Chapitre 1

Mots et motifs

1.1 Définitions usuelles

Soit Σ un *alphabet* fini, c'est-à-dire un ensemble fini dont les éléments sont appelés *lettres* (les alphabets que nous considèrerons seront toujours finis). Le *monoïde libre* engendré par Σ, noté Σ^*, est l'ensemble des suites finies d'éléments de Σ, ou *mots*, y compris le *mot vide* ε, muni de l'opération de *concaténation* (ou simplement *produit*) qui consiste à mettre deux mots bout à bout. Ainsi, si $\Sigma = \{a, b\}$, le produit des deux mots $u = abaab$ et $v = bba$ est $uv = abaabbba$.

La *longueur* d'un mot u est le nombre de lettres qui composent ce mot, et on la note $|u|$. On a bien sûr $|\varepsilon| = 0$, $|x| = 1$ si $x \in \Sigma$, et $|uv| = |u| + |v|$ pour deux mots quelconques u et v. Pour tout mot u, on note $\mathrm{Alph}\,(u)$ l'ensemble des lettres de Σ qui apparaissent effectivement dans u.

Si u et v sont deux mots de Σ^*, on dit que u est un *facteur* de v s'il existe deux autres mots x et y tels que $v = xuy$. Si de plus x est le mot vide, le mot u est dit *préfixe* de v (*préfixe propre* si y n'est pas vide); si $y = \varepsilon$, le mot u est dit *suffixe* de v (*suffixe propre* si $x \neq \varepsilon$).

Un *langage* est un ensemble de mots, c'est-à-dire une partie (finie ou infinie) de Σ^*. On note $F(L)$ l'ensemble de tous les facteurs de tous les éléments d'un langage L; un langage tel que $F(L)$, qui a la propriété qu'un facteur d'un de ses éléments est encore un de ses éléments, est dit *langage factoriel*.

L'*image miroir* d'un mot $x = x_1x_2 \ldots x_m$ est le mot, que l'on note $x^\sim$, obtenu en inversant l'ordre des lettres de x, à savoir $x^\sim = x_m x_{m-1} \ldots x_1$. Les mots x tels que $x = x^\sim$ sont appelés *palindromes*.

Un *mot infini* $\mathbf{u}$ sur Σ est une suite $(u_n)_{n\in\mathbb{N}^*}$ à valeurs dans Σ, et on note Σ^ω l'ensemble des mots infinis sur Σ. Le monoïde Σ^* agit à gauche sur Σ^ω de la manière suivante: si $x = x_1x_2 \ldots x_l$ est un mot de longueur l et $\mathbf{u}$ est un mot infini, alors $\mathbf{v} = x\mathbf{u}$ est défini par la suite $(v_n)_{n\in\mathbb{N}^*}$ avec $v_n = x_n$ pour $1 \leqslant n \leqslant l$ et $v_n = u_{n+l}$ pour $l < n$. On peut étendre aux mots infinis les notions de facteur, préfixe et suffixe : un mot fini u est facteur de $\mathbf{v}$ s'il existe x fini et $\mathbf{y}$ infini tels que $\mathbf{v} = xu\mathbf{y}$, c'est un préfixe si $x = \varepsilon$; un mot infini $\mathbf{u}$ est suffixe de $\mathbf{v}$ s'il existe x fini tel que $\mathbf{v} = x\mathbf{u}$. On note $F(\mathbf{u})$ l'ensemble de tous les facteurs du mot infini $\mathbf{u}$.

On peut également définir des *mots biinfinis*, qui formellement sont les orbites sous l'action du *décalage* σ d'éléments de $\Sigma^{\mathbb{Z}}$: si $x\colon n \mapsto x(n)$ est un élément de $\Sigma^{\mathbb{Z}}$, $\sigma(x)$ est défini par $\sigma(x)(n) = x(n+1)$ et représente le même mot biinfini que x. On note ${}^{\omega}\Sigma^{\omega}$ l'ensemble des mots biinfinis sur Σ.

On dit qu'une application $f\colon \Sigma^* \to \Sigma'^*$ est un *morphisme de monoïdes libres* si $f(uv) = f(u)f(v)$ pour tous mots u et v de Σ^*. Un morphisme est uniquement déterminé par l'image des lettres de Σ. Si aucune lettre n'a pour image le mot vide, on dit que le morphisme f est *non-effaçant*. Si les images de toutes les lettres sont de longueur au moins 2, on dit que f est *partout croissant*. Enfin, si toutes les images de lettres ont même longueur, on dit que f est *uniforme*.

Pour qu'un morphisme f soit injectif, il faut et il suffit d'une part que deux lettres distinctes aient des images distinctes (on dit alors que f est *injectif sur les lettres*), et d'autre part que l'ensemble $C = f(\Sigma)$ des images de lettres forme un code sur Σ', c'est-à-dire que tout mot de Σ'^* s'écrive d'au plus une manière comme produit d'éléments de C. On dira parfois que f est un code pour dire que f est injectif.

1.2 Motifs

Pour parler de motifs, nous avons besoin d'un autre alphabet, Δ. Les lettres de Δ sont appelées *variables* alors que celles de Σ sont des *constantes*. Pour les distinguer facilement, nous emploierons des minuscules pour les constantes et des majuscules pour les variables.

Un *motif sans constante*, ou simplement *motif*, est un élément du monoïde Δ^*. Un motif est donc un mot dont toutes les lettres sont des variables.

Étant donnés un mot $u \in \Sigma^*$ et un motif $p \in \Delta^*$, on dit que *u contient p* (ou contient une *occurrence* de p) s'il existe un morphisme $h\colon \Delta^* \to \Sigma^*$ non-effaçant (c'est-à-dire que l'image d'une variable n'est jamais le mot vide) tel que $h(p)$ est facteur de u. Dans le cas contraire, on dit que *u évite p*. Par exemple, le motif $AABBA$ est contenu dans les mots $abbaabb$ (avec $h(A) = b$ et $h(B) = a$) et $cbcabcabababcaca$ (avec $h(A) = bca$ et $h(B) = ba$) mais est évité par $aaabaabbbaaabaaaa$. Cette définition s'applique aussi si u est un mot infini. On peut également l'étendre aux langages : on dit qu'un langage $L \subset \Sigma^*$ contient p si l'un des éléments de L contient p, et que L évite p si tous ses éléments évitent p.

1.3 Évitabilité

1.3.1 Définition

Le mot *évitabilité* regroupent plusieurs notions voisines : dans tous les cas, on dit qu'un motif est évitable s'il existe un langage infini qui l'évite, les différences résidant dans les langages autorisés.

Proposition 1.1. *Soit Σ un alphabet fini et p un motif. Les trois assertions suivantes sont équivalentes :*

(i) pour tout entier n, il existe un mot $u \in \Sigma^*$ de longueur supérieure ou égale à n tel que u évite p;

(ii) il existe un langage infini $L \in \mathscr{P}(\Sigma^*)$ qui évite p;

(iii) il existe un mot infini $\mathbf{u} \in \Sigma^\omega$ qui évite p.

