UQAC
Université du Québec
à Chicoutimi

Comment une évaluation des contes générés par intelligence artificielle à l'aide de la *Story Grammar Theory* peut-elle nourrir la création d'un dispositif narratif interactif adapté aux enfants ?

Par Daniel Lavoie

Mémoire présenté à l'Université du Québec à Chicoutimi
en vue de l'obtention du grade de Maîtrise ès arts (M. A.) en art

Québec, Canada

À mes enfants, Amélie, Guillaume et Charles,
à qui j'ai raconté tant d'histoires à partir de trois mots.
Vous étiez mes premiers critiques, et les meilleurs.

À Guillaume Couture,
mon ami parti trop tôt, qui avait lui aussi choisi l'UQAC pour sa maîtrise.
Tu aurais été fier. Ce travail est aussi à ta mémoire.

À Jean-Jacques Tremblay,
professeur au BASA, collègue et ami.
C'est lui qui a semé l'idée de cette maîtrise.
Je l'ai menée à terme en son souvenir.

RÉSUMÉ

La recherche de ce mémoire a émergé du projet *3 mots*, une expérience immersive reposant sur la génération automatisée de contes audio-visuels immersif destiné aux enfants. Dans ce contexte, la présente étude se concentre exclusivement sur la composante centrale du dispositif : la génération textuelle par intelligence artificielle, considérée comme le socle narratif d'une future expérience complète.

Très peu de travaux portent spécifiquement sur la génération de contes pour enfants par IA. Cette rareté, combinée aux limites observées dans les récits produits automatiquement, soulève une question centrale : dans quelle mesure les modèles de langage génératif parviennent-ils à produire des contes cohérents, plausibles et engageants pour un jeune public ?

En suivant une approche de practice-led research, ce mémoire évalue un corpus de cent contes générés par quatre modèles d'IA (GPT-4o, Claude Sonnet, LLaMA 3 et Mistral) à partir de triplets de mots-clés. L'analyse s'effectue en deux temps : d'abord à l'aide de la Story Grammar Theory (SGT), qui permet de vérifier la présence et l'organisation des composantes narratives classiques, puis à l'aide d'une grille composite développée spécifiquement pour cette recherche, évaluant la cohérence, la plausibilité interne et l'engagement narratif des récits générés.

Les résultats montrent que les modèles performent bien sur le plan structurel (score SGT moyen de 89 %), mais présentent des lacunes sur le plan qualitatif (score composite moyen de 83 %). L'analyse croisée révèle une très faible corrélation entre les deux outils, confirmant qu'une structure narrative complète ne garantit pas la qualité d'un conte pour enfants. En ce sens, une typologie de treize catégories d'incohérences narratives récurrentes a été développée à partir de ces observations afin de suppléer aux outils actuels les qualités de l’intelligence humaine. Il s’avère important de développer des outils automatisés permettant d’éduquer de futurs outils en fonction de critères narratifs rigoureux.

En somme, ce mémoire contribue à poser les bases méthodologiques d'une évaluation rigoureuse des récits générés par IA destinés à l'enfance, tout en ouvrant des pistes concrètes pour le développement d'outils de génération narrative mieux adaptés au jeune public.

TABLE DES MATIÈRES

RÉSUMÉ ..... **ii**
TABLE DES MATIÈRES ..... **iii**
LISTE DES TABLEAUX ..... **vi**
LISTE DES FIGURES ..... **vii**
LISTE DES ACRONYMES ..... **viii**
INTRODUCTION GÉNÉRALE **1**
1 CHAPITRE 1 : PROBLÉMATIQUE **7**
1.1 Contexte 7
1.2 Domaine de recherche 8
1.2.1 La narrativité générative par intelligence artificielle 8
1.2.2 La narrativité interactive numérique 9
1.2.3 La narrativité immersive 9
1.2.4 Positionnement de cette recherche 10
1.2.5 Précision épistémologique 10
1.3 Hypothèse de recherche 11
1.4 Question de recherche 11
1.5 Objectifs de recherche 12
2 CHAPITRE 2 : REVUE DE LITTÉRATURE **15**
2.1 Introduction de la revue de littérature 15
2.2 Méthodologie de la revue 16
2.3 Trois formes de narrativité 17
2.3.1 Narrativité interactive numérique 18
2.3.2 Narrativité immersive 18
2.3.3 Narrativité générative 19
2.4 Cartographie du corpus 20
2.4.1 Répartition par axe narratif 20
2.4.2 Répartition par catégorie 21
2.4.3 Répartition par sous-catégorie 23
2.4.4 Répartition par axes de sous-catégorie 24
2.4.5 Distribution temporelle du corpus 25
2.4.6 Personnes autrices majeures du corpus 26
2.5 Ressources spécialisées et corpus complémentaires 27
2.6 Cadres théoriques pour l'analyse narrative 28
2.6.1 Structure narrative pour les contes pour enfants 28
2.6.2 Modèles d'analyse narrative : Propp et Fabula 31
2.7 La *Story Grammar Theory* et ses variantes 32
2.8 Synthèse de la revue de littérature 34
3 CHAPITRE 3 : MÉTHODOLOGIE **36**
3.1 Cadre méthodologique général et posture du chercheur-créateur 36
3.1.1 Approche méthodologique mixte 37

3.1.2 De l'Analyseur 17 à la grille composite : une trajectoire méthodologique 38
3.1.3 Méthode itérative et réflexive 39
3.1.4 Générativité et intelligence artificielle 40
3.1.5 Structure méthodologique en deux volets 41
3.2 Calibrage des paramètres utilisés 42
3.3 Phase d'expérimentation technique et calibration des générateurs 47
3.4 Méthode pour atteindre les objectifs spécifiques 49
3.4.1 Objectif 1 : évaluation d'un corpus de contes à partir de la Story Grammar Theory 49
3.4.2 Objectif 2 : conception et expérimentation d'une grille composite adaptée aux contes générés 51
3.4.3 Objectif 3 : développement d'une typologie des incohérences narratives 55
3.5 Synthèse de la méthodologie 56
4 CHAPITRE 4 : RÉSULTATS DE LA RECHERCHE 58
4.1 Objectif 1 : évaluer un corpus de contes à partir de la Story Grammar Theory 59
4.1.1 Résultats de l'évaluation SGT 59
4.1.2 Constats et limites de cette évaluation 64
4.1.3 Identifier les limites de la SGT en contexte automatisé 65
4.2 Objectif 2 : concevoir et expérimenter une grille composite adaptée aux contes générés 69
4.2.1 Résultats de la grille composite 69
4.2.2 Comparaison des lectures humaine et assistée par IA 71
4.2.3 Limites de l'atteinte de l'objectif et de la double lecture 72
4.3 Objectif 3 : développer une typologie des incohérences narratives 73
4.3.1 Présentation de la typologie 73
4.3.2 Répartition et analyse des incohérences 75
4.3.3 Limites de la typologie 79
4.4 Analyse croisée SGT / composite 80
4.4.1 Concordances et divergences 81
4.4.2 Pourquoi les deux grilles divergent ? 83
4.5 Observations transversales 84
4.5.1 Récurrence des personnages 84
4.5.2 Lieux de départ récurrents 86
4.5.3 Lieux secondaires ou absence de changement de décor 87
4.5.4 Objets agentifs 87
4.5.5 Présence et rôle des dialogues dans les contes générés 88
4.5.6 Présence des morales dans les contes 90
4.5.7 Fiche comparative des modèles IA 92
4.5.8 Biais stylistiques et scénaristiques récurrents 94
4.5.9 Impact sur la diversité narrative 96

4.5.10 Les incipits comme révélateurs stylistiques 96
4.5.11 Le biais linguistique et culturel 97
4.5.12 Le déséquilibre des données d'entraînement 100
4.5.13 Ethnocentrisme algorithmique et zone de sécurité narrative 101
4.5.14 Absence de cadre temporel et biais stylistiques 102
4.5.15 La normalisation narrative comme biais structurel 103
4.5.16 Synthèse des observations transversales 104
5 CHAPITRE 5 : DISCUSSION ET ATTEINTE DES OBJECTIFS 105
5.1 Analyse des résultats 105
5.2 Atteinte des trois objectifs spécifiques 106
5.3 Complémentarité des approches d'évaluation 108
5.4 Principale contribution de la recherche 109
5.5 Limites de la présente recherche 112
5.6 Pistes de recherche futures 113
5.7 CONCLUSION 121
6 BIBLIOGRAPHIE 124
7 ANNEXES 128

LISTE DES TABLEAUX

TABLEAU 2.3.3 - TYPOLOGIE DES NARRATIVITES : DEFINITION ET APPLICATIONS 20
TABLEAU 2.6.1 - STRUCTURES NARRATIVES POUR LES CONTES POUR ENFANTS 30
TABLEAU 2.6.2 - MODÈLES D'ANALYSE NARRATIVE: PROPP ET FABULA 31
TABLEAU 3.2C - PARAMÉTRAGES SELON LES MODÈLES SPÉCIFIQUES ANALYSÉS 44
TABLEAU 3.4.2 - VALEUR AJOUTÉE DE LA GRILLE COMPOSITE 54
TABLEAU 4.3.2A - PERTINENCE ET RÔLE DES OBJETS AGENTIFS 75
TABLEAU 4.3.2C - SYNTHÈSE DES PROFILS (FORCES/FAIBLESSES) 79
TABLEAU 4.5.7A - COMPARAISON ENTRE LES MODÈLES 93
TABLEAU 4.5.7B - FORCES ET FAIBLESSES DES GRILLES D'ANALYSE PAR MODÈLE 94
TABLEAU 4.5.11 - STRATÉGIE POUR COMBLER LE BIAIS LINGUISTIQUE ET CULTUREL 99
TABLEAU 5.6C - EXEMPLES DE LA RÈGLE DU LIEU 119
TABLEAU 5.6D - EXEMPLE DE STRATÉGIE ALÉATOIRE 120
TABLEAU 5.6E - EXEMPLE DE DIVERSIFICATION D'AMORCES NARRATIVES 120

LISTE DES FIGURES

FIGURE I.1 - CONCEPT ART DU DISPOSITIF IMMERSIF INTERACTIF 2
FIGURE I.2 - SCHÉMA TECHNIQUE DU DISPOSITIF IMMERSIF INTERACTIF 3
FIGURE 2.4.1 - RÉPARTITION PAR AXE NARRATIF 21
FIGURE 2.4.2 - RÉPARTITION PAR CATÉGORIE 22
FIGURE 2.4.3 - RÉPARTITION PAR TOP SOUS-CATÉGORIE 23
FIGURE 2.4.4 - RÉPARTITION PAR TOP 5 AXES DE SOUS-CATÉGORIE 24
FIGURE 2.4.5 - DISTRIBUTION DES ANNÉES PAR AXE NARRATIF 26
FIGURE 2.4.6 - TOP 5 AUTEURS PAR CONTRIBUTION 27
FIGURE 3.2A - PARAMÉTRAGES DE CONFIGURATION POUR LES CONTES 43
FIGURE 3.2B - PROTOCOLES STANDARDISÉS: LOGIQUE, ÉQUILIBRÉE ET CRÉATIVE 44
FIGURE 3.4.1A – PHASE DE RECHERCHE 50
FIGURE 3.4.1B – PHASE DE CRÉATION 51
FIGURE 4.1.1A - SGT: MOYENNE PAR COMPOSANTE 61
FIGURE 4.1.1B - SGT: DISTRIBUTION DES CONTES SELON LE SCORE OBTENU 62
FIGURE 4.1.1C - CLASSEMENT PAR TRANCHE DE COMPLÉTUDE 63
FIGURE 4.1.1D - BIAIS INDUITS PAR LA GRILLE SGT 64
FIGURE 4.2.1 - PERFORMANCE MOYENNE SELON LA GRILLE COMPOSITE 70
FIGURE 4.3.1 - RÉPARTITION DES INCOHÉRENCES SELON LES CATÉGORIES 74
FIGURE 4.3.2B - RADAR DES INCOHÉRENCES PAR MODÈLES 77
FIGURE 4.4.1A - CONCORDANCES ET DIVERGENCES ENTRE LA SGT ET LA GRILLE COMPOSITE 82
FIGURE 4.4.1B - CONCORDANCES ET DIVERGENCES ENTRE LA SGT ET LA GRILLE COMPOSITE 83
FIGURE 4.5.1A - TOP 10 DES PERSONNAGES PRINCIPAUX 85
FIGURE 4.5.1B - TOP 10 DES PERSONNAGES SECONDAIRES 85
FIGURE 4.5.2 - TOP 10 DES LIEUX DE DÉPART 87
FIGURE 4.5.4 - TOP 10 DES AGENTS AGENTIFS 88
FIGURE 4.5.5A - RÉPARTITION DES CONTES AVEC ET SANS DIALOGUES 89
FIGURE 4.5.5B - RÉPARTITION DES CONTES AVEC ET SANS DIALOGUES PAR MODÈLE 89
FIGURE 4.5.6A - RÉPARTITION DES CONTES AVEC ET SANS MORALE 91
FIGURE 4.5.6B - RÉPARTITION DES CONTES AVEC ET SANS MORALE PAR MODÈLE 92
FIGURE 5.6A - PIPELINE NARRATIF EXPÉRIMENTAL 116
FIGURE 5.6B - PROTOTYPE NARRATIF MULTIMODAL 116

LISTE DES ACRONYMES

API : Interface de programmation
IA : Intelligence artificielle
LLM : Grands modèles de langage
SGT : Story Grammar Theory

INTRODUCTION GÉNÉRALE

Situé dans le domaine de la recherche-création en art numérique, à partir d'une étude d'un corpus de 100 contes générés artificiellement et analysés selon la *Story Grammar Theory* puis une grille composite développée spécifiquement pour cette recherche, ce mémoire vise à évaluer la qualité narrative des récits produits par intelligence artificielle destinés aux enfants, et à poser les bases méthodologiques et critiques d'un futur dispositif narratif interactif adapté à l'enfance.

L'intelligence artificielle connaît actuellement ce que Mustafa Suleyman (2023, p. 68), cofondateur de DeepMind et actuel *Chief Executive Officer* (CEO) de Microsoft AI, qualifie de tournant historique : " *Nous approchons d'un point d'inflexion marqué par l'arrivée de ces technologies avancées, le plus profond de l'histoire.*"

Depuis l'émergence des modèles de langage génératifs comme ChatGPT, la création narrative automatisée connaît une croissance fulgurante. Cette transformation touche particulièrement la littérature jeunesse, où la capacité des intelligences artificielles à produire des récits cohérents, plausibles et engageants soulève à la fois des promesses et des interrogations (Rooein et al., 2025). Dans quelle mesure ces systèmes peuvent-ils générer des contes qui respectent non seulement les structures narratives classiques, mais qui captent aussi l'imaginaire des enfants ? Comment distinguer les récits qui « remplissent toutes les cases » structurelles de ceux qui touchent vraiment leur jeune public ?

Ce mémoire explore ces questions en adoptant une double posture : celle du chercheur qui analyse rigoureusement un corpus, et celle du créateur qui imagine les conditions d'un dispositif narratif futur. Il s'inscrit dans une perspective de recherche-création où l'analyse critique nourrit directement la conception d'outils et de méthodes adaptés à la génération de contes interactifs pour enfants.

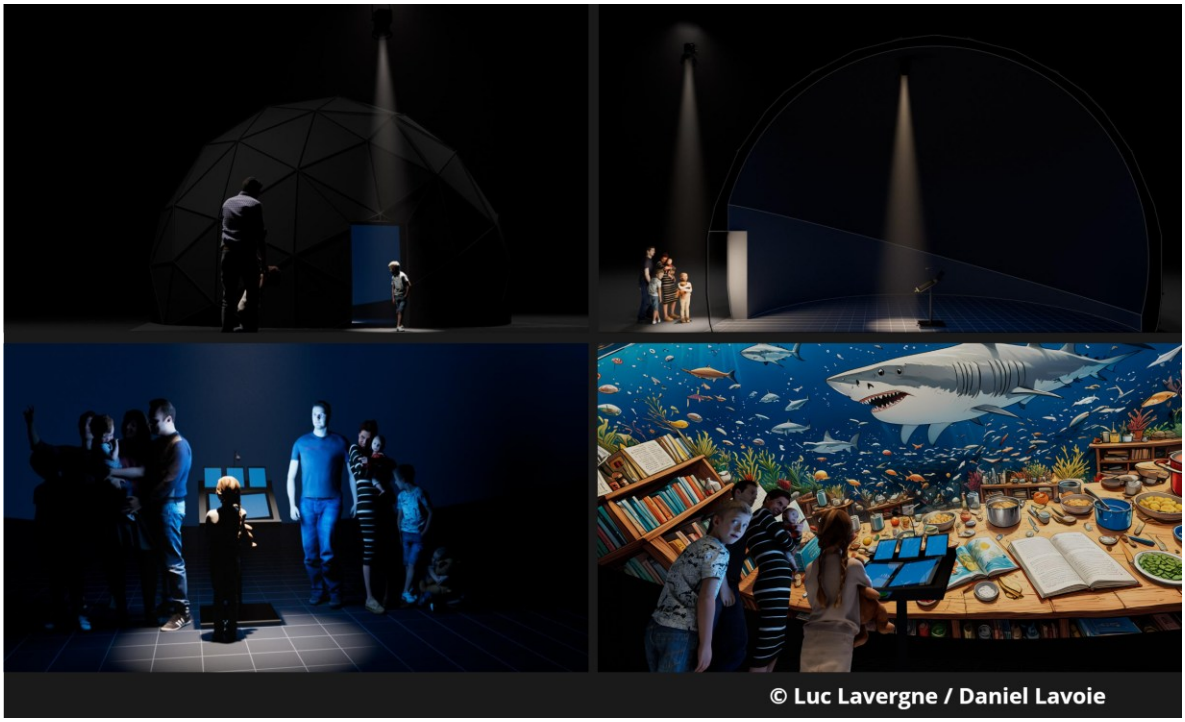


Figure I.1 - Concept art du dispositif immersif interactif

Ce projet de recherche est en amont de la création d'un dispositif narratif multimodal où l'enfant deviendrait cocréateur d'histoires. Plutôt que de recevoir passivement des récits préfabriqués, il pourrait y participer par la parole, le dessin, le geste ou même un objet personnel. L'intelligence artificielle servirait alors de partenaire créatif, capable de générer des récits personnalisés tout en respectant les contraintes de cohérence, de plausibilité et d'engagement propres à la littérature jeunesse.

Cette recherche se situe à l'intersection de trois formes de narrativité contemporaine : la narrativité générative (création automatique de récits par intelligence artificielle), la narrativité interactive numérique (où la personne utilisatrice influence le déroulement de l'histoire) et la narrativité immersive (expériences sensorielles et spatiales). Le schéma technique illustre cette position hybride en montrant comment le projet croise plusieurs domaines : intelligence artificielle, création numérique, conception interactive et narration.

Il est important de préciser que ce prototype constitut le contexte de création au sein duquel la présente recherche est ancrée. Le prototype technique tel que représenté aux figures I.1 et I.2 ont été développés en collaboration avec un autre étudiant à la maîtrise, Luc Lavergne. Ensemble, nous avons conçu une architecture d'intelligence artificielle complète, avec plusieurs preuves de concept qui démontrent et valident le fonctionnement global du système. De mon côté, je me suis concentré sur la portion centrale du projet : la génération du texte narratif. C'est cette dimension qui est au cœur de ma recherche. En ce sens, elle s'inscrit dans une approche de practice-led research, où le processus de création sert de moteur à la réflexion. Le mémoire consiste principalement en une activité d'évaluation des technologies génératives, dans la perspective d'un éventuel retour à la création. Plutôt que de chercher à produire une œuvre terminée, j'ai choisi d'approfondir la question de la narrativité elle-même. Le travail présenté dans ce mémoire vient ainsi poser les bases théoriques et conceptuelles d'un futur prototype narratif interactif complet.

Plus précisément, **cette recherche se situe exclusivement dans le domaine de la narrativité générative textuelle**, laissant délibérément de côté les dimensions visuelles, sonores et immersives du dispositif.

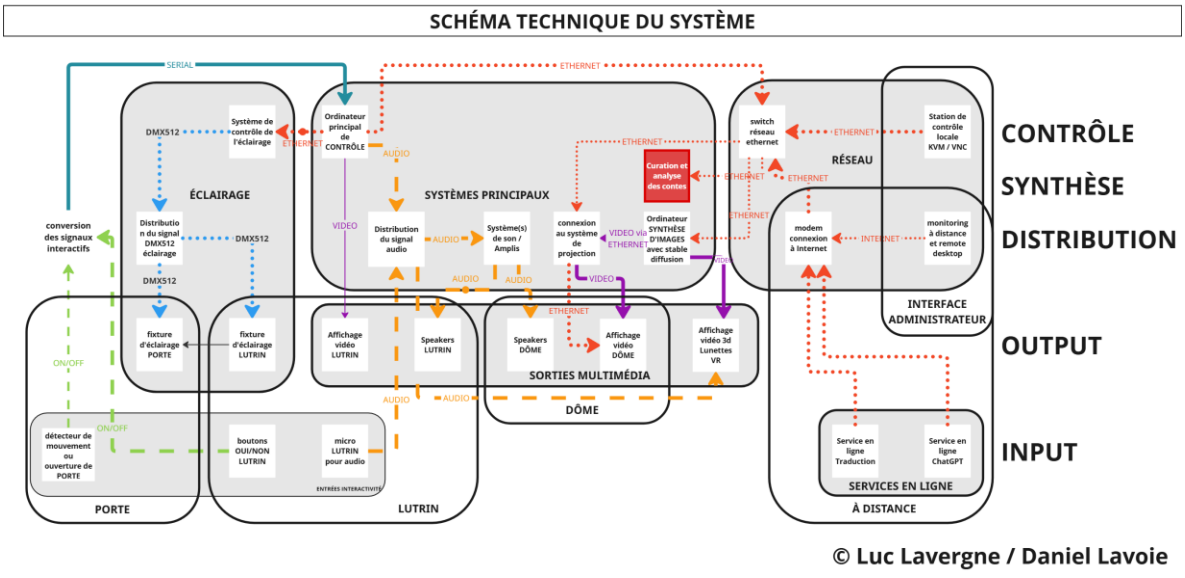


Figure I.2 - Schéma technique du dispositif immersif interactif

Tel que représenté à la figure I.2, le prototype repose sur une architecture modulaire organisée en quatre zones fonctionnelles interconnectées. La zone *Input* capte les requêtes des utilisateurs via des interfaces publiques et des services en ligne à distance, incluant notamment ChatGPT qui génère les contes initiaux ainsi que des services de traduction. Ces contenus sont transmis aux systèmes principaux, qui constituent le cœur de la synthèse : on y trouve l'orchestrateur de traitements et contrôle, les modules de génération et distribution du son, ainsi que les systèmes de synthèse Ethernet et vidéo. Les contenus validés sont acheminés vers la zone Distribution, qui assure le formatage pour le dispositif immersif, la synchronisation et la transmission vers les dispositifs de projection. Enfin, la zone *Contrôle* supervise l'ensemble du système en temps réel via le réseau, les stations de contrôle local et les outils de monitoring.

