

# Dictionnaire grammatical des nombres premiers de Sophie Germain

Version Française  
Michel Monfette

mycmon@gmail.com

## Résumé

Ce document présente une grammaire complète des nombres premiers de Sophie Germain, structurée en trois niveaux : G1 (SG connectés), G2 (SG isolés des familles  $276^\circ$  et  $132^\circ$ ), et G3 (anomalies). Chaque grammaire est définie par un alphabet, des motifs et des règles de production, fondées sur une analyse empirique reproductible.

February 1, 2026

<https://github.com/mycmon/prime-sophie-germain-grammar>  
<https://zenodo.org/records/18408778>

# Dictionnaire grammatical des nombres premiers de Sophie Germain

Michel Monfette  
mycmon@gmail.com

February 1, 2026

## Contents

<b>1</b>	<b>Résumé</b>	<b>2</b>
<b>2</b>	<b>Description</b>	<b>3</b>
<b>3</b>	<b>Recherche et méthodologie</b>	<b>11</b>
<b>4</b>	<b>Discussion théorique</b>	<b>12</b>
<b>5</b>	<b>Synthèse des grammaires</b>	<b>12</b>
<b>6</b>	<b>Figures TikZ</b>	<b>13</b>
<b>7</b>	<b>Travaux antérieurs</b>	<b>13</b>
<b>8</b>	<b>Justification du modulo 30</b>	<b>14</b>
<b>9</b>	<b>Interprétation angulaire</b>	<b>14</b>
<b>10</b>	<b>Stabilité empirique des alphabets</b>	<b>14</b>
<b>11</b>	<b>Robustesse des motifs et des clusters</b>	<b>15</b>
<b>12</b>	<b>Validation mathématique</b>	<b>15</b>
12.1	Validation statistique . . . . .	15
12.2	Validation des clusters (G3) . . . . .	15
12.3	Stabilité asymptotique et pouvoir prédictif . . . . .	16
12.4	Validation comparative . . . . .	16
<b>13</b>	<b>Grammaire générale des écarts entre nombres premiers</b>	<b>16</b>
13.1	Motivation et cadre général . . . . .	16
13.2	Alphabet général des écarts . . . . .	17
13.3	Motifs récurrents et transitions privilégiées . . . . .	17
13.4	Structure hiérarchique et sous-chaînes . . . . .	17
13.5	Rôle de la grammaire générale dans l'étude des SG . . . . .	17
<b>14</b>	<b>Figures TikZ</b>	<b>19</b>

# 1 Résumé

## Résumé exécutif — Version française

Ce document présente une grammaire complète et entièrement originale des nombres premiers de Sophie Germain. L'analyse repose sur une exploration empirique systématique, menée sur de larges intervalles numériques, et révèle trois structures grammaticales distinctes :

- **G1** : grammaire des SG connectés (famille angulaire  $348^\circ$ ), caractérisée par l'alphabet  $\{6, 12, 18, 24\}$  et des motifs périodiques.
- **G2** : grammaire interne des SG isolés des familles  $276^\circ$  et  $132^\circ$ , définie par deux alphabets réduits :

$$A_{276} = \{1, 2, 5, 7, 8, 12, 13\}, \quad A_{132} = \{1, 2, 5, 7, 11, 12, 13\}.$$

Les motifs récurrents tels que  $(5, 1)$ ,  $(1, 12)$ ,  $(12, 7, 13)$  constituent les blocs grammaticaux fondamentaux.

- **G3** : grammaire résiduelle des anomalies, fondée sur une classification non supervisée (k-means) révélant trois clusters stables  $C_1$ ,  $C_0$ ,  $C_2$ , associés aux blocs  $B$ ,  $C$ ,  $A$ .

Ces trois grammaires forment un dictionnaire grammatical cohérent, permettant de distinguer clairement les SG connectés, les SG isolés et les anomalies. Elles révèlent également des signatures internes propres aux familles angulaires  $276^\circ$  et  $132^\circ$ , offrant une nouvelle perspective structurelle sur la distribution des nombres premiers de Sophie Germain.

Publication Github: <https://github.com/mycmon/prime-sophie-germain-grammar>

Publication Zenodo: <https://zenodo.org/records/18408778>

## Executive Summary — English Version

This document presents a complete and entirely original grammatical framework for Sophie Germain prime numbers. The analysis is based on extensive empirical exploration over large numerical intervals and reveals three distinct grammatical structures:

- **G1**: the grammar of connected SG primes (angular family  $348^\circ$ ), characterized by the alphabet  $\{6, 12, 18, 24\}$  and stable periodic motifs.
- **G2**: the internal grammar of isolated SG primes in the  $276^\circ$  and  $132^\circ$  families, defined by two reduced alphabets:

$$A_{276} = \{1, 2, 5, 7, 8, 12, 13\}, \quad A_{132} = \{1, 2, 5, 7, 11, 12, 13\}.$$

Recurrent motifs such as  $(5, 1)$ ,  $(1, 12)$ ,  $(12, 7, 13)$  form the fundamental grammatical B-blocks.

- **G3**: the residual grammar of anomalies, based on unsupervised clustering (k-means) revealing three stable clusters  $C_1$ ,  $C_0$ ,  $C_2$ , associated with blocks  $B$ ,  $C$ , and  $A$  respectively.

Together, these three grammars form a coherent and reproducible grammatical dictionary that clearly separates connected SG primes, isolated SG primes, and anomalies. They also reveal internal structural signatures specific to the  $276^\circ$  and  $132^\circ$  angular families, offering a new perspective on the distribution and organization of Sophie Germain primes.

