est ramseyen si tout nœud

- est une feuille étiquetée par une lettre, ou
- a exactement deux fils,
- tous ses fils ont pour valeur le même idempotent de $S$.

# FORÊTS DE FACTORISATION

On considère $S = (\mathbb{Z}/5\mathbb{Z}, +)$ et le mot '210232300322002'.




Un arbre est ramseyen si tout nœud

- est une feuille étiquetée par une lettre, ou
- a exactement deux fils,
- tous ses fils ont pour valeur le même idempotent de $S$.

**Th(Simon 90, C07):** Tout mot sur un semigroupe fini $S$ admet un arbre de factorisation ramseyen de hauteur au plus $3|S|$.

## Présentation par expression régulières

**Th:** Soit $S$ un semigroupe fini et un morphisme de semigroupe $\varphi : A^+ \to S$. Il existe une expression régulière ramseyenne $e$ qui s'évalue en $A^*$.

Expression régulière Ramseyenne:
$L^*$ autorisé uniquement si $\varphi(L) = \{e\}$ pour $e = e^2 \in S$

# Présentation par expression régulières

**Th:** Soit $S$ un semigroupe fini et un morphisme de semigroupe $\varphi : A^+ \to S$.
Il existe une expression régulière ramseyenne $e$ qui s'évalue en $A^*$.

Expression régulière Ramseyenne:
$L^*$ autorisé uniquement si $\varphi(L) = \{e\}$ pour $e = e^2 \in S$

**Exemple:** $(S, .) = (\mathbb{Z}/2\mathbb{Z}, +)$.

$$(0 + 10^*1)^* \quad + \quad 0^*1(0 + 10^*1)^*$$

# Présentation par expression régulières

**Th:** Soit $S$ un semigroupe fini et un morphisme de semigroupe $\varphi : A^+ \to S$.
Il existe une expression régulière ramseyenne $e$ qui s'évalue en $A^*$.

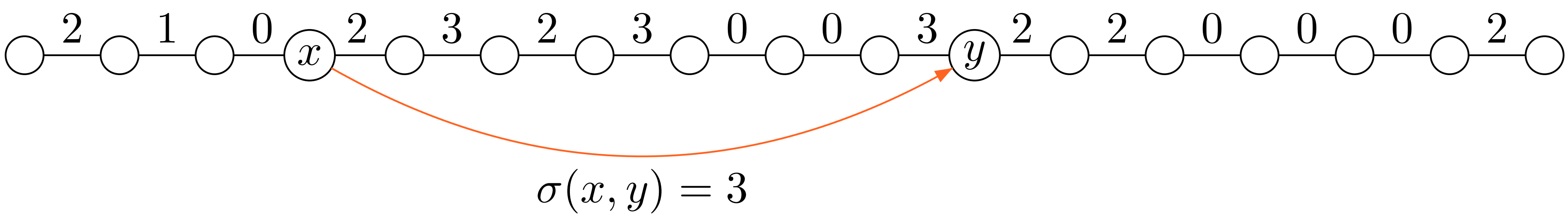
Expression régulière Ramseyenne:
$L^*$ autorisé uniquement si $\varphi(L) = \{e\}$ pour $e = e^2 \in S$

**Exemple:** $(S, .) = (\mathbb{Z}/2\mathbb{Z}, +)$.

$$(0 + 10^*1)^* \quad + \quad 0^*1(0 + 10^*1)^*$$

**Utilisation:** Effectuer une preuve par récurrence sur l'expression régulière.

# Présentation par découpage

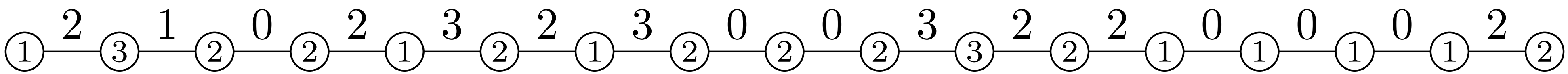



$\alpha$ ordre linéaire (total)
**Def:** Un étiquetage additif $\sigma$ est une application de $\alpha^2$ dans $S$ (semigroupe), $\sigma(x,y)$ défini ssi $x < y$, tq:

$$\forall x < y < z \in \alpha, \qquad \sigma(x,z) = \sigma(x,y).\sigma(y,z)$$

**Exemple:** $u$ mot de longueur $n$, $\alpha = ([0,n],<)$, $\sigma(x,y) = \varphi(u_{]x,y]})$.

# Présentation par découpage

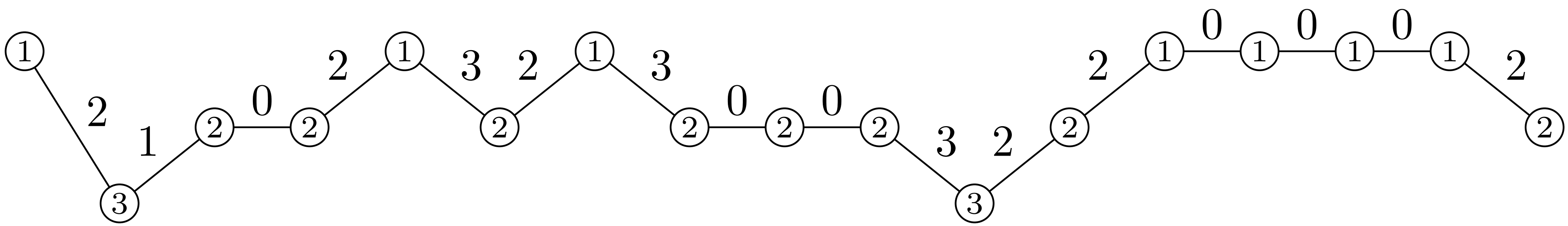


$\alpha$ ordre linéaire (total)
**Def:** Un étiquetage additif $\sigma$ est une application de $\alpha^2$ dans $S$ (semigroupe), $\sigma(x,y)$ défini ssi $x < y$, tq:

$$\forall x < y < z \in \alpha, \qquad \sigma(x,z) = \sigma(x,y).\sigma(y,z)$$

**Exemple:** $u$ mot de longueur $n$, $\alpha = ([0,n],<)$, $\sigma(x,y) = \varphi(u_{]x,y]})$.
**Def:** Un découpage est une application $s : \alpha \to [1,N]$ ($N$=profondeur).

# Présentation par découpage




$\alpha$ ordre linéaire (total)

**Def:** Un étiquetage additif $\sigma$ est une application de $\alpha^2$ dans $S$ (semigroupe), $\sigma(x, y)$ défini ssi $x < y$, tq:

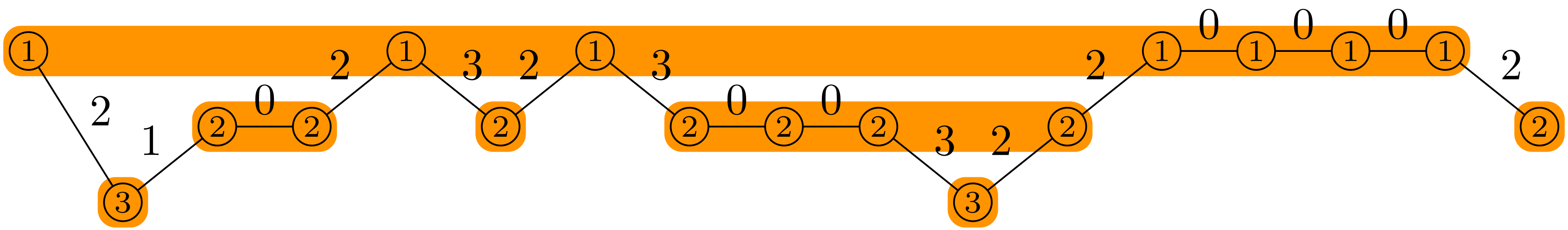
$$\forall x < y < z \in \alpha, \qquad \sigma(x, z) = \sigma(x, y).\sigma(y, z)$$