Mon projet de générateur de contes pour enfants s'articule autour d'un processus itératif impliquant deux zones du système. La génération initiale du texte s'effectue via ChatGPT dans la zone *Input* (services en ligne à distance). Le conte produit est ensuite transmis au module de curation. Celui-ci analyse des contes situés dans les systèmes principaux et examine le texte pour identifier les incohérences narratives, les problèmes de structure ou tout élément inapproprié pour le jeune public. Un flux bidirectionnel (représenté par la flèche rouge en pointillés) permet au système de renvoyer le conte à ChatGPT pour révision et réécriture. Ce cycle itératif se répète jusqu'à ce que le conte soit jugé cohérent, approprié et de qualité suffisante. Le texte validé est alors intégré à la chaîne de production multimédia pour créer l'expérience immersive complète destinée aux enfants.

Dans le cadre de ce mémoire, je concentre ma recherche exclusivement sur le module de curation et d’analyse des contes (représenté en rouge dans le schéma technique de la figure I.2). Cette recherche se limite donc au processus itératif de validation narrative qui s'opère dans les systèmes principaux, entre la génération initiale du texte par ChatGPT (zone Input) et son intégration dans la chaîne de production multimédia. Bien que le système complet comprenne d'autres composantes essentielles, mon analyse porte spécifiquement sur la validation et l'amélioration des textes narratifs destinés aux enfants. Il faut donc préciser que toute analyse des autres formes multimodales

(images, éclairage, son, projection) ainsi que les mécanismes de génération initiale et de distribution finale sont exclues de la présente recherche.

Tel que le mémoire démontrera dans la portion de la revue de littérature, l'analyse d'un large corpus de publications scientifiques révèle un vide critique : très peu de travaux portent spécifiquement sur la génération de contes pour enfants par intelligence artificielle. La majorité des recherches concernent les jeux vidéo, la réalité virtuelle ou d'autres formes interactives qui ne ciblent pas le jeune public. Cette rareté confirme la pertinence de la présente recherche, qui traite les récits générés par IA comme matériaux de réflexion artistique plutôt que comme objets strictement techniques, tout en centrant son attention sur les enjeux spécifiques de l'enfance.

Le mémoire se déploie en cinq chapitres qui s'enchaînent de manière logique et progressive. Le premier chapitre pose la problématique de recherche, formule l'hypothèse centrale et définit les objectifs du travail de recherche. Il pose les bases conceptuelles de cette recherche en situant la génération de contes pour enfants dans le contexte plus large des transformations narratives induites par l'intelligence artificielle. Le deuxième chapitre présente une revue de littérature cartographiant les trois formes de narrativité identifiées et confirmant l'originalité du projet. Le troisième chapitre détaille la méthodologie d'évaluation appliquée à un corpus de 100 contes générés par quatre modèles d'intelligence artificielle différents. Le quatrième chapitre analyse les résultats en trois sections principales correspondant aux trois objectifs, suivi d'observations transversales et d'une discussion sur les approches d'évaluation. Enfin, le cinquième chapitre présente la contribution de cette recherche, démontre l'atteinte des objectifs spécifiques, identifie les limites méthodologiques, et ouvre sur les pistes de recherche futures. Ces cinq chapitres s'enchaînent de manière logique et progressive, depuis l'ancrage théorique jusqu'aux perspectives de création.

L'originalité de cette recherche tient dans son triple apport : méthodologique, par le développement d'outils d'analyse adaptés aux récits générés pour enfants; empirique, par l'identification de biais narratifs encore peu documentés ; et créatif, par l'esquisse d'un dispositif interactif ancré dans les constats de l'analyse critique.

Dans un souci éthique de transparence méthodologique, je précise ici l'usage de l'intelligence artificielle dans le cadre de ce mémoire. J'ai eu recours à des outils d'IA pour des tâches spécifiques : la sélection et l'organisation du corpus de textes analysés, la génération des données du corpus lui-même (les contes produits par ChatGPT), l'élaboration des scripts Python servant d'outils d'analyse des données, ainsi que la révision linguistique (correction grammaticale, amélioration de la clarté et de la structure du texte). J'ai également utilisé l'IA comme outil de dialogue pour clarifier ma pensée, notamment pour la collecte et la classification de concepts ainsi que l'organisation de mes idées. En revanche, l'ensemble de la réflexion, l'élaboration du plan de recherche (avec mon directeur de maitrise), l'analyse des résultats et la formulation des conclusions demeurent le fruit de mon travail. Les idées, concepts, interprétations et positionnements exposés dans ce mémoire sont les miens.

1 CHAPITRE 1 : PROBLÉMATIQUE

1.1 Contexte

Dans son ouvrage sur la convergence des médias, Henry Jenkins (2011) annonçait déjà l'émergence de nouvelles formes narratives capables de complexifier les récits en multipliant les chemins possibles. Il écrivait :

> *"We are seeing the emergence of new story structures, which create complexity by expanding the range of narrative possibility rather than pursuing a single path with a beginning, middle, and end."*

Si Jenkins parlait surtout de transmédia et de culture participative, son idée annonçait déjà une partie des changements que l'on vit aujourd'hui avec l'intelligence artificielle. Les contes pour enfants reposent encore majoritairement sur une structure narrative classique, articulée autour d'une situation initiale, d'un développement (nœud narratif) et d'un dénouement menant à une situation finale. Depuis l'arrivée de modèles génératifs comme ChatGPT, la création narrative automatisée a connu une croissance spectaculaire (Kumar, 2025). La capacité des intelligences artificielles à produire textes, histoires et dialogues transforme en profondeur des domaines comme l'éducation, le divertissement et la création numérique. La vitesse d'adoption est sans précédent : ChatGPT a atteint un million d'utilisateurs en cinq jours, alors qu'Instagram a mis 2,5 mois et Netflix 3,5 ans pour franchir ce seuil (Exploding, 2025).

Dans ce contexte fulgurant, la capacité de l'intelligence artificielle (IA) à générer des récits cohérents, plausibles et engageants pour un jeune public soulève des enjeux majeurs, notamment en matière de cohérence narrative, de contrôle sémantique, de maîtrise des contraintes structurelles et de prévisibilité des enchaînements narratifs (Concepción et al., 2016). Ce mémoire repose sur le constat que l'automatisation de la narration ouvre des perspectives inédites pour la création immersive en art

numérique. En contrepartie, il est nécessaire de déployer une réflexion rigoureuse et approfondie sur les critères d'évaluation de la qualité narrative (R. Li, 2024).

Le domaine de l'intelligence artificielle est actuellement en pleine effervescence dans tous les domaines de la société. Le seul secteur spécialisé de la génération de récits par IA évolue presque chaque jour et révèle à la fois une grande diversité et des régularités inattendues. La littérature analysée au chapitre suivant met en évidence plusieurs biais récurrents dans les récits générés : des biais de répétition, liés aux données d'entraînement disponibles, des biais de conformité structurelle, associés aux formes narratives conventionnelles, ainsi que des biais de plausibilité. Ils se manifestent notamment par l'émergence de personnages stéréotypés, de situations répétées et de structures narratives proches (Rettberg & Wigers, 2025; Rooein et al., 2025). C'est à partir de ces constats que se sont dessinés les premiers questionnements qui ont orienté ce mémoire. Ces observations s'inscrivent dans un champ de recherche plus vaste : celui de la narrativité générative par intelligence artificielle.

1.2 Domaine de recherche

En ce sens, le domaine de recherche se situe au croisement de trois formes de narrativité contemporaine qui redéfinissent les manières de raconter et d'expérimenter les histoires à l'ère numérique : la narrativité générative, la narrativité interactive numérique et la narrativité immersive (voir section 2.3 pour un développement complet de ces concepts).

1.2.1 La narrativité générative par intelligence artificielle

La narrativité générative désigne la capacité des systèmes d'intelligence artificielle à produire automatiquement des récits cohérents à partir d'amorces ou de contraintes définies. Plusieurs chercheurs ont déjà exploré ces potentialités. Mark Riedl, par exemple, a travaillé sur la narration interactive et les systèmes intelligents (2013). Zhuohan Xie, Trevor Cohn et Jey Han Lau (2023) ont

analysé les structures narratives produites par des modèles de génération textuelle avancés. Dans le même esprit, Swartjes et Theune (2006) ont mis en lumière le rôle des modèles de fabula dans l'émergence de récits cohérents, tandis que Carmichael et Mould (2014) se sont intéressés à la manière dont la cohérence narrative peut être maintenue dans des environnements interactifs.

Dans le contexte de la littérature jeunesse, la génération procédurale d'histoires par des modèles de langage me semble ouvrir un terrain particulièrement prometteur, permettant de créer des récits qui peuvent soutenir des objectifs éducatifs, ludiques ou émotionnels adaptés à l'enfance.

1.2.2 La narrativité interactive numérique

La narrativité interactive numérique place la personne utilisatrice au cœur de l'évolution du récit. Les recherches dans ce domaine montrent que cette direction permet d'imaginer des récits qui ne suivent pas nécessairement une logique linéaire, où les décisions de l'utilisateur peuvent modifier directement le déroulement de l'histoire. Des travaux récents documentent des dispositifs et des cadres pour structurer et analyser des récits interactifs et adaptatifs (Aylett & Louchart, 2003; Koenitz, 2015; Roth & Koenitz, 2017; Ryan, 2006).

1.2.3 La narrativité immersive

La narrativité immersive se distingue par sa capacité à plonger la personne participante dans un environnement sensoriel ou émotionnel intense, que ce soit par la réalité virtuelle, la réalité augmentée, les interfaces multisensorielles ou des dispositifs narratifs spatialisés. Marie-Laure Ryan, notamment avec *Narrative as Virtual Reality 2*, explore la frontière entre récit et expérience vécue. D'autres auteurs examinent comment la spatialisation, la temporalité subjective et les modalités perceptives façonnent une nouvelle grammaire du récit.

1.2.4 Positionnement de cette recherche

Ce projet de mémoire s'inscrit dans cette dynamique de convergence, en examinant spécifiquement comment la génération automatisée de contes par IA peut nourrir la conception d'un dispositif narratif interactif et immersif adapté à l'enfance.

L'intégration de la génération d'images à partir de textes enrichit encore ces perspectives, permettant d'associer à chaque récit généré une composante visuelle singulière qui stimule l'imaginaire des jeunes lecteurs. Cette convergence entre texte, image et interaction ouvre de nouvelles pistes pour repenser la narration immersive et créer des contes personnalisés et profondément engageants.

À l'échelle institutionnelle, des organismes comme ARDIN *(Association for Research in Digital Interactive*), dédiée à l'avancement de la narration interactive numérique, et IDC *(Interaction Design and Children*), centrée sur la conception de technologies interactives adaptées à l'enfance, offrent des ressources précieuses pour encadrer la recherche et stimuler l'innovation

1.2.5 Précision épistémologique

Cette recherche ne relève pas de la linguistique et encore moins de l'ingénierie des modèles de langage. Elle prend place dans une perspective de recherche-création en art numérique, où les récits générés par intelligence artificielle sont traités comme matériaux de réflexion artistique plutôt que comme objets strictement techniques. Il ne s'agit pas d'optimiser les algorithmes, mais de comprendre leurs usages et de les mettre au service d'une perspective créative particulière adaptée aux formes narratives propres à l'enfance. C'est à partir de ce positionnement que je formule mon hypothèse de recherche.

1.3 Hypothèse de recherche

Ce mémoire part de l'hypothèse que l'analyse de contes générés par intelligence artificielle, lorsqu'elle s'appuie sur des outils structurés comme la *Story Grammar Theory* (SGT), (Stein & Glenn, 1977) et qu'elle se prolonge par une lecture plus fine de leur cohérence et de leur plausibilité, permet de mieux comprendre jusqu'où ces récits respectent les structures fondamentales et dans quelle mesure ils peuvent susciter un véritable engagement.

La démarche se déploie en deux temps. Dans un premier temps, j'examine si les composantes narratives essentielles (cadre, déclencheur, tentative, résolution, etc.) sont présentes et bien articulées selon les critères de la SGT. Dans un second temps, je dresse un inventaire des incohérences et des tendances observées, puis je cherche, lorsque c'est pertinent, à les rattacher à des cadres conceptuels issus de la narratologie, de la linguistique et de la théorie de la réception. Cette posture inductive et réflexive signifie que les cadres ne sont pas posés d'avance, mais mobilisés comme des ressources au moment où l'analyse en révèle la nécessité.

L'objectif de cette recherche n'est donc pas d'optimiser les modèles d'IA, mais de comprendre ce qu'ils produisent et de développer des outils d'analyse adaptés aux enjeux de la narration enfantine. J'envisage l'intelligence artificielle comme un agent de création partielle, avec lequel il devient possible de coconstruire des récits immersifs, structurés et porteurs de sens.

1.4 Question de recherche

Comme le rappellent Bruneau et Villeneuve dans *Traiter de recherche-création en art*, « une question de recherche, c'est avant tout une phrase avec un point d'interrogation » (Bruneau & Villeneuve, 2007). Celle-ci doit guider le chercheur dans sa recherche et délimiter la terminologie est le terrain

d'investigation. Le domaine de la narrativité immersive automatisé est immense, le présent mémoire se concentre donc sur un seul modèle d'évaluation et énonce dans quelle optique la recherche sera réalisée :

Dans ce mémoire, ma question s'énonce de la façon suivante :

« Dans une perspective recherche pour la création en art numérique, comment l'évaluation de contes générés par intelligence artificielle à partir de la Story Grammar Theory de Stein et Glenn (1977) peut-elle alimenter la conception d'outils d'analyse et de réflexion destinés à soutenir le développement d'un dispositif narratif immersif pour enfants ? »

Cette question guide une démarche en deux phases complémentaires :

Une première phase d'évaluation structurée des récits générés par IA à partir de la grille canonique de la *Story Grammar Theory* (SGT), afin d'en dégager les forces, les limites et le niveau de structuration narrative ; une seconde phase d'analyse pour la conception d'une grille composite expérimentale. Cette grille permet d'évaluer plus finement la cohérence, la plausibilité et l'engagement des contes générés, et ouvre la voie à la conception d'un prototype de dispositif narratif interactif destiné aux jeunes publics.

Ce projet se situe à la jonction entre analyse structurée et analyse réflexive pour la création, et propose des outils et des méthodologies pour mieux intégrer la narration générative dans le champ de la création d'expérience numérique (XN) destinée à l'enfance.

1.5 Objectifs de recherche

Dans cette continuité, ce mémoire cherche à comprendre comment l'intelligence artificielle peut générer des récits cohérents et engageants pour un jeune public, en évaluant la qualité narrative de contes produits par différents modèles de langage. Les objectifs sont atteints par l'analyse structurée des récits générés à l'aide d'outils narratologiques établis, et par la conception d'une grille composite adaptée à l'enfance, dans une perspective de recherche-création en art numérique. Il faut mentionner que cette grille, une fois produite, n'est pas évaluée par l'intelligence artificielle mais m'offre un cadre d'analyse qualitatif me permettant d'évaluer la cohérence du récit.

Objectif général :

La recherche vise à développer une grille d'analyse qualitative permettant d'évaluer la qualité narrative des contes générés par intelligence artificielle, dans une perspective de recherche-création en art numérique orientée vers le jeune public.

Objectifs spécifiques :

Je désire contribuer à la recherche en suivant trois objectifs spécifiques. Ceux-ci ont permis de structurer la recherche et limiter les outils méthodologiques pour répondre à la question posée :

1. **Évaluer** un corpus de contes générés par différents modèles d'intelligence artificielle à partir de la Story Grammar Theory (Stein & Glenn, 1977), afin d'identifier les forces et limites de ce cadre théorique dans un contexte de génération automatisée.
2. **Concevoir** et expérimenter une grille d'analyse composite, adaptée aux récits pour enfants générés par IA, en combinant une lecture humaine et une lecture assistée par intelligence artificielle.
3. **Développer** une typologie des incohérences et des structures narratives récurrentes observées dans le corpus, en vue d'alimenter la réflexion sur la conception d'outils d'analyse et de dispositifs narratifs immersifs destinés aux enfants.

Il est important de rappeler que cette démarche ne cherche pas à créer un dispositif narratif complet ; elle en pose plutôt les bases critiques et méthodologiques pour mieux évaluer les biais et logiques de création de textes par intelligence artificielle. Cependant, avant d'utiliser une telle technologie, je considère essentiel, sur le plan éthique, de développer des outils critiques permettant de comprendre les mécanismes en jeu plutôt que de se laisser porter par la technologie existante.

Pour ancrer cette démarche dans les connaissances existantes et identifier les outils narratologiques pertinents, le chapitre 2 cartographie le champ de recherche en narrativité générative et situe la Story Grammar Theory dans les travaux actuels.

2 CHAPITRE 2 : REVUE DE LITTÉRATURE

J'ai effectué cette revue de littérature afin de situer le domaine général, d'établir les grandes tendances et de proposer un état de l'art cohérent en lien avec mon objet de recherche.

La liste du matériel complet utilisé pour la revue de littérature (qui s'étend de septembre 2022 à août 2025) comprenant l'ensemble des informations pertinentes aux différents classements se retrouve sous ces deux liens. J'ai généré des permaliens afin de partager les données de la revue de littérature. Il est possible de consulter la liste complète des références[1] et les données et catégories d'analyse de la revue de littérature[2] en ligne.

2.1 Introduction de la revue de littérature

Force est de constater qu'il s'agit d'un domaine de recherche dynamique reposant sur une imposante littérature. Cette revue de littérature a donc été limitée aux articles permettant d'offrir une synthèse permettant soutenir la double démarche du mémoire soit : situer l'usage de la *Story Grammar Theory* (SGT) dans les travaux existants afin de préparer l'analyse d'un corpus de contes générés par IA ; nourrir la conception d'une grille d'analyse hybride adaptée à la narrativité jeunesse et interactive.

En m'appuyant sur une veille scientifique continue réalisée pendant l'année 2025, j'examine les recherches en narrativité générative, en narration interactive et en conception immersive spécifiquement adaptée à l'enfance. L'objectif est de dégager les tendances et méthodes qui soutiennent cette démarche réflexive. La sélection des sources a été guidée par leur pertinence pour ce double ancrage méthodologique et créatif.

Évaluer la qualité narrative de ces récits suppose de s'appuyer sur des cadres structurants qui permettent d'en analyser les composantes. Dans cette optique, la *Story Grammar Theory* sera

[1] Lavoie, D. (2025). revue_litterature_all_hyperlinks_v01. Zenodo. Adresse : https://tinyurl.com/3c3u7vyd
[2] Lavoie, D. (2025). revue_litterature_analyses_v02 [Data set]. Adresse : https://tinyurl.com/yv2m6saa

mobilisée pour étudier le corpus, tandis que la grille hybride viendra compléter et affiner cette démarche.

Dans les sections suivantes, je cartographie le corpus (axes, catégories et sous-catégories), je formule des observations transversales, puis j'examine l'évolution temporelle des publications afin de situer ces tendances. Pour amorcer ce travail, je présente d'abord la méthodologie de la revue.

2.2 Méthodologie de la revue

En premier lieu, je précise que je ne suis ni linguiste ni spécialiste de l'analyse littéraire ou philosophique. Je ne prétends pas produire une étude exhaustive en narratologie. Mon intention est de mener une recherche-création en art numérique où la narration générée par IA devient un terrain à explorer, à critiquer et à transformer. Mon but est de m'en inspirer pour créer un outil adapté à l'enfance. Ma posture est à la fois inductive et réflexive : je ne pars pas d'un cadre théorique figé, je m'appuie sur les constats de l'analyse pour mobiliser, quand c'est utile, certains cadres conceptuels. Ces cadres deviennent des appuis pour construire une grille évolutive et critique, capable d'évaluer la qualité narrative des contes générés.

Pour structurer cette revue, un corpus de 397 articles, livres et références a été constitué à partir de bases scientifiques pertinentes. Chaque article a d'abord été parcouru et trié manuellement selon sa pertinence potentielle au regard des objectifs du projet. Seuls les textes offrant un apport clair à la compréhension de la narration générative, de l'interactivité numérique ou des enjeux liés à l'enfance ont été retenus, tandis que d'autres champs comme les applications de jeu vidéo ou certaines approches techniques trop éloignées ont été exclus.

Chaque article retenu a ensuite été évalué à l'aide de l'Analyseur 17, un système d'évaluation conçu spécifiquement pour cette maîtrise. Développé initialement sous forme de script Python, cet outil s'est heurté à la complexité des textes académiques : formats hétérogènes, erreurs de classification, interprétation constante des critères d'évaluation. Il a donc été stabilisé comme prompt structuré, plus

stable et adaptable, combinant extraction automatique et analyse qualitative. Dans sa forme finale, il génère pour chaque publication une fiche d'analyse complète selon 17 critères méthodologiques : 1-titre, 2-auteur(s), 3-date, 4-résumé, 5-arguments principaux, 6-méthodes, 7-mots-clés, 8-analyse de l'abstract, 9-références importantes, 10-contributions novatrices, 11-limites, 12-synthèse, 13-figures, 14-analyse critique, 15-références bibliographiques, 16-classement End Note et 17-pertinence pour le mémoire.

En suivant une approche quantitative, chaque critère reçoit un score pondéré, permettant de calculer une note finale (0-100%) qui reflète la pertinence de l'article pour la recherche. Cette note apparaît à la fois dans la grille d'évaluation et dans le nom du fichier généré (exemple : "*Storylets you want them* 84.5%.docx"). L'outil classe également chaque article selon une taxonomie de 6 catégories principales et leurs sous-catégories, facilitant les regroupements thématiques ultérieurs (voir section 3.9). Les résultats de cette évaluation structurée ont ensuite servi à établir des regroupements thématiques et à produire les graphiques présentés dans la section 2.3.