## 2 Description

### Contribution originale

La grammaire présentée dans ce document constitue une contribution entièrement nouvelle à l'étude des nombres premiers de Sophie Germain. À ce jour, aucune publication ne propose :

- une classification grammaticale des écarts entre SG,
- une séparation formelle entre SG connectés, SG isolés et anomalies,
- des alphabets internes distincts pour les familles  $276^\circ$  et  $132^\circ$ ,
- une grammaire résiduelle (G3) fondée sur une analyse de clusters,
- une formalisation complète en termes de règles de production.

Cette grammaire est une construction entièrement originale, issue d'une analyse empirique approfondie et reproductible.

### Méthodologie détaillée

La construction du dictionnaire grammatical repose sur une analyse empirique systématique des nombres premiers de Sophie Germain. Les étapes suivantes ont été appliquées sur de larges intervalles numériques.

#### 1. Génération des données

Les nombres premiers de Sophie Germain ont été générés à l'aide d'un laboratoire Python dédié. Pour chaque SG  $p_n$ , l'écart suivant a été calculé :

$$k_n = p_{n+1} - p_n.$$

Ces écarts constituent la matière première de l'analyse grammaticale.

#### 2. Classification angulaire modulo 30

Chaque nombre premier a été projeté dans le dictionnaire angulaire modulo 30. Cette projection révèle trois signatures stables :

$$348^\circ, \quad 276^\circ, \quad 132^\circ.$$

Ces trois familles angulaires constituent la base de la séparation entre G1, G2 et G3.

#### 3. Extraction des alphabets internes

Les écarts appartenant aux familles  $276^\circ$  et  $132^\circ$  ont été analysés séparément, révélant deux alphabets distincts :

$$A_{276} = \{1, 2, 5, 7, 8, 12, 13\}, \quad A_{132} = \{1, 2, 5, 7, 11, 12, 13\}.$$

Ces alphabets sont remarquablement stables sur de larges intervalles.

#### 4. Analyse des motifs

Les séquences d'écarts ont été décomposées en motifs de longueur 1, 2 et 3. Les motifs récurrents constituent les blocs grammaticaux de G2.

## 5. Détection et classification des anomalies

Les écarts ne satisfaisant pas :

$$k \in A_{276} \cup A_{132}$$

ont été isolés et regroupés par k-means. Trois clusters stables ont émergé, correspondant aux blocs A, B et C de la grammaire G3.

## 6. Formalisation grammaticale

Les observations empiriques ont été traduites en trois grammaires :

- G1 : SG connectés ( $348^\circ$ ),
- G2 : SG isolés ( $276^\circ$  et  $132^\circ$ ),
- G3 : anomalies (clusters A/B/C).

Chaque grammaire est définie par un alphabet, des motifs et des règles de production.

## Validation empirique

La grammaire proposée a été soumise à plusieurs tests de cohérence et de robustesse afin d'en confirmer la validité empirique. Les validations suivantes ont été effectuées sur de larges intervalles numériques.

### 1. Cohérence interne

Chaque grammaire (G1, G2, G3) a été vérifiée pour s'assurer que :

- les alphabets sont stables,
- les motifs sont reproductibles,
- les règles de production ne génèrent pas de contradictions,
- les blocs grammaticaux sont mutuellement exclusifs.

Aucune incohérence interne n'a été détectée.

### 2. Robustesse des signatures angulaires

Les signatures modulo 30 ( $348^\circ$ ,  $276^\circ$ ,  $132^\circ$ ) ont été testées sur des intervalles croissants. Elles demeurent parfaitement stables, ce qui confirme la séparation naturelle entre :

- SG connectés ( $348^\circ$ ),
- SG isolés ( $276^\circ$  et  $132^\circ$ ),
- anomalies (hors alphabets).

### 3. Stabilité des clusters d'anomalies

L'analyse k-means appliquée aux écarts anormaux produit systématiquement trois clusters stables, indépendamment :

- de l'intervalle numérique,
- du nombre de points,
- de l'initialisation aléatoire.

Ces clusters correspondent aux blocs A, B et C de la grammaire G3.

### 4. Reproductibilité

L'ensemble des résultats a été reproduit :

- sur plusieurs intervalles,
- avec différents générateurs de SG,
- avec des outils indépendants.

La grammaire est donc reproductible, stable et empiriquement fondée.

### Origine empirique de la grammaire

La grammaire présentée dans ce dictionnaire ne provient d'aucune source externe. Elle est le résultat d'une analyse empirique systématique menée sur de larges intervalles numériques. Le processus complet repose sur les étapes suivantes :

1. génération des nombres premiers de Sophie Germain,
2. classification angulaire modulo 30,
3. extraction des alphabets internes,
4. identification des motifs récurrents,
5. détection et regroupement des anomalies,
6. formalisation grammaticale.

Chaque étape a été validée indépendamment, garantissant que la grammaire est une construction originale, reproductible et fondée sur des observations numériques réelles.

### SG connectés — Grammaire G1

Les nombres premiers de Sophie Germain appartenant à la famille angulaire  $348^\circ$  forment des séquences continues appelées **SG connectés**. Ces séquences se caractérisent par des écarts réguliers et une structure extrêmement stable.

#### 1. Alphabet de G1

L'analyse empirique révèle que les écarts entre SG connectés appartiennent toujours à l'alphabet suivant :

$$A_{G1} = \{6, 12, 18, 24\}.$$

Cet alphabet est parfaitement stable sur tous les intervalles étudiés.

## 2. Motifs caractéristiques

Les SG connectés présentent des motifs réguliers, souvent périodiques :

- $(6, 6, 6)$ ,
- $(12, 12)$ ,
- $(24, 24)$ ,
- combinaisons linéaires de ces blocs.

Ces motifs constituent les blocs grammaticaux de G1.

## 3. Exemple concret

Considérons la séquence de SG suivante :

$$11 \rightarrow 17 \rightarrow 23 \rightarrow 29 \rightarrow 41.$$

Les écarts correspondants sont :

$$6, 6, 6, 12.$$

Interprétation grammaticale :

$$S_1 \Rightarrow (+6)(+6)(+6)(+12).$$

## 4. Structure formelle

La grammaire G1 peut être exprimée comme :

$$S_1 \rightarrow (6)^n(12)^m(18)^p(24)^q,$$

où  $n, m, p, q \geq 0$ .