Preuve : Il est clair que *(i)* et *(ii)* sont équivalentes : sur un alphabet fini, un langage est infini si et seulement s'il contient des mots arbitrairement longs.

Si *(iii)* est vraie, l'ensemble $F(\mathbf{u})$ des facteurs du mot infini $\mathbf{u}$ est un langage infini qui évite p, d'où *(ii)*. Réciproquement, si p est évité par un langage infini L, soit $\mathbf{u}$ un mot infini dans l'adhérence de L, c'est-à-dire que tout préfixe de $\mathbf{u}$ est préfixe d'une infinité d'éléments de L (un tel mot se construit aisément par récurrence); comme tous les préfixes de $\mathbf{u}$ évitent p, il en est de même de $\mathbf{u}$. ■

L'alphabet Σ étant fixé, on dit qu'un motif p est *évitable sur* Σ s'il vérifie les trois assertions équivalentes de la proposition 1.1. Dans le cas contraire, on dit naturellement que p est *inévitable sur* Σ.

Quand un motif p est évitable sur un alphabet Σ, il est évidemment évitable sur tous les alphabets de même cardinal. Si on ne veut pas préciser Σ, on dira dans ce cas que p est *k-évitable*, où k est le cardinal de Σ.

Si on ne veut pas non plus préciser k, on dira que p est *évitable*, tout court. On dira donc que p est *inévitable* s'il est inévitable sur tout alphabet fini.

1.3.2 Indice d'évitabilité

Il est clair que les différentes notions définies ci-dessus sont hiérarchisées de la manière suivante :

$$\text{2-évitable} \Longrightarrow \text{3-évitable} \Longrightarrow \text{4-évitable} \Longrightarrow \cdots \Longrightarrow \text{évitable}$$

et inversement

$$\text{inévitable} \Longrightarrow \cdots \Longrightarrow \text{4-inévitable} \Longrightarrow \text{3-inévitable} \Longrightarrow \text{2-inévitable}$$

ce qui permet de définir l'*indice d'évitabilité* d'un motif p, noté $\alpha(p)$: si p est évitable, c'est le plus petit entier k tel que p est k-évitable et $(k-1)$-inévitable; si p est inévitable, $\alpha(p) = +\infty$ par convention.

1.3.3 Exemples

Si l'alphabet Σ n'a qu'un seul élément, tous les motifs sont inévitables sur Σ, puisqu'un mot $w \in \Sigma^*$ contient le motif $p \in \Delta^*$ dès que $|w| \geqslant |p|$. On supposera désormais qu'il y a au moins deux lettres distinctes dans Σ.

Si Σ a au moins deux éléments, il existe des motifs évitables sur Σ, par exemple $ABABA$ qui est évité par le mot infini de Thue et Morse. L'indice d'évitabilité de ce motif est donc $\alpha(ABABA) = 2$.

Le motif AA est 2-inévitable, car tout mot de longueur au moins 4 sur un alphabet à deux lettres contient un carré. Nous verrons plus loin qu'il est 3-évitable : $\alpha(AA) = 3$.

Par contre, le motif ABA est k-inévitable pour tout k, car tout mot de longueur $2k+1$ sur un alphabet à k lettres contient une lettre a répétée au moins 3 fois, donc contient un facteur de la forme axa où x est un mot non vide : ainsi ABA est un exemple de motif inévitable, $\alpha(ABA) = \infty$.

1.3.4 $\mathscr{L}$-évitabilité

Nous verrons plus loin qu'il peut être intéressant de considérer d'autres restrictions que la taille de l'alphabet. Plaçons-nous donc d'emblée dans le cadre le plus général possible : soit $\mathscr{L}$ une famille quelconque de langages; on dit qu'un motif p est *$\mathscr{L}$-évitable* s'il existe un langage infini $L \in \mathscr{L}$ qui évite p.

On retrouve les définitions précédentes pour des valeurs particulières de $\mathscr{L}$. Ainsi, un motif évitable sur Σ est $\mathscr{P}(\Sigma^*)$-évitable, un motif k-évitable est $\mathscr{P}(\Sigma_k^*)$-évitable avec $\Sigma_k = \{0, 1, 2, \ldots, k-1\}$, et un motif évitable est $\mathscr{L}$-évitable avec $\mathscr{L} = \bigcup_{k=1}^{\infty} \mathscr{P}(\Sigma_k^*)$.

On peut aussi essayer de prendre pour $\mathscr{L}$ des classes de langages usuelles. Par exemple, si $\mathscr{L} = \mathrm{RAT}(\Sigma)$ est l'ensemble des langages rationnels sur Σ, aucun motif n'est $\mathscr{L}$-évitable car d'après le lemme de l'étoile tout langage rationnel infini contient un sous-langage de la forme uv^*w (avec $v \neq \varepsilon$), donc contient comme facteur l'image de n'importe quel motif par le morphisme qui à chaque variable associe le mot v (voir la proposition 1.5). La situation est similaire avec la classe des langages algébriques. Comme nous le verrons plus loin, la théorie des L-systèmes fournit des exemples plus intéressants.

1.4 Premières propriétés

Commençons par énumérer quelques propriétés élémentaires des motifs.

1.4.1 Symétries

Quand on connaît les propriétés d'évitabilité d'un certain motif, on peut facilement, par des arguments de symétrie, déduire que plusieurs autres motifs ont les mêmes propriétés. Ainsi :

Proposition 1.2. *Si un motif p est $\mathscr{L}$-évitable, alors pour toute permutation σ de l'alphabet Δ le motif $\sigma(p)$ est également $\mathscr{L}$-évitable. En fait, p et $\sigma(p)$ sont évités par les mêmes mots.* ■

et

Proposition 1.3. *Si un motif p est $\mathscr{L}$-évitable et si la classe $\mathscr{L}$ est stable par image miroir (ce qui est le cas de toutes les classes qu'on utilisera), alors l'image miroir $p^\sim$ de p est également $\mathscr{L}$-évitable.* ■

1.4.2 Divisibilité

On dit qu'un motif $p \in \Delta^*$ *divise* un autre motif $q \in \Delta^*$ s'il existe un morphisme non-effaçant $h\colon \Delta^* \to \Delta^*$ tel que $h(p)$ est facteur de q; en d'autres termes, si q, vu comme un mot et non comme un motif, contient le motif p. Il est alors clair que

Proposition 1.4. *Si un motif p est $\mathscr{L}$-évitable et divise un motif q, alors ce dernier est également $\mathscr{L}$-évitable.* ■

Par exemple, sachant que AA est 3-évitable, cette proposition permet de conclure immédiatement que tous les motifs qui contiennent des carrés, par exemple $ABCBABCB$, sont également 3-évitables.