**Exemple:** $u$ mot de longueur $n$, $\alpha = ([0, n], <)$, $\sigma(x, y) = \varphi(u_{]x,y]})$.

**Def:** Un découpage est une application $s : \alpha \to [1, N]$ ($N$=profondeur).

**Def:** $x \sim_s y$ si $s(x) = s(y)$ et pour tout $z \in [\min(x, y), \max(x, y)]$, $s(z) \geq s(x)$.

# Présentation par découpage



$\alpha$ ordre linéaire (total)

**Def:** Un étiquetage additif $\sigma$ est une application de $\alpha^2$ dans $S$ (semigroupe), $\sigma(x,y)$ défini ssi $x < y$, tq:

$$\forall x < y < z \in \alpha, \qquad \sigma(x,z) = \sigma(x,y).\sigma(y,z)$$

**Exemple:** $u$ mot de longueur $n$, $\alpha = ([0,n],<)$, $\sigma(x,y) = \varphi(u_{]x,y]})$.

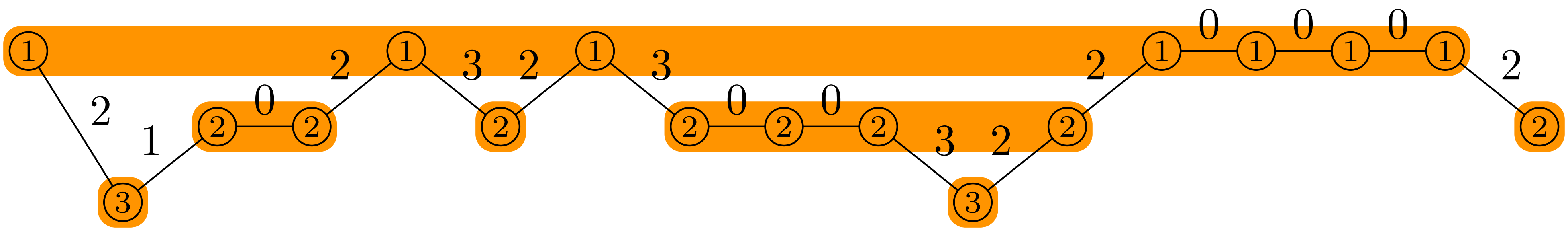
**Def:** Un découpage est une application $s : \alpha \to [1,N]$ ($N$=profondeur).

**Def:** $x \sim_s y$ si $s(x) = s(y)$ et pour tout $z \in [\min(x,y), \max(x,y)]$, $s(z) \geq s(x)$.

**Def:** Un découpage $s$ est ramseyen pour $\sigma$

$\forall x < y, x' < y'. \quad (x \sim_s y \sim_s x' \sim_s y') \to c(x,y) = c(x,y)^2 = c(x',y')$

# Présentation par découpage



$\alpha$ ordre linéaire (total)

**Def:** Un étiquetage additif $\sigma$ est une application de $\alpha^2$ dans $S$ (semigroupe), $\sigma(x,y)$ défini ssi $x<y$, tq:

$$\forall x<y<z \in \alpha, \qquad \sigma(x,z)=\sigma(x,y).\sigma(y,z)$$

**Exemple:** $u$ mot de longueur $n$, $\alpha=([0,n],<)$, $\sigma(x,y)=\varphi(u_{]x,y]})$.

**Def:** Un découpage est une application $s:\alpha\to[1,N]$ ($N$=profondeur).

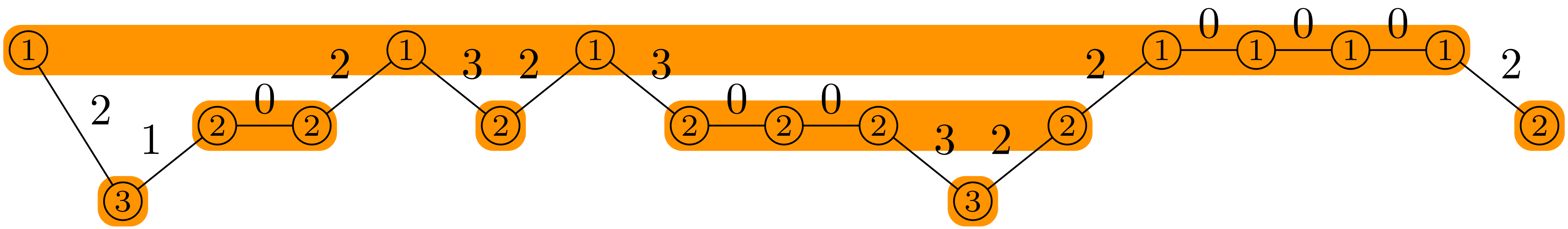
**Def:** $x\sim_s y$ si $s(x)=s(y)$ et pour tout $z\in[\min(x,y),\max(x,y)]$, $s(z)\geq s(x)$.

**Def:** Un découpage $s$ est ramseyen pour $\sigma$

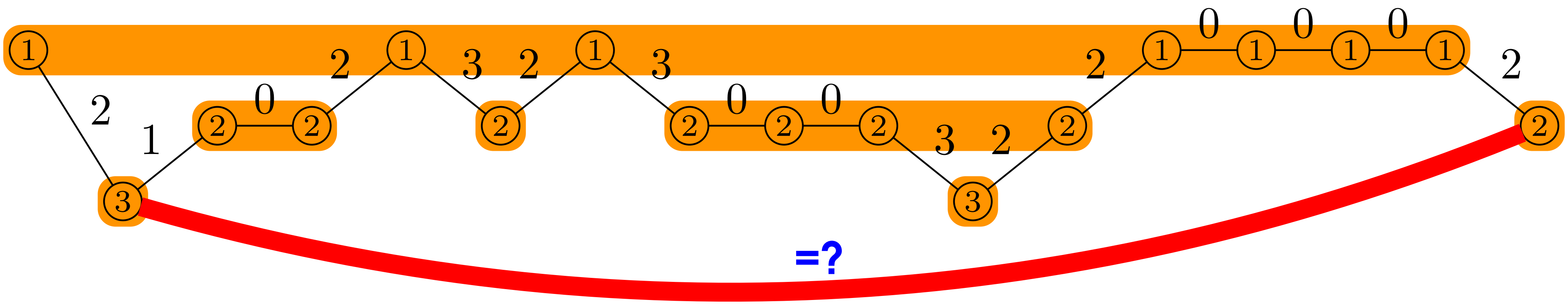
$\forall x<y, x'<y'. \quad (x\sim_s y\sim_s x'\sim_s y')\to c(x,y)=c(x,y)^2=c(x',y')$

**Th(C07):** Soit $\alpha$ fini, $S$ un semigroupe fini et $\sigma$ étiquetage additif de $\alpha$ par $S$. Il existe un découpage $s$ ramseyen pour $\sigma$ de hauteur $|S|$.