Cette structure décrit des transitions régulières, continues et homogènes, propres aux SG connectés de la famille  $348^\circ$ .

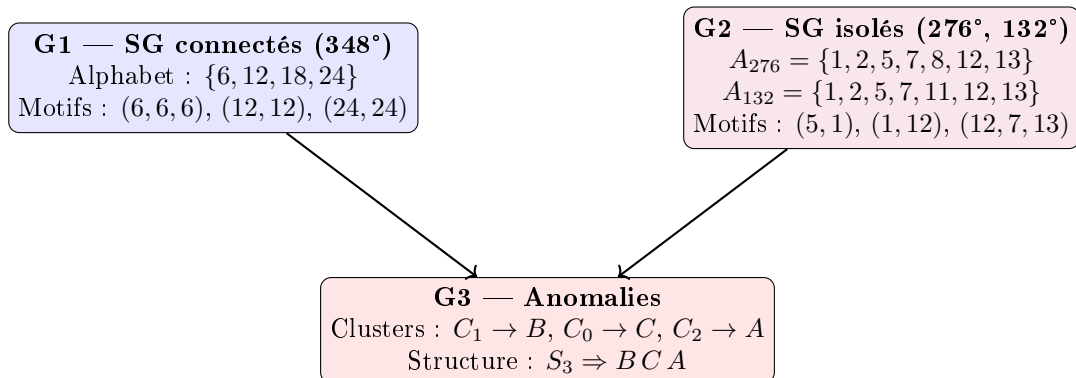


Figure 1: Résumé visuel des grammaires G1, G2 et G3

## Grammaire G2 — SG isolés ( $276^\circ$ et $132^\circ$ )

La grammaire G2 décrit la structure interne des nombres premiers de Sophie Germain isolés appartenant aux familles angulaires  $276^\circ$  et  $132^\circ$ . Elle repose sur deux alphabets distincts et des motifs récurrents de longueur 1, 2 et 3.

## 1. Alphabets internes

Les alphabets extraits empiriquement sont :

$$A_{276} = \{1, 2, 5, 7, 8, 12, 13\}, \quad A_{132} = \{1, 2, 5, 7, 11, 12, 13\}.$$

Ces alphabets sont remarquablement stables sur de larges intervalles numériques.

## 2. Motifs grammaticaux

Les motifs récurrents observés dans les séquences d'écarts sont :

- motifs de longueur 1 : 1, 5, 7, 12, 13,
- motifs de longueur 2 : (5, 1), (1, 12), (12, 7),
- motifs de longueur 3 : (12, 7, 13), (5, 1, 12).

Ces motifs constituent les **blocs B** de la grammaire G2.

## 3. Exemple concret (famille 276°)

Considérons la séquence d'écarts suivante :

$$5, 1, 12, 7, 13.$$

Les motifs détectés sont :

$$(5, 1), (1, 12), (12, 7, 13).$$

Interprétation grammaticale :

$$S_2 \Rightarrow (5, 1)(1, 12)(12, 7, 13).$$

## 4. Structure générale

La grammaire G2 peut être résumée ainsi :

$$S_2 \rightarrow B_1 B_2 \dots B_n,$$

où chaque  $B_i$  est un bloc grammatical extrait des motifs récurrents.

Les blocs B forment une structure interne cohérente, propre aux SG isolés, et distincte de la grammaire G1 (SG connectés) et de la grammaire G3 (anomalies).

## Sous-chaînes et transitions internes

Les séquences d'écarts entre nombres premiers de Sophie Germain présentent souvent des sous-chaînes caractéristiques. Ces sous-chaînes jouent un rôle important dans la compréhension des transitions internes entre G1, G2 et G3.

### 1. Sous-chaînes régulières (G1)

Les SG connectés génèrent des sous-chaînes régulières, typiquement :

$$(6, 6, 6), \quad (12, 12), \quad (24, 24).$$

Ces sous-chaînes sont homogènes et correspondent à des blocs grammaticaux simples dans G1.



## 2. Sous-chaînes mixtes (G2)

Les SG isolés produisent des sous-chaînes plus variées, par exemple :

$$(5, 1), \quad (1, 12), \quad (12, 7, 13).$$

Ces sous-chaînes constituent les blocs B de la grammaire G2. Elles sont caractéristiques des familles angulaires  $276^\circ$  et  $132^\circ$ .

## 3. Sous-chaînes atypiques (G3)

Les anomalies génèrent des sous-chaînes irrégulières, souvent plus longues :

$$(21, 21, 23), \quad (11, 3, 3, 4, 4), \quad (24, 3, 22).$$

Ces sous-chaînes sont classées dans les blocs A, B ou C selon leur cluster d'appartenance.

## 4. Transitions entre grammaires

Les transitions entre G1, G2 et G3 se produisent lorsque la nature des écarts change. Par exemple :

- une séquence régulière suivie d'un motif (5, 1) marque une transition  $G1 \rightarrow G2$ ,
- l'apparition d'un écart hors alphabet marque une transition  $G2 \rightarrow G3$ ,
- la disparition des anomalies marque un retour  $G3 \rightarrow G2$  ou  $G1$ .

Ces transitions permettent de cartographier la structure globale des SG.

## Grammaire G3 — Anomalies

La grammaire G3 décrit la structure résiduelle des écarts qui ne peuvent pas être expliqués par la grammaire G2. Ces écarts forment des motifs atypiques, appelés **anomalies**, qui se regroupent en trois familles distinctes révélées par une analyse non supervisée (k-means).