On notera que la relation de divisibilité est une relation d'ordre sur Δ^*, et que pour cet ordre l'indice d'évitabilité est une fonction décroissante de Δ^* dans $\mathbb{N} \cup \{\infty\}$

1.4.3 Langages répétitifs et motifs

Citons enfin une propriété très simple, mais qui est un argument décisif dans plusieurs preuves (nous nous en sommes d'ailleurs déjà servi en 1.3.4). On dit qu'un langage L est *répétitif* si $F(L)$ contient des puissances arbitraires, c'est-à-dire si pour tout entier n il existe un mot non vide u tel que u^n est facteur d'au moins un élément de L.

Proposition 1.5. *Si L est un langage répétitif et p un motif, alors L n'évite pas p.*

Preuve : Soit n la longueur du motif p : par hypothèse, il existe un mot non vide u tel que u^n est facteur d'un mot v de L. Soit alors h le morphisme de Δ^* dans Σ^* qui à toute variable $X \in \Delta$ associe le mot $u \in \Sigma^*$. Comme u n'est pas vide, h est non-effaçant. Le mot v contient $h(p) = u^n$ comme facteur, donc L n'évite pas p. ∎

1.5 Motifs avec constantes, formules

Dans cette section nous présentons trois généralisations possibles des motifs, qui peuvent d'ailleurs se combiner à volonté.

1.5.1 Motifs avec constantes

Tout d'abord, nous pouvons ajouter des lettres constantes dans le motif : un *motif avec constantes* est un élément de $(\Delta \cup \Sigma)^*$. On dit qu'un mot w contient un motif avec constantes p s'il existe un morphisme $h : (\Delta \cup \Sigma)^* \to \Sigma^*$ non-effaçant et laissant les éléments de Σ^* inchangés tel que $h(p)$ est facteur de u.

Comme l'alphabet Σ est attaché au motif p, on ne peut plus parler de motif évitable ou k-évitable, mais seulement de motif évitable sur Σ, ce qui n'est d'ailleurs pas très intéressant : si p contient au moins une constante a, et si Σ contient au moins un autre élément b, alors p est évité par les mots b^m pour tout entier m donc est évitable sur Σ.

Les motifs avec constantes présentent toutefois un intérêt quand il s'agit de construire des mots évitant des motifs. En particulier, ils constituent un passage obligé dans notre algorithme pour décider si un DOL-langage évite un motif (cf. chapitre 6).

1.5.2 Ensembles de motifs

Il est parfois utile de dire que des motifs sont simultanément évitables. Pour cela, nous dirons qu'un mot w contient un ensemble de motifs $P \in \mathscr{P}(\Delta^*)$ s'il contient au moins l'un des éléments de P. On définit de même les ensembles de motifs évitables sur Σ, k-évitables, évitables, $\mathscr{L}$-évitables, etc.

On notera par exemple qu'un ensemble fini de motifs est évitable si et seulement si chacun de ses éléments est évitable. Cette condition est évidemment nécessaire, montrons qu'elle est suffisante : soit $P = \{p_1, \ldots, p_m\}$ un ensemble de motifs. On suppose que chacun des motifs p_i est évitable, c'est-à-dire qu'il existe un mot infini $\mathbf{u}_i = u_{i0}u_{i1}u_{i2}\ldots$ sur un alphabet Σ_i qui évite p_i. On considère alors le mot infini $\mathbf{v} = v_0 v_1 v_2 \dots$ sur l'alphabet $\Sigma = \Sigma_1 \times \cdots \times \Sigma_m$ défini par $v_n = (u_{1n}, u_{2n}, \dots, u_{mn})$. Il est clair que $\mathbf{v}$ évite chacun des motifs p_i, car si $\mathbf{v}$ avait un facteur de la forme $h(p_i)$, alors $\mathbf{u}_i$ aurait un facteur de la forme $(\pi \circ h)(p_i)$ où π est la projection de Σ sur Σ_i. Évidemment, l'alphabet Σ est bien plus grand que les alphabets Σ_i, donc ce raisonnement ne s'applique pas à l'évitabilité sur un alphabet fixé. Par exemple, l'ensemble de motifs $\{AAA, AABBA\}$ est 2-inévitable, bien que chacun de ses éléments soit 2-évitable, comme nous le verrons plus loin (2.4).

On remarquera que les ensembles de mots de [51] sont un cas particulier d'ensembles de motifs avec constantes.

1.5.3 Formules

Formules simples

Kirby Baker a proposé une autre extension des motifs : les formules. Baker, McNulty et Taylor considèrent dans [5] le motif

$$U_\Delta = ABWBCXCAYBAZAC$$

qui est 4-évitable mais 3-inévitable. Ils font remarquer que ce qui est important dans ce motif, ce sont les facteurs AB, BC, CA, BA et AC, et non l'ordre dans lequel ils interviennent ni ce qui les sépare. On peut exprimer cela en disant que le motif U_Δ est associé à la formule $f = AB.BC.CA.BA.AC$, où on a remplacé toutes les variables *isolées* de p (i.e. qui n'apparaissent qu'une fois) par des points.

Une *formule simple* est la donnée d'un ensemble fini de motifs sur un alphabet Δ, et on la note conventionnellement en juxtaposant les motifs qui la composent, dans un ordre quelconque, séparés par des points. En d'autres termes, l'ensemble des formules simples est le monoïde commutatif engendré par les motifs non vides, dont la loi ., que nous appellerons *superposition des formules*, vérifie $f.f = f$ pour toute formule simple f.

On dit alors qu'un mot w contient la formule simple $f = p_1.p_2.\dots.p_m$ s'il existe un morphisme non-effaçant $h \colon \Delta^* \to \Sigma^*$ tel que les mots $h(p_1)$, $h(p_2)$, ..., $h(p_m)$ soient *tous* des facteurs de w. On peut alors parler de formule simple évitable, etc.

Il faut bien distinguer cette définition de celle, donnée dans le paragraphe précédent, d'un mot contenant un ensemble de motifs. La différence tient à un quantificateur : un mot w contient l'ensemble de motifs $\{p_1, \dots, p_m\}$ s'il existe un morphisme non-effaçant h tel que *l'un des mots* $h(p_1)$, ..., $h(p_m)$ soit facteur de w.

L'algèbre des formules

Il est naturel de chercher à combiner ces deux notions qui semblent très proches : nous appellerons *formule* un ensemble fini de formules simples. Nous noterons désormais les ensembles de motifs de manière additive : une formule est donc une somme (finie) de formules simples. L'ensemble des formules $\mathscr{F}$, que nous appellerons *algèbre des formules* (même si ce n'est pas à proprement parler une algèbre, puisque l'addition n'est pas une loi de groupe), est muni de deux lois de composition interne commutatives et associatives $+$ et $.$ (on définit le produit de deux sommes de façon à avoir la distributivité). Nous noterons comme d'habitude 0 l'élément neutre de $+$, qui correspond à l'ensemble de motifs vide, et est évité par tous les mots, et 1 l'élément neutre de $.$, qui, comme le motif vide (auquel nous l'identifierons bientôt) est contenu dans tous les mots.

