Très bien. Je vais partir à la recherche de travaux académiques et publications mathématiques traitant de la représentation des nombres premiers comme chaînes (notamment en base 16), de critères de divisibilité basés sur des structures périodiques, et de toute tentative de formalisation catégorique ou symbolique de ce type.

Je te prépare un rapport avec les papiers et concepts pertinents, et une synthèse qui fait le pont avec ton théorème cyclique. Je te préviens dès que c'est prêt.

Représentation des nombres premiers comme chaînes en base 16 – Recherches académiques

Travaux académiques pertinents

- **Sur un problème de Gelfond : la somme des chiffres des nombres premiers** – *Christian Mauduit, Joël Rivat, Annals of Mathematics, 2010.* Ce travail célèbre résout une conjecture de 1968 d'Alexandre Gelfond en prouvant que les nombres premiers, pris en base 10, ont une somme des chiffres paire aussi fréquemment qu'impaire. Autrement dit, la parité de la somme des chiffres décimaux d'un nombre premier est aléatoire en moyenne, ce qui suggère l'absence de biais simple dans les motifs digitaux des nombres premiers. Les auteurs utilisent des méthodes d'analyse harmonique et de combinatoire pour aboutir à ce résultat, ouvrant la voie à d'autres recherches sur la distribution des chiffres des suites arithmétiques.
- **On the Recognition of Primes by Automata** – *J. Hartmanis, H. Shank, Journal of ACM, 1968.* Cet article fondateur étudie la reconnaissance des nombres premiers par des automates finis. Il démontre que le langage des nombres premiers (exprimés en base 2 ou 10) a une complexité en états exponentielle, c'est-à-dire qu'aucun automate fini déterministe de taille sous-exponentielle ne peut reconnaître l'ensemble des nombres premiers. Ce résultat, obtenu en utilisant des arguments de théorie des nombres (notamment des progressions arithmétiques de premiers), indique qu'il n'existe pas de *critère local simple* (de type motif fini sur la chaîne de chiffres) pour caractériser la primalité. En d'autres termes, traiter un nombre comme une *chaîne de symboles* et déterminer sa primalité requiert nécessairement une mémoire croissant exponentiellement avec la longueur du nombre, soulignant la complexité structurelle du langage des nombres premiers.
- **Automaticity IV: sequences, sets, and diversity** – *Jeffrey Shallit, Journal de Théorie des Nombres de Bordeaux, 1996.* Dans ce quatrième volet de ses travaux sur l'« automaticité » des suites, Shallit approfondit les résultats de Hartmanis et Shank. Il démontre en particulier que la *diversité automatique* de la suite des nombres premiers est exponentiellement grande, renforçant l'idée qu'aucune machine à états finis (même non déterministe avec certaines relaxations) ne peut reconnaître efficacement les nombres premiers. Ce résultat utilise la notion d'automaticité (mesurant la taille minimale d'un automate qui approche la reconnaissance de la suite) et conclut que cette taille croît exponentiellement pour la suite des nombres premiers.
- **Lower bounds for the state complexity of probabilistic languages and the language of prime numbers** – *Nathanaël Fijalkow, arXiv 2019 (Journal of Logic and Computation, à paraître).* Ce travail récent prolonge l'étude en examinant des modèles d'automates plus puissants (automates alternants infinis). Fijalkow établit que même pour un automate *alternant* (qui peut effectuer des choix disjonctifs et conjonctifs lors de la lecture du mot), le langage des nombres premiers en base 2 nécessite un nombre d'états au moins linéaire en la longueur du mot. Il renforce ainsi le résultat de 1968 (qui montrait une borne exponentielle pour les automates déterministes) en prouvant une borne inférieure non triviale dans un modèle plus général. Cette preuve fait appel à des résultats avancés de théorie des nombres (théorème de Dirichlet et conjectures sur les progressions arithmétiques de premiers) pour construire des entrées nécessitant de nombreux états. Cela confirme que, même avec des logiques multi-états sophistiquées (alternantes), la primalité ne peut être capturée sans une complexité importante.
- **Integer partitions detect the primes** – *William Craig, Jan-Willem van Ittersum, Ken Ono, PNAS 121(39), 2024.* Cet article introduit une approche complètement nouvelle et *symbolique* pour caractériser

les nombres premiers, en exploitant les *partitions entières*. Les auteurs montrent que les nombres premiers sont exactement les solutions d'une infinité d'équations polynomiales faisant intervenir des fonctions de partition (à la MacMahon). Par exemple, ils exhibent une formule du type: pour tout entier $n \geq 2$, n est premier si et seulement si