Ils font également apparaître trois formes principales de narrativité, développées à la section suivante : la narrativité interactive numérique, la narrativité immersive et la narrativité générative.

2.3 Trois formes de narrativité

Trois grandes tendances se dégagent de l'analyse des articles : la narrativité générative, la narrativité interactive numérique et la narrativité immersive. Ces formes, bien qu'interconnectées, se distinguent par leur relation à la personne utilisatrice, aux technologies mobilisées et à la structure du récit.

Cette revue de littérature ne cherche pas seulement à brosser un portrait des approches en narration générative, interactive ou immersive. Elle s'inscrit aussi dans une posture de recherche-création, où les sources servent autant à nourrir la réflexion qu'à façonner, pas à pas, les outils mobilisés dans ce mémoire. Chaque lecture a contribué à bâtir une compréhension sensible et critique de la

narrativité enfantine à l'ère de l'intelligence artificielle. Pour amorcer ce parcours, j'aborde d'abord la narrativité interactive numérique, qui met en lumière le rôle central de la personne utilisatrice dans l'évolution du récit.

2.3.1 Narrativité interactive numérique

La narrativité interactive numérique rassemble les travaux qui placent la personne utilisatrice au cœur de l'évolution du récit. Szilas (1999) a montré que cette façon permet d'imaginer des récits qui ne suivent pas nécessairement une logique linéaire. Les recherches de Charles, Mead et Cavazza (2001) montrent également que les décisions de la personne utilisatrice peuvent modifier directement le déroulement de l'histoire.

Les sous-catégories EndNote associées à cet axe sont : Narration interactive, Histoires adaptatives, Logiciels d'IA et Pipeline narratif. Des travaux de Hartmut Koenitz, Ruth Aylett, Sandy Louchart, Chris Roth, Chris Hargood et Marie-Laure Ryan documentent des dispositifs et des cadres pour structurer et analyser des récits interactifs et adaptatifs.

2.3.2 Narrativité immersive

La narrativité immersive se distingue par sa capacité à plonger la personne participante dans un environnement sensoriel ou émotionnel intense, que ce soit par la réalité virtuelle, la réalité augmentée, les interfaces multisensorielles ou des dispositifs narratifs spatialisés. Les sous-catégories EndNote associées à cet axe sont : Expérience immersive et Supports multimodaux.

Les articles recensés traitent de la présence, de l'incarnation (*embodiment*), du rôle du corps dans la réception narrative, ainsi que des effets psychologiques et physiologiques de l'immersion. Parmi les contributions clés figure Marie-Laure Ryan, notamment avec *Narrative as Virtual Reality 2*, où elle

explore la frontière entre récit et expérience vécue. D'autres auteurs examinent comment la spatialisation, la temporalité subjective et les modalités perceptives façonnent une nouvelle grammaire du récit (Aylett, 2013; Koenitz, 2010; Lescop, 2019; Xu et al., 2018). À côté de ces expériences immersives, la narrativité générative s'intéresse à la production même du récit par des systèmes d'intelligence artificielle.

2.3.3 Narrativité générative

La narrativité générative rassemble les travaux qui portent sur la création automatique de récits par des intelligences artificielles ou des programmes automatisés. Ces récits naissent de modèles de langage, de simulateurs narratifs ou d'algorithmes guidés par des prompts ou des graphes d'histoire (structure arborescente à la « livre dont vous êtes le héros »).

Ce champ de recherche est en pleine effervescence, surtout depuis l'arrivée des grands modèles de langage (LLM) entre 2019 et 2025.

Parmi les auteurs influents de cet axe, on retrouve notamment des travaux portant sur la cohérence structurelle, l'adaptation thématique, la qualité stylistique et la réception émotionnelle dans un contexte automatisé (Gervás et al., 2005; Meng et al., 2022; Young & Riedl, 2003).

Type de narrativité	Définition	Applications
Narrativité générative	Utilisation de l'IA pour générer des récits de manière autonome, avec possibilité d'ajustements humains.	Jeux vidéo, littérature interactive, outils d'aide à l'écriture.
Narrativité interactive numérique	Permet aux utilisateurs d'influencer activement le déroulement de l'histoire par leurs choix.	Jeux vidéo narratifs, livres interactifs, applications éducatives.
Narrativité immersive	Plonge le participant dans l'univers narratif, souvent via la réalité virtuelle ou augmentée, pour une expérience sensorielle et émotionnelle.	Réalité virtuelle, installations artistiques, théâtre interactif.

Tableau 2.3.3 - Typologie des narrativités : définition et applications

En résumé, le tableau ci-dessus regroupe les trois formes de narrativité. La section 2.4 montre ensuite leur répartition dans le corpus, par axes, catégories et sous-catégories.

2.4 Cartographie du corpus

Cette section présente la cartographie du corpus selon différents niveaux d'analyse. Les sous-sections détaillent successivement la répartition par axe narratif (2.4.1), par catégorie (2.4.2), par sous-catégorie (2.4.3), par axes de sous-catégorie (2.4.4), par distribution temporelle (2.4.5) ainsi que les personnes autrices majeures du corpus (2.4.6).

2.4.1 Répartition par axe narratif

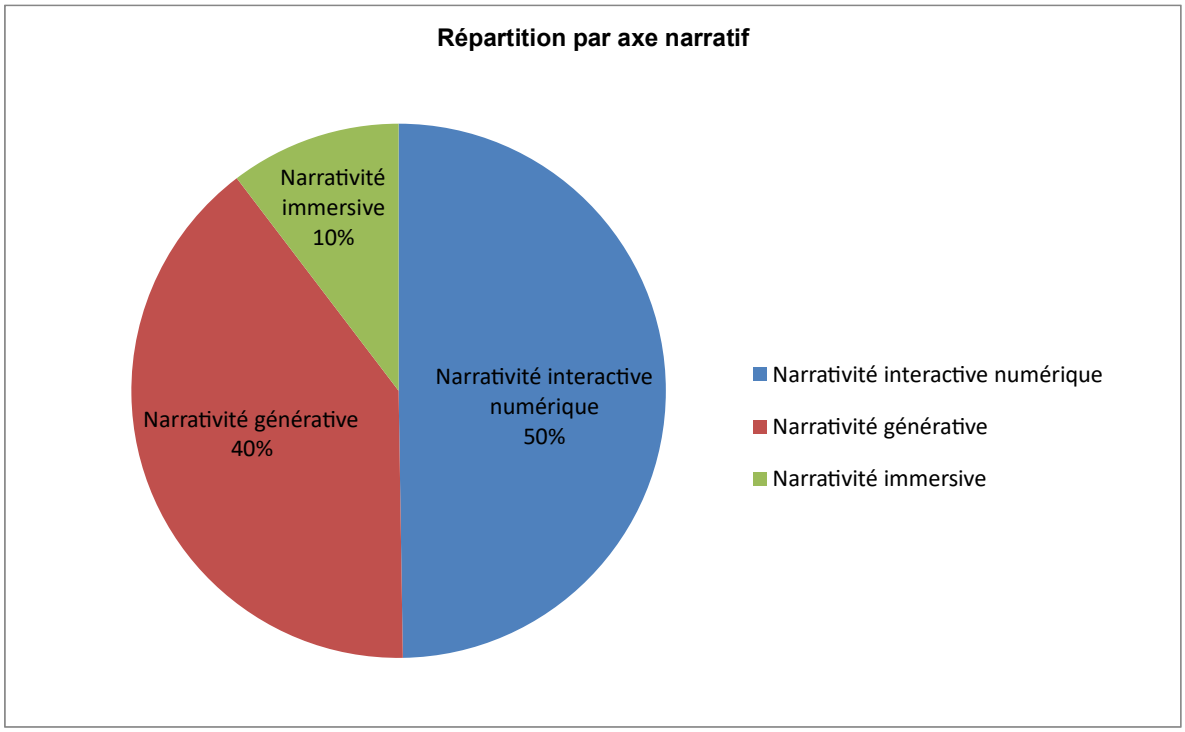


Figure 2.4.1 - Répartition par axe narratif

La répartition par axe narratif montre un corpus dominé par la narrativité interactive numérique (50 %), suivie de la narrativité générative (40 %) et, plus marginalement, de la narrativité immersive (10 %). Cette distribution met en évidence le poids accordé aux récits interactifs, tout en montrant la progression des approches génératives et la place encore limitée de l'immersion.

2.4.2 Répartition par catégorie

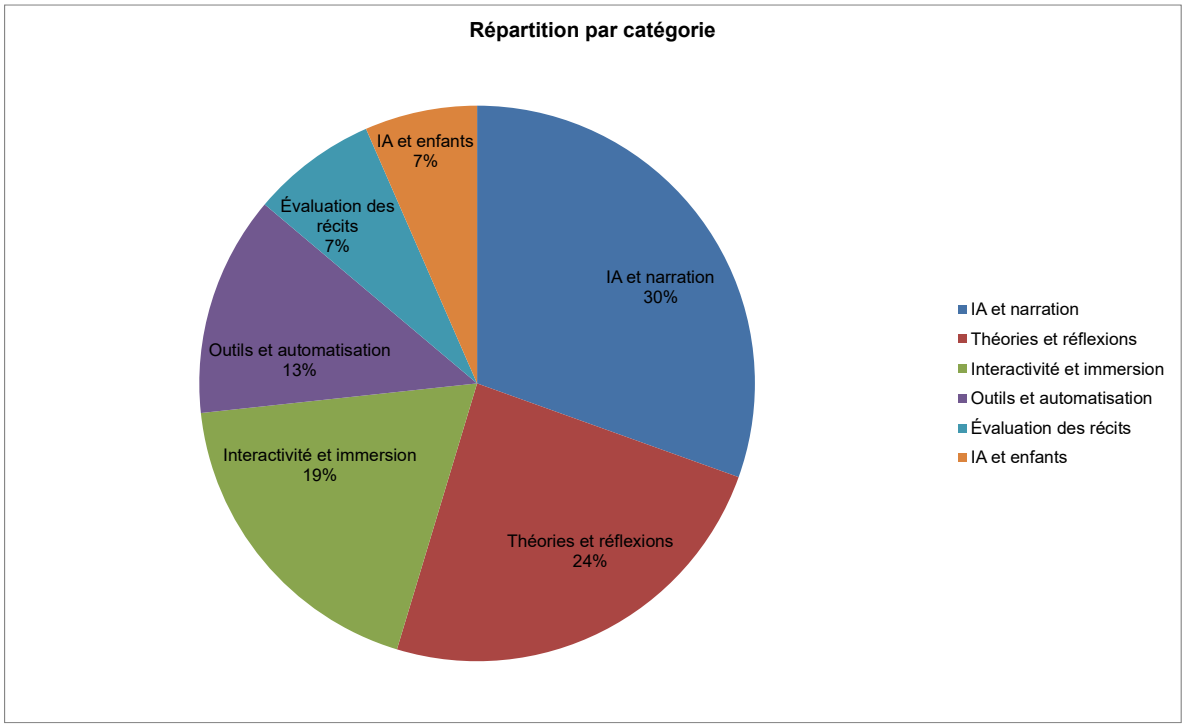


Figure 2.4.2 - Répartition par catégorie

La répartition par catégories principales met en évidence la place dominante des recherches sur l'IA et la narration (30 %), suivies par celles liées aux théories et réflexions (24 %) et à l'interactivité et l'immersion (19 %). Les catégories outils et automatisation (13 %) et évaluation des récits (7 %) occupent une position intermédiaire, traduisant un intérêt réel mais moins marqué. La catégorie IA et enfants (7 %) reste marginale, ce qui confirme la rareté des travaux portant spécifiquement sur les jeunes publics.

Dans l'ensemble, cette répartition montre un champ structuré autour de pôles théoriques et narratifs solides, tout en laissant apparaître des pistes encore peu explorées pour l'enfance et les approches multimodales.

- IA et enfants (histoires adaptées, impact éducatif, supports multimodaux)
- Théories et réflexions (éthique, enjeux futurs, théories narratives)
- Outils et automatisation (comparaison d'outils, pipelines de génération narrative, logiciels d'IA)
- Interactivité et immersion (narration interactive, expérience immersive, adaptativité des histoires)
- Évaluation des récits (cohérence, crédibilité, méthodes d'évaluation)

- IA et narration (contes générés, collaboration humain-IA, modèles narratifs, optimisation)

2.4.3 Répartition par sous-catégorie

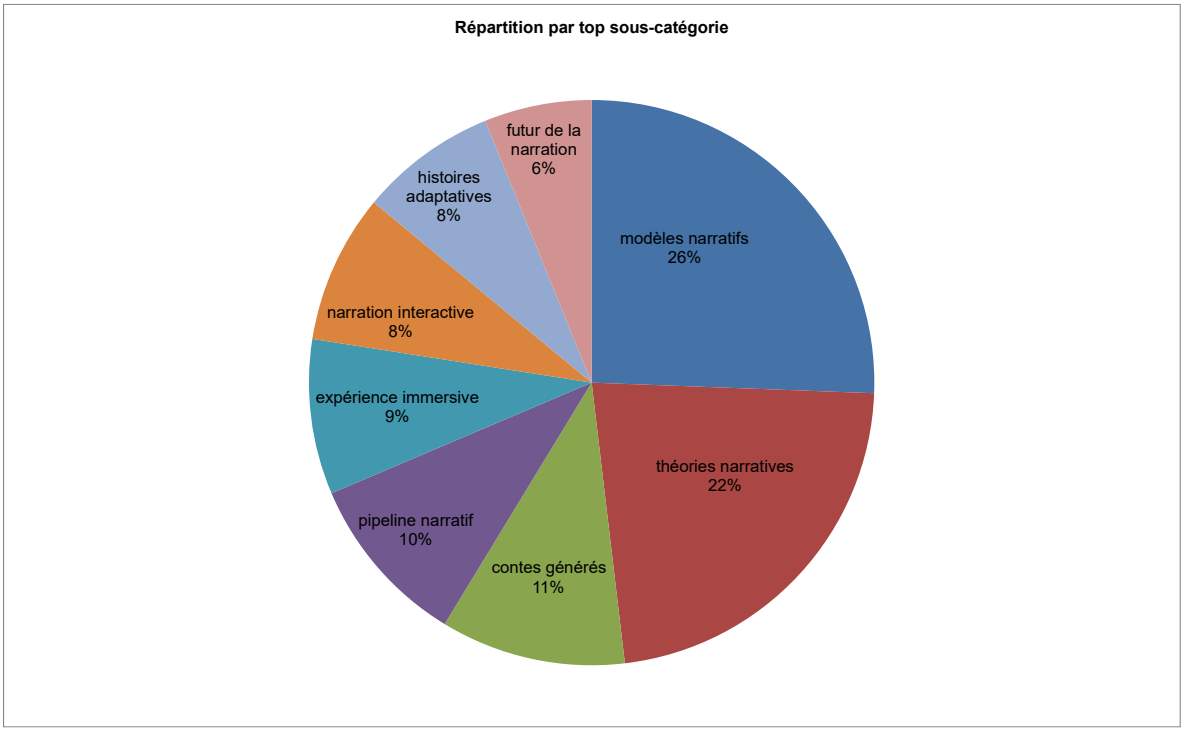


Figure 2.4.3 - Répartition par top sous-catégorie

L'analyse par sous-catégorie met en évidence une forte concentration de travaux autour des modèles narratifs (26 %) et des théories narratives (22 %), qui représentent ensemble près de la moitié du corpus.

Des pôles secondaires apparaissent ensuite, tels que les contes générés (11 %), le pipeline narratif (10 %) et l'expérience immersive (9 %). Les sous-catégories narration interactive (8 %), histoires adaptatives (8 %) et futur de la narration (6 %) occupent une place plus réduite mais non négligeable.

Cette répartition confirme le rôle central accordé aux cadres théoriques et aux modèles formels, tout en faisant ressortir des approches plus appliquées qui demeurent encore en marge.

2.4.4 Répartition par axes de sous-catégorie

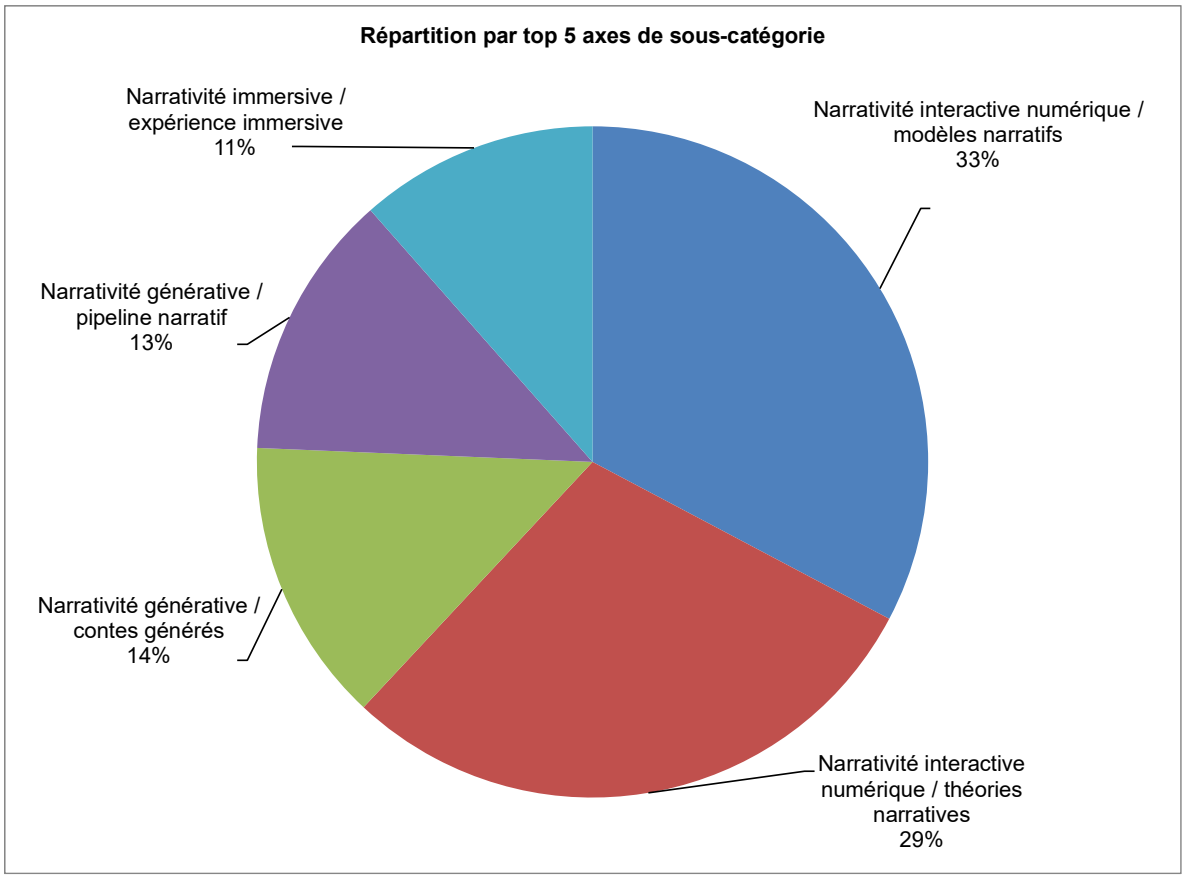


Figure 2.4.4 - Répartition par top 5 axes de sous-catégorie

Pour mieux montrer la dynamique entre les catégories principales et leurs déclinaisons, le graphique ci-dessus présente la répartition des sous-catégories au sein de chaque axe narratif. On voit que les modèles narratifs (33 %) et les théories narratives (29 %), regroupés dans l'axe « IA et narration », forment le noyau dur de la recherche actuelle. L'axe "Narrativité générative" se divise entre les contes générés (14 %) et le pipeline narratif (13 %), tandis que l'axe "Immersion" s'appuie principalement sur l'expérience immersive (11 %).

Cette cartographie souligne les liens entre axes principaux et sous-catégories, en mettant en lumière les zones de concentration et les approches émergentes. Elle permet ensuite de situer l'évolution de ces recherches dans le temps (2.4.5), avant de mettre en évidence les personnes autrices les plus représentées dans le corpus (2.4.6).

2.4.5 Distribution temporelle du corpus

L'évolution du corpus montre que l'intérêt pour la narrativité générative et interactive reste marginal avant 2000, puis s'intensifie clairement entre 2000 et 2009, avec une forte poussée de la narrativité interactive numérique et une croissance régulière de la narrativité générative. La décennie 2010-2019 marque un véritable pic d'activité pour l'ensemble des axes, avant une reconfiguration récente entre 2020 et 2025.

Il faut toutefois préciser comment lire la présence de certaines publications dans les axes narratifs dès le début du XXe siècle. Par exemple, Freytag's *Technique of the Drama* (Freytag, 1900) apparaît ici sous l'étiquette de narrativité interactive numérique. Bien sûr, Freytag ne parlait pas d'ordinateurs ni d'intelligence artificielle : son ouvrage relève de la théorie dramatique classique. Mais en proposant une structuration du récit qui met en valeur la progression et la participation du lecteur/spectateur, ce texte devient un antécédent théorique de ce que l'on appelle aujourd'hui la narrativité interactive.

Autrement dit, certains jalons anciens sont reclassés rétrospectivement dans les catégories actuelles, non pas parce qu'ils étaient « numériques » à l'époque, mais parce qu'ils éclairent les bases conceptuelles qui permettront plus tard à la narrativité numérique et interactive de se développer.

Cette façon de classer me permet de montrer que les idées associées à l'interactivité ou à la structure participative d'un récit précèdent de loin l'arrivée des technologies numériques. En ce sens, les pointes de la courbe avant les années 2000 doivent être comprises comme des repères conceptuels plutôt que comme de véritables travaux numériques.

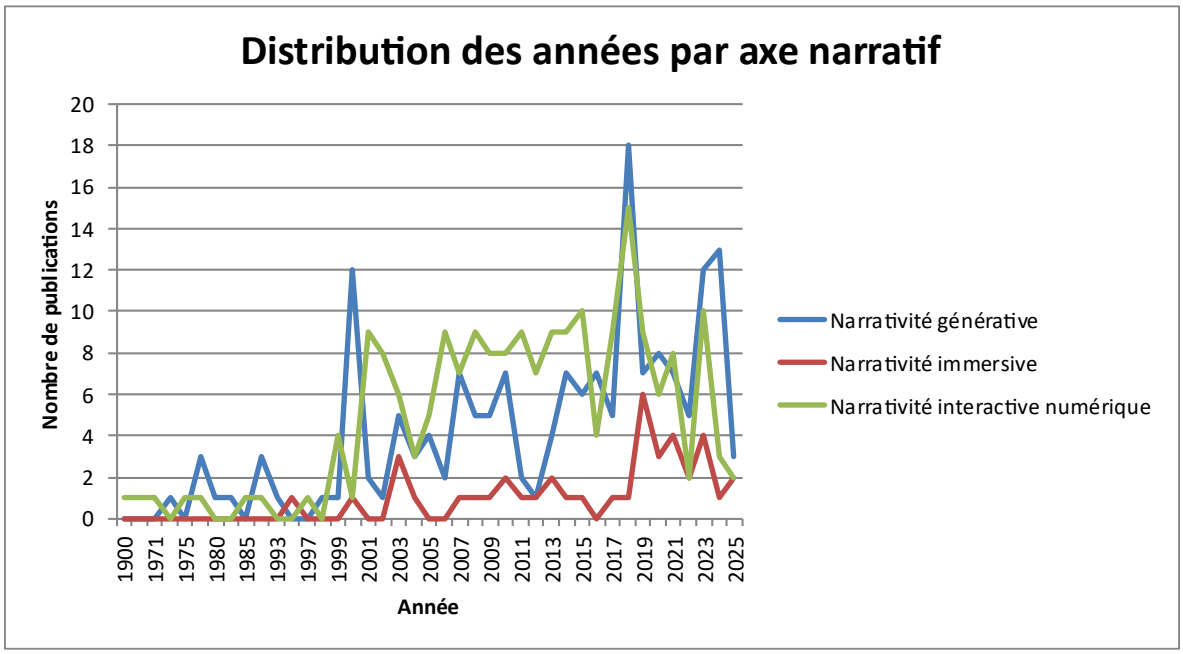


Figure 2.4.5 - Distribution des années par axe narratif

2.4.6 Personnes autrices majeures du corpus

En complément de la répartition thématique, il est pertinent d'identifier les auteurs et autrices les plus représentés dans le corpus. Ce repérage met en lumière les figures majeures qui structurent la recherche en narrativité générative, interactive et immersive. Il s'agit ici d'un classement quantitatif, basé sur la fréquence des publications retenues.

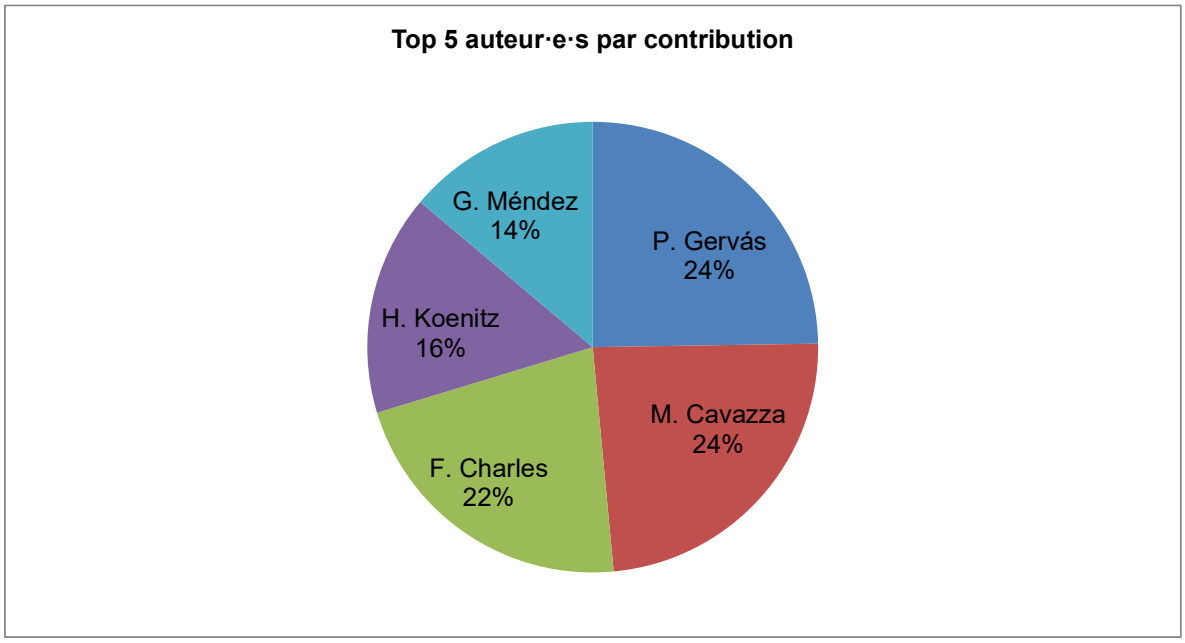


Figure 2.4.6 - Top 5 auteurs par contribution

Ces résultats confirment l'importance de quelques figures phares : Pablo Gervás (24 %) et Marc Cavazza (24 %) occupent la première place, suivis de Fred Charles (22 %). Hartmut Koenitz (16 %) et Gonzalo Méndez (14 %) complètent ce top 5. Leurs travaux constituent des repères importants qui contribuent à définir le champ.

2.5 Ressources spécialisées et corpus complémentaires

En complément de la revue principale, j'ai aussi exploré des ressources issues de conférences et de réseaux spécialisés, pour approfondir les liens entre interactivité, enfance et narration numérique.

L'*Association* for *Research in Digital Interactive Narratives* (ARDIN) regroupe les recherches les plus récentes sur la narration interactive numérique. Elle organise chaque année *l'International Conference on Interactive Digital Storytelling* (ICIDS). Mon adhésion récente m'a donné accès à l'ensemble des actes des conférences de 2001 à 2023. J'ai aussi assisté à une conférence en ligne organisée par ARDIN, où j'ai pu poser quelques questions à Hartmut Koenitz, une figure majeure du

domaine. Ces échanges indirects, même ponctuels, ont enrichi ma compréhension des enjeux contemporains liés à la narrativité interactive.

L'*Association for Computing Machinery* (ACM) organise chaque année la conférence *Interaction Design and Children* (IDC), qui constitue une source précieuse d'articles scientifiques, axés sur la conception d'interfaces centrées sur l'enfant, l'apprentissage interactif et les technologies éducatives. Malgré le fait que les conférences ACM IDC portent davantage sur l'interaction que sur la narration à proprement parler, les articles qui en sont issus offrent des perspectives utiles pour concevoir des expériences immersives adaptées aux jeunes publics.

En complément, je me suis aussi procuré un abonnement d'un an à la plateforme Academia. Grâce à des filtres thématiques précis, je reçois régulièrement une sélection d'articles liés à mes champs d'intérêt. Cette veille quotidienne me permet de suivre l'actualité des recherches sur la narrativité générative et interactive, et d'alimenter en continu la réflexion menée dans ce mémoire.

Ces apports confirment la vitalité du domaine tout en révélant des zones encore peu explorées, en particulier celles qui concernent l'enfance. La conclusion qui suit propose une synthèse de ces constats et souligne la pertinence de la démarche adoptée dans ce mémoire.

2.6 Cadres théoriques pour l'analyse narrative

Avant de présenter la méthodologie d'évaluation appliquée au corpus de contes générés, il est nécessaire de situer les cadres théoriques qui structurent cette recherche. Cette section présente les modèles narratologiques qui serviront d'outils d'analyse, en particulier la *Story Grammar Theory* qui constitue le socle de l'évaluation structurelle des récits.

2.6.1 Structure narrative pour les contes pour enfants

Une structure narrative solide est essentielle pour capter l'attention des enfants et rendre un conte à la fois engageant et instructif. Les composantes fondamentales de la structure narrative soutiennent la plausibilité de l'histoire. Cette structure varie selon l'âge ciblé et selon le type de conte.

Pour les enfants de 0 à 3 ans

Structure	Description	Étapes
Simple et répétitive	Mise sur la répétition pour créer un sentiment de familiarité.	1. Introduction simple 2. Répétition d'un motif 3. Conclusion rassurante
Basique en trois parties	Adaptée à l'attention limitée, avec un développement très simple.	1. Introduction d'un concept 2. Un événement simple 3. Conclusion rapide
De la découverte	Stimule la curiosité et l'apprentissage par l'exploration.	1. Présentation d'un élément 2. Exploration/découverte 3. Conclusion avec apprentissage
Rythmique et sonore	Utilise des rimes et des rythmes pour captiver.	1. Introduction avec rimes/sons 2. Développement rythmique 3. Conclusion rythmique
Imagerie	Basée sur des images pour compenser le langage limité.	1. Introduction avec image 2. Succession d'images 3. Image finale marquante

Pour les enfants de 4 à 8 ans

Structure	Description	Étapes
Classique "5 actes"	Structure traditionnelle avec un conflit et une résolution.	1. Introduction 2. Problème 3. Action 4. Résolution 5. Conclusion
Classique "3 actes"	Structure traditionnelle avec un conflit et une résolution.	1. Exposition 2. Conflit 3. Résolution
Circulaire	L'histoire se termine où elle a commencé, créant un cycle.	1. Introduction similaire à la conclusion 2. Développement 3. Conclusion en boucle
Interactive	Permet aux enfants de choisir la direction de l'histoire.	1. Introduction 2. Points de choix 3. Divers chemins et fins

Pour les enfants de 6 à 12 ans		
Structure	**Description**	**Étapes**
En miroir	Les événements de la première partie se reflètent dans la seconde.	1. Introduction 2. Série d'événements 3. Événements miroirs 4. Conclusion
En emboîtement	Histoires dans l'histoire, avec des niveaux de narration.	1. Histoire principale 2. Histoires secondaires 3. Conclusion principale
Non linéaire	Présente des événements dans un ordre non chronologique.	1. Événements non chronologiques 2. Flashbacks/Forwards 3. Conclusion
De la quête	Accent sur le voyage ou l'aventure du héros.	1. Introduction du héros 2. Présentation de la quête 3. Voyages et épreuves 4. Accomplissement 5. Retour/Conclusion
Thématique	Autour d'un thème central ou d'une leçon morale.	1. Introduction du thème 2. Exploration du thème 3. Conclusion liée au thème
En trois actes	Très courante dans la littérature et le cinéma.	1. Introduction des personnages 2. Conflit 3. Résolution du conflit
De Frank Daniel Moore	Basée sur une progression dramatique en cinq étapes.	1. Introduction 2. Montée 3. Point culminant 4. Chute 5. Résolution

Tableau 2.6.1 - Structures narratives pour les contes pour enfants

Note : Les tableaux proposés dans la Figure 2.6.1 sont issus d'une élaboration expérimentale dans le cadre de cette recherche. Ils s'inspirent de principes narratologiques généraux (par ex. la Story Grammar Theory) et de ma pratique de création et d'enseignement. Ils ne prétendent pas constituer une typologie universelle, mais visent à fournir un outil comparatif et opérationnel pour l'analyse des contes générés par IA.

En général, un conte pour enfants s'organise autour d'une structure narrative relativement stable. Cette organisation correspond au schéma narratif classique, souvent décrit en cinq étapes : la situation initiale, l'élément perturbateur, les péripéties, la résolution et la situation finale (Larivaille, 1974). Dans le cadre de cette recherche, ces étapes peuvent être décrites de la manière suivante :

1. Introduction : présentation du ou des personnages principaux et du contexte initial.

2. Problème ou élément déclencheur : survenue d'un événement perturbateur qui rompt l'équilibre initial.

3. Actions et péripéties : tentatives successives pour surmonter l'obstacle.

4. Résolution : événement ou action qui permet de résoudre le problème.

5. Conclusion : retour à un nouvel équilibre, souvent accompagné d'une morale ou d'un apprentissage.

Ce type de structure se retrouve dans de nombreuses formes de contes, qu'ils soient merveilleux, animaliers, de voyage ou de quête, mythologiques, réalistes, humoristiques ou initiatiques.

2.6.2 Modèles d'analyse narrative : Propp et Fabula

Parmi les modèles d'analyse narrative, celui de Vladimir Propp (1970) occupe une place historique majeure. Il a mis en évidence 31 fonctions récurrentes (par exemple : interdiction, don, victoire) et 7 personnages types (héros, méchant, faux héros, etc.), qui interagissent selon un schéma stable. Le modèle est ancien et souvent jugé rigide, mais il reste un repère pour les premières formalisations de la structure narrative.

Le concept de fabula, issu du formalisme russe, désigne la suite des événements d'une histoire dans leur ordre chronologique « réel ». Il s'oppose au *sjuzhet*, c'est-à-dire l'ordre narratif choisi par le récit. Cette distinction a été largement reprise par des narratologues contemporains, notamment Marie-Laure Ryan, qui l'a appliquée à l'étude des récits interactifs et numériques dans *Narrative as Virtual Reality 2* (2015).

Aspect	Modèle Propp	Concept de Fabula
Origine	Développé par Vladimir Propp (Morphologie du conte, 1928)	Introduit par Boris Tomashevsky (formalisme russe), repris par de nombreux narratologues
Objet d'étude	Structure des contes de fées et identification de fonctions narratives récurrentes	Chronologie réelle des événements d'un récit (histoire) par opposition à l'ordre de présentation (récit)
Structure clé	31 fonctions narratives + 7 rôles types (héros, méchant, adjuvant, etc.)	Séquence linéaire et causale des événements de l'histoire
But	Décrire un schéma universel récurrent dans les contes merveilleux	Distinguer la structure profonde (fabula) de la structure racontée (sjuzhet)
Utilisation	Analyse des contes traditionnels et identification de motifs récurrents	Étude de la temporalité et de la causalité dans les récits

Tableau 2.6.2 - Modèles d'analyse narrative: Propp et Fabula

Le modèle de Propp et le concept de fabula aident à comprendre certaines dimensions fondamentales de la construction narrative. La *Story Grammar Theory*, développée dans un contexte éducatif et cognitif, propose un cadre plus structuré. Bien qu'elle ait été conçue bien avant l'essor des récits générés par intelligence artificielle, ses critères se prêtent aujourd'hui à l'évaluation de ce type de productions. Avant de présenter son utilisation dans cette recherche, il est pertinent d'examiner brièvement les principaux modèles qui en sont issus.

2.7 La *Story Grammar Theory* et ses variantes

Depuis la fin des années 1970, plusieurs versions de la *Story Grammar Theory* (SGT) ont vu le jour, chacune apportant ses nuances à la description de la structure narrative. Ces modèles visent à identifier les composantes essentielles d'un récit et leurs relations, afin de mieux comprendre comment les histoires sont construites, comprises et mémorisées.

Stein et Glenn (1977) proposent une structure en sept composantes, largement utilisée pour l'étude des récits jeunesse: *Setting* (mise en place), *Initiating Event* (événement déclencheur), *Internal Response* (réaction interne), *Goal* (objectif), *Attempt* (tentative), *Outcome* (résultat) et *Reaction* (réaction finale). Ce modèle, clair et opérationnel, sert de référence principale dans ce mémoire.

Mandler et Johnson (1977) démontrent une organisation hiérarchique des épisodes narratifs : les éléments sont regroupés en segments reliés par des liens de cause à effet, ce qui met en lumière l'architecture globale d'un récit et la façon dont les événements s'imbriquent. Cette méthode souligne l'importance des relations causales dans la compréhension et la mémorisation des histoires.

Rumelhart (1975) décrit la narration comme un réseau de propositions reliées par des liens logiques, adoptant une perspective issue de la linguistique computationnelle. Ce modèle propositionnel a influencé de nombreux travaux de modélisation et d'analyse automatique des récits, notamment en sciences cognitives et en informatique.

Applebee (1980) propose une synthèse des recherches sur la compréhension et la production de récits chez les enfants. Il montre comment les enfants acquièrent progressivement les bases du récit : présenter un personnage, exposer clairement un problème, cheminer vers une résolution et intégrer une morale. Ces apports permettent de situer l'évaluation des contes générés en fonction des compétences narratives attendues selon l'âge.

Trabasso et Van den Broek (1985) se concentrent sur la structure causale des événements : un récit cohérent repose sur la solidité des liens qui relient actions, intentions et conséquences. Cette

perspective complète la SGT en offrant un critère de plausibilité et de cohérence interne allant au-delà de la simple présence formelle des composantes narratives.

Ces modèles reposent sur une base commune de composantes narratives, mais ils varient par leur niveau de détail et leur angle d'analyse. Dans ce mémoire, je m'appuie principalement sur la structure de Stein et Glenn (1977), dont le modèle en sept composantes offre un cadre clair, opérationnel et largement utilisé dans la recherche sur la compréhension des récits. D'autres travaux (Mandler & Johnson, Rumelhart, Applebee, Trabasso & van den Broek) proposent des perspectives complémentaires : segmentation en épisodes, grammaire narrative, développement des compétences chez l'enfant, relations causales, mais ils ne sont pas explorés en profondeur ici.

La Story Grammar Theory, telle qu'adoptée dans ce mémoire, s'inscrit dans une tradition de recherches qui ont affiné et diversifié les manières de décrire la structure des récits. La grille complète est reproduite en **Annexe A**.

Malgré sa solidité comme outil d'analyse structurelle, la SGT présente des limites importantes dans un contexte de génération automatique. Conçue pour des productions humaines, elle ne permet pas d'évaluer les glissements de ton, les erreurs d'agentivité, la richesse des dialogues, l'accessibilité lexicale pour le jeune public, ni la cohérence fonctionnelle des composantes. Elle vérifie la présence des éléments narratifs, mais pas leur qualité d'usage. Ces limites justifient le développement d'un outil complémentaire, présenté au chapitre 3.

2.8 Synthèse de la revue de littérature

Cette revue de littérature met en lumière un corpus riche, diversifié et en constante évolution. Il révèle à la fois la complexité des récits générés ou assistés par intelligence artificielle et leur potentiel créatif, éducatif et expérientiel.

Sur près de 400 documents recensés, très peu répondent directement à ma question de recherche. La majorité des travaux portent sur les jeux vidéo, la réalité virtuelle ou d'autres formes immersives, tandis que les recherches spécifiquement axées sur la génération de contes pour enfants demeurent extrêmement rares. Ce constat confirme la pertinence et la nécessité de cette recherche, qui cherche justement à combler ce vide.

Au-delà de cette rareté, la lecture du corpus a nourri ma phase réflexive. Elle m'a permis d'explorer différentes approches, de saisir les perspectives critiques et prospectives des chercheurs, et de m'inspirer pour imaginer comment concevoir un dispositif narratif interactif adapté à l'enfance. Cette dimension réflexive, typique de la recherche-création, trouve un prolongement direct dans les discussions et perspectives du chapitre 5.

L'analyse quantitative du corpus révèle un champ en pleine recomposition, marqué par la montée en puissance de la narrativité générative et interactive depuis les années 2000. Les travaux recensés montrent une concentration importante autour des modèles narratifs et des théories structurelles (près de 50 % du corpus), traduisant une volonté persistante de formaliser le récit, même à l'ère de l'automatisation. Cette tendance souligne que, malgré la prolifération des outils d'intelligence artificielle, la recherche continue de s'appuyer sur des cadres narratologiques hérités pour penser la cohérence, la causalité et l'expérience du récit. À l'inverse, les publications portant sur l'enfance et sur les formes immersives demeurent marginales (moins de 10 % chacune), ce qui met en évidence un angle mort significatif : celui de la conception de récits adaptés aux jeunes publics dans des environnements numériques interactifs.

Ces résultats quantitatifs me permettent de situer ma recherche à l'intersection de deux zones peu explorées : la narrativité générative appliquée aux contes pour enfants et l'analyse critique des récits produits par IA. En ce sens, le corpus ne sert pas seulement à dresser un état des lieux, mais à repérer les écarts, les manques et les tensions qui justifient la démarche expérimentale du mémoire. Cette lecture statistique se prolonge donc par une lecture interprétative et créative : elle oriente la conception de la grille d'analyse composite et prépare l'expérimentation des contes générés à travers la *Story Grammar Theory*.

En somme, la constitution de ce corpus a permis d'identifier trois grandes formes de narrativité (générative, interactive et immersive) qui ont servi à cartographier le champ et à situer la Story Grammar Theory (SGT) dans un ensemble plus large. Afin de circonscrire le champ de la recherche, j'ai retenu ce cadre conceptuel, reconnu par les pairs et la communauté scientifique, afin d'évaluer la cohérence des contes générés.

3 CHAPITRE 3 : MÉTHODOLOGIE

La démarche de recherche de ce mémoire est mixte, à la fois rigoureuse et réflexive. Elle s'inscrit dans une perspective de recherche-création en art numérique, où analyse critique et fabrication théorique se renforcent mutuellement. Elle prépare le terrain pour les résultats de recherche (la génération de contes automatisés) et l'analyse des résultats issus de l'application conjointe des grilles SGT et composite sur le corpus.

3.1 Cadre méthodologique général et posture du chercheur-créateur

Ce chapitre présente la méthode utilisée pour explorer la plausibilité narrative dans les récits générés par intelligence artificielle. Il fait le lien entre la phase de création du projet et l'analyse du corpus, en montrant comment la pratique et la réflexion se complètent dans une démarche de recherche-création.

La question de la plausibilité narrative dans les textes générés par IA, apparue dans le cadre d'une première phase de création, a orienté toute la suite du travail. En considérant l'intelligence artificielle comme un véritable partenaire créatif, j'ai entrepris une exploration systématique de la génération de contes pour enfants afin d'en évaluer la cohérence, la structure et la crédibilité.

La posture globale de la recherche s'inscrit dans une approche de *creative practice research*, telle que définie par Batty et Zalipour (2024) et Smith (2009), où la recherche et la création forment un même mouvement de production de connaissance. Dans cette perspective, la création ne vient pas illustrer la recherche, elle en est la source. Mon travail relève plus précisément du *practice-led research* (2009): c'est en créant et en évaluant les contes que j'ai développé mes outils d'analyse et généré mes constats. Cette approche met l'accent sur la pratique comme moteur de la recherche, mais avec pour objectif principal de générer des connaissances théoriques. La pratique artistique

sert ici à informer ou à enrichir des théories existantes plutôt qu'à produire de nouvelles connaissances à travers l'acte créatif lui-même.

Dans la continuité de Biggs et Büchler (2007), cette recherche conserve la même rigueur méthodologique que la recherche académique traditionnelle, tout en adaptant ses critères à la spécificité de la pratique artistique. La démarche vise ainsi à concilier l'exigence scientifique avec la liberté exploratoire propre à la création.

Ce mémoire s'inscrit dans un parcours de recherche-création en trois temps. La première phase a consisté à développer un prototype d'IA narrative. La deuxième phase, qui correspond au présent mémoire, se concentre sur l'évaluation systématique de la narrativité générée à travers l'analyse d'un corpus de 100 contes et le développement d'outils critiques adaptés (grille composite, typologie des incohérences). Cette phase combine recherche (par l'évaluation des modèles d'IA avec la SGT et la grille composite) et création (par la génération automatisée d'un corpus narratif et la conception d'outils d'analyse). La troisième phase, qui demeure prospective, mobilisera les résultats de cette recherche pour concevoir un dispositif narratif interactif complet destiné aux enfants. Le mémoire constitue donc un pont entre une première exploration créative et un futur projet de création informé par la recherche.

3.1.1 Approche méthodologique mixte

Cette recherche adopte une approche méthodologique mixte, combinant des méthodes quantitatives et qualitatives. La dimension quantitative se manifeste dans la génération systématique de 100 contes à partir de triplets de mots-clés, et dans l'évaluation mesurable via la SGT et la grille composite avec attribution de scores. La dimension qualitative se déploie à travers l'observation réflexive des récits, l'analyse interprétative des incohérences narratives, et l'évaluation critique des modèles d'IA en relation avec les cadres théoriques narratologiques. Cette approche semi-quantitative permet de

concilier la rigueur de l'analyse structurée avec la sensibilité nécessaire à l'évaluation de la qualité narrative destinée aux enfants.

Dans ce projet de maîtrise, je ne suis jamais un observateur extérieur. Je travaille comme chercheur-créateur, en m'impliquant dans toutes les étapes : concevoir le dispositif, générer les contes, les lire, les analyser et inventer des outils pour mieux comprendre ce qui en ressort. Cette posture me permettra de combiner deux dimensions : la rigueur d'une analyse structurée et l'attention sensible aux conditions concrètes dans lesquelles ces récits sont produits et pourraient être reçus par des enfants.

3.1.2 De l'Analyseur 17 à la grille composite : une trajectoire méthodologique

Face à un corpus de plus de 400 publications, j'ai développé l'Analyseur 17, un système d'évaluation systématique qui m'a confronté pour la première fois aux questions qui allaient structurer toute ma recherche : comment définir des critères d'évaluation rigoureux sans rigidifier le processus ? Comment équilibrer automatisation et jugement humain ? Comment transformer une masse d'informations en connaissance structurée ?

L'Analyseur 17 m'a aussi appris l'importance de la pondération. Tous les critères ne pèsent pas le même poids : la pertinence pour la question de recherche (15 %) n'a pas la même valeur que la richesse des références (10 %). Ces pondérations reposaient sur mon propre jugement de chercheur-créateur : en comparant les scores générés avec ma lecture personnelle des articles, j'ai calibré les pondérations jusqu'à ce que les résultats reflètent fidèlement ma perception de la pertinence de chaque publication pour ma recherche. Cette hiérarchisation des critères, que j'ai dû ajuster plusieurs fois, m'a préparé à concevoir plus tard la pondération des incohérences dans la grille composite (fautes majeures vs mineures).

C'est à partir de cette expérience que se développera la grille composite pour les contes générés. Elle ne viendra pas d'un modèle théorique figé, mais de mon expérience directe avec les récits. Je prendrai des notes, regrouperai des observations, corrigerai, ajusterai, parfois même changerai d'avis

sur mes propres critères. Peu à peu, la grille deviendra plus qu'un outil technique. Elle sera une manière de comprendre ce que les récits générés par IA peuvent vraiment montrer : leurs forces, leurs faiblesses, leurs répétitions, et parfois des trouvailles inattendues.

À chaque lecture, je chercherai à identifier le décalage entre ce qu'on attend d'un bon conte pour enfants et ce que produisent les modèles. C'est ce décalage qui me guidera. Je ne chercherai pas à imposer une norme stricte, mais à mettre en place un cadre souple, capable de montrer autant les erreurs que les particularités.

Je conçois la grille composite comme un objet hybride : à la fois une méthode, un outil d'analyse et une façon de réfléchir. Elle me permettra de faire le pont entre le langage des modèles d'IA et ce qu'on attend d'un récit destiné aux enfants : cohérence, engagement et crédibilité.

3.1.3 Méthode itérative et réflexive

La méthode adoptée est itérative : elle alterne entre phases de création, d'observation et de réflexion critique. Concrètement, la phase de création consistait à générer des contes via les modèles et à concevoir les outils d'analyse ; la phase d'observation, à lire et coder les récits à l'aide des grilles SGT et composite ; la phase de réflexion critique, à consigner dans MIRO mes constats, ajustements et questionnements au fil de l'analyse. La recherche ne se limite pas à observer les récits générés ; elle les questionne, les code, les mesure et les met en dialogue avec les cadres théoriques existants afin de mieux comprendre ce que signifie « raconter une histoire » dans un contexte d'intelligence artificielle.

Mon parcours se déploiera sur deux plans complémentaires. D'un côté, je mènerai un travail analytique : constituer un corpus de cent contes générés à partir de mots-clés, les évaluer avec la SGT et avec la grille composite, comparer les résultats et en tirer des constats. De l'autre, je suivrai une démarche plus expérientielle : entrer moi-même dans ces histoires, juger ce qui fonctionne ou

non (la force d'une morale, la clarté d'un dialogue, la crédibilité d'un personnage) et construire peu à peu une typologie critique qui m'aidera à mieux comprendre la qualité narrative des textes.

Cette démarche m'amènera aussi à mettre les outils eux-mêmes à l'épreuve. En croisant les résultats de la SGT et de la grille composite, je ne mesurerai pas seulement la valeur des contes, je questionnerai aussi ce que ces outils révèlent et ce qu'ils laissent dans l'ombre quand on essaie de définir ce qu'est un bon récit pour enfants. Les choix subjectifs que je ferai au fil de l'analyse ne seront pas des biais à effacer : ils feront partie intégrante d'une recherche-création, où l'acte d'évaluer est toujours lié à une voix, une expérience et un projet plus large.

Tout au long de cette démarche, j'ai utilisé MIRO comme un journal de bord et un dépôt central. J'y ai colligé mes réflexions, mes hésitations, les reformulations de grille, ainsi que des tableaux, des résumés et même des extraits de mes échanges avec ChatGPT et Claude. Ces notes ont été relues périodiquement afin d'identifier des patterns récurrents, de repérer les tensions entre mes attentes et les résultats obtenus, et d'ajuster les critères de la grille composite en conséquence. Le résultat de cette réflexion critique se concrétise dans les ajustements méthodologiques documentés tout au long du chapitre 3. Cet espace m'a permis de suivre le fil du processus et d'en dégager la logique d'ensemble.

Cette méthodologie reconnaît que ma subjectivité de chercheur-créateur fait partie du processus. Plutôt que de l'écarter, je l'assume comme un élément qui nourrit la production de connaissances. Mon analyse avance donc en dialogue constant avec les outils numériques et les récits produits.

3.1.4 Générativité et intelligence artificielle

Dans le cadre de ce mémoire, la générativité désigne la capacité des modèles à produire des textes narratifs cohérents à partir d'amorces définies. Ces textes reposent sur des modèles entraînés sur de vastes corpus, qui apprennent des régularités de style, de vocabulaire et de structure. Les

modèles d'IA récents génèrent le texte pas à pas, en choisissant à chaque étape la suite de mots la plus probable en fonction de ce qui a déjà été écrit.

À mon sens, plusieurs limites demeurent : maintenir la cohérence à long terme, assurer la plausibilité des événements et des personnages, éviter les ruptures de ton et dépasser la simple variabilité pour atteindre une vraie originalité. Dans ce projet, l'interrogation de différents modèles à partir de triplets de mots-clés a mis en évidence que chaque système adopte sa propre logique de construction narrative, avec des régularités repérables dans les contes générés.

3.1.5 Structure méthodologique en deux volets

Le processus d'évaluation s'articule en deux grands volets méthodologiques :

Volet 1 : Évaluation structurée des récits générés par IA

L'objectif ici est d'examiner un corpus de 100 contes générés à partir de modèles de langage (GPT-4o, Claude 3, LLaMA 3 et Mistral). Ces quatre modèles ont été retenus pour leur accessibilité, leur représentativité des principales familles de LLM disponibles au moment de la recherche, et la diversité de leurs architectures, permettant ainsi une comparaison significative des logiques de génération narrative. L'analyse repose sur deux grilles successives. D'abord, la *Story Grammar Theory* (SGT) de Stein et Glenn (1977) est utilisée pour déterminer si les composantes fondamentales de la structure narrative classique sont présentes dans les textes générés (voir section 2.5.3 pour une présentation détaillée de ce modèle et de ses variantes).

Ensuite, une grille composite a été élaborée à partir d'un examen plus approfondi du corpus, afin de relever les erreurs et incohérences et de tenter de les codifier. Cette grille permet une évaluation plus nuancée de la cohérence, de la plausibilité et de l'engagement narratif, en tenant compte des spécificités du jeune public.

Volet 2 : Conception d'un outil critique expérimental

Ce volet correspond à l'objectif de création. Il s'agit de concevoir une grille analytique adaptée aux récits générés par IA et orientée vers une éventuelle intégration dans un dispositif narratif interactif pour enfants. Ce dispositif, même s'il ne fait pas partie de la réalisation de ce mémoire, constitue l'arrière-plan réflexif qui a guidé la conception de la grille. La grille se veut donc à la fois un outil d'évaluation rigoureux et une étape vers la création d'un environnement immersif et narratif.

La double lecture, humaine et assistée par IA, s'inscrira dans cette même logique. En confrontant mes propres évaluations à celles produites par les modèles, j'observerai à la fois des convergences et des écarts. Cela me permettra de mieux cerner les limites de chaque approche et d'explorer ce que pourrait être une collaboration entre humain et machine dans l'analyse critique de contes.

3.2 Calibrage des paramètres utilisés

Pour rendre les résultats comparables et pertinents, un calibrage précis des paramètres de génération a été effectué avant la production des contes. La plateforme ChatGPT a d'abord servi de point de départ pour explorer les paramètres disponibles (température, longueur maximale, format de sortie, structure narrative, tonalité adaptée aux enfants, etc.) et tester leur influence sur la qualité des récits. L'objectif n'était pas de produire uniquement avec ce modèle, mais de disposer d'un référentiel technique permettant d'émuler ou d'adapter ces paramètres auprès d'autres grands modèles de langage (Claude, Mistral, Meta), en tenant compte de leurs contraintes propres.

Un dictionnaire de mots adaptés au jeune public a été conçu spécifiquement pour ce projet (voir section 3.3 pour les détails de sa conception et des ajustements effectués). Les mots-clés servant à amorcer les histoires ont été tirés au hasard à partir de ce lexique, garantissant une pertinence thématique tout en conservant un élément d'imprévisibilité. J'ai volontairement conservé des doublons pour voir si les modèles allaient, ou non, générer des structures similaires à partir des mêmes déclencheurs.

Pour soutenir ce calibrage, plusieurs outils Python sur mesure ont été développés : un module de génération automatisée intégrant les paramètres définis, un système d'injection de dictionnaire pour uniformiser les amorces, un extracteur automatique de métadonnées narratives (longueur, structure, tonalité), et un gestionnaire de sorties en lots pour accélérer la production du corpus en formats .docx, .xlsx et .json.

Ces outils ont permis une production à grande échelle, un corpus suffisant et une documentation précise des conditions de génération.

Les trois graphiques ci-dessous illustrent la logique du calibrage :

1. **Paramètres spécifiques à configurer** : vue d'ensemble des réglages clés.

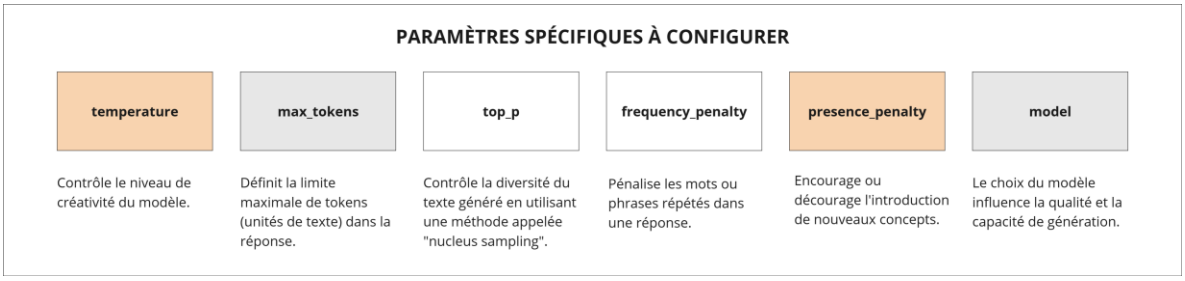


Figure 3.2a - Paramétrages de configuration pour les contes

Les valeurs de chaque paramètre correspondent aux réglages par défaut de ChatGPT, adaptés ensuite aux autres modèles selon leurs possibilités techniques respectives, comme l'illustre le tableau comparatif (Figure 3.2c)

2. **Protocoles standardisés** : démonstration de l'influence de chaque combinaison de paramètres sur le style narratif (logique, équilibré, créatif). Le prompt dans l'exemple ne correspond pas au prompt exact utilisé dans la génération du corpus de 100 contes.

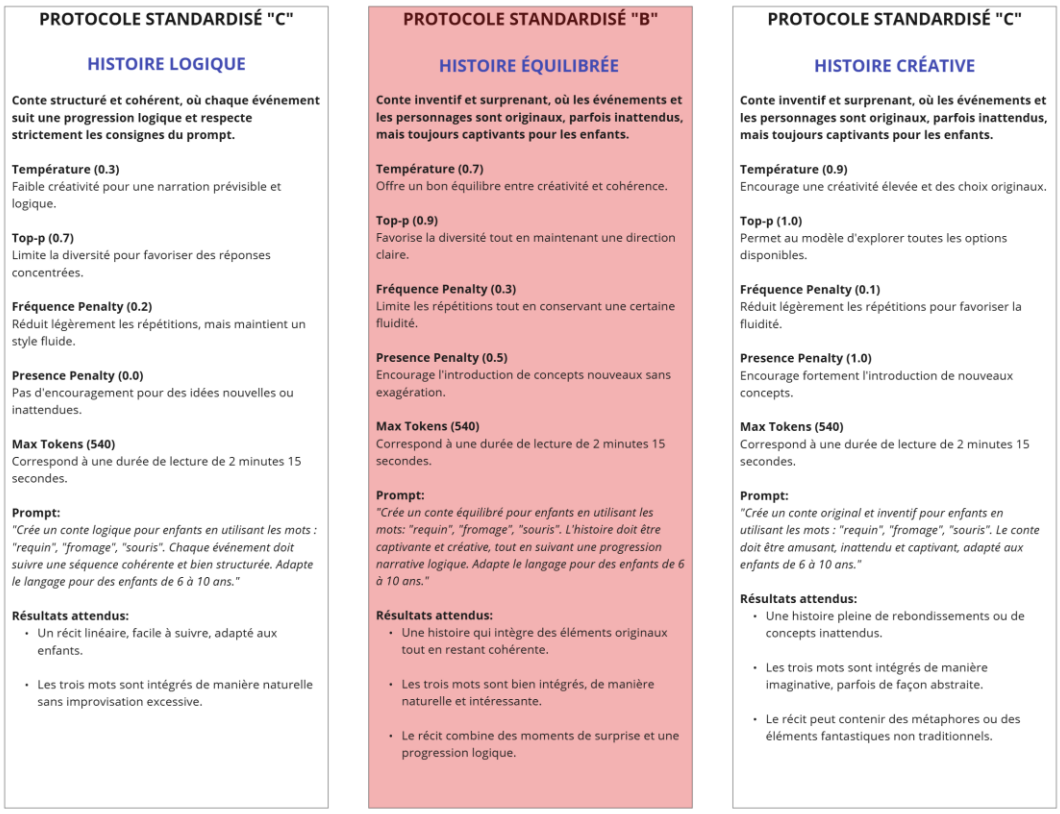


Figure 3.2b - Protocoles standardisés: logique, équilibrée et créative

3. **Tableau comparatif par modèle** : synthèse des possibilités et limitations techniques de chaque LLM testé.

Générateur	ChatGPT	Meta (LLaMA)	Mistral AI	Claude
Modèle	gpt-4o	meta-llama-3-70b-instruct	mistral-large-latest	claude-sonnet-4-20250514
Temperature	Oui	Oui	Oui	Oui
Max Tokens	Oui	Oui	Oui	Oui
Top-p	Oui	Oui	Oui	Non
Frequency Penalty	Oui	Simulable	***simulé dans le prompt***	Non
Presence Penalty	Oui	Simulable	***simulé dans le prompt***	Non
Model Choix	Oui	Oui	Oui	Oui

Tableau 3.2c - Paramétrages selon les modèles spécifiques analysés

3.2.1 Critères de sélection des modèles

Avant de procéder au calibrage des paramètres, une étape essentielle a consisté à sélectionner les modèles de langage à inclure dans le protocole. Le choix s'est appuyé sur quatre critères complémentaires permettant de constituer un corpus comparable et représentatif.

Le premier critère était l'accessibilité technique : les modèles devaient être disponibles via API publique avec une documentation claire, permettant leur utilisation dans un cadre de recherche académique. Le deuxième critère concernait le contrôle des paramètres : chaque modèle devait permettre d'harmoniser les paramètres (voir Figure 3.2a) de génération (température, max_tokens, top-p, etc.) afin d'assurer des comparaisons équitables. Le troisième critère visait la diversité des architectures : les modèles proviennent de différents laboratoires (OpenAI, Anthropic, Meta, Mistral AI), offrant une variété d'approches dans la génération de texte. Enfin, le quatrième critère était la capacité narrative : ces modèles devaient avoir démontré des capacités de génération de texte créatif dans des contextes similaires.

Sur la base de ces critères, quatre modèles ont été retenus : GPT-4o (OpenAI), Claude Sonnet 4 (Anthropic), LLaMA 3 70B Instruct (Meta) et Mistral Large (Mistral AI). Cette sélection permet de travailler avec un échantillon cohérent représentant différentes approches de génération, tout en maintenant une rigueur dans l'ajustement et l'harmonisation des paramètres entre les moteurs testés.

Certains modèles comme Gemini (Google), DeepSeek (Hangzhou DeepSeek AI) ou Grok (xAI) n'ont pas été intégrés, soit en raison de contraintes liées aux paramètres disponibles, soit par choix stratégique visant à se concentrer sur un nombre limité de modèles offrant un contrôle technique suffisant pour la comparaison.

3.2.2 Ajustement du nombre de jeton (*tokens*) selon le modèle

Mes tests montrent que, pour un même max_tokens, la longueur varie d'un modèle à l'autre. Claude Sonnet 4 écrit plus long et descriptif ; GPT-4o, plus court. Pour viser une durée de lecture d'environ deux minutes par conte, j'ai fixé max_tokens à 800 pour GPT-4o et à 650 pour Claude Sonnet 4. Cela ne les uniformise pas complètement, mais rapproche les longueurs et rend la comparaison plus équitable.

Ajuster chaque moteur reste toutefois compliqué. Avec les mêmes réglages et la même consigne, certains dépassent ou raccourcissent la durée visée. Le style joue aussi : des phrases plus longues, davantage de dialogues ou des répétitions peuvent faire varier le temps de lecture. Des répétitions peuvent encore survenir malgré les consignes. Je les ai notées sans corriger les textes, afin de refléter fidèlement les sorties des modèles.

Ces réglages m'ont permis d'identifier les paramètres clés à aligner entre moteurs. Sur cette base, j'ai retenu un ensemble restreint de modèles populaires et comparables techniquement. Cette sélection sert de fondement aux tests exploratoires décrits dans la section suivante.

Lors des phases de calibration, un comportement particulier a été observé avec le modèle Claude (Anthropic) lorsque la génération d'un conte était interrompue avant sa conclusion, par exemple, en atteignant la limite de tokens définie. Contrairement à GPT-4o ou Mistral, qui reprennent le texte qui reprennent le texte là où il s'était arrêté, Claude régénère souvent non seulement la phrase incomplète, mais aussi une ou plusieurs phrases déjà écrites, voire tout un paragraphe. Cette stratégie, probablement destinée à rétablir la cohérence narrative, entraîne des duplications partielles ou totales de la fin du récit.

Pour éviter ces doublons et préserver l'intégrité des textes, un module spécifique a été intégré au pipeline Python. Ce mécanisme compare la fin du texte initial et le début de la complétion, même en cas de coupures ou de variations mineures de ponctuation, et supprime toute portion redondante. Un avertissement est ensuite inscrit dans les métadonnées du conte, assurant la traçabilité et avec mention dans les métadonnées pour validation ultérieure si nécessaire.

Ces ajustements et optimisations techniques ont jeté les bases des tests exploratoires menés avant la production complète du corpus, permettant d'évaluer systématiquement la capacité des générateurs à produire des récits cohérents et adaptés au jeune public.

3.3 Phase d'expérimentation technique et calibration des générateurs

Avant de lancer la production des cent contes, j'ai mené plus d'une douzaine de tests exploratoires. Un script Python pilotait ces essais en appelant directement les interfaces de programmation (API) de chaque modèle (GPT-4o, Claude Sonnet, Mistral, Meta), avec des prompts identiques et des paramètres harmonisés (langue française, température/top-p, max_tokens). L'objectif était de recréer les mêmes conditions d'un moteur à l'autre et de comparer leurs sorties et anticiper les écarts de longueur et de style propres à chaque générateur.

Les essais ont permis d'ajuster le modèle interprétatif sur plusieurs plans. D'abord, calibrer les paramètres de génération (longueur, température, niveau de créativité, gestion des dialogues, etc.) afin de garantir une cohérence entre les différents modèles. Ensuite, vérifier la solidité narrative des histoires, en s'assurant que les débuts soient clairs, que la progression soit fluide et que les fins soient fonctionnelles et satisfaisantes. Enfin, observer la variabilité des réponses à partir d'un même triplet de mots-clés et identifier les régularités stylistiques ou structurelles propres à chaque modèle.

Au fil de ces tests, plusieurs ajustements majeurs ont été nécessaires. La révision du dictionnaire pour enfants constituait un premier enjeu : le lexique initial comptait environ 1300 mots, mais les premières générations ont révélé la présence d'adjectifs, de concepts ou de termes trop complexes pour un jeune public. Après deux itérations, la liste a été ramenée à environ 1100 mots, avec un accent mis sur des termes simples, concrets et accessibles. La gestion de la longueur des récits posait également un problème : certains modèles, plus « bavards », dépassaient le cadre prévu et nécessitaient une augmentation du nombre de tokens pour garantir une fin cohérente. D'autres, au contraire, produisaient des textes plus courts et plus précis, ce qui a permis d'ajuster les paramètres pour chaque moteur. Enfin, bien que cette phase de calibrage n'ait pas éliminé toutes les

imperfections, elle a permis de réduire considérablement les risques de coupures narratives ou de récits inachevés lors de la génération finale.

Entre novembre 2024 et juillet 2025, j'ai mené une phase d'expérimentation intensive comprenant 175 versions du script Python et générant 156 fichiers de travail (53 contes de test, ainsi que des fichiers de configuration JSON et des tableaux Excel de suivi). Cette phase itérative a permis de documenter plusieurs observations importantes.

La capacité générative s'est révélée variable selon les modèles : les quatre moteurs (GPT-4o, Claude Sonnet, Mistral, Meta) produisaient tous des intrigues complètes à partir de trois mots-clés, mais avec des différences marquées en termes de longueur et de style narratif. Certains modèles nécessitaient des ajustements spécifiques de paramètres pour garantir une fin cohérente. J'ai également observé une convergence vers des motifs narratifs "sécuritaires" : les récits générés réutilisaient massivement des archétypes récurrents. Par exemple, le prénom "Léo" apparaissait avec une fréquence anormalement élevée dans les premières générations, tout comme l'incipit classique "Il était une fois". Les quêtes suivaient presque systématiquement une structure linéaire avec un héros bienveillant et une résolution positive.

Les premières versions du générateur produisaient régulièrement des contes inachevés ou avec des fins abruptes, révélant des problèmes de finalisation narrative. Cela a nécessité l'ajout progressif de contraintes explicites dans les prompts (ex: "Termine ton histoire avec une phrase complète et une conclusion claire. Ne laisse pas l'histoire inachevée.") et l'ajustement des paramètres max_tokens pour chaque modèle. Enfin, même lorsque les fins étaient structurellement cohérentes, elles manquaient souvent de profondeur ou de richesse thématique, justifiant un suivi attentif lors de la génération finale du corpus de 100 contes.

Cette phase exploratoire a constitué une étape charnière entre la préparation technique et la production finale, posant les bases d'un protocole robuste et adapté aux particularités de chaque modèle.

3.4 Méthode pour atteindre les objectifs spécifiques

Cette section explique comment je vais atteindre les trois objectifs que j'ai posés dans le premier chapitre. Chaque objectif représente une étape du travail, mais ensemble ils forment un parcours cohérent d'analyse rigoureuse et de démarches créatives. Mon approche se déploie en deux temps: une phase d'analyse structurée avec des outils établis, puis une phase plus créative où je conçois mes propres outils d'évaluation adaptés aux récits générés par IA. Les sections qui suivent montrent concrètement comment je vais m'y prendre.

3.4.1 Objectif 1 : évaluation d'un corpus de contes à partir de la Story Grammar Theory

La première phase de ma recherche consiste à évaluer la qualité narrative d'un corpus de cent contes pour enfants générés par intelligence artificielle, à partir de triplets de mots choisis. Il s'agit de poser un regard critique sur les forces et les limites des récits produits par différents LLM, dans une perspective à la fois analytique, réflexive et exploratoire.

Cette phase s'inscrit pleinement dans une posture de recherche, puisqu'elle repose sur une méthode d'analyse rigoureuse appuyée sur un cadre théorique établi : la Story Grammar Theory (SGT). L'objectif est de mesurer, pour chaque conte généré, le respect des principales composantes narratives traditionnellement reconnues dans la littérature jeunesse. Ce protocole permettra de voir si les modèles d'IA peuvent produire des récits cohérents, structurés et accessibles pour des enfants de 6 à 10 ans. Une logique similaire a récemment été proposée par Wilke et Berov (2018), qui montrent comment l'analyse par unités fonctionnelles permet d'évaluer la cohérence interne d'un récit en dépassant la simple vérification de la présence formelle de ses composantes.

La grille SGT utilisée s'appuiera sur les sept composantes fondamentales du modèle :

- Le cadre initial (*setting*),
- L'événement déclencheur (*initiating event*),
- La réaction interne du protagoniste (*internal response*),
- L'objectif poursuivi (*goal*),

- La tentative (*attempt*),
- Le résultat (*outcome*),
- La réaction finale (*reaction*).

Chaque composante sera notée selon une échelle simple : 0 point si elle est absente, 0,5 point si elle est implicite ou floue, et 1 point si elle est présentée clairement.

J'analyserai manuellement l'ensemble des cent contes, en tenant compte de la lisibilité des structures narratives, de leur fluidité et de leur cohérence interne. Une démarche comparable est proposée par Rowe, McQuiggan, Robison et Lester (2009), qui ont développé Storyeval, un cadre empirique spécifiquement conçu pour évaluer la qualité narrative de récits générés.

Les figures 3.4.1a et 3.4.1b offrent une vue d'ensemble du processus méthodologique adopté pour cette recherche, articulé en deux volets complémentaires : ne phase d'analyse rigoureuse (RECHERCHE), couvrant la revue de littérature, l'évaluation des modèles d'IA via la Story Grammar Theory (SGT), la construction de la grille composite et la cartographie dans MIRO, et une phase d'expérimentation créative (CRÉATION), couvrant la réutilisation du corpus IA, le développement et l'application de la grille composite, la synthèse dans MIRO et les perspectives futures.

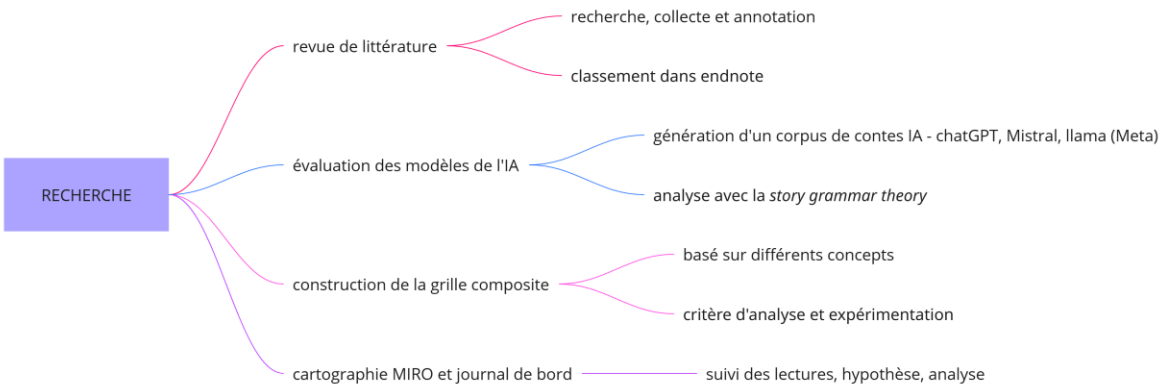


Figure 3.4.1a – Phase de recherche

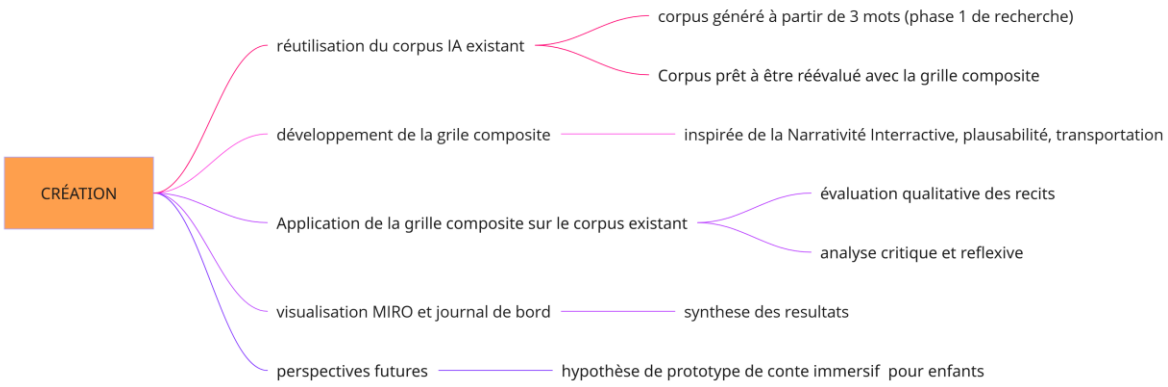


Figure 3.4.1b – Phase de création

3.4.2 Objectif 2 : conception et expérimentation d'une grille composite adaptée aux contes générés

À la suite de l'analyse du corpus à l'aide de la Story Grammar Theory (SGT), il apparaîtra nécessaire de concevoir un outil d'évaluation complémentaire, mieux adapté aux spécificités des récits générés par intelligence artificielle. L'objectif est de concevoir une grille composite qui évalue plus finement la qualité des contes générés, en tenant compte des attentes du jeune public.

La construction de la grille sera effectuée pas une série d'essais-erreurs, et en toute réflexivité, à partir d'un sous-ensemble de contes déjà analysés. Un premier ensemble de critères émergera de l'observation directe des récits.

Les catégories se définiront progressivement au fil de l'analyse, jusqu'à former un ensemble de treize critères. Elles couvriront différents types d'incohérences, allant des aspects matériels, spatiaux et temporels aux dimensions lexicales, narratives et morales.

Chaque conte du corpus sera relu et évalué avec la grille composite. Là où la SGT se limite à vérifier la présence des composantes structurelles, la grille composite utilisera une pondération par critère et produira un score global en pourcentage. Ce score permettra de situer chaque récit et de repérer plus précisément les incohérences rencontrées. Kybartas (2023), propose une approche comparable, en montrant comment une analyse quantitative peut compléter l'étude qualitative des

récits. Cela confirme l'intérêt d'une grille hybride adaptée aux contes générés par intelligence artificielle.

Cette grille ne cherche pas à remplacer la SGT, mais à la compléter et à l'enrichir dans une perspective de recherche-création. Elle constitue un outil critique, sensible aux particularités des textes générés, et capable de nourrir des discussions sur les conditions de production narrative automatisée, les attentes du jeune public, et les possibilités de collaboration homme-machine dans le champ de la création littéraire. La version intégrale de cette grille, avec un exemple d'application, se trouve en **Annexe B.**

Le développement d'une grille composite est justifié par la nécessité d'évaluer des contes générés par IA destinés aux enfants. Elle répond à un besoin simple : distinguer les textes qui « remplissent toutes les cases » sur le plan structurel, mais qui ne captent pas vraiment l'attention de l'enfant, de ceux qui, même simples et courts, parviennent à être clairs, engageants et mémorables.

Les outils existants pour analyser les récits se divisent généralement en deux grandes familles :

Les grilles de narratologie (exemple : la *Story Grammar Theory*). Elles décrivent les étapes clés d'un récit : cadre, déclencheur, réactions, but, actions, résultat, conclusion. Elles sont souvent utilisées pour analyser la façon dont les enfants comprennent ou produisent des histoires. Ces grilles confirment la présence d'une structure claire, mais elles ne disent pas si le récit capte l'attention, touche l'enfant ou reste mémorable.

Les mesures automatiques en traitement du langage, comme BLEU (Papineni et al., 2002), ROUGE (Lin, 2004) ou BERTScore (Zhang et al., 2020). Elles comparent surtout un texte généré à un texte de référence. D'autres indicateurs existent, par exemple pour la cohérence locale (Barzilay & Lapata, 2008), la diversité des sorties et la détection des répétitions (J. Li et al., 2016). Plus récemment, MAUVE (Pillutla et al., 2023) compare directement des textes humains et générés pour donner un

score entre 0 et 1. Tous ces outils sont utiles pour un diagnostic technique, mais aucun ne dit si un conte « se tient », s'il est clair ou s'il plaît à un enfant.

Pour ce mémoire, la grille composite sera élaborée en travaillant directement sur des récits jeunesse. Elle ne sera pas pensée comme un outil universel, mais comme une réponse à un problème précis : comment évaluer la qualité narrative des contes générés par IA pour l'enfance. Elle proposera une lecture plus fine que les grilles classiques, en relevant les incohérences, les répétitions et les biais propres aux modèles, tout en tenant compte des codes narratifs du conte court.

Ce que fera la grille, concrètement La grille évaluera la cohérence du récit du début à la fin. Elle repèrera les pièges fréquents des modèles comme les amorces génériques, les lieux vagues, les prénoms récurrents, les morales plaquées, les objets magiques inutiles et les incohérences de logique. Elle pourra être appliquée à des contes isolés, mais permettra aussi de faire ressortir des tendances et biais quand on l'utilisera sur un corpus entier, comme ici avec 100 récits. Elle proposera également des pistes d'amélioration simples et concrètes, telles que varier les débuts, inventer des toponymes, éviter la double morale, ou introduire un mini-choix pour le héros.

À l'inverse, la grille ne sera pas une simple liste à cocher : elle observera comment les éléments fonctionnent réellement dans le récit. Elle ne sera pas non plus une mesure abstraite, mais sera conçue pour des textes destinés aux enfants, même si ce point comporte une part de subjectivité. Enfin, elle ne reposera pas sur une impression vague : son codage s'appuiera sur des critères explicites et reproductibles.

La grille a été construite en trois temps. D'abord, une première version a été esquissée à partir des lacunes identifiées dans la SGT lors de l'analyse du corpus, en listant les types d'incohérences non couverts par ce modèle. Ensuite, ces observations ont été regroupées et hiérarchisées en catégories, ajustées au fil de la lecture d'une vingtaine de contes pilotes. Enfin, les critères ont été pondérés selon leur gravité et testés sur l'ensemble du corpus avant d'être stabilisés dans leur forme finale.

La grille composite reliera trois dimensions : la structure du récit, la plausibilité narrative et l'observation des biais propres aux modèles génératifs. Elle permettra de diagnostiquer les faiblesses

et de proposer des améliorations concrètes. Sa valeur tiendra à son ancrage clair : évaluer les contes générés par IA pour la jeunesse, afin d'alimenter la réflexion critique et la conception de futurs dispositifs narratifs adaptés.

Élément	Valeur ajoutée
Angle d'analyse	Grille hybride centrée sur la littérature jeunesse, qui relie structure narrative et critères adaptés aux enfants.
Corpus	Analyse comparative de 100 contes générés par IA (et non d'un seul texte isolé).
Public ciblé	Spécifiquement les récits jeunesse : lisibilité, clarté, intérêt, émotions accessibles.
Critères	Treize catégories opérationnelles (cohérence, plausibilité, créativité/engagement), adaptées aux codes du conte court.
Méthodologie	SGT + grille composite, codage reproductible, analyse par modèle, plus pistes de bonification.
Utilité pratique	Règles concrètes de bonification (varier les amorces, diversifier les prénoms, inventer des toponymes, éviter les doubles morales, introduire des mini-choix pour le héros).

Tableau 3.4.2 - Valeur ajoutée de la grille composite

Une fois la grille composite conçue, je l'expérimenterai sur l'ensemble du corpus de cent contes générés par intelligence artificielle. Cette phase constituera un pivot méthodologique : elle me permettra de tester la solidité de l'outil et de documenter en détail la qualité narrative des textes.

Contrairement à l'analyse SGT, qui reposera sur une seule lecture humaine, j'appliquerai la grille composite à travers une double lecture croisant évaluation humaine et évaluation assistée par IA :

Dans une première lecture humaine, je remplirai manuellement la grille pour chaque conte, en attribuant une note aux treize critères d'évaluation retenus (ex. : incohérence matérielle, incohérence spatiale, incohérence temporelle, morale incohérente, usage des mots-clés, etc.).

Dans une seconde lecture assistée, j'utiliserai GPT-4o et Claude Sonnet conjointement : GPT-4o identifie les incohérences dans chaque conte, et Claude Sonnet confirme, complète ou nuance ces

observations. Les résultats des deux modèles sont combinés pour former une seule évaluation automatisée.

Chaque grille remplie produira un score pondéré final, exprimé en pourcentage. Ce système n'a pas fait l'objet d'une validation psychométrique formelle : il repose sur des critères explicites, définis et ajustés de façon itérative au fil de l'analyse, dans une posture assumée de recherche-création.

Cette double lecture servira plusieurs objectifs : évaluer la fiabilité de la grille en comparant les résultats entre évaluation humaine et évaluation assistée par IA, identifier les écarts de perception entre le jugement humain et celui produit par les modèles, et mesurer l'influence du modèle génératif utilisé (GPT-4o, Claude, Mistral, Meta) sur la structure, la cohérence et la lisibilité des contes produits, en comparant les scores moyens et la distribution des incohérences par modèle.

Cette expérimentation rigoureuse renforcera la validité de la grille composite comme outil d'analyse, tout en posant les bases d'une future typologie des erreurs narratives. Elle ouvrira aussi une réflexion sur la coévaluation humain/IA, une pratique encore émergente dans le champ de la création numérique, comme le montrent des travaux récents (Haase & Pokutta, 2024)

La prochaine section explore justement cette réflexion, en proposant une synthèse des constats tirés de cette phase d'expérimentation et en ouvrant sur les perspectives de création qui en découlent.

3.4.3 Objectif 3 : développement d’une typologie des incohérences narratives

L'expérimentation de la grille composite sur les cent contes générés visera à mettre en évidence des incohérences récurrentes qui ne peuvent être pleinement saisies par une grille d'analyse traditionnelle. Ces observations mèneront à la création progressive d'une typologie des incohérences narratives propres à la génération automatisée.

Cette typologie naîtra d'un travail inductif : au fil de l'analyse manuelle des contes, je noterai systématiquement toutes les anomalies, erreurs et patterns que j'observerai. Je les documenterai dans un journal de bord, sans catégorisation préétablie.

La construction de cette grille suivra une démarche inductive : pour chaque conte analysé, je prendrai des notes détaillées sur tout ce qui accroche, détonne ou semble problématique du point de vue narratif. Une fois l'analyse du corpus complétée, je regrouperai ces observations pour faire émerger des catégories. Ce travail de catégorisation se fera par itération : regroupement des observations similaires, identification de patterns récurrents, création de catégories distinctes.

Je développerai également un système de pondération pour différencier les incohérences selon leur degré de gravité. Certaines erreurs nuisent gravement à la compréhension du récit (incohérences majeures), tandis que d'autres sont mineures. Cette pondération permettra d'affiner l'évaluation globale de chaque conte.

Cette typologie prolongera le travail de la grille composite. Là où la grille permet de mesurer et de comparer les récits à partir de critères précis, la typologie visera à regrouper et comprendre les failles narratives propres aux textes générés.

La typologie complète, avec ses catégories finales et son système de pointage, sera présentée au chapitre 4 (Résultats).

3.5 Synthèse de la méthodologie

Cette phase méthodologique, centrée sur l'évaluation des contes générés par intelligence artificielle, permettra de poser des fondations solides pour la suite du projet. L'approche adoptée se construira progressivement en réponse aux limites observées lors de l'évaluation initiale. Elle combinera deux volets complémentaires : une analyse structurée avec la Story Grammar Theory (SGT) appliquée à

cent contes, puis la création et l'expérimentation d'une grille composite adaptée aux récits générés par IA.

Chacun des objectifs spécifiques jalonnera cette progression : évaluer les contes à l'aide d'un cadre narratif classique (SGT) afin d'établir un diagnostic initial de leurs forces et faiblesses, identifier les limites de la SGT face aux particularités des textes automatisés, concevoir une grille d'analyse composite pour saisir la diversité, la complexité et les imprécisions propres aux récits génératifs, expérimenter cette grille sur l'ensemble du corpus en intégrant une stratégie de double lecture humaine et assistée par IA, et développer une typologie des incohérences narratives tout en réfléchissant à la posture du chercheur-créateur à travers la fabrication d'un outil critique, situé entre analyse structurée et intuition créative.

4 CHAPITRE 4 : RÉSULTATS DE LA RECHERCHE

Ce chapitre présente les résultats de la recherche visant à développer une grille d'analyse adaptée aux contes générés par intelligence artificielle, à partir de cent récits produits selon des triplets de mots-clés prédéfinis (la liste complète figure en **Annexe F**). Il constitue le cœur empirique du mémoire et s'appuie sur deux grilles d'évaluation complémentaires : la Story Grammar Theory (SGT) et la grille composite élaborée dans une démarche de recherche-création. Les résultats sont organisés selon les trois objectifs spécifiques de la recherche.

L'analyse repose sur un double dispositif. La grille SGT permet de vérifier si les récits contiennent les principales étapes d'une structure narrative classique. La grille composite, quant à elle, offre une lecture plus fine : elle met en évidence les incohérences, les maladresses et les manques qui traversent les textes générés, qu'ils soient logiques, stylistiques, lexicaux, structurels ou liés à la caractérisation et à la morale.

La SGT a été appliquée uniquement par moi, puisque les évaluations automatisées (GPT-4o et Claude Sonnet) donnaient presque toujours des scores trop élevés pour être vraiment exploitables. À l'inverse, la grille composite a été testée de trois façons : ma propre lecture, puis celles de GPT-4o et de Claude Sonnet. Les résultats ont ensuite été mis en commun avec une pondération, afin de comparer les analyses et vérifier la robustesse de l'outil.

Les résultats de ce chapitre visent à présenter les tendances générales observées, tant sur le plan de la structure narrative que sur la qualité globale des récits. Ils mettent en évidence les écarts et les complémentarités entre les deux grilles d'analyse, en soulignant ce que chacune permet de révéler et ce qu'elle laisse en marge. Enfin, ces résultats sont interprétés dans une perspective à la fois critique et créative, en questionnant la capacité réelle des intelligences artificielles à produire des récits adaptés à l'enfance et en esquissant des pistes pour la conception de dispositifs narratifs plus attentifs à ces enjeux.

L’analyse ne se réduit donc pas à un simple décompte. Elle cherche à comprendre ce que ces récits révèlent, non seulement des limites actuelles des modèles de langage, mais aussi des critères implicites qui façonnent notre idée d’un « bon conte » pour enfants.

4.1 Objectif 1 : évaluer un corpus de contes à partir de la Story Grammar Theory

Ce premier objectif visait à évaluer la qualité narrative de cent contes générés par quatre modèles d'intelligence artificielle (GPT-4o, Claude Sonnet 4, Mistral et Meta LLaMA 3) à l'aide de la Story Grammar Theory (SGT). Cette section présente d'abord les résultats de cette évaluation, puis identifie les constats et limites observés.

4.1.1 Résultats de l’évaluation SGT

L'évaluation des cent contes avec la SGT permet d'examiner leur structure de base, selon les critères décrits à la section 3.

Les résultats présentés ici donnent donc un portrait de la cohérence structurelle des contes. Ils montrent où les modèles réussissent à respecter la logique d’un récit, et où ils ont plus de difficulté.

Dans la suite du chapitre, je commence par les résultats de l’évaluation SGT (4.2), avant de passer à la grille composite (4.3), qui permet d’aller plus loin dans l’examen des incohérences et de la qualité globale des récits. L’analyse croisée (4.4) fait ressortir les grandes tendances, puis je propose des observations transversales (4.5) avant d’ouvrir sur une analyse réflexive (4.6) et de conclure le chapitre (4.7).

L'analyse des cent contes avec la grille de la Story Grammar Theory (SGT) permet d'établir la distribution des scores par composante en vérifiant, une à une, les sept composantes classiques d'une histoire : le cadre initial (setting), l'événement déclencheur (initiating event), la réaction interne du personnage (internal response), l'objectif poursuivi (goal), la tentative (attempt), le résultat (outcome) et la réaction finale (reaction).

Chaque élément a été noté simplement :

- **1 point** quand la composante est claire, bien présente et fonctionne dans le récit,
- **0,5 point** quand elle est un peu floue, implicite ou ambiguë,
- **0 point** quand elle est absente ou impossible à identifier.

Plusieurs tendances générales se dégagent de l'analyse des sept composantes, révélant des résultats contrastés. Le cadre initial (49,5 %) constitue la composante la plus faible du corpus. Il faut toutefois nuancer ce résultat : dans les contes pour enfants, le début est souvent volontairement vague ou stylisé, avec le fameux « Il était une fois... », sans qu'on précise toujours le lieu ou le moment. Cette ouverture floue ne gêne pas la compréhension, au contraire, elle favorise l'universalité et laisse l'enfant projeter son imaginaire. Bref, une note partielle au setting ne veut pas forcément dire que l'histoire est mal construite : c'est aussi un trait caractéristique du genre narratif. La SGT est utile pour compter et comparer, mais elle reste aveugle à ce type de conventions narratives. L'événement déclencheur (99,5 %) est presque toujours présent et bien formulé : les modèles réussissent clairement à lancer une action. La réaction interne (93 %) est très présente, avec des émotions ou réflexions des personnages apparaissant dans la grande majorité des récits. L'objectif (93,5 %) est bien identifié, souvent lié directement au déclencheur. La tentative (95 %) est généralement claire et distincte des autres étapes. Le résultat (97,5 %) est presque toujours présent, avec une résolution identifiable. Enfin, la réaction finale (91,5 %) est souvent incluse, parfois brièvement, mais assez pour boucler l'histoire.

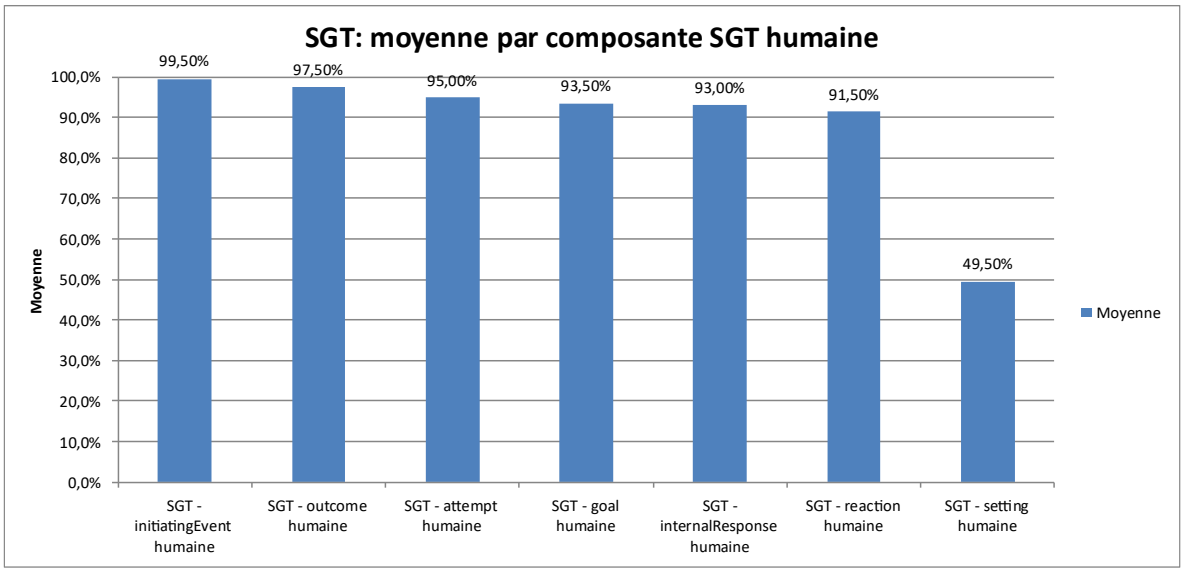


Figure 4.1.1a - SGT: Moyenne par composante

Dans l'ensemble, les composantes narratives sont bien couvertes. La seule vraie faiblesse est le cadre initial, mais encore une fois, c'est autant une question de grille qu'une caractéristique normale des contes jeunesse.

Après avoir examiné chaque composante séparément, j'ai établi un classement narratif des récits selon leur complétude SGT pour voir comment les contes tenaient dans leur ensemble. J'ai donc regroupé les récits selon leur score global SGT (maximum de 7 points). Ce classement donne une idée de la capacité des modèles à générer des histoires complètes selon la structure narrative classique.

La répartition des récits selon leur score total révèle une forte concentration dans le haut du tableau : 72 % obtiennent entre 6 et 7 points. Autrement dit, presque toutes les composantes attendues (cadre, événement, réaction, etc.) sont là. Mais comme on le verra plus loin, avoir toutes les pièces ne garantit pas forcément que l'histoire soit cohérente ou adaptée à un jeune public.

Un deuxième groupe, soit 21 % des contes, obtient entre 5 et 6 points. Ces récits sont globalement complets, mais une ou deux parties restent floues ou sous-développées. Plus rares, 5 % des histoires tombent entre 4 et 5 points : la base est là, mais fragilisée par plusieurs manques. Finalement, c'est 2 % des contes qui descendent sous les 4 points, sans jamais tomber en dessous de 2,5.

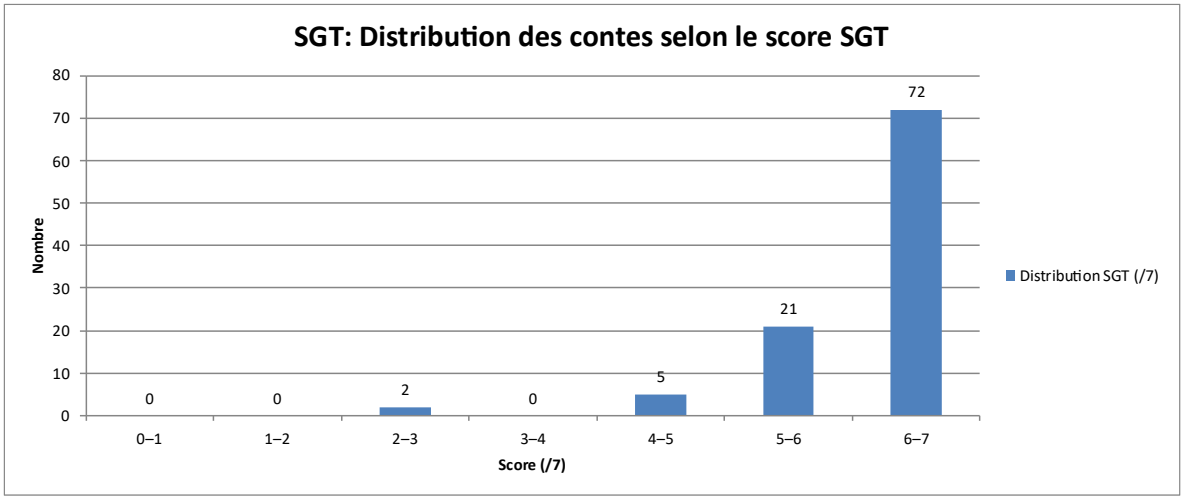


Figure 4.1.1b - SGT: distribution des contes selon le score obtenu

Pour établir un classement qualitatif par tranche de complétude et rendre les résultats plus parlants, j'ai regroupé les récits en cinq paliers :

- 6,5 à 7 points : récits très complets, avec toutes les composantes bien articulées.
- 5,5 à 6,5 points : récits complets, mais avec une ou deux parties plus faibles.
- 4 à 5,5 points : récits partiels, avec des manques visibles.
- 2,5 à 4 points : récits incomplets, avec des trous importants dans la structure.
- 0 à 2,5 points : aucun conte de ce type dans le corpus.

Ces paliers servent à regrouper les résultats en catégories interprétables et à comparer la distribution des scores entre les différents modèles testés.

En élargissant la perspective, on constate que **95 % des récits atteignent au moins 5/7**, et que **84 % franchissent le seuil de 6/7**, ce qui illustre le resserrement des scores dans le haut du classement.

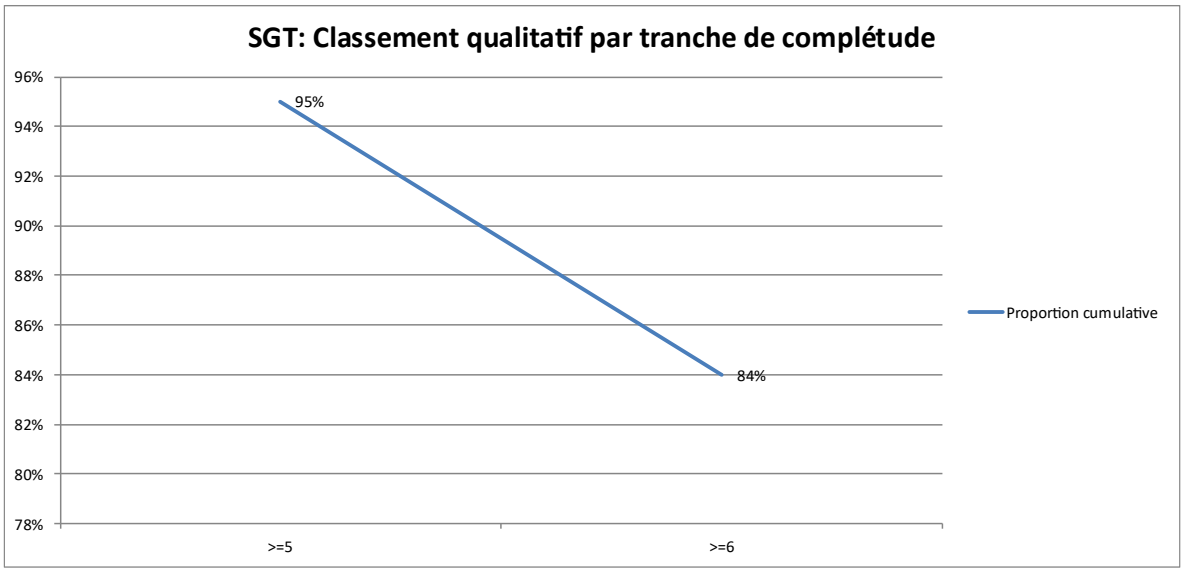


Figure 4.1.1c - Classement par tranche de complétude

Certains récits présentent un brouillage entre les composantes narratives de la SGT, ce qui complique la lecture. Par exemple, l'objectif peut se fondre directement dans le déclencheur : « Zoé trouve une clé étrange et part en voyage », sans préciser clairement ce qu'elle cherche à accomplir. Le résultat peut également se limiter à une action finale déconnectée : « Léo franchit la montagne et alla jouer dans la rivière », sans que le but initial soit résolu. Enfin, la réaction finale peut être réduite à une formule automatique : « Ils rentrèrent chez eux et furent heureux », qui clôture l'histoire sans réelle transformation. Ces cas montrent que la présence des composantes ne garantit pas toujours leur fonction narrative : la structure est là, mais elle peut rester superficielle.

La réaction interne obtient une moyenne élevée (93 %), mais cette apparente réussite masque une réduction ponctuelle de la dimension affective et réflexive : certains récits passent rapidement sur l'intériorité du personnage. Trois cas fréquents illustrent cette limite. D'abord, la réaction peut être absente ou trop factuelle, comme dans « Léa vit que la forêt avait disparu. Elle se mit à courir. » Ensuite, l'émotion peut être réduite à une étiquette générique : « Il était content de rentrer chez lui. » Enfin, on observe parfois une absence de transformation : « Après cette aventure, elle retourna à l'école comme si de rien n'était. » Ces exemples, même minoritaires, montrent que les émotions et les apprentissages ne sont pas toujours exploités pour enrichir l'histoire.

Au-delà de ces observations ponctuelles, la SGT comporte également des biais structurels induits par sa propre conception : elle valorise surtout la conformité à une séquence canonique, sans prendre en compte le style, le rythme ou la richesse du texte. Elle peut donc surévaluer des récits complets, mais pauvres narrativement, et pénaliser des histoires plus libres, mais cohérentes. C'est cette limite qui a motivé la mise en place de la grille composite, mieux adaptée pour capter la diversité et la qualité des récits générés.

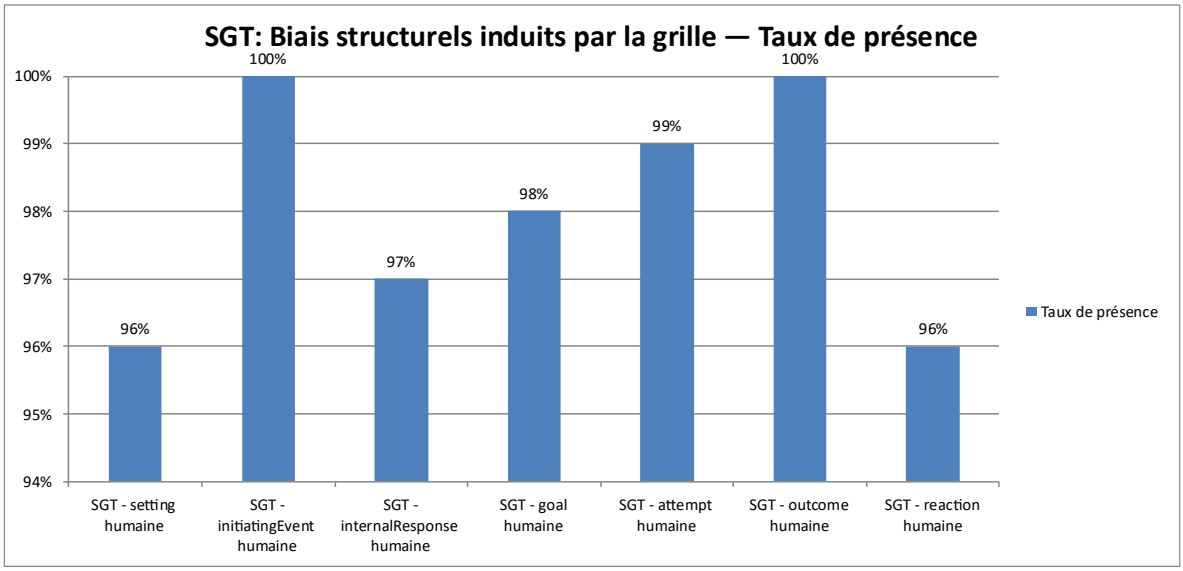


Figure 4.1.1d - Biais induits par la grille SGT

4.1.2 Constats et limites de cette évaluation

Le score SGT moyen de l'ensemble du corpus est de 89 %, ce qui indique une bonne conformité structurelle globale des récits générés. Les modèles testés (GPT-4o, Claude, Meta, Mistral) arrivent assez bien à produire des histoires qui suivent la structure classique d'un conte telle que la décrit la SGT : un cadre de départ, un événement déclencheur, une réaction interne, un objectif, une tentative, un résultat et, souvent, une réaction finale. Sur le plan formel, on obtient donc des récits qui semblent

complets. Dans le corpus, cinq des sept composantes SGT atteignent ou dépassent 93 % en moyenne.

Mais cette solidité reste de surface. Les histoires manquent souvent de tension, de vraies transformations des personnages ou de progression émotive. L'action s'enchaîne correctement, mais mécaniquement, comme si les événements se suivaient sans vrai moteur intérieur. Il y a bien une narration, mais elle reste faible du point de vue de l'expérience de lecture.

Cette première évaluation révèle aussi ce que la SGT ne peut pas évaluer : ses limites méthodologiques pour ce type de corpus. La grille vérifie que chaque composante est présente, mais elle ne juge pas comment cette composante joue son rôle dans le récit. Par exemple, un résultat peut clore l'histoire sans résoudre l'objectif initial, ou une réaction finale peut être générique (« et tout le monde fut heureux ») sans vraie transformation du personnage.

De plus, la SGT ne tient pas compte du style (voix, rythme, musicalité), de la dimension émotionnelle (émotions, pensées, apprentissages), ni de la manière dont les personnages sont construits, crédibles ou attachants. Elle n'intègre pas non plus l'accessibilité lexicale, la gestion émotionnelle ou l'interaction via les dialogues, des points pourtant cruciaux pour un jeune public.

Ces limites ont confirmé la nécessité de développer un outil d'évaluation complémentaire, mieux adapté aux particularités des récits générés par IA et aux attentes du jeune public. C'est ce qui motive l'objectif suivant : identifier systématiquement les limites de la SGT en contexte automatisé.

4.1.3 Identifier les limites de la SGT en contexte automatisé

Cette section vise à identifier systématiquement les limites de la Story Grammar Theory lorsqu’elle est appliquée à des récits générés automatiquement. Elle présente d’abord les limites structurelles observées, puis examine les limites méthodologiques plus larges de la SGT pour ce type de corpus.

Limites structurelles identifiées

Pour l'interprétation des écarts, les divergences observées entre la SGT et la grille composite montrent que la présence formelle des composantes narratives (SGT) n'équivaut pas à leur fonction narrative ni à leur pertinence pour un lectorat jeunesse. Dans le corpus, cinq des sept composantes SGT atteignent ou dépassent 93% en moyenne, tandis que le *setting* (49,5%) et la *Reaction* (réaction finale) (91,5%) obtiennent des scores plus faibles. Parallèlement, la grille composite met encore en évidence plusieurs faiblesses fréquentes :

- Construction simpliste (88 %)
- Absence de caractérisation (55 %)
- Dialogues absents/peu fonctionnels (55 %)
- Incohérences narratives (53 %)

Autrement dit, la structure est là, mais la qualité d'usage de cette structure varie.

Un problème récurrent apparaît lorsque la case est cochée, mais que le sens narratif fait défaut : la SGT vérifie que chaque « case » est cochée, mais elle ne juge pas comment cette composante joue son rôle. Plusieurs cas problématiques illustrent cette limite. D'abord, un résultat peut clore l'histoire sans résoudre l'objectif initial. Par exemple, dans un conte où le héros voulait sauver son amie, l'histoire se termine simplement par « Il rentra chez lui et alla dormir », sans jamais revenir à la mission. Ensuite, un objectif peut être absorbé par le déclencheur ou la tentative, rendant le but flou. C'est le cas dans « Le loup apparut et elle s'enfuit », où l'action immédiate prend toute la place, mais on ne sait jamais clairement ce que cherche à accomplir le personnage. Enfin, une réaction finale peut être générique (« ...et tout le monde fut heureux ») sans vraie transformation, comme dans l'histoire d'un enfant peureux qui affronte un dragon, mais qui termine sans montrer qu'il a appris ou changé.

Ces cas créent un effet de surface : le récit semble complet, mais son arc narratif reste faible. C'est précisément ce que la grille composite sanctionne, en s'attardant sur la cohérence et la pertinence d'ensemble. Cette distinction rejoint d'ailleurs l'analyse de Riedl et Bulitko (2013), qui rappellent que la génération narrative ne peut pas se limiter à vérifier la présence formelle de composantes, mais doit aussi considérer leur fonction et leur pertinence dans un cadre interactif.

La SGT rencontre également des cas limites avec les récits non-agentifs : certains contes pourraient, en théorie, mettre en scène un héros qui n'agit pas vraiment par lui-même, par exemple un personnage endormi, absent ou même mort. Dans un tel cas, le protagoniste ne ressentirait rien, n'aurait pas d'objectif clair et n'entreprendrait pas d'action volontaire. L'histoire avancerait surtout grâce à des événements extérieurs ou à d'autres personnages.

La SGT pourrait quand même identifier un protagoniste, un déclencheur et parfois une résolution. Mais, elle ne saurait pas vraiment traiter l'absence de but, de réaction interne ou d'action consciente, puisque ce sont des éléments qu'elle considère comme essentiels. Ce genre de situation, même si elle ne s'est pas présentée dans mon corpus, montre bien une limite de la SGT quand elle est appliquée à des récits atypiques.

Sur le plan de l'analyse, cette double évaluation révèle la complémentarité des deux approches. La SGT reste utile comme échafaudage : elle garantit une ossature narrative reconnaissable. La grille composite, quant à elle, sert de test d'usage en mettant en évidence les incohérences, les faiblesses de construction et les points qui influencent l'engagement et la lisibilité pour un jeune public. Les deux grilles sont donc complémentaires : la première valide la structure, la seconde teste la valeur narrative telle qu'elle ressort dans le cadre de cette recherche-création.

Limites méthodologiques de la SGT

Au-delà des limites structurelles, la SGT présente des limites méthodologiques plus fondamentales lorsqu'on l'applique à des contes générés pour enfants.

Trois dimensions narratives essentielles échappent à la SGT : le style, les émotions et la construction des personnages. La grille ne tient pas compte du style (voix, rythme, musicalité), de la dimension émotionnelle (émotions, pensées, apprentissages), ni de la manière dont les personnages sont construits, crédibles ou attachants. C'est précisément dans ces zones que les contes générés révèlent leurs plus grandes faiblesses : absence de caractérisation des personnages (55 %),

dialogues rares ou mécaniques (55 %), et récits répétitifs ou simplistes (88 %). Ces défauts n'empêchent pas d'obtenir un bon score avec la SGT, mais ils réduisent l'engagement du jeune lecteur. C'est là que la grille composite devient essentielle, puisqu'elle met en évidence ce que la SGT laisse dans l'ombre.

Ces éléments n’empêchent pas un bon score SGT, mais réduisent l’engagement du jeune lecteur, et c’est là que la grille composite « voit » ce que la SGT ne voit pas.

Sur le plan de l'adaptation au public jeunesse, la SGT présente des lacunes importantes : elle n'intègre ni l'accessibilité lexicale, ni la gestion émotionnelle, ni l'interaction via les dialogues, des points pourtant cruciaux en jeunesse. La morale n'est pas exigée en soi : son absence n'est pas pénalisée. Cependant, une morale contradictoire ou répétée est traitée comme incohérence (catégorie « morale incohérente ou contradictoire », env. 10 % des récits du corpus analysé). Là encore, la grille composite évalue la cohérence fonctionnelle, pas la conformité à une norme attendue.

On note une forte prégnance de l'illusion de la cohérence narrative. Certains contes paraissent cohérents parce qu’ils sont fluides et bien écrits. Mais dès qu’on les lit attentivement, des problèmes apparaissent : des résolutions trop faciles (catégorie « construction simpliste »), des glissements de sens (catégorie « incohérence narrative ») ou des enchaînements qui ne collent pas (catégorie « incohérence temporelle » ou « spatiale »). Ce décalage rejoint les analyses de Ryan (2015), qui rappelle que l’immersion ne repose pas seulement sur la fluidité grammaticale, mais sur la capacité d’un récit à engager le lecteur dans un monde crédible. La grille composite a justement permis de pointer ces failles, en montrant que la présence des éléments narratifs ne suffit pas : il faut aussi juger de leur cohérence et de leur fonction dans le récit.

Ce point est essentiel : les modèles génératifs excellent à donner l’apparence d’un récit solide, mais leurs histoires manquent souvent de cohérence et de consistance quand on les évalue avec un regard plus exigeant.

4.2 Objectif 2 : concevoir et expérimenter une grille composite adaptée aux contes générés

Cette section décrit l'expérimentation de la grille composite appliquée à l'ensemble du corpus selon un protocole de double lecture croisant évaluation humaine et évaluation assistée par IA. Elle présente les résultats obtenus, compare les deux lectures effectuées, puis met en lumière les limites de cette approche.

4.2.1 Résultats de la grille composite

Après avoir appliqué la Story Grammar Theory (SGT), j'ai utilisé la grille composite sur l'ensemble des cent contes (voir section 3.4.2 pour la description complète de cet outil). Elle prend en compte des aspects qualitatifs comme la caractérisation des personnages, la pertinence des dialogues et l'intégration des mots-clés de départ.

Pour organiser les résultats, j'ai divisé l'analyse en cinq volets :

- **Présentation globale des scores** : une vue d'ensemble de la distribution des scores du corpus, avec la moyenne générale et les écarts entre modèles.
- **Répartition des incohérences par catégories** : la fréquence des treize types d'incohérences relevés et leur importance relative.
- **Présence et rôle des objets agentifs** : un regard sur la place narrative des objets, qu'ils soient magiques, utiles ou simplement décoratifs.
- **Tensions entre erreurs majeures et fautes mineures** : la différence entre les incohérences lourdes qui nuisent vraiment à la compréhension et les petites maladresses.
- **Corrélations possibles entre modèles et types d'incohérences** : une comparaison des quatre LLM étudiés (Claude, GPT-4o, Mistral et Meta) pour dégager leurs tendances et points faibles récurrents.

L'idée est donc de montrer non seulement à quelle fréquence ces problèmes surviennent, mais aussi leur impact réel sur la cohérence, la plausibilité et l'attrait narratif des contes destinés aux enfants.

L'application de la grille composite aux cent contes permet de présenter une vue d'ensemble des scores et de brosser un portrait plus nuancé de leur qualité narrative. Contrairement à la SGT, qui se limite à vérifier la présence des grandes étapes d'un récit, la grille composite tient compte de treize

types d'incohérences, chacune pondérée selon sa gravité et son effet sur la compréhension. Le score final est exprimé en pourcentage inversé : 100 % équivaut à un récit sans incohérences, et chaque erreur détectée fait descendre la note. Des exemples concrets de bons **(Annexe D)** et de mauvais contes **(Annexe E)** sont présentés en annexe.

En analysant les tendances générales, la moyenne du corpus est de **82,8 %.** Dans le cadre de cette recherche, un score supérieur à **80 %** a été retenu comme seuil indicatif d'une performance satisfaisante, ce qui place l'ensemble du corpus juste au-dessus de ce repère. Mais les écarts entre modèles sont importants : Claude se démarque en tête avec **87,6 %**, tandis que Meta ferme la marche avec **73,9 %**. Certains contes frôlent la perfection (≥ 95 %), alors que d'autres chutent sous les 65 %, alourdis par des incohérences majeures. Si nous comparons avec les modèles génératifs d'intelligence artificiel, voici l'ordre qui ressort de l'analyse :

1. **Claude Sonnet - 87,6 %** : résultats stables et cohérents, très peu d'erreurs graves.
2. **Mistral - 85,4 %** : de bons récits, mais souvent jugés « simplistes », avec une profondeur limitée.
3. **ChatGPT (GPT-4o) - 84 %** : proche de Mistral, mais plus variable d'un conte à l'autre selon les mots-clés.
4. **Meta (LLaMA 3) - 73,9 %** : beaucoup plus inégal, avec des récits qui accumulent les incohérences matérielles et une intégration faible des mots-clés.

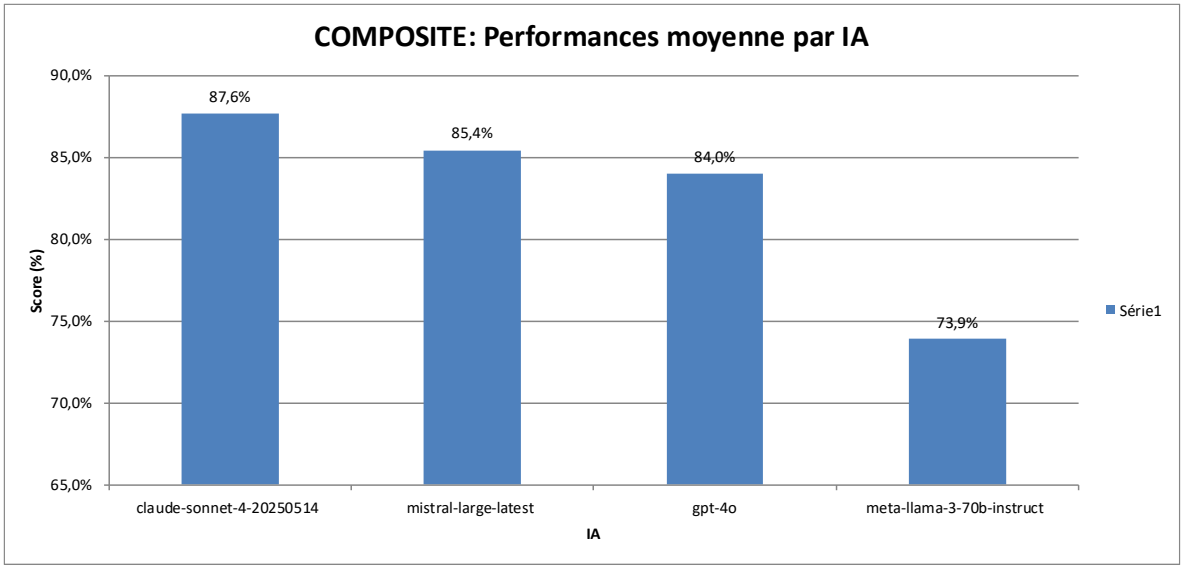


Figure 4.2.1 - Performance moyenne selon la grille composite

Pour bien lire ce score composite, il faut comprendre qu'il ne reflète pas seulement le nombre d'erreurs, mais aussi leur nature et leur poids dans l'histoire. Quelques petites maladresses (par exemple, une répétition légère ou l'absence de dialogue) n'empêchent pas un conte de garder un bon résultat. En revanche, une incohérence lourde, comme une rupture temporelle qui désorganise tout le récit, peut faire plonger la note. Cela évite les faux positifs : un conte qui a toutes ses étapes peut quand même être mal noté s'il est truffé de contradictions internes, tandis qu'une histoire simple mais claire peut obtenir un bon score.

En ce qui concerne les relations entre le score global versus le score « maîtrise » deux lectures des résultats ont été faites :

- Le **score global**, basé sur 11 des 13 catégories de la grille composite (les deux catégories exclues étant l'intégration des mots-clés et la présence de dialogues, réservées au score maîtrise), sert de référence dans toutes les analyses statistiques.
- **Le score maîtrise**, plus spécifique au projet, met l'accent sur deux critères essentiels à l'interactivité enfant/IA : l'intégration réelle des mots-clés et la présence de dialogues significatifs, qui rendent les personnages plus vivants et permettent aux enfants de se projeter.

Le premier donne une mesure neutre et académique. Le second reflète ma posture de chercheur-créateur, en pensant déjà à un dispositif interactif où l'enfant est cocréateur. L'écart entre les deux scores n'est donc pas une contradiction : c'est un indicateur complémentaire qui montre deux regards sur la même production.

Pour garder la lecture claire et uniforme, je présente dans ce chapitre uniquement le score global. Le score maîtrise, lui, est documenté (**Annexe H**), afin d'illustrer comment certaines histoires, très bien notées en général, perdent en pertinence lorsqu'elles sont replacées dans le contexte particulier de la cocréation enfant/IA.

4.2.2 Comparaison des lectures humaine et assistée par IA

J'ai voulu tester la grille composite de trois façons différentes : en l'appliquant moi-même sur les cent contes, puis en demandant à GPT-4o et à Claude Sonnet de faire la même chose. L'idée était de voir si la grille donnait des résultats cohérents, peu importe qui l'utilise, et de repérer les biais possibles. Dans l'ensemble, les deux lectures se rejoignent sur les mêmes problèmes récurrents : construction simpliste, absence de caractérisation, dialogues mécaniques. Mais quand on regarde les détails, des divergences intéressantes apparaissent.

Un phénomène inattendu est apparu pendant l'exercice : Claude Sonnet manifestait un biais favorable récurrent envers les contes générés par son propre modèle. Ce biais revenait suffisamment souvent pour que je doive, à plusieurs reprises, lui signaler qu'il se trompait et qu'il évaluait excessivement positivement ses propres textes. Chaque fois, Claude reconnaissait qu'il avait été trop favorable en raison de son modèle et s'excusait de ce biais. Malgré ces ajustements ponctuels, le réflexe avait tendance à revenir. Il ne s'agissait pas d'un problème généralisé à tous les contes, mais bien du nombre élevé d'interventions nécessaires pour le lui faire remarquer. GPT-4o, de son côté, s'est montré beaucoup plus constant et neutre.

4.2.3 Limites de l'atteinte de l'objectif et de la double lecture

La double lecture valide la robustesse générale de la grille, mais elle a aussi ses limites. Ma propre évaluation reste subjective. Je suis à la fois celui qui a créé la grille et celui qui l'utilise. Ce double « chapeau » colore forcément mes jugements : j'évalue les contes selon mes propres attentes et ma sensibilité narrative. Faire tester la grille par d'autres évaluateurs humains aurait permis de mieux mesurer sa reproductibilité. Les évaluations automatisées ont aussi leurs angles morts. Claude Sonnet favorisait les contes de sa propre famille de modèles, ce qui rappelle qu'aucun outil d'analyse n'est parfaitement neutre. J'ai corrigé ce biais par pondération, mais ça limite quand même la portée des comparaisons entre modèles.

4.3 Objectif 3 : développer une typologie des incohérences narratives

Le troisième objectif consistait à développer une typologie des incohérences narratives récurrentes observées dans les contes générés. Cette section présente d'abord la typologie développée, puis analyse la répartition des incohérences, et identifie les limites de cette catégorisation.

4.3.1 Présentation de la typologie

La typologie développée regroupe treize catégories d'incohérences narratives sous trois grandes dimensions : cohérence narrative, plausibilité narratologique, et créativité / originalité / engagement. Ces catégories n'ont pas été fixées a priori : elles ont émergé de façon itérative au fil de la lecture du corpus, par regroupement progressif des types d'incohérences observées, puis stabilisées avant l'évaluation systématique de l'ensemble des cent contes. Chaque catégorie est définie précisément et pondérée selon sa gravité (majeure : 10 %, modérée : 6 %, mineure : 4 %.

Au sujet de la répartition des incohérences par catégories, l’un des grands avantages de la grille composite est qu’elle ne se contente pas de vérifier la structure d’une histoire. Elle permet aussi de repérer et de classer les incohérences présentes dans les contes générés. Chaque récit du corpus a donc été évalué à partir de treize catégories précises, couvrant aussi bien les ruptures logiques que les maladresses stylistiques ou l’absence de certaines composantes attendues.

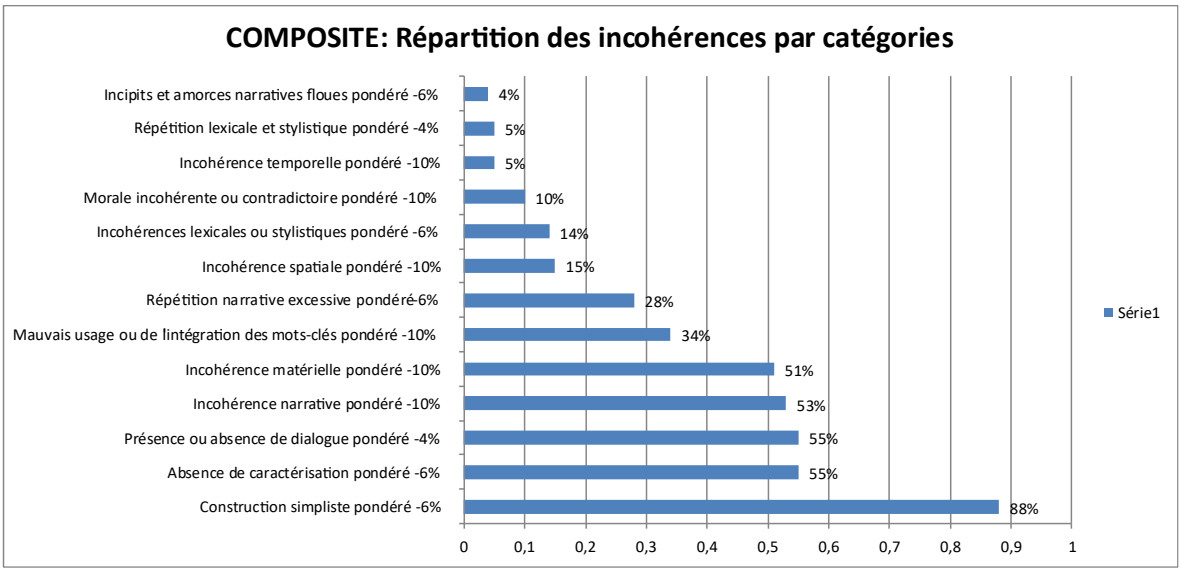


Figure 4.3.1 - Répartition des incohérences selon les catégories

Dans le contexte de mon analyse, un classement des incohérences les plus fréquentes a été relevé.

L'analyse des 100 contes met en évidence un ensemble de régularités :

- **Construction simpliste (88 %)** : la majorité des récits suivent une trame trop linéaire, sans véritable rebondissement, ce qui réduit la richesse de l'expérience narrative.
- **Absence de caractérisation (55 %)** : beaucoup de personnages n'ont ni nom, ni traits distinctifs, ni intentions claires, ce qui limite l'attachement du lecteur.
- **Présence ou absence de dialogue (55 %)** : les dialogues sont rares ou trop mécaniques, réduits à transmettre des faits plutôt qu'à créer une interaction vivante.
- **Incohérence narrative (53 %)** : des enchaînements d'événements sans lien clair, des résolutions arbitraires ou des progressions illogiques.
- **Incohérence matérielle (51 %)** : objets impossibles, contradictions physiques ou éléments matériels incohérents.
- **Mauvais usage ou intégration faible des mots-clés (34 %)** : les mots-clés imposés sont parfois utilisés en « décoration », détournés ou carrément oubliés.
- **Répétition narrative excessive (28 %)** : retour mécanique des mêmes structures ou séquences, qui donnent une impression de stagnation.
- **Incohérence spatiale (15 %)** : déplacements incohérents, lieux décrits qui sont contradictoires.
- **Incohérences lexicales ou stylistiques (14 %)** : maladresses de vocabulaire ou de ton, qui détonnent avec le contexte ou le public cible.
- **Morale incohérente ou contradictoire (10 %)** : morale répétée, absente ou en décalage avec le déroulement du récit.
- **Incohérence temporelle (5 %)** : confusion dans la chronologie, durées irréalistes ou ruptures de temps.
- **Répétition lexicale et stylistique (5 %)** : utilisation redondante de certains mots ou tournures.
- **Incipits et amorces narratives floues (4 %)** : débuts trop vagues, sans ancrage clair qui situe l'histoire.

La pondération hiérarchise les incohérences selon leur gravité : toutes ces incohérences ne pèsent pas le même poids dans le calcul du score. Les fautes majeures (comme une incohérence narrative ou matérielle) font chuter plus fortement la note qu'une répétition lexicale légère, considérée comme mineure. Cette pondération permet de mieux refléter l'impact réel de chaque erreur sur la qualité globale du récit.

4.3.2 Répartition et analyse des incohérences

En lisant les contes, un élément revenait régulièrement qui mérite attention : la présence et le rôle des objets agentifs. Ces objets ne sont pas indispensables à la réussite d'une histoire, mais quand ils apparaissent, ils changent souvent la dynamique du récit. Parfois, ils soutiennent l'action, ajoutent une surprise ou une tension. D'autres fois, au contraire, ils affaiblissent l'intrigue parce qu'ils deviennent trop attendus ou clichés. L'important n'est donc pas leur simple présence, mais la manière dont ils influencent la progression de l'histoire.

Catégorie d'objet	Définition	Caractéristiques typiques	Exemple tiré du corpus
Objet agentif magique	Objet doté d'un pouvoir autonome qui agit par lui-même ou influence directement l'action.	- Possède une volonté ou déclenche une action sans intervention humaine - Souvent porteur de magie ou d'esprit	**Plume magique** qui exauce les vœux sans manipulation
Objet agentif instrumentalisé	Objet utilisé activement par un personnage pour atteindre un but narratif.	- Le personnage s'en sert volontairement - Joue un rôle actif dans la quête ou la résolution	**Clé ancienne** utilisée pour ouvrir un coffre secret
Objet passif	Objet présent dans l'histoire, mais sans rôle narratif structurant.	- Sert d'accessoire - N'est ni déclencheur ni support de l'action	**Écharpe bleue** décrite dans une scène mais sans impact
Aucun objet agentif	Aucun objet ne joue de rôle narratif déterminant.	- L'histoire progresse sans support matériel distinct - L'intrigue repose sur les personnages seuls	Un récit de voyage initié par la curiosité d'un enfant

Tableau 4.3.2a - Pertinence et rôle des objets agentifs

L'analyse globale des données a permis de dégager quatre grandes configurations, identifiées par regroupement inductif des patterns observés lors de la lecture du corpus, puis validées par leur récurrence à travers les cent contes :

1. **Objet magique autonome** : porteur d'un pouvoir ou d'un esprit propre, il agit par lui-même.
2. **Objet instrumentalisé** : utilisé volontairement par un personnage pour atteindre un objectif.
3. **Objet passif** : présent dans l'histoire mais sans véritable impact sur l'intrigue.
4. **Absence d'objet agentif** : l'action repose uniquement sur les personnages.

Ces cas montrent que les objets agentifs ne sont pas centraux, mais qu'ils apportent une couleur particulière au récit. Leur usage, surtout quand il est magique, peut à la fois dynamiser ou affaiblir l'histoire. Dans le cadre de cette recherche, cette observation sert surtout à enrichir la compréhension qualitative des contes, sans en faire un critère strict d'évaluation.

Une distinction importante s'impose entre erreurs majeures et fautes mineures : toutes les incohérences ne pèsent pas du même poids dans une histoire. Certaines, comme une incohérence narrative ou matérielle, cassent complètement la compréhension et brisent l'immersion. D'autres, comme une petite maladresse de style ou un incipit un peu flou, restent mineures et ne nuisent pas vraiment au déroulement du récit.

Dans ma méthode, chaque conte part d'un score parfait de 100 %. À ce score, j'applique une pénalité pour chaque incohérence relevée, en fonction de sa gravité :

- **Faute majeure** (–10 %) : par exemple, une incohérence matérielle, spatiale, temporelle ou narrative.
- **Faute modérée** (–6 %) : par exemple, un incipit ou une amorce floue, une absence de caractérisation, une construction simpliste, une répétition narrative.
- **Faute mineure** (–4 %) : par exemple, une répétition lexicale et stylistique.

La note finale correspond donc au score de départ moins la somme de toutes les pénalités. Ça veut dire qu'un seul gros problème peut faire chuter rapidement une histoire, tandis qu'une accumulation de petites maladresses entraîne une baisse plus progressive.

Exemple concret : un conte avec 1 faute majeure, 2 fautes modérées et 1 faute mineure se retrouve avec 74 % (100-10-6-6-4).

L'analyse révèle une variabilité significative entre les modèles d'IA, chacun présentant des faiblesses qui reviennent avec une certaine fréquence. Claude Sonnet présente peu d'incohérences majeures, mais tend à moraliser un peu trop directement et à uniformiser ses structures. Mistral maintient une

bonne cohérence de surface, mais produit beaucoup d'histoires jugées « simplistes », souvent dépourvues de dialogues vivants. GPT-4o affiche un profil assez équilibré, mais plus instable : les résultats varient beaucoup selon les mots-clés utilisés. Meta (LLaMA 3) génère davantage d'incohérences matérielles et narratives, avec une intégration des mots-clés plus faible que les autres modèles.

L'analyse comparative des récits générés par les quatre modèles (Claude Sonnet, GPT-4o, Mistral et Meta LLaMA 3) révèle des corrélations possibles entre chaque modèle et certains types d'incohérences. Chaque moteur présente des profils bien distincts, avec ses forces, mais aussi ses faiblesses récurrentes, qu'on peut relier à des catégories spécifiques d'incohérences.