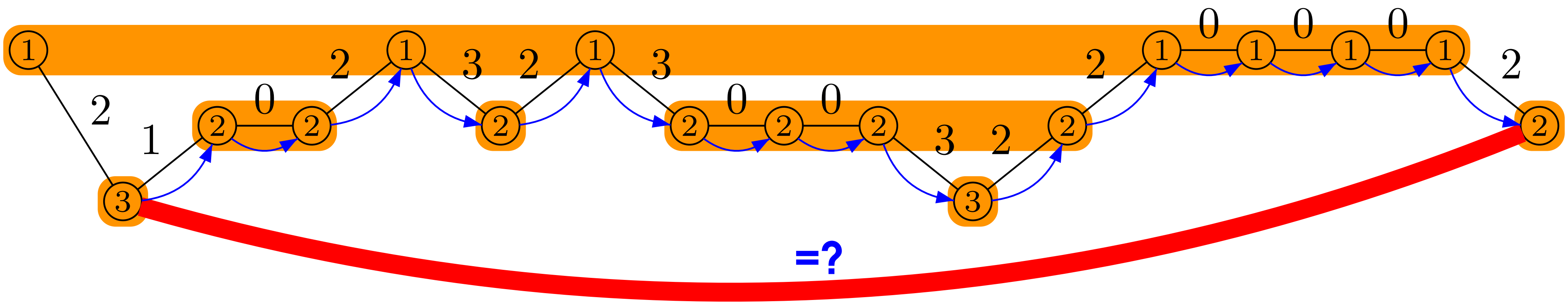
# DÉCOUPAGE: STRUCTURE D'ACCELERATION



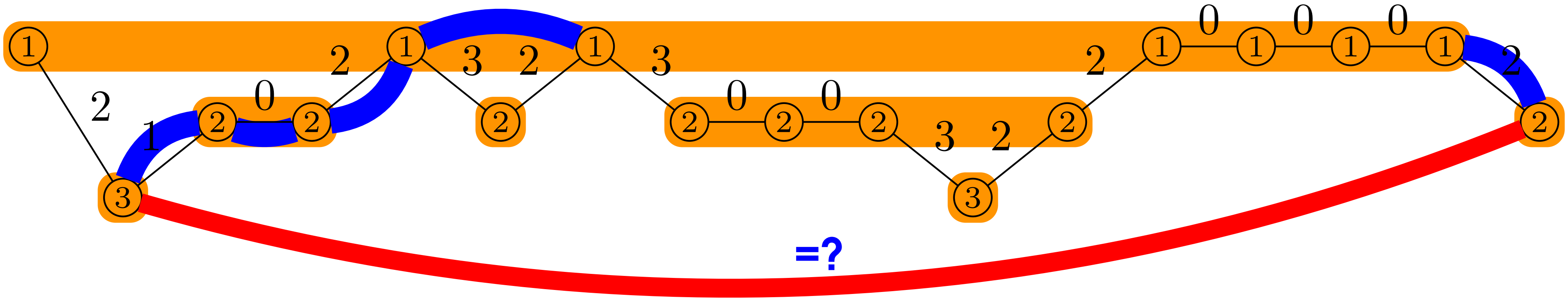
# DÉCOUPAGE: STRUCTURE D'ACCELERATION



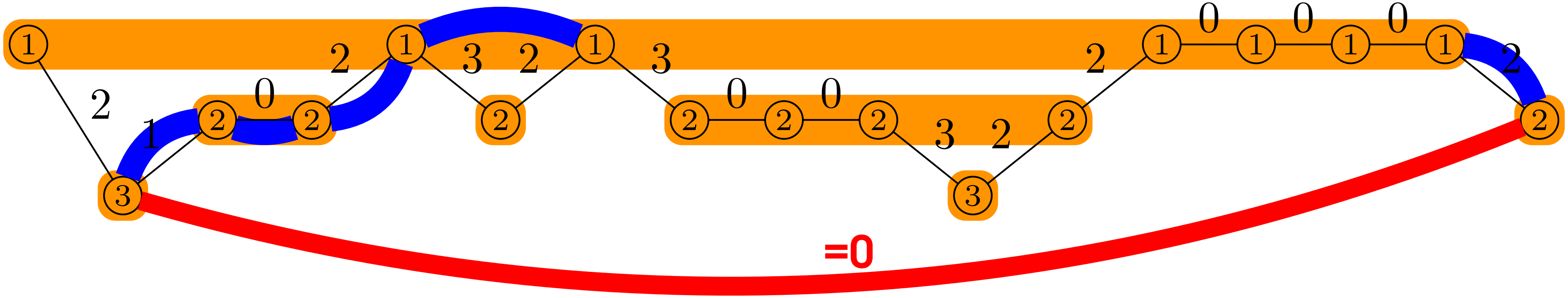
# DÉCOUPAGE: STRUCTURE D'ACCELERATION



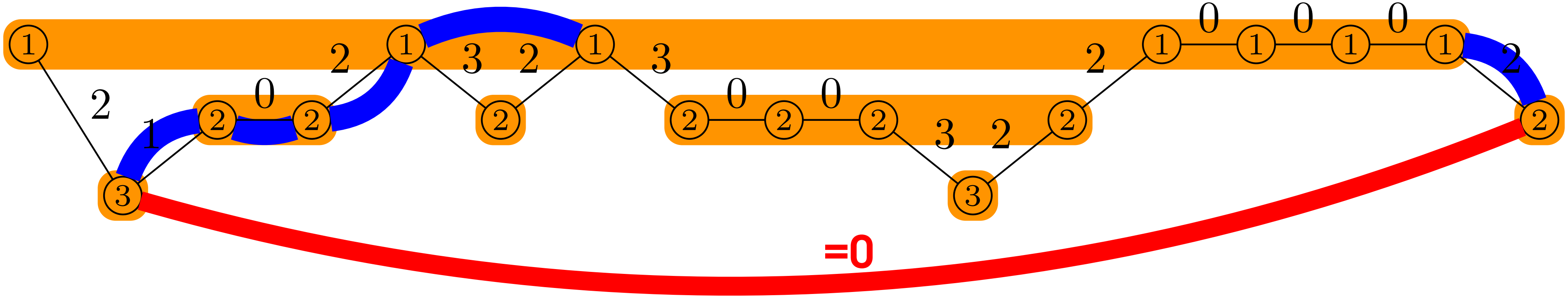