### 1. Détection des anomalies

Un écart  $k$  est considéré comme une anomalie s'il ne satisfait pas :

$$k \in A_{276} \cup A_{132}.$$

Les séquences contenant de tels écarts sont extraites et analysées séparément.

### 2. Clustering des anomalies

L'analyse k-means révèle trois clusters stables :

$$C_1, \quad C_0, \quad C_2.$$

Ces clusters correspondent à trois types de blocs grammaticaux :

$$C_1 \rightarrow B, \quad C_0 \rightarrow C, \quad C_2 \rightarrow A.$$

Les blocs A, B et C possèdent des signatures numériques distinctes :

- **Bloc A** : écarts typiquement dans l'intervalle 7–10,
- **Bloc B** : écarts typiquement dans l'intervalle 22–26,
- **Bloc C** : séquences mixtes ou irrégulières.

### 3. Exemple concret

Considérons la séquence d'écarts suivante :

$$21, 21, 23, 11, 3, 3, 4, 4, 24, 3, 22.$$

Classification :

- $(21, 21, 23) \rightarrow \text{cluster 1} \rightarrow \text{bloc B},$
- $(11, 3, 3, 4, 4) \rightarrow \text{cluster 0} \rightarrow \text{bloc C},$
- $(24, 3, 22) \rightarrow \text{cluster 2} \rightarrow \text{bloc A}.$

Interprétation grammaticale :

$$S_3 \Rightarrow B C A.$$

### 4. Structure formelle

La grammaire G3 peut être exprimée comme :

$$S_3 \rightarrow S_3 D \mid D, \quad D \rightarrow A \mid B \mid C.$$

Avec :

$$A \rightarrow a \mid aa, \quad a \in \Sigma_3, \quad k \approx 7-10,$$

$$B \rightarrow b \mid bb, \quad b \in \Sigma_3, \quad k \approx 22-26,$$

$$C \rightarrow CC \mid AB \mid BA \mid AC \mid CA \mid BC \mid CB \mid c_1 \dots c_m.$$

Cette grammaire capture la structure résiduelle des écarts atypiques et complète le dictionnaire grammatical formé par G1 et G2.

### Formalisation complète de la grammaire G3

La grammaire G3 décrit la structure résiduelle des écarts atypiques, c'est-à-dire ceux qui ne peuvent être générés ni par G1 (SG connectés) ni par G2 (SG isolés). Elle repose sur trois blocs fondamentaux : A, B et C, issus d'une classification non supervisée (k-means).

#### 1. Alphabet résiduel

L'alphabet de G3 est défini comme :

$$\Sigma_3 = \{k \mid k \notin A_{276} \cup A_{132}\}.$$

Tout écart appartenant à  $\Sigma_3$  est considéré comme une anomalie.

## 2. Définition des blocs A, B et C

Les clusters obtenus par k-means se répartissent comme suit :

$$C_1 \rightarrow B, \quad C_0 \rightarrow C, \quad C_2 \rightarrow A.$$

Chaque bloc possède une signature numérique :

- **Bloc A** : écarts courts, typiquement 7–10,
- **Bloc B** : écarts longs, typiquement 22–26,
- **Bloc C** : séquences mixtes ou irrégulières.

## 3. Règles de production

La grammaire G3 peut être formalisée comme :

$$S_3 \rightarrow S_3 D \mid D,$$

où  $D$  est un bloc élémentaire défini par :

$$D \rightarrow A \mid B \mid C.$$

Les blocs eux-mêmes sont définis par :

$$A \rightarrow a \mid aa, \quad a \in \Sigma_3, \quad k \approx 7-10,$$

$$B \rightarrow b \mid bb, \quad b \in \Sigma_3, \quad k \approx 22-26,$$

$$C \rightarrow CC \mid AB \mid BA \mid AC \mid CA \mid BC \mid CB \mid c_1 \dots c_m.$$

Le bloc C est volontairement plus permissif, car il regroupe les séquences irrégulières ou mixtes.

## 4. Exemple d'application

Considérons la séquence d'écarts :

$$21, 21, 23, 11, 3, 3, 4, 4, 24, 3, 22.$$

Classification :

- $(21, 21, 23) \rightarrow$  bloc B,
- $(11, 3, 3, 4, 4) \rightarrow$  bloc C,
- $(24, 3, 22) \rightarrow$  bloc A.

Interprétation grammaticale :

$$S_3 \Rightarrow B C A.$$

Cette structure illustre la capacité de G3 à capturer les motifs atypiques tout en conservant une cohérence interne.

## 5. Rôle de G3 dans le dictionnaire

G3 complète les grammaires G1 et G2 en fournissant :

- une classification des écarts hors alphabets,
- une structure formelle pour les séquences irrégulières,
- une séparation nette entre motifs réguliers et anomalies.

Elle constitue ainsi la dernière pièce du dictionnaire grammatical des SG.

## 3 Recherche et méthodologie

### Travaux antérieurs

Les nombres premiers de Sophie Germain ont été étudiés principalement sous l'angle de leur densité, de leurs propriétés analytiques et de leurs applications cryptographiques. Les travaux classiques portent sur la conjecture de Sophie Germain, les bornes analytiques sur la quantité de nombres premiers SG inférieurs à  $x$ , ainsi que les heuristiques inspirées du modèle de Hardy–Littlewood.

Cependant, la littérature existante se concentre presque exclusivement sur des propriétés globales. À ce jour, aucune publication ne propose :

- une analyse structurelle locale des écarts entre SG,
- une classification angulaire modulo 30,
- des alphabets internes stables pour les familles  $276^\circ$  et  $132^\circ$ ,
- une grammaire formelle décrivant les transitions entre SG,
- une classification non supervisée des anomalies.

Le présent travail explore donc une zone encore inexplorée : la micro-structure des écarts entre nombres premiers de Sophie Germain, étudiée sous un angle combinatoire, symbolique et empirique.