En cas d'ambiguïté, nous préciserons $\mathscr{F} = \mathscr{F}(\Delta, \Sigma)$ pour parler des formules à variables dans Δ avec constantes dans Σ, ou $\mathscr{F} = \mathscr{F}(\Delta) = \mathscr{F}(\Delta, \emptyset)$ pour les formules sans constante.

Équivalence de formules

L'un des intérêts majeurs des formules est qu'il existe des manières de les transformer sans changer leurs propriétés d'évitabilité. Pour cela, nous définissons une relation d'équivalence $\sim$ sur l'algèbre des formules $\mathscr{F}(\Delta, \Sigma)$ (avec $\Sigma = \emptyset$ pour les formules sans constante), appelée simplement *équivalence*, qui a la propriété de conserver la k-évitabilité.

Soit Δ' un sous-ensemble de Δ. On dit qu'une formule f divise une formule f' *en fixant* Δ' si toute formule simple de f', vue comme un mot sur l'alphabet $\Delta \cup \Sigma \cup \{.\}$, contient f, vue comme une formule à variables dans $\Delta \setminus \Delta'$ et à constantes dans $\Sigma \cup \Delta'$, avec la restriction supplémentaire que la « lettre » $.$ ne peut pas apparaître dans l'image d'une variable.

En d'autres termes, si $f = \sum_{i=1}^{m} \prod_{p \in P_i} p$ et $f' = \sum_{j=1}^{m'} \prod_{q \in P'_j} q$, on dit que f divise f' en fixant Δ' si et seulement si pour tout $j \in [1, m']$, il existe $h \colon (\Delta \cup \Sigma)^* \to (\Delta \cup \Sigma)^*$, morphisme non-effaçant qui fixe $\Delta' \cup \Sigma$ tel que

$$\exists i \in [1, m], \forall p \in P_i, \exists q \in P'_j, h(p) \text{ est facteur de } q \ .$$

Il faut bien faire attention que, si dans une formule le signe $+$ peut être lu « ou » puisque « w contient $f_1 + f_2$ » signifie « w contient f_1 ou w contient f_2 », sa signification est inversée à droite de la relation de divisibilité : « f divise $f_1 + f_2$ » signifie « f divise f_1 et f divise f_2 ».

Si Σ' est un alphabet contenant Σ et $h \colon \Delta'^* \to \Sigma'^*$ un morphisme non-effaçant, on appelle *spécialisation de f par h* et on note $h(f)$ la formule de $\mathscr{F}(\Delta \setminus \Delta', \Sigma')$ obtenue en remplaçant dans f toute occurrence d'une variable

$X \in \Delta'$ par son image $h(X)$. Il est clair qu'un mot de Σ'^* contient f si et seulement s'il contient l'une de ses spécialisations : si $[f] = [f]_{\Sigma'}$ désigne l'ensemble des mots de Σ'^* qui contiennent une occurrence de f, on a

$$[f]_{\Sigma'} = \bigcup_{h\colon \Delta' \to \Sigma'^+} [h(f)]_{\Sigma'} \ .$$

Pour tout Δ', on définit une relation $\sim_{\Delta'}$ comme la plus petite relation d'équivalence vérifiant les propriétés suivantes, où f et f' sont deux formules :

(i) si σ est une permutation de Δ fixant Δ', alors $\sigma(f) \sim_{\Delta'} f$;

(ii) si $f = \varepsilon$ est la formule vide, $f \sim_{\Delta'} 1$.

(iii) si $f \sim_{\Delta'} f'$, alors $f + f'' \sim_{\Delta'} f' + f''$;

(iv) si $f \sim_{\Delta'} f'$, et si $(\mathrm{Alph}\,(f) \cup \mathrm{Alph}\,(f')) \cap \mathrm{Alph}\,(f'')$ est inclus dans $\Delta' \cup \Sigma$, alors $f.f'' \sim_{\Delta'} f'.f''$;

(v) si f divise f' en fixant Δ', alors $f + f' \sim_{\Delta'} f$;

(vi) si f divise f' en fixant Δ', et si $\mathrm{Alph}\,(f) \cap \mathrm{Alph}\,(f')$ est inclus dans $\Delta' \cup \Sigma$, alors $f.f' \sim_{\Delta'} f'$;

(vii) si $\Delta'' \subset \Delta'$ et $f \sim_{\Delta'} f'$, alors $f \sim_{\Delta''} f'$;

(viii) si p et p' sont des motifs et Z est une variable qui n'est ni dans Δ', ni dans $\mathrm{Alph}\,(p) \cup \mathrm{Alph}\,(p')$, alors $p.p' \sim_{\Delta'} pZp'$.

Si $f \sim_{\Delta'} f'$, on dira que f *est équivalente à* f' *en fixant* Δ'. Si $\Delta' = \emptyset$, on dit simplement que f *est équivalente à* f' et on note $f \sim f'$.

Les propriétés *(i)* à *(vii)* sont assez naturelles; toute la difficulté est de comprendre quelles variables doivent être fixées. La propriété *(viii)* est plus originale : elle traduit le fait que quand une variable n'apparaît qu'une seule fois dans un motif, elle sépare ce dernier en deux parties qui, dans une occurrence du motif, peuvent apparaître séparées par n'importe quoi, presque comme dans la formule composée de ces deux parties, sauf que pour la formule l'ordre des parties n'est pas imposé et elles peuvent même se superposer. Nous allons voir que cette différence n'est pas gênante. Montrons d'abord le lemme technique suivant :

Lemme 1.1. *Soient Δ' une partie de Δ et Σ' un alphabet contenant Σ. Soient f et f' deux formules équivalentes en fixant Δ'. Alors, pour tout morphisme non-effaçant $h\colon \Delta'^* \to \Sigma'^*$, et pour tout entier M, il existe un entier n tel que*

$$\left([h(f)]_{\Sigma'} \cap \Sigma'^{\leqslant M}\right)^n \subset [h(f')]_{\Sigma'}$$

où $\Sigma'^{\leqslant M}$ est l'ensemble des mots de Σ'^* de longueur inférieure ou égale à M.

Preuve : Il suffit de vérifier que cette assertion est compatible avec chacune des propriétés définissant $\sim_{\Delta'}$, et avec la transitivité.

Pour *(i)* et *(ii)*, il est clair que $[h(\sigma(f))] = [h(f)]$ et que $[h(\varepsilon)] = [h(1)] = \Sigma'^*$ donc l'assertion est vérifiée avec $n = 1$.