# DÉCOUPAGE: STRUCTURE D'ACCELERATION



# DÉCOUPAGE: STRUCTURE D'ACCELERATION



# DÉCOUPAGE: STRUCTURE D'ACCELERATION



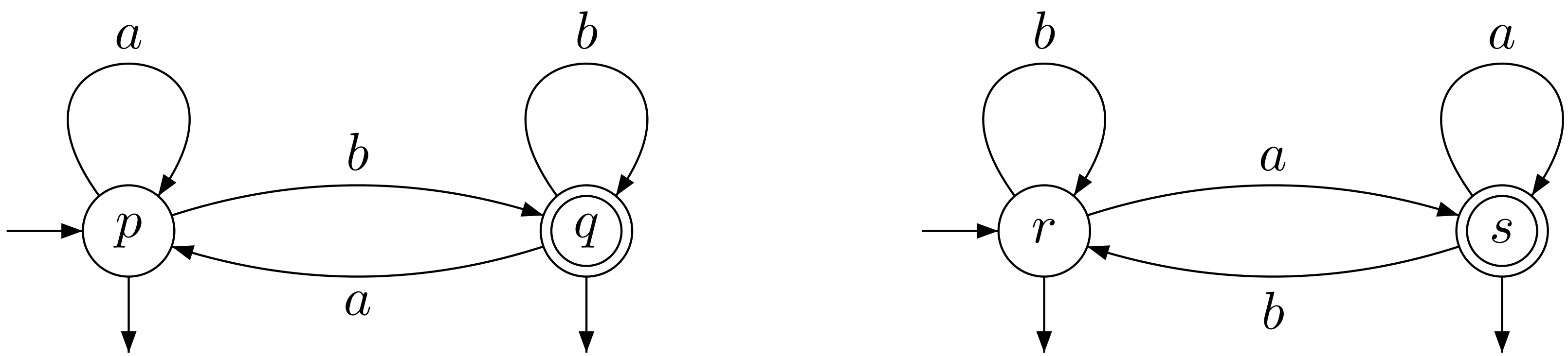

**Conclusion:** Calculer $\sigma(x, y)$ pour tout $x, y$ est linéaire dans la profondeur du découpage, et indépendant de la longueur du mot.