(3 n^3 - 13 n^2 + 18 n - 8) M_{1}(n) + (12 n^2 - 120 n + 212) M_{2}(n) - 960 M_{3}(n) = 0,

où M_1, M_2, M_3 sont des fonctions de partition bien connues. Ce résultat étonnant fournit *une infinité de nouveaux critères exacts* de primalité, fort différents des tests usuels basés sur la factorisation ou les tests probabilistes. Il établit un pont inédit entre la théorie additive (partitions) et la structure multiplicative des entiers, ouvrant la porte à de nouvelles recherches sur les *invariants formels* caractérisant la primalité. Comme le note Ono, c'est « comme si notre travail donnait une infinité de nouvelles définitions du mot *nombre premier* ».

- **Repetends maximaux et nombres premiers « full reptend »** – *Divers travaux en théorie des nombres, 20e siècle.* Un **repetend** est la période de la fraction décimale $1/p$. Un nombre premier *full reptend* (ou *long prime*) en base b est un premier p pour lequel $1/p$ a une période maximale $p-1$ dans cette base. Ceci équivaut à dire que b est une racine primitive modulo p. Par exemple, 7 est un premier à période maximale en base 10 car $1/7 = 0.\overline{142857}$ a une période de $6 = 7-1$ chiffres. De même, en base 8 on peut montrer que 5 est full reptend (période $4 = 5-1$). Ces travaux relient la représentation d'un premier en base b à la notion de générateur du groupe multiplicatif $(\mathbb{Z}/p\mathbb{Z})^*$. Une conjecture d'Artin stipule qu'il existe une proportion constante (environ 37,39% en base 10) de nombres premiers pour lesquels b est racine primitive, donc à repetend maximal. Bien que non prouvé en général, on pense qu'il y a une infinité de nombres premiers à période maximale pour chaque base. *N.B.:* En base 16 spécifiquement, aucun premier impair ne peut avoir de période $p-1$ car 16 est une puissance parfaite (2^4) : ainsi $16^k \equiv 1 \pmod{p}$ au plus tard pour $k = \frac{p-1}{2}$, le repetend hexadécimal de $1/p$ divisant $(p-1)/2$ au maximum. En effet, 16 étant un carré modulo p, la période de $1/p$ en base 16 est au plus la moitié de $p-1$.
- **Nombres premiers de formes spéciales (Mersenne, répunits, etc.)** – De nombreux travaux se sont intéressés à des familles de nombres premiers possédant *des représentations particulières.* Les **nombres de Mersenne** $(2^p - 1)$ sont ceux dont l'écriture binaire est composée uniquement de 1 (par exemple $2^5 - 1 = 31 = 11111_2$). Ils sont premiers pour certains exposants p (47 cas connus à ce jour) et leur structure binaire permet un test de primalité dédié (test de Lucas-Lehmer). De même, les **répunits décimaux** (nombres écrits uniquement avec des '1' en base 10, comme 11111) ont fait l'objet de recherches : quelques-uns sont premiers (le plus grand connu a plus de 100000 chiffres en base 10). Ces objets relient critères de divisibilité et primalité : par exemple, si p divise $111\ldots1$ (k fois) en base 10, alors $10^k \equiv 1 \pmod{p}$. En base 16, un nombre composé uniquement de chiffres 'F' (15 en décimal) correspond à $16^k - 1$; si un premier p divise ce nombre, on a $16^k \equiv 1 \pmod{p}$ (d'où la connexion avec l'ordre multiplicatif mentionnée ci-dessus). Bien que la plupart de ces familles (Mersenne, répunits, cycliques) ne représentent qu'une infinité très clairsemée de premiers, elles fournissent des *critères symboliques* élégants restreints à des formes particulières – par exemple, le critère de Lucas-Lehmer exploite la structure binaire cyclique des Mersenne. Ces travaux illustrent comment des motifs dans la chaîne de chiffres (suite de 1, suite de F, etc.) peuvent être exploités pour la primalité, mais seulement dans des cas très spécifiques.

Synthèse des idées majeures

1. Complexité et absence de motif simple universel : Les résultats issus de la théorie des automates montrent qu'il n'existe pas de *motif fini répétitif* ou de règle locale sur la représentation en base (quelle qu'elle soit) permettant de reconnaître tous les nombres premiers. Hartmanis et Shank ont établi dès 1968 que la propriété "être premier" n'est pas un langage régulier ni même de complexité inférieure à exponentielle en termes d'automate déterministe. Intuitivement, cela signifie qu'aucune suite de tests cycliques sur les chiffres (comme on le fait pour les critères de divisibilité usuels) ne pourra distinguer infailliblement les nombres premiers des composés. Même en autorisant des calculs logiques plus élaborés (automates alternants, logique multi-états), la primalité reste *structurellement complexe* et nécessite une puissance de calcul croissante au

moins linéaire avec la taille de l'entrée. Ceci rejoint l'intuition que les nombres premiers, considérés comme des chaînes de symboles, ne présentent pas de régularité exploitable simple – en dehors bien sûr de l'élimination triviale des multiples de petites bases (par exemple, en base 10, tout premier > 5 doit finir par 1,3,7 ou 9, ce qui n'est qu'une condition nécessaire mod 2 et 5). En somme, **la distribution des chiffres des nombres premiers "ressemble au hasard"** à bien des égards, une idée appuyée par Mauduit et Rivat qui ont montré l'absence de biais dans la somme des chiffres. On conjecture même que d'autres propriétés "aléatoires" valent : par exemple, on ignore si la suite des chiffres décimaux des nombres premiers est statistiquement équidistribuée (normale) ou si, disons, il existe une infinité de nombres premiers n'utilisant qu'un alphabet restreint de chiffres.

2. Critères modulaires cycliques et divisibilité : Bien que aucune *règle de divisibilité* ne caractérise la primalité de façon générale, on peut combiner de nombreux critères modulaires pour filtrer les composés. Les critères classiques (somme des chiffres pour 3 ou 9, alternance pour 11, etc.) s'étendent à d'autres bases et moduli, mais ils ne détectent que la divisibilité par des petits nombres particuliers. En théorie, pour un nombre n, si l'on veut tester qu'il n'a **aucun** petit facteur, on pourrait appliquer en parallèle ces règles pour toutes bases/facteurs jusqu'à une certaine limite. Cependant, cela revient essentiellement à une variante de l'élimination par crible. Par exemple, en base 16, un nombre premier (sauf 2) doit finir par un chiffre hexadécimal impair (1, 3, 5, 7, 9, B, D ou F), faute de quoi il serait divisible par 2. De plus, s'il finit par 5 (digit A_{16}), il serait divisible par 5 en base 10, donc par 5 en valeur. Mais au-delà de ces éliminations évidentes, les *sommes pondérées cycliques* des chiffres n'apportent pas de critère simple pour la primalité en général – elles sont toutefois utiles pour des bases particulières et des moduli fixes. Par exemple, comme le note John Cook, les mêmes astuces de "découper le dernier chiffre, le combiner linéairement avec le reste" pour tester la divisibilité par 7 ou 17 marchent en base 16 comme en base 10. Mais pour être premier, un nombre doit échapper *simultanément* à tous ces tests de divisibilité pour chaque petit modulateur – ce qui complexifie rapidement le processus et ne fournit pas de *pattern* global concis.

3. Approches structurelles et invariants formels : Face à l'absence de motif simple dans les représentations, les chercheurs ont exploré des **caractérisations plus structurelles** de la primalité. L'approche par les *partitions entières* (Craig, Ono et al. 2024) en est un exemple marquant, fournissant un invariant algébrique sophistiqué qui s'annule exactement pour les nombres premiers. De même, le test AKS (2002, pas cité ci-dessus) avait déjà proposé une caractérisation polynomiale de la primalité via l'identité $(x+1)^n \equiv x^n + 1 \pmod{n}$ pour les entiers n premiers. Ces résultats s'inscrivent dans la quête de **critères formels généraux**, aux antipodes des simples motifs de chiffres : ils traduisent la primalité en une propriété algébrique (polynomiale ou combinatoire) équivalente. Ces critères sont souvent coûteux à vérifier directement, mais ils apportent une compréhension conceptuelle nouvelle – par exemple, le critère via les partitions relie la primalité à des congruences sur des fonctions purement additives. Par ailleurs, la notion de *full reptend primes* relie la représentation en base *à l'ordre multiplicatif* : si un nombre premier p possède une expansion décimale maximale, cela signifie que 10 engendre le groupe multiplicatif mod p. Cette approche lie un invariant structurel (l'ordre de 10 mod p) à un motif décimal (la longueur de la période de $1/p$). Elle ne caractérise qu'une partie des nombres premiers (ceux pour lesquels un certain base est génératrice) et repose sur des conjectures comme celle d'Artin pour son aspect quantitatif, mais illustre une fois de plus la connexion entre *patterns* en base et propriétés modulaires profondes.

4. Base 16 et particularités mod 16 : La base hexadécimale (16) offre un cas intéressant car 16 est une puissance de 2. Cela signifie que les critères de divisibilité y sont dominés par la factorisation par 2. Tout nombre premier impair écrit en base 16 doit finir par un chiffre impair (comme mentionné) et, plus généralement, ne peut avoir que des digits satisfaisant les restrictions modulaires usuelles (pas de facteur 2 ou 5 trivial). Cependant, la structure $16^k \equiv 1 \pmod{p}$ évoquée par l'utilisateur renvoie à la *périodicité* de l'expansion de $1/p$ en base 16. Si $16^k \equiv 1 \pmod{p}$, alors p divise $16^k - 1$, c'est-à-dire le nombre hexadécimal composé de k chiffres 'F'. La longueur k du repetend hexadécimal de $1/p$ est justement le plus petit k vérifiant cette congruence. Comme noté, k divise au plus $(p-1)/2$ car 16 est un carré mod p. Ainsi, *étudier la représentation hexadécimale des inverses mod p revient à étudier l'ordre de 16 mod p.* Peu de travaux ciblent spécifiquement la base 16, mais on peut en déduire que la moitié des nombres premiers environ ont un repetend hexadécimal de longueur $(p-1)/2$ (si 2 est générateur modulo p, ce qui est fréquent heuristiquement), tandis que d'autres ont des périodes plus courtes divisant $(p-1)/4$, etc., selon la valuation 2-adique de $p-1$. En

pratique, cela n'a pas donné lieu à un *critère de primalité* exploitable, mais c'est un exemple de structure modulaire (périodicité de puissances de 16) associée à un motif de la chaîne (répétition de 'F' en hexadécimal).

Pistes ouvertes et angles morts

Malgré ces avancées, plusieurs questions liées aux *motifs symboliques des nombres premiers* demeurent ouvertes :

- **Nombres premiers à alphabet restreint :** On **ignore** s'il existe une infinité de nombres premiers ne contenant pas un certain chiffre dans une base donnée. Par exemple, existe-t-il une infinité de nombres premiers en base 10 n'utilisant pas le chiffre '7' ? Cette question, liée à la distribution des premiers dans les suites digitales, reste non résolue. De même, la question des **nombres premiers palindromiques** (qui se lisent de même dans les deux sens, comme 131 en décimal ou FEF_{16} en hexadécimal) n'est pas tranchée : aucune preuve ne garantit qu'il y en ait infiniment (les données numériques suggèrent que oui, mais sans certitude théorique).
- **Normalité et pseudo-aléatoire des chiffres des premiers :** On conjecture que la suite infinie obtenue en écrivant tous les nombres premiers les uns à la suite des autres est *normal* (c'est-à-dire que chaque motif de m chiffres y apparait avec la fréquence attendue 16^{-m} en base 16, par exemple). Ce serait une forme forte de "pseudo-aléatoire" des nombres premiers. Actuellement, cela dépasse nos techniques : même montrer l'infime biais ou corrélation dans la suite des chiffres des premiers est ardu. Les travaux de Mauduit-Rivat ont réussi pour la **somme** des chiffres (qui est une fonction linéaire des digits), mais pour des propriétés plus complexes (comme la normalité ou l'absence de certains motifs digitaux), on est réduit à des conjectures.
- **Approches catégoriques et logiques :** À ce jour, il n'existe pas de *formalisme catégorique* reconnu qui apporterait un éclairage majeur sur la primalité via la représentation en chaînes. Les nombres premiers sont bien sûr les éléments *irrédicibles* dans l'anneau $\mathbb{Z}$, ce qui peut se formuler dans le langage des catégories (primes comme objets atomiques en théorie des anneaux ou comme *objets initiaux* de certaines catégories fibrées), mais cela reste une reformulation abstraite de leur définition algébrique. De même, en logique du premier ordre (théorie des nombres), la propriété "n est premier" est définissable, mais très complexe – on sait par exemple qu'aucune formule simple de la logique de Peano ne capte exactement l'ensemble des nombres premiers sans quantificateurs non bornés. L'absence de structure répétitive simple se reflète dans ces constats logiques. Une *piste ouverte* consiste à rechercher d'éventuelles symétries ou invariants dans des structures algébriques plus générales (par exemple, des codes correcteurs, des réseaux, des automates cellulaires) qui "cacheraient" la primalité. Jusqu'à présent, aucun code ou automate simple ne génère la suite des nombres premiers de façon propre (sauf à incorporer implicitement un test de division). Cela reste un terrain spéculatif, entre théorie des nombres et informatique théorique.
- **Combinaisons de critères pour des tests de primalité :** Sur le plan plus appliqué, on pourrait chercher à exploiter les *motifs modulaires* de façon plus systématique. Par exemple, l'article d'Ono et al. fournit une infinité de critères polynomiaux exacts. Peut-on en déduire un algorithme pratique ? Pour l'instant, ces formules sont surtout d'intérêt théorique (elles n'offrent pas de gain de complexité par rapport aux tests comme Miller-Rabin ou ECPP). De même, les critères cycliques (repunits, etc.) sont utilisés de manière ciblée pour trouver des grands premiers (on teste $2^p - 1$, $10^n \pm 1$, etc., avec des algorithmes spécialisés). Un **défi ouvert** est de trouver d'autres familles de nombres à représentation structurée susceptibles de produire beaucoup de premiers, ou de combiner plusieurs invariants pour réduire l'espace de recherche de premiers. Par exemple, existe-t-il des représentations en base mixtes ou des codes sur l'expansion hexadécimale qui permettraient de filtrer plus finement les candidats ? À ce jour, aucune méthode de ce genre n'a surpassé les approches analytiques classiques.

En conclusion, la modélisation des nombres premiers comme chaînes de symboles a révélé surtout *ce qu'ils ne sont pas* : ni un langage régulier simple, ni une suite possédant des motifs périodiques évidents. Les nombres premiers semblent dispersés parmi les entiers de façon quasi aléatoire du point de vue de leurs représentations, hormis les contraintes triviales (divisibilité de base). Les quelques *îlots de structure* identifiés – périodes de fractions $1/p$, critères polynomiaux, formes spéciales – montrent qu'il faut souvent monter

en abstraction (automates enrichis, fonctions de partition, etc.) pour capturer la primalité. Cela reflète la profonde nature des nombres premiers, à la fois simples à définir et étonnamment difficiles à caractériser par des motifs fins, ce qui continue de motiver des recherches à l'interface de la théorie des nombres, de l'informatique et de la logique.

Sources: Les références citées en évidence tout au long du texte (par exemple [29], [10], [23]...) renvoient aux travaux académiques listés dans la première section, où figurent titres, auteurs et liens correspondants.