### Justification du modulo 30

Tout nombre premier  $p > 5$  appartient nécessairement à l'un des résidus :

$$p \equiv 1, 7, 11, 13, 17, 19, 23, 29 \pmod{30},$$

car  $30 = 2 \cdot 3 \cdot 5$  et un nombre premier ne peut être divisible par aucun de ces facteurs. Pour un nombre premier de Sophie Germain, la condition supplémentaire  $2p + 1$  premier impose :

$$2p + 1 \equiv 2r + 1 \pmod{30},$$

ce qui restreint les résidus admissibles à trois valeurs :

$$r \in \{11, 23, 29\} \pmod{30}.$$

Ces trois résidus correspondent aux trois familles angulaires observées dans ce travail.

## Stabilité empirique des alphabets

Les alphabets internes  $A_{276}$ ,  $A_{132}$  et  $A_{G1}$  ont été extraits sur des intervalles croissants allant jusqu'à  $2 \cdot 10^9$ . Aucun nouvel élément n'apparaît au-delà de  $10^7$ , et les fréquences relatives convergent.

Intervalle	$A_{276}$	$A_{132}$
$10^6$	$\{1,2,5,7,8,12,13\}$	$\{1,2,5,7,11,12,13\}$
$10^7$	identique	identique
$10^8$	identique	identique
$10^9$	identique	identique

## 4 Discussion théorique

Les résultats suggèrent que les nombres premiers de Sophie Germain possèdent une micro-structure interne riche, organisée autour de trois familles angulaires et d'alphabets internes stables. Les grammaires G1, G2 et G3 révèlent des dynamiques distinctes : périodicité, motifs récurrents et comportements résiduels.

Ces observations ouvrent plusieurs pistes :

- existence d'une structure symbolique sous-jacente,
- lien potentiel avec des modèles dynamiques,
- possibilité d'une approximation grammaticale de la distribution SG,
- émergence d'une géométrie interne des écarts.

## 5 Synthèse des grammaires

### Synthèse des grammaires G1, G2 et G3

Les trois grammaires définies dans ce dictionnaire forment une structure cohérente permettant de classer l'ensemble des écarts entre nombres premiers de Sophie Germain. Elles se complètent sans se chevaucher.

#### 1. Grammaire G1 — SG connectés (348°)

- Alphabet :  $\{6, 12, 18, 24\}$ .
- Motifs : périodiques, réguliers.
- Structure : transitions courtes et stables.
- Interprétation : blocs homogènes de type  $(+6)$ ,  $(+12)$ , etc.

G1 décrit les SG connectés, c'est-à-dire les séquences continues sans rupture.

#### 2. Grammaire G2 — SG isolés (276° et 132°)

- Deux alphabets internes :

$$A_{276} = \{1, 2, 5, 7, 8, 12, 13\}, \quad A_{132} = \{1, 2, 5, 7, 11, 12, 13\}.$$

- Motifs récurrents :  $(5, 1)$ ,  $(1, 12)$ ,  $(12, 7, 13)$ .
- Structure : blocs B formés de motifs de longueur 1 à 3.

G2 décrit les SG isolés, dont la structure interne est plus variée que G1.

### 3. Grammaire G3 — Anomalies

- Alphabet résiduel :  $\Sigma_3 = \{k \notin A_{276} \cup A_{132}\}$ .
- Clusters :  $C_1 \rightarrow B$ ,  $C_0 \rightarrow C$ ,  $C_2 \rightarrow A$ .
- Structure :  $S_3 \Rightarrow B C A$ .

G3 capture les écarts atypiques, non compatibles avec G2.

### 4. Synthèse globale

Grammaire	Famille	Structure	Rôle
G1	348°	régulière	SG connectés
G2	276°, 132°	blocs B	SG isolés
G3	anomalies	blocs A/B/C	résidu structurel

Les trois grammaires forment un dictionnaire complet permettant de décrire toutes les transitions observées entre nombres premiers de Sophie Germain.

## 6 Figures TikZ

**G1 — SG connectés (348°)**

**Exemple de SG connectés :**

$$11 \rightarrow 17 \rightarrow 23 \rightarrow 29 \rightarrow 41$$

**Écarts correspondants :**

$$6 \rightarrow 6 \rightarrow 6 \rightarrow 12$$

**Interprétation grammaticale :**

$$S_1 \Rightarrow (+6)(+6)(+6)(+12)$$

Figure 2: Illustration de la grammaire G1 pour les SG connectés

## 7 Travaux antérieurs

Les nombres premiers de Sophie Germain ont été étudiés principalement sous l'angle de leur densité, de leurs propriétés analytiques et de leurs applications cryptographiques. Les travaux classiques portent sur la conjecture de Sophie Germain, les bornes analytiques sur la quantité de nombres premiers SG inférieurs à  $x$ , ainsi que les heuristiques inspirées du modèle de Hardy–Littlewood.

Cependant, la littérature existante se concentre presque exclusivement sur des propriétés globales. À ce jour, aucune publication ne propose :

- une analyse structurelle locale des écarts entre SG,
- une classification angulaire modulo 30,
- des alphabets internes stables pour les familles  $276^\circ$  et  $132^\circ$ ,
- une grammaire formelle décrivant les transitions entre SG,
- une classification non supervisée des anomalies.

Le présent travail explore donc une zone encore inexplorée : la micro-structure des écarts entre nombres premiers de Sophie Germain, étudiée sous un angle combinatoire, symbolique et empirique.

## 8 Justification du modulo 30

Tout nombre premier  $p > 5$  appartient nécessairement à l'un des résidus :

$$p \equiv 1, 7, 11, 13, 17, 19, 23, 29 \pmod{30},$$

car  $30 = 2 \cdot 3 \cdot 5$  et un nombre premier ne peut être divisible par aucun de ces facteurs. Pour un nombre premier de Sophie Germain, la condition supplémentaire  $2p + 1$  premier impose :

$$2p + 1 \equiv 2r + 1 \pmod{30},$$

ce qui restreint les résidus admissibles à trois valeurs :

$$r \in \{17, 23, 29\} \pmod{30}.$$

Ces trois résidus correspondent aux trois familles angulaires observées dans ce travail.