# Automate de distance (min,+)

**Automate de distance:** automate non-déterministe $\mathcal{A}$, dont on identifié un ensemble d'états $E$.

**Sémantique:** $|\rho|_E$ = nombre de fois qu'un état de $E$ est vu dans $\rho$.
Pour tout mot $u$, $\mathcal{A}(u)$ est le minimum de $|\rho|_E$ pour $\rho$ exécution sur $u$.

**Exemple:** $\mathcal{A}(u) = \min(|u|_a, |u|_b)$




**Problème (limitedness):** Étant donné un automate de distance $\mathcal{A}$, est-ce que la fonction $\mathcal{A} : \Sigma^* \to \mathbb{N}$ est bornée?

**Th (Hashigushi, Simon, Leung):** Le problème de 'limitedness' est décidable.

**Utilisation:** Déterminer la hauteur d'étoile d'un langage régulier (Hashigushi, Kirsten).

# VARIANTE INFINIE

**Def:** Un ordre linéaire est complet si toute partie majorée a une borne sup et toute partie minorée a une borne inf.

**Th(C07):** Soit $\alpha$ ordre linéaire complet, $S$ un semigroupe fini et $\sigma$ étiquetage additif de $\alpha$ par $S$.
Il existe un découpage $s$ ramseyen pour $\sigma$ de hauteur $3|S|$.

# Variante infinie

**Def:** Un ordre linéaire est complet si toute partie majorée a une borne sup et toute partie minorée a une borne inf.

**Th(C07):** Soit $\alpha$ ordre linéaire complet, $S$ un semigroupe fini et $\sigma$ étiquetage additif de $\alpha$ par $S$.
Il existe un découpage $s$ ramseyen pour $\sigma$ de hauteur $3|S|$.

**Application:** Complémentation des automates sur les ordres linéaires dispersés dénombrables **(Carton&Rispal)**.

# Variante déterministe

**Question:** Est-il possible de calculer la factorisation de manière déterministe (en lisant le mot de la gauche vers la droite) ?

**Réponse:** Impossible!

# VARIANTE DÉTERMINISTE

**Question:** Est-il possible de calculer la factorisation de manière déterministe (en lisant le mot de la gauche vers la droite) ?
**Réponse:** Impossible!

**Solution:** Utiliser une factorisation 'ramseyenne avant':
$\forall x < y, x' < y'. \quad (x \sim_s y \sim_s x' \sim_s y') \rightarrow \sigma(x,y).\sigma(x',y') = c(x,y)$.

**Th(C07):** Soit $\alpha$ fini, $S$ un semigroupe fini et $\sigma$ étiquetage additif de $\alpha$ par $S$. Il existe un découpage $s$ ramseyen avant pour $\sigma$ de hauteur $|S|$ tel que

- $s(x)$ ne dépend que de $\sigma|_{\leq x}$,
- $s(x)$ est calculable par automate fini.

# Variante déterministe

**Question:** Est-il possible de calculer la factorisation de manière déterministe (en lisant le mot de la gauche vers la droite) ?
**Réponse:** Impossible!

**Solution:** Utiliser une factorisation 'ramseyenne avant':
$\forall x < y, x' < y'. \quad (x \sim_s y \sim_s x' \sim_s y') \to \sigma(x,y).\sigma(x',y') = c(x,y)$.

**Th(C07):** Soit $\alpha$ fini, $S$ un semigroupe fini et $\sigma$ étiquetage additif de $\alpha$ par $S$. Il existe un découpage $s$ ramseyen avant pour $\sigma$ de hauteur $|S|$ tel que

- $s(x)$ ne dépend que de $\sigma|_{\leq x}$,
- $s(x)$ est calculable par automate fini.