Pour *(iii)*, en supposant l'assertion vraie pour f et f', on voit qu'elle reste vraie avec la même valeur de n pour $f+f''$ et $f'+f''$. En effet, on a $[h(f+f'')] = [h(f)] \cup [h(f'')]$, donc un produit de n éléments de $[h(f+f'')]$ de longueur inférieure ou égale à M est soit dans $[h(f'')]$ (dès que l'un des éléments est dans cet ensemble), soit dans $\left([h(f)] \cap \Sigma'^{\leqslant M}\right)^n$, donc par hypothèse dans $[h(f')]$: dans les deux cas il est dans $[h(f'+f'')]$.

Pour *(iv)*, le raisonnement est similaire, la condition sur les alphabets garantissant que $[h(f.f'')] = [h(f)] \cap [h(f'')]$ et que $[h(f'.f'')] = [h(f')] \cap [h(f'')]$.

Pour *(v)*, on a $[h(f')] \subset [h(f)]$, donc $[h(f+f')] = [h(f)] \cup [h(f')] = [h(f)]$ et l'assertion est vraie avec $n = 1$.

Pour *(vi)*, on a $[h(f')] \subset [h(f)]$, donc (grâce à la condition sur les alphabets) $[h(f.f')] = [h(f)] \cap [h(f')] = [h(f')]$, et on conclut de même.

Pour *(vii)*, soit $h' \colon \Delta''^* \to \Sigma'^*$ un morphisme non-effaçant et M un entier. On a $[h'(f)] = \bigcup_{h''} [h''(h'(f))]$, où h'' décrit l'ensemble (infini) des morphismes non-effaçants de $(\Delta' \setminus \Delta'')^*$ dans Σ'^*, et de même pour f'; mais comme on s'intéresse à l'ensemble fini $[h'(f)] \cap \Sigma'^{\leqslant M}$, il suffit de considérer un ensemble fini $H = \{h''_1, \ldots, h''_q\}$ de morphismes pour avoir

$$[h'(f)] \cap \Sigma'^{\leqslant M} = \bigcup_{h'' \in H} [h''(h'(f))] \cap \Sigma'^{\leqslant M} \quad .$$

À chacun de ces morphismes $h''_i \in H$, on associe un morphisme non-effaçant $h_i \colon \Delta'^* \to \Sigma'^*$ tel que $h_i(f) = h''_i(h'(f))$ et $h_i(f') = h''_i(h'(f'))$. On construit par récurrence des suites (M_i) et (n_i) pour $1 \leqslant i \leqslant q$ comme suit : $M_1 = M$; pour tout i, par hypothèse, il existe un entier n_i tel que l'assertion concernant $\sim_{\Delta'}$ soit vérifiée pour h_i et M_i. On pose alors $M_{i+1} = n_i M_i$, et on vérifie que

$$\left(\bigcup_{1 \leqslant j \leqslant i} [h_j(f)] \cap \Sigma'^{\leqslant M}\right)^{M_{i+1}/M} \subset \bigcup_{1 \leqslant j \leqslant i} \left([h_j(f)] \cap \Sigma'^{\leqslant M_j}\right)^{n_j} \quad .$$

Soit alors $n = \prod_{1 \leqslant i \leqslant q} n_i$. Tout mot de $\left([h'(f)] \cap \Sigma'^{\leqslant M}\right)^n$ est dans au moins un des ensembles $\left([h_i(f)] \cap \Sigma'^{\leqslant M_i}\right)^{n_i}$, donc dans $[h''_i(h'(f'))]$, qui est inclus dans $[h'(f')]$.

Pour *(viii)*, l'assertion est clairement vérifiée si on prend $f = pZp'$ et $f' = p.p'$ avec $n = 1$, puisque $[pZp'] \subset [p.p']$. Dans l'autre sens, c'est-à-dire pour $f = p.p'$ et $f' = pZp'$, soit $n = 1 + 2\#\left([p.p'] \cap \Sigma'^{\leqslant M}\right)$ ($\#E$ désigne le cardinal d'un ensemble E). Parmi n éléments de $[p.p'] \cap \Sigma'^{\leqslant M}$, au moins deux non consécutifs seront donc égaux. Tout élément de $\left([h(p.p')] \cap \Sigma'^{\leqslant M}\right)^n$ contient donc un facteur de la forme uvu, où $u \in [h(p.p')]$ et $v \in \Sigma'^{+}$, donc contient une occurrence de $h(pZp')$, ce qui permet de conclure.

Enfin, la transitivité est conservée : supposons l'assertion vraie pour les formules f et f', et pour les formules f' et f''. Étant donnés h et M, on obtient un entier n pour f et f'. Soit alors M' la longueur maximale des mots de $\left([h(f)] \cap \Sigma'^{\leqslant M}\right)^{n}$. Appliquée à h et M', l'assertion pour f' et f'' donne un entier n', et on a

$$\left([h(f)] \cap \Sigma'^{\leqslant M}\right)^{n''} \subset [h(f'')]$$

avec $n'' = nn'$. ■

Nous pouvons maintenant démontrer le résultat principal :

Théorème 1.1. *Soient f et f' deux formules équivalentes, et Σ' un alphabet. Si f' est évitable sur Σ', alors f est aussi évitable sur Σ'.*

Preuve : Supposons que f est inévitable sur Σ'. D'après le lemme 1.1, appliqué à $\Delta' = \emptyset$, pour tout entier M il existe un entier n tel que $\left([f] \cap \Sigma'^{\leqslant M}\right)^{n} \subset [f']$. Puisque f est inévitable sur Σ', on peut choisir l'entier M suffisamment grand pour que tout mot de longueur M contienne au moins une occurrence de f. Tout mot de longueur nM est donc dans $\left([f] \cap \Sigma'^{\leqslant M}\right)^{n}$, donc contient f', ce qui veut dire que f' est inévitable sur Σ'. ■

Ce théorème ne s'applique pas à la $\mathscr{L}$-évitabilité pour une classe de langages quelconque, car il n'est pas possible de borner la longueur des occurrences de f. Toutefois, on peut l'étendre au cas des classes de langages *uniformément récurrents*, c'est-à-dire aux classes $\mathscr{L}$ telles que pour tout langage $L \in \mathscr{L}$ et pour tout mot $u \in L$, il existe un entier M tel que tout élément de longueur M de $F(L)$ contient u comme facteur.