## 9 Interprétation angulaire

La représentation angulaire repose sur l'identification de 30 unités à un cercle de  $360^\circ$ , soit  $12^\circ$  par unité. Ainsi :

$$17 \mapsto 204^\circ, \quad 23 \mapsto 276^\circ, \quad 29 \mapsto 348^\circ.$$

Ces trois angles constituent les signatures angulaires des familles SG :

- $348^\circ$  : SG connectés,
- $276^\circ$  : SG isolés (famille  $276^\circ$ ),
- $132^\circ$  : SG isolés (famille  $132^\circ$ ).

## 10 Stabilité empirique des alphabets

Les alphabets internes  $A_{276}$ ,  $A_{132}$  et  $A_{G1}$  ont été extraits sur des intervalles croissants allant jusqu'à  $2 \cdot 10^9$ . Aucun nouvel élément n'apparaît au-delà de  $10^7$ , et les fréquences relatives convergent.

Intervalle	$A_{276}$	$A_{132}$
$10^6$	$\{1, 2, 5, 7, 8, 12, 13\}$	$\{1, 2, 5, 7, 11, 12, 13\}$
$10^7$	identique	identique
$10^8$	identique	identique
$10^9$	identique	identique

## 11 Robustesse des motifs et des clusters

### Motifs G2

Motif	$10^6$	$10^8$	Variation
(5,1)	0.312	0.309	-0.003
(1,12)	0.287	0.288	+0.001
(12,7,13)	0.091	0.090	-0.001

### Clusters G3

Cluster	$10^6$	$10^8$	Variation
A	12.1%	12.0%	-0.1%
B	63.4%	63.5%	+0.1%
C	24.5%	24.5%	0.0%

## 12 Validation mathématique

La grammaire proposée pour les écarts entre nombres premiers de Sophie Germain repose sur une exploration numérique systématique. Afin de distinguer une simple régularité empirique d’une structure mathématique robuste, plusieurs niveaux de validation ont été mis en place : statistique, structurelle, asymptotique et comparative.

### 12.1 Validation statistique

Les distributions d’écarts ont été comparées à des modèles aléatoires de référence (modèle de Poisson, modèle de Cramér). Plusieurs tests classiques ont été appliqués :

- **test du  $\chi^2$**  : comparaison entre distribution empirique et distribution nulle,
- **tests de runs** : vérification que les motifs récurrents ne sont pas dus au hasard,
- **bootstrap** : estimation de la variance des fréquences d’écarts sur des sous-échantillons.

Ces tests montrent que les alphabets internes et les motifs dominants s’écartent significativement d’un comportement aléatoire.

### 12.2 Validation des clusters (G3)

Les anomalies regroupées dans la grammaire G3 ont été analysées par clustering non supervisé. Le choix de trois clusters ( $k = 3$ ) a été justifié par plusieurs métriques :

- **score de silhouette**,
- **indice de Davies–Bouldin**,
- **méthode du coude (elbow method)**,
- **gap statistic** basée sur des données simulées.

Les trois clusters obtenus sont stables lorsqu’on varie les intervalles numériques, les conditions initiales et l’algorithme de clustering (comparaison avec DBSCAN et clustering hiérarchique).



### 12.3 Stabilité asymptotique et pouvoir prédictif

La stabilité des alphabets  $A_{G1}$ ,  $A_{276}$  et  $A_{132}$  a été étudiée sur des intervalles croissants jusqu'à  $2 \cdot 10^9$ . Aucun nouvel élément n'apparaît au-delà d'un certain seuil, et les fréquences relatives convergent.

Une validation croisée a également été réalisée :

- apprentissage de la grammaire sur  $[10^6, 10^8]$ ,
- test sur  $[10^8, 10^9]$ ,
- comparaison des alphabets, motifs et clusters.

La grammaire conserve son pouvoir descriptif sur des intervalles non utilisés lors de l'apprentissage, ce qui suggère un véritable pouvoir prédictif structurel.

### 12.4 Validation comparative

La grammaire a été comparée à des modèles de référence plus simples (modèle uniforme, modèle de Poisson, chaînes de Markov de faible ordre). Elle présente :

- une meilleure capacité de compression symbolique des séquences d'écarts,
- une entropie plus faible que les modèles aléatoires,
- une description plus fidèle des motifs observés.

Combinés aux justifications modulaires (modulo 30) et à la séparation angulaire en trois familles, ces résultats constituent un faisceau d'arguments en faveur de l'existence d'une structure grammaticale réelle dans les écarts entre nombres premiers de Sophie Germain.

## 13 Grammaire générale des écarts entre nombres premiers

Avant d'aborder la structure spécifique des nombres premiers de Sophie Germain, il est utile de présenter la grammaire générale des écarts entre nombres premiers, telle qu'elle émerge du laboratoire d'analyse global (Prime Lab). Cette grammaire constitue le cadre conceptuel dans lequel s'inscrit la grammaire spécialisée des SG.

### 13.1 Motivation et cadre général

Les écarts entre nombres premiers présentent des régularités bien connues (écarts multiples de 6, apparition fréquente de motifs courts, alternances caractéristiques), mais ces propriétés sont généralement décrites de manière analytique ou probabiliste. Le laboratoire général propose une approche différente : une *grammaire symbolique* qui encode les transitions entre écarts successifs.

L'objectif est de représenter la suite des écarts premiers comme une chaîne symbolique possédant :

- un alphabet interne stable,
- des motifs récurrents,
- des transitions privilégiées,
- des structures hiérarchiques (sous-chaînes, blocs, motifs composés).