**Application1:** Déterminisation des automates de Büchi en automates à parité (McNaughton 66).

# Variante déterministe

**Question:** Est-il possible de calculer la factorisation de manière déterministe (en lisant le mot de la gauche vers la droite) ?
**Réponse:** Impossible!

**Solution:** Utiliser une factorisation 'ramseyenne avant':
$\forall x < y, x' < y'. \quad (x \sim_s y \sim_s x' \sim_s y') \to \sigma(x,y).\sigma(x',y') = c(x,y)$.

**Th(C07):** Soit $\alpha$ fini, $S$ un semigroupe fini et $\sigma$ étiquetage additif de $\alpha$ par $S$. Il existe un découpage $s$ ramseyen avant pour $\sigma$ de hauteur $|S|$ tel que

- $s(x)$ ne dépend que de $\sigma|_{\leq x}$,
- $s(x)$ est calculable par automate fini.

**Application1:** Déterminisation des automates de Büchi en automates à parité (McNaughton 66).

**Application2** (Accélération sur les arbres):
Toute expressions MSO $\Phi(x_1, \dots, x_N)$ est équivalente à une expression FO $\phi(x_1, \dots, x_N)$ qui utilise des formules MSO unaires $\Psi(x)$.

# VARIANTE DÉTERMINISTE

**Question:** Est-il possible de calculer la factorisation de manière déterministe (en lisant le mot de la gauche vers la droite) ?
**Réponse:** Impossible!

**Solution:** Utiliser une factorisation 'ramseyenne avant':
$\forall x < y, x' < y'. \quad (x \sim_s y \sim_s x' \sim_s y') \to \sigma(x,y).\sigma(x',y') = c(x,y)$.

**Th(C07):** Soit $\alpha$ fini, $S$ un semigroupe fini et $\sigma$ étiquetage additif de $\alpha$ par $S$. Il existe un découpage $s$ ramseyen avant pour $\sigma$ de hauteur $|S|$ tel que

- $s(x)$ ne dépend que de $\sigma|_{\leq x}$,
- $s(x)$ est calculable par automate fini.

**Application1:** Déterminisation des automates de Büchi en automates à parité (McNaughton 66).

**Application2** (Accélération sur les arbres):
Toute expressions MSO $\Phi(x_1, \ldots, x_N)$ est équivalente à une expression FO $\phi(x_1, \ldots, x_N)$ qui utilise des formules MSO unaires $\Psi(x)$.

**Application3:** Une structure est MSO-interprétable dans un arbre ssi elle est FO-interprétable dans un arbre.

# RÉCAPITULATIF

**Théorème de Ramsey infini:**
Complémentation des automates sur les mots de longueur $\omega$.

# RÉCAPITULATIF

**Théorème de Ramsey infini:**
Complémentation des automates sur les mots de longueur $\omega$.

**Théorème de forêts de factorisation:**
Technique d'accélération
Automates de distance
+ Caractérisation de logiques
+ Complémentation des langage $\omega BS$

# Récapitulatif

**Théorème de Ramsey infini:**
Complémentation des automates sur les mots de longueur $\omega$.

**Théorème de forêts de factorisation:**
Technique d'accélération
Automates de distance
+ Caractérisation de logiques
+ Complémentation des langage $\omega BS$

**Variante pour ordres linéaires infinis:**
Complémentation des ordres infinis dispersés
+ Théorie des boréliens de $\mathbb{R}$ (en cours).

# RÉCAPITULATIF

**Théorème de Ramsey infini:**
Complémentation des automates sur les mots de longueur $\omega$.

**Théorème de forêts de factorisation:**
Technique d'accélération
Automates de distance
+ Caractérisation de logiques
+ Complémentation des langage $\omega BS$

**Variante pour ordres linéaires infinis:**
Complémentation des ordres infinis dispersés
+ Théorie des boréliens de $\mathbb{R}$ (en cours).

**Variante déterministe:**
Déterminisation de McNaughton
"Équivalence" MSO/FO sur les arbres.
+ ...