Introduction

La combinatoire des mots a déjà une longue histoire : au début de ce siècle (entre 1906 et 1914), le mathématicien norvégien Axel Thue publia une série d'articles sur les problèmes combinatoires soulevés par l'étude des suites de symboles. Deux d'entre eux [61, 62] en particulier traitaient des répétitions de facteurs dans les suites, et des moyens d'éviter de telles répétitions. Ainsi, il a montré qu'il est possible de construire une suite infinie sur un alphabet à trois lettres qui ne contienne pas de facteur répété deux fois consécutivement : on dit que c'est une suite sans carré.

Par la suite, les résultats de Thue furent plusieurs fois redécouverts, dans des buts divers. Alors que Thue avait fait cette étude « pour le développement des sciences logiques » et sans application en vue, Aršon [3] construisit une suite sans carré en 1937 pour résoudre un problème d'algèbre et Marston Morse et Gustav Hedlund [41] s'intéressèrent à ce problème dans les années 1940 en vue de l'étude des propriétés de surfaces de courbure négative. Citons également des applications aux algèbres universelles, à la théorie des groupes, à la théorie ergodique, mais aussi en physique et bien sûr en informatique où le problème de la recherche de motifs dans un texte est un sujet d'étude important.

De nombreux auteurs ont étudié les mots et les morphismes sans carré ou sans cube [1, 7, 8, 20, 22, 27, 29, 32, 33, 37, 49, 60, 64]. En 1979, Bean, Ehrenfeucht et McNulty [6] et indépendamment Zimin [65] proposèrent de généraliser l'étude de répétitions à celle de motifs arbitraires, formés de blocs de différents types disposés dans un ordre précis. Ils ont montré qu'on peut caractériser les motifs qui sont évités par un mot infini sur un alphabet fini mais non précisé; dans le cas où l'alphabet est fixé, aucune caractérisation n'est connue à ce jour. En 1989, Baker, McNulty et Taylor [5] ont poursuivi l'étude de l'évitabilité des motifs sur un alphabet fixé, en donnant entre autres un exemple de motif évitable sur un alphabet à quatre lettres mais pas sur un alphabet plus petit.

L'étude des motifs peut être vue comme un cas particulier de celle des ensembles de mots évitables : étant donné un ensemble de mots I, existe-t-il une suite infinie dont aucun facteur n'est dans cet ensemble, autrement dit qui *évite* I ? Le cas des ensembles finis est déjà riche en problèmes intéressants (voir par exemple [51]); toutefois, la décidabilité ne pose pas de problème dans ce cas, car l'ensemble des mots qui évitent I est un langage rationnel. Pour parler de décidabilité quand I est infini, il est nécessaire de se donner une description finie de I : un motif n'est autre que la description d'un ensemble infini de mots d'un certain type, et le motif est évitable si et seulement si l'ensemble qu'il décrit l'est.

Une des caractéristiques essentielles des motifs est le rôle important que jouent les morphismes, d'une part dans leur définition (l'ensemble de mots décrit par un motif p est l'ensemble des images morphiques de p), et surtout dans la construction de suites qui les évitent. Aussi, l'étude des motifs est très liée à celle des morphismes, notamment quand ceux-ci sont itérés, ce qui la rapproche de celle des L-systèmes (cf. [54]).

Dans cette thèse, nous nous proposons de poursuivre l'étude des motifs évitables sur un alphabet fixé. Le premier chapitre est consacré à une présentation des diverses notions utiles, des propriétés simples des motifs, des résultats désormais classiques de [5, 6, 65] ainsi que de quelques extensions de la notion de motif évitable.

Au chapitre 2, nous achevons la classification des motifs binaires entreprise par Peter Roth [53]. Ce chapitre permet de se rendre compte d'un certain nombre de difficultés posées par la détermination pratique de l'indice d'évitabilité d'un motif, et qui ont motivé les recherches exposées dans les chapitres suivants.

Ainsi, c'est le statut particulier des motifs binaires $ABAAB$ et $AABBA$ qui nous a amené à introduire la DOL-évitabilité et la HDOL-évitabilité, qui font l'objet du chapitre 3. Nous y étudions les motifs qui peuvent être évités par des langages engendrés par certains types de DOL-systèmes, ce qui nous permet d'ajouter un critère supplémentaire à la classification des motifs binaires.

Le chapitre 4 expose un résultat obtenu avec Peter Roth : pour un alphabet de variables donné, il n'y a qu'un nombre fini de motifs qui sont inévitables sur l'alphabet à deux lettres; cela généralise le résultat obtenu par Roth pour les motifs à deux variables.

Dans le chapitre 5, nous nous intéressons à un problème assez différent des précédents : cette fois, pour un motif bien précis (le chevauchement), nous cherchons à énumérer les mots binaires qui l'évitent en fonction de leur longueur. Nous prouvons que cette fonction d'énumération, dont on sait qu'elle a une croissance polynomiale, oscille en fait entre deux polynômes de degrés distincts. Pour cela, nous construisons un automate et des relations de récurrence explicites. Nous évaluons aussi précisément la croissance en moyenne de cette fonction.

Nous revenons ensuite aux L-systèmes dans le chapitre 6, où nous montrons comment décider si un HDOL-langage donné évite un motif, à condition que le HDOL-système vérifie certaines propriétés. Cet algorithme s'avère extrêmement utile pour classifier les motifs binaires ou autres, car il automatise des preuves souvent très longues. Il est de plus possible qu'il permette d'obtenir d'autres résultats de décidabilité concernant les motifs.

En mettant au point cet algorithme, nous avons été amené à étudier certaines propriétés des morphismes circulaires, qui ont fait l'objet de plusieurs travaux récents [40, 42] et qui semblent utiles pour de nombreux problèmes. Le chapitre 7 montre comment ils peuvent servir à résoudre un problème sur les suites de complexité ultimement affine posé par Jean-Paul Allouche.

En guise de conclusion, nous énumérons quelques questions ouvertes dont nous pensons qu'elles constituent des voies de recherche à explorer dans le domaine des motifs.

Enfin, nous donnons en annexe l'état actuel de la classification des motifs ternaires : un grand nombre de motifs 2-inévitables sont identifiés, ce qui est suffisant pour déterminer leur longueur maximale. Cependant il reste encore un certain nombre de motifs qui sont vraisemblablement 2-évitables mais pour lesquels rien n'est démontré. En ce qui concerne la 3-évitabilité, la situation est beaucoup plus favorable puisque sur les cinq motifs non traités par Silke Nilgens [47] il n'en reste plus qu'un, avec un mot infini conjecturé.