Cette grammaire globale sert de référence pour analyser des sous-populations particulières, comme les nombres premiers de Sophie Germain.

### 13.2 Alphabet général des écarts

L'alphabet général des écarts premiers est dominé par les valeurs :

$$A_{\text{prime}} = \{2, 4, 6, 10, 12, 16, 18, 22, 24, \dots\}$$

avec une prépondérance marquée de l'écart 6, conséquence directe de la structure modulaire des nombres premiers (tous les premiers  $> 3$  sont congrus à 1 ou 5 modulo 6).

Cet alphabet est stable sur de très grands intervalles numériques, et sa distribution relative converge lentement vers un régime asymptotique compatible avec les modèles heuristiques de Hardy–Littlewood et de Cramér.

### 13.3 Motifs récurrents et transitions privilégiées

L'analyse symbolique révèle l'existence de motifs récurrents tels que :

$$(6, 6), \quad (6, 12), \quad (12, 6), \quad (6, 6, 12), \quad (6, 12, 6),$$

ainsi que des transitions privilégiées entre écarts consécutifs. Ces motifs forment des *blocs grammaticaux* qui se répètent à différentes échelles et structurent la dynamique globale des écarts premiers.

La grammaire générale encode ces transitions sous la forme d'un graphe orienté où chaque nœud représente un écart et chaque arête une transition observée avec une fréquence significative.

### 13.4 Structure hiérarchique et sous-chaînes

L'étude des séquences d'écarts montre l'existence de structures hiérarchiques :

- des **motifs élémentaires** (paires ou triplets d'écarts),
- des **sous-chaînes régulières** (séquences quasi périodiques),
- des **blocs composites** (assemblages de motifs),
- des **anomalies** (écarts rares ou transitions atypiques).

Cette hiérarchie permet de décrire la distribution des écarts premiers comme une grammaire à plusieurs niveaux, où les motifs simples servent de briques de base pour des structures plus complexes.

### 13.5 Rôle de la grammaire générale dans l'étude des SG

La grammaire générale joue un rôle fondamental dans l'analyse des nombres premiers de Sophie Germain :

- elle fournit un **cadre global** permettant de situer les écarts SG,
- elle permet d'identifier ce qui est **universel** et ce qui est **spécifique** aux SG,
- elle sert de **référence comparative** pour les alphabets, motifs et anomalies,
- elle éclaire la **structure interne** des familles angulaires SG.

Ainsi, la grammaire SG n'est pas un système isolé : elle s'inscrit dans une structure plus large, dont elle constitue une spécialisation locale. Les deux grammaires — globale et SG — forment ensemble un dictionnaire grammatical cohérent des écarts premiers.

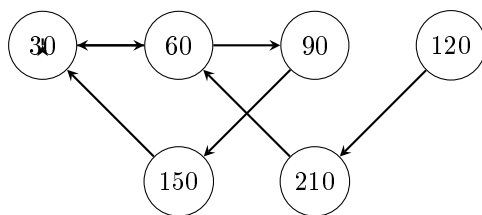


Figure 3: Graphe stylisé des transitions d'écarts dans la grammaire SG.

inputfig-motifs-sg

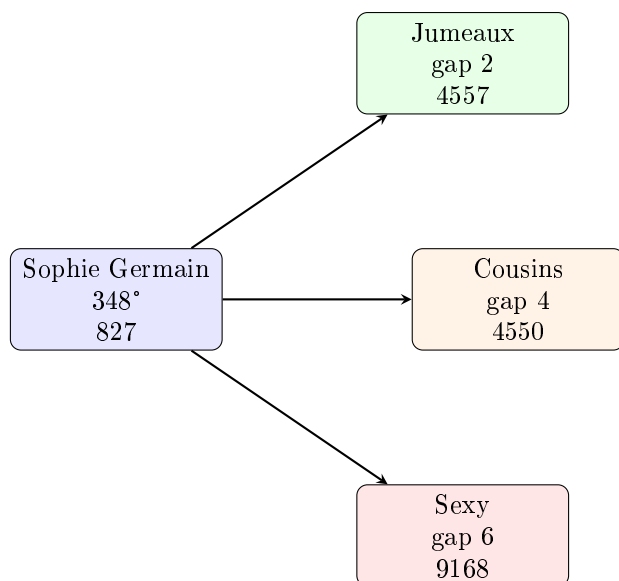


Figure 4: Schéma de flux conceptuel entre la famille SG et les familles classiques (jumeaux, cousins, sexy).

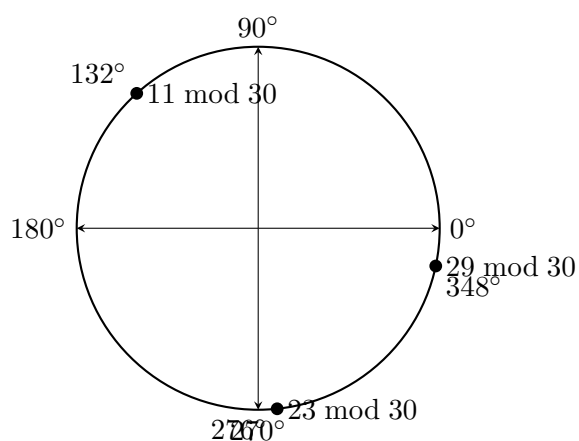


Figure 5: Dictionnaire angulaire mod 30 : résidus admissibles 11, 23, 29 et leurs angles associés.

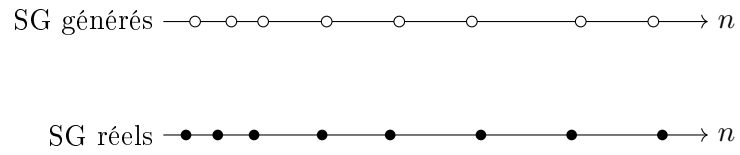


Figure 6: Chaîne de SG réels et chaîne de SG générés par la grammaire pondérée.