Théorème 1.2. *Soient f et f' deux formules équivalentes, et $\mathscr{L}$ une classe de langages uniformément récurrents. Si f' est $\mathscr{L}$-évitable, alors f est aussi $\mathscr{L}$-évitable.*

Preuve : Supposons que f est $\mathscr{L}$-inévitable. Soit $L \in \mathscr{L}$ un langage infini : il existe donc un mot $u \in L$ qui contient f. Comme L est uniformément récurrent, il existe donc un entier M tel que tout élément de longueur M de $F(L)$ contient u comme facteur, et donc contient f. D'après le lemme 1.1, il existe un entier n tel que $\left([f] \cap \Sigma'^{\leqslant M}\right)^{n} \subset [f']$. Tout élément de $F(L)$ de longueur nM contient donc f', donc L n'évite pas f'. ■

1.5.4 Répétitions

Les répétitions sont une autre forme de régularité, qui a été largement étudiée [21, 43, 48]. On dit qu'un mot w contient une répétition d'exposant α s'il existe deux facteurs non vides u et p de w tels que p est une période de u (c'est-à-dire que u est facteur d'une puissance de p) et que $\frac{|u|}{|p|} = \alpha$.

Le cas le plus simple et celui des exposants entiers, et il s'exprime aisément en termes de motifs : un mot contient une répétition d'exposant n si et seulement il contient le motif A^n.

Un autre cas qui se ramène aux motifs et celui des *chevauchements* : on dit qu'un mot contient un chevauchement s'il contient une répétition d'exposant strictement plus grand que 2. Or une répétition d'exposant strictement compris entre 2 et 3 est de la forme $u = xyxyx$, où x et y sont deux mots non vides, donc un mot contient un chevauchement si et seulement s'il contient l'un des motifs AAA et $ABABA$, c'est-à-dire s'il contient la formule $AAA + ABABA$.

Le cas des exposants fractionnaires sort du cadre des motifs. Cependant, certains résultats sur les répétitions peuvent être utiles à l'étude des motifs. Par exemple, tout mot contenant le motif $ABCBABC$ contient une répétition d'exposant strictement supérieur à $\frac{3}{2}$: en effet, un tel mot contient un facteur de la forme $xyzyxyz$ où x, y et z sont des mots non vides, c'est-à-dire une répétition de période $xyzy$ et d'exposant $\frac{|xyzyxyz|}{|xyzy|} = \frac{3}{2} + \frac{|xz|}{2|xyzy|} > \frac{3}{2}$. Comme Pansiot [48] a montré qu'il existe un mot infini sur un alphabet à 4 lettres sans répétition d'exposant strictement supérieur à $\frac{7}{5}$, on en déduit que $ABCBABC$ est 4-évitable (voir aussi [18]).

1.6 Variantes

Nous mentionnons ici quelques façons de modifier la définition de l'évitabilité (ou k-évitabilité, $\mathscr{L}$-évitabilité, etc.) d'un motif. Nous n'avons pas étudié plus en détail ces variantes, mais il est possible que les méthodes développées dans cette thèse conduisent à des résultats intéressants les concernant.

1.6.1 Évitabilité commutative

On dit qu'un mot $u \in \Sigma^*$ *contient commutativement* un motif $p \in \Delta^*$, que l'on notera $p = X_1X_2\ldots X_m$ où les X_i sont des variables, s'il existe m mots u_1, u_2, ..., u_m non vides tels que $u_1u_2\ldots u_m$ est facteur de u et que, pour tout couple (i, j) tel que $X_i = X_j$, les mots u_i et u_j ne diffèrent que par une permutation de leurs lettres. Par exemple, le mot $bbbaaababbbaaa$ contient commutativement $AABBA$, avec $u_1 = baa$, $u_2 = aba$, $u_3 = b$, $u_4 = b$, $u_5 = baa$. On définit ensuite les motifs *commutativement k-évitables*, etc. de manière habituelle, en remplaçant *contient* par *contient commutativement*.

Il est clair que si u contient p au sens habituel, alors il contient aussi p commutativement. Tout motif k-évitable commutativement est donc aussi k-évitable, mais la réciproque n'est pas vraie (par exemple AA n'est pas commutativement 3-évitable). Pour cette raison, James Currie [17] appelle ces motifs *fortement k-évitables.*

L'origine de cette définition est le problème des *carrés abéliens*, posé par Erdös [28, p. 240] : le motif AA est-il commutativement 4-évitable ? La réponse positive à ce problème a finalement été donnée par Keränen [34] au moyen d'un morphisme uniforme de longueur 85 (voir aussi [23, 27, 49]).

Currie pose également la question suivante : existe-t-il un algorithme pour décider si un motif donné est commutativement évitable (sur un alphabet arbitraire), de la même manière que l'algorithme de Zimin permet de décider si un motif est évitable. Or il se trouve que les seuls motifs non commutativement évitables que nous connaissions sont les motifs inévitables. Il serait donc intéressant d'essayer d'en construire d'autres; s'il s'avérait que ce n'est pas possible, l'algorithme de Zimin permettrait donc de décider également l'évitabilité commutative.

1.6.2 l-occurrences

Entringer, Jackson et Schatz [27] ont construit une suite binaire dont les seuls carrés sont de longueur inférieure ou égale à 4. De manière générale, on peut se demander, quand un motif est k-inévitable, si cela est dû simplement au fait qu'un certain nombre d'images du motif forment un ensemble inévitable de mots, ou s'il y a une raison plus globale. Il s'agit donc d'affaiblir la notion de k-évitabilité en ne considérant que les occurrences suffisamment longues du motif.

Quand l'alphabet des variables a deux éléments ou plus, il y a deux manières de procéder. La première consiste à ignorer toutes les occurrences $h(p)$ du motif p dont la longueur totale est inférieure à un entier l donné; ceci n'a à notre connaissance jamais été étudié. La seconde, encore plus tolérante, consiste à ne considérer que les morphismes h tels que l'image de toute variable soit de longueur au moins l. Dans ces conditions, Peter Roth [52] appelle le mot $h(p)$ une l-occurrence de p. On notera que des images de p arbitrairement longues peuvent être éliminées de cette manière, pour peu que l'image de l'une des variables soit courte.

On peut donc appeler *faiblement k-évitable* un motif p pour lequel il existe un entier l et un mot infini sur l'alphabet à k lettres sans l-occurrence de p. Peter Roth a montré que tout motif évitable était faiblement k-évitable pour tout $k \geqslant 2$, $k = 2$ en particulier, ce qui fournit une nouvelle caractérisation des motifs évitables. D'autres questions restent ouvertes : par exemple, pour un motif p donné, quelle est la plus petite valeur de l pour laquelle les l-occurrences de p sont 2-évitables ?

Dans les deux sections précédentes, nous avons décrit un certain nombre d'extensions possibles. Toutefois, il semble que les problèmes essentiels se posent déjà dans le cas le plus simple : la k-évitabilité des motifs ordinaires. Dans les chapitres suivants, nous nous y intéresserons en priorité, mais parfois il sera plus avantageux pour cela de travailler dans le cadre de l'une des généralisations.

1.7 Évitabilité sur un alphabet non fixé

Le problème de l'évitabilité sur un alphabet non fixé a été étudié en détail par Zimin [65] et Bean, Ehrenfeucht et McNulty [6]. En particulier, ils ont montré qu'on pouvait décider si un motif donné est évitable ou non. Nous présentons ici cet algorithme en termes de graphes d'adjacence, d'après [5].

1.7.1 Quelques propriétés

La proposition suivante illustre ce que peut apporter le fait que la taille de l'alphabet n'est pas fixée :

Proposition 1.6. *Un ensemble fini de motifs est évitable si et seulement si chacun de ses éléments pris individuellement est évitable.*

Preuve : Soit $P = \{p_1, \ldots, p_m\}$ un ensemble fini de motifs. Si l'un des p_i est inévitable, il est clair que P est également inévitable. Supposons maintenant que tous les p_i sont évitables : il existe donc des alphabets Σ_1, ..., Σ_m (ici Σ_i ne désigne pas l'alphabet à i éléments !) et des mots infinis $\mathbf{u}_1 \in \Sigma_1^\omega$, ..., $\mathbf{u}_m \in \Sigma_m^\omega$ tels que $\mathbf{u}_i$ évite p_i. On considère alors le mot infini $\mathbf{u} \in \Sigma^\omega$, avec $\Sigma = \Sigma_1 \times \cdots \times \Sigma_m$, tel que la projection sur ${\Sigma_i}^\omega$ de $\mathbf{u}$ soit égale à $\mathbf{u}_i$: ce mot infini évite simultanément tous les p_i, donc évite P. ■

La proposition qui suit est une conséquence immédiate de l'algorithme exposé plus bas :

Proposition 1.7. *Tout motif inévitable contient une variable isolée (c'est-à-dire une variable qui n'apparaît qu'une seule fois).* ■

Elle permet de montrer que

Proposition 1.8. *Un motif inévitable comprenant n variables différentes est de longueur au plus $2^n - 1$.*

Preuve : La preuve est par récurrence sur n. Le seul motif inévitable formé avec une seule variable est A, de longueur $1 = 2^1 - 1$. Supposons maintenant la proposition démontrée pour les alphabets de variables de taille inférieure à n.

Soit p un motif comprenant n variables différentes. D'après la proposition 1.7, l'une de ses variables X n'apparaît qu'une fois. On a alors $p = p_1 X p_2$, où p_1 et p_2 sont des motifs inévitables qui contiennent au plus $n-1$ variables différentes. Par hypothèse de récurrence, $|p_1| \leqslant 2^{n-1}-1$ et $|p_2| \leqslant 2^{n-1}-1$, donc $|p| = |p_1| + 1 + |p_2| \leqslant 2^n - 1$. ■

Cette borne $2^n - 1$ est optimale, car on peut construire des motifs de cette longueur par le procédé suivant : soit $Z_1 = A_1$ et pour tout $n > 1$, $Z_n = Z_{n-1} A_n Z_{n-1}$ (les Z_n sont parfois appelés *sesquipuissances*, ou *motifs de Zimin*, et ont été introduits par Aršon [3]). Il est clair que l'alphabet des variables de Z_n a n éléments, et que la longueur de Z_n est $2^n - 1$.

Proposition 1.9. *Les motifs Z_n sont tous inévitables.*

Preuve : Cette propriété est évidente quand on la voit du point de vue des formules : en effet, $Z_n = Z_{n-1} A_n Z_{n-1}$ est équivalent à $Z_{n-1}.Z_{n-1}$, puisque la variable A_n est isolée. Mais cette dernière formule est elle-même égale à Z_{n-1}. En itérant cette opération, on voit que tous les Z_n sont équivalents, au sens des formules, à Z_1, et même au motif vide, qui est bien entendu inévitable. ■

Enfin, en combinant les propositions 1.6 et 1.8, on obtient le résultat surprenant suivant :

Proposition 1.10. *Pour un alphabet de variables Δ donné, il existe un alphabet Σ et un mot infini sur Σ qui évite tous les motifs évitables de Δ^*.*

Preuve : L'ensemble des motifs évitables de Δ^* est infini : pour pouvoir appliquer la proposition 1.6, il faut se ramener à un ensemble fini. On se limite donc aux motifs évitables de longueur au plus 2^n, où n est le cardinal de Δ, et on peut alors construire un mot infini $\mathbf{u}$ qui les évite tous simultanément au moyen de la proposition 1.6. Si p est un motif évitable de longueur supérieure à 2^n, soit q un facteur de longueur 2^n de p : d'après la proposition 1.8, q est évitable, donc est évité par $\mathbf{u}$, donc p est aussi évité par $\mathbf{u}$: on a montré que $\mathbf{u}$ évite tous les motifs évitables. ■

1.7.2 L'algorithme de Zimin-Bean-Ehrenfeucht-McNulty

Passons maintenant au résultat principal de Zimin [65] et Bean, Ehrenfeucht, McNulty [6] :

Théorème 1.3. *Un motif est évitable si et seulement s'il est irréductible.*

L'*irréductibilité* d'un motif, définie ci-dessous à la manière de [5], est une propriété décidable.

Soit $p \in \Delta^*$ un motif. On appelle *graphe d'adjacence* de p le graphe biparti $AG(p)$, dont les sommets sont étiquetés par deux copies de Δ, que nous noterons Δ^L et Δ^R quand nous aurons besoin de les distinguer, et dont les arêtes sont les (X^L, Y^R) pour tout XY facteur de p. Ainsi, le graphe d'adjacence de $ABACBA$ (figure 1.1) comporte 6 sommets et 4 arêtes.

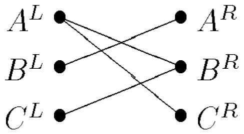

FIG. 1.1 - *Le graphe d'adjacence de $ABACBA$*

Une partie non vide F de Δ est dite *libre* (pour p) s'il n'existe pas de chemin dans $AG(p)$ allant d'un sommet de gauche X^L à un sommet de droite Y^R avec X et Y dans F. Pour déterminer les ensembles libres, il est donc utile de repérer les composantes connexes de $AG(p)$. Dans l'exemple ci-dessus, il y a deux composantes connexes et les ensembles libres sont $\{A\}$ et $\{B\}$.

On dit que le motif *p se réduit en p' en une étape* s'il existe un ensemble de variables F libre pour p tel que p' soit obtenu en effaçant dans p les variables de F. On dit que p se réduit en p' s'il existe une suite de réductions en une étape partant de p et arrivant à p'. Enfin, on dit que p est *réductible* s'il existe une réduction de p au motif vide, et *irréductible* sinon.

Dans l'exemple ci-dessus, $p = ABACBA$ se réduit en BCB si on efface l'ensemble libre $\{A\}$, et BCB lui-même se réduit en C puis en ε. Le motif p est donc réductible. Toutefois, si on avait commencé par effacer l'ensemble libre $\{B\}$, on aurait obtenu le motif irréductible $AACA$: pour s'assurer qu'un motif est irréductible, il faut donc explorer récursivement toutes les possibilités de réduction et vérifier qu'aucune ne conduit au motif vide, ce qui donne l'algorithme de la figure 1.2, que nous appellerons pour simplifier *algorithme de Zimin.*

Il convient de noter qu'il est parfois indispensable d'effacer des ensembles libres autres que des singletons, par exemple pour le motif

$$p = ABACA'BADABA'CA'BA'$$

dû à Bean, Ehrenfeucht et McNulty, pour lequel l'ensemble à effacer est $\{A, A'\}$.

1.8 Motifs verrouillés

Certains motifs ont un comportement particulièrement simple pour l'algorithme exposé ci-dessus : il s'agit des motifs dont le graphe d'adjacence est

Algorithme *réduction*
Entrée : un motif p
Sortie : un booléen, indique si p est réductible

Début
    Calculer le graphe d'adjacence de p
    Calculer l'ensemble $\mathscr{F}$ des ensembles libres pour p
    Pour tout $F \in \mathscr{F}$:
        Calculer p', motif obtenu en effaçant les variables de F dans p
        Appeler *réduction* avec pour entrée p'
        Si la réponse est *vrai*, rendre *vrai*
    Rendre *faux*
Fin.

FIG. 1.2 - *L'algorithme de Zimin*

connexe, et pour lesquels aucune réduction n'est donc possible. De tels motifs sont dits *verrouillés*. Un motif p est verrouillé si et seulement si, quelles que soient les variables X et Y qui apparaissent effectivement dans p, il existe un entier $m \geqslant 1$ et $2m$ variables $X_1, ..., X_m, Y_1, ..., Y_m$ telles que $X_1 = X$, $Y_m = Y$, et que les $2m-1$ motifs X_iY_i (pour $1 \leqslant i \leqslant m$) et $X_{i+1}Y_i$ (pour $1 \leqslant i < m$) soient facteurs de p.

Baker, McNulty et Taylor [5] ont montré que les motifs verrouillés sont tous 4-évitables. Plus précisément :

Théorème 1.4. *Les motifs verrouillés sont tous évités par le mot infini* $\Omega_{2,1}$, *point fixe du morphisme suivant :*

$$\begin{array}{rrcl} h_{2,1}: & \{a,b,c,d\} & \longrightarrow & \{a,b,c,d\} \quad . \\ & a & \longmapsto & ab \\ & b & \longmapsto & cb \\ & c & \longmapsto & ad \\ & d & \longmapsto & cd \end{array}$$

■

Le mot $\Omega_{2,1}$ a été introduit par Dean [20]. C'est ainsi que Baker, McNulty et Taylor prouvent que le motif $U_\Delta = ABWBCXCAYBAZAC$ est 4-évitable, puisqu'il est verrouillé. Pour ce motif, il est plus naturel de parler de la formule équivalente $f_\Delta = AB.BC.CA.BA.AC$, qui n'a que trois variables distinctes. On dira encore que cette formule est verrouillée : on peut en effet construire le graphe d'adjacence d'une formule simple, avec une arête entre X^L et Y^R

quand XY est facteur de l'un des éléments de la formule. Pour la formule f_Δ, on obtient le graphe d'adjacence de la figure 1.3.

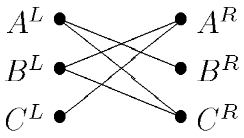

FIG. 1.3 - *Le graphe d'adjacence de f_Δ*

On a alors les propriétés suivantes :

Proposition 1.11. *Si le motif $p = p_0Y_1p_1 \dots p_{m-1}Y_mp_m$, où les variables Y_i sont isolées, est verrouillé, alors la formule équivalente $f = p_0.p_1.\dots.p_{m-1}.p_m$ est verrouillée.*

Preuve : Le graphe d'adjacence de p diffère de celui de f par les sommets Y_i^L et Y_i^R; de chacun de ces sommets part au plus une arête, donc cela ne peut pas faire diminuer le nombre de composantes connexes du graphe. Cela peut par contre en ajouter : ainsi $p = AABCA$ n'est pas verrouillé alors que $f = AA.A$ est verrouillée. ∎

Proposition 1.12. *Un motif p est verrouillé si et seulement si la formule simple f dont les éléments sont les facteurs de longueur 2 de p (dite* 2-hachage *de p) est verrouillée.*

Preuve : Cette proposition découle immédiatement du fait que le graphe d'adjacence de p dépend uniquement de l'ensemble des facteurs de longueur 2 de p. ∎

Plus généralement, on appelle *k-hachage* d'une formule simple f la formule simple dont les éléments sont les facteurs de longueur inférieure ou égale à k des composantes de f, dans laquelle on supprime les éléments qui sont facteurs d'autres éléments plus grands. Pour $k = 2$ et si f est un motif, cette définition coïncide avec celle de la proposition 1.12, sauf pour le motif réduit à une variable qui de toutes manières n'est pas verrouillé. Par exemple, le 3-hachage de $ABCB.BC.BA$ est $ABC.BCB.BA$. Le k-hachage d'une formule divise toujours cette formule, donc est *a priori* plus difficile à éviter. Il est donc raisonnable, si on cherche par exemple à construire des motifs évitables qui ne soient pas 4-évitables (l'existence de tels motifs est un problème ouvert, cf [5]), de considérer des k-hachages pour un entier k petit. Toutefois, $k = 2$ ne convient pas :

Proposition 1.13. *Tout 2-hachage est soit 4-évitable, soit inévitable.*

Preuve : Soit f un 2-hachage; si f a des composantes de longueur 1, il s'agit nécessairement de variables isolées, donc si on les supprime on obtient une formule équivalente. La formule f est donc entièrement caractérisée par son graphe d'adjacence. S'il existe une restriction de $AG(f)$ qui soit connexe (on ne considère que les restrictions à des ensembles de sommets du type $\Delta'^L \bigcup \Delta'^R$ où Δ' est une partie de l'alphabet des variables), alors f est divisible par une formule verrouillée, donc est 4-évitable. Si ce n'est pas le cas, il y a au moins un ensemble libre. On applique alors l'algorithme de Zimin à f, sachant que le graphe d'adjacence obtenu après effacement d'un ensemble libre F est tout simplement la restriction au complémentaire de F du graphe de départ : ce graphe a également la propriété qu'aucune de ses restrictions n'est connexe, donc on peut recommencer et éliminer des ensembles libres dans n'importe quel ordre, jusqu'à éliminer tous les sommets du graphe : f est donc inévitable. ∎

Pour $k \geqslant 3$, il n'est plus possible d'utiliser le même genre d'arguments, car l'élimination d'un ensemble libre peut faire apparaître de nouvelles arêtes dans le graphe, et que toutes les suites de réductions ne sont pas nécessairement équivalentes.