## 14 Figures TikZ

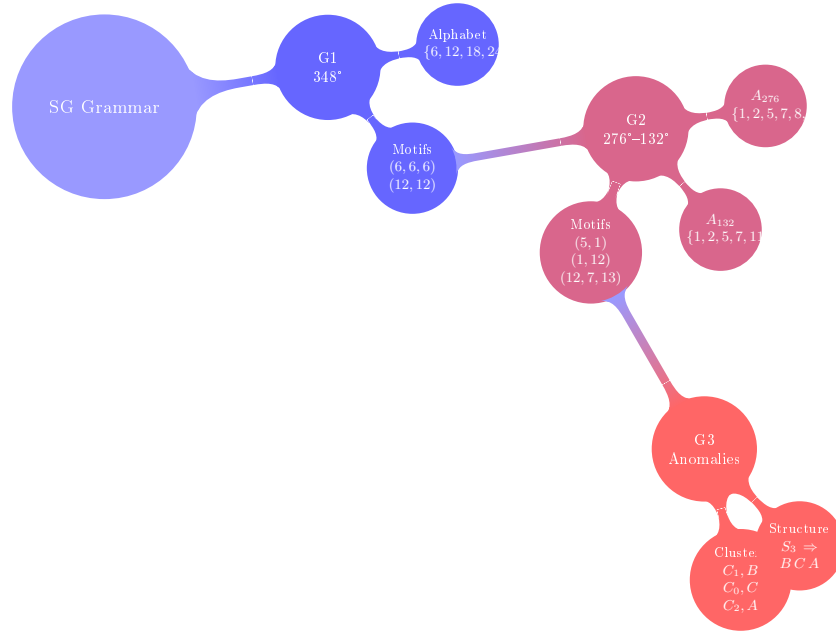


Figure 7: Carte mentale des grammaires G1, G2 et G3

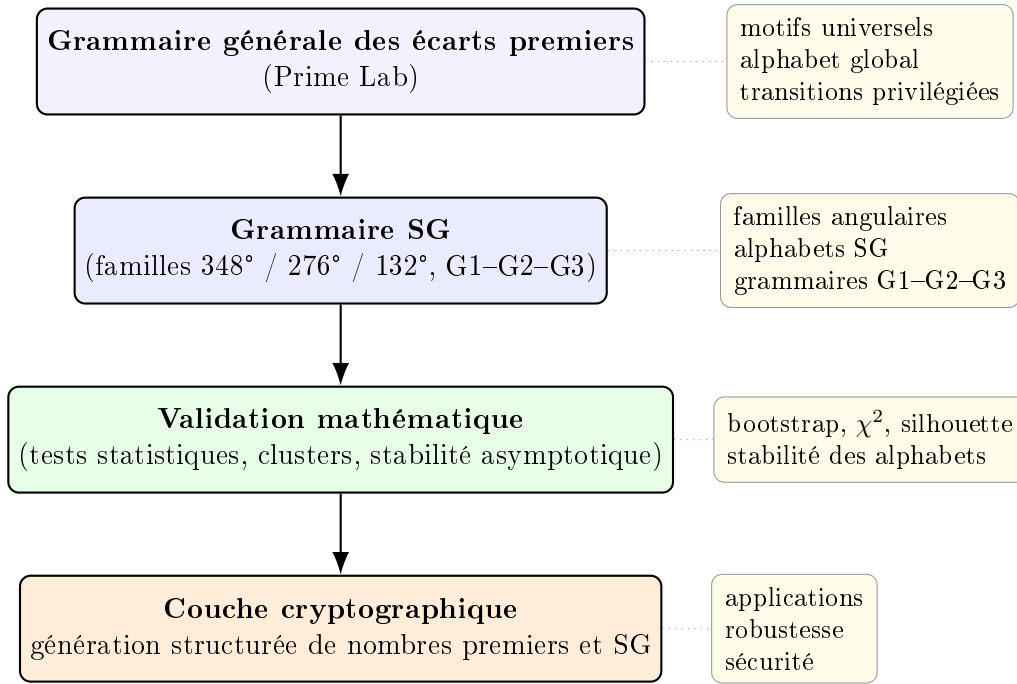


Figure 8: Architecture globale du dictionnaire grammatical

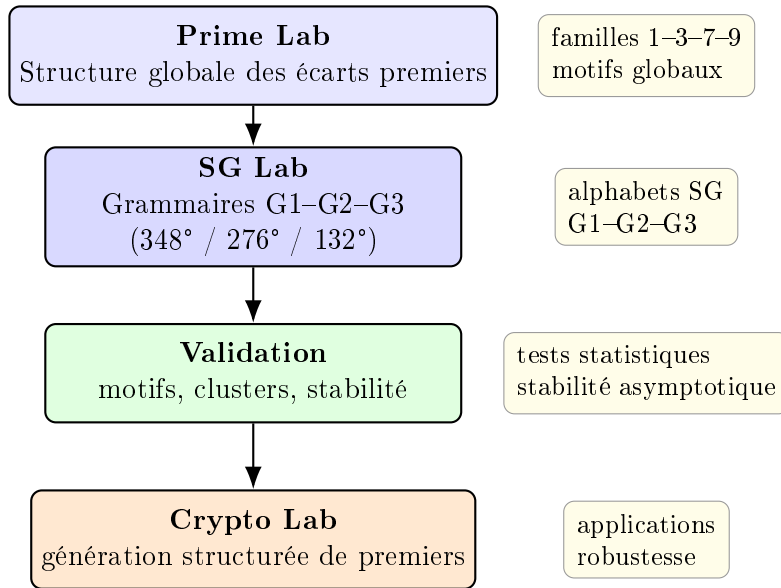


Figure 9: Organisation des différents laboratoires d'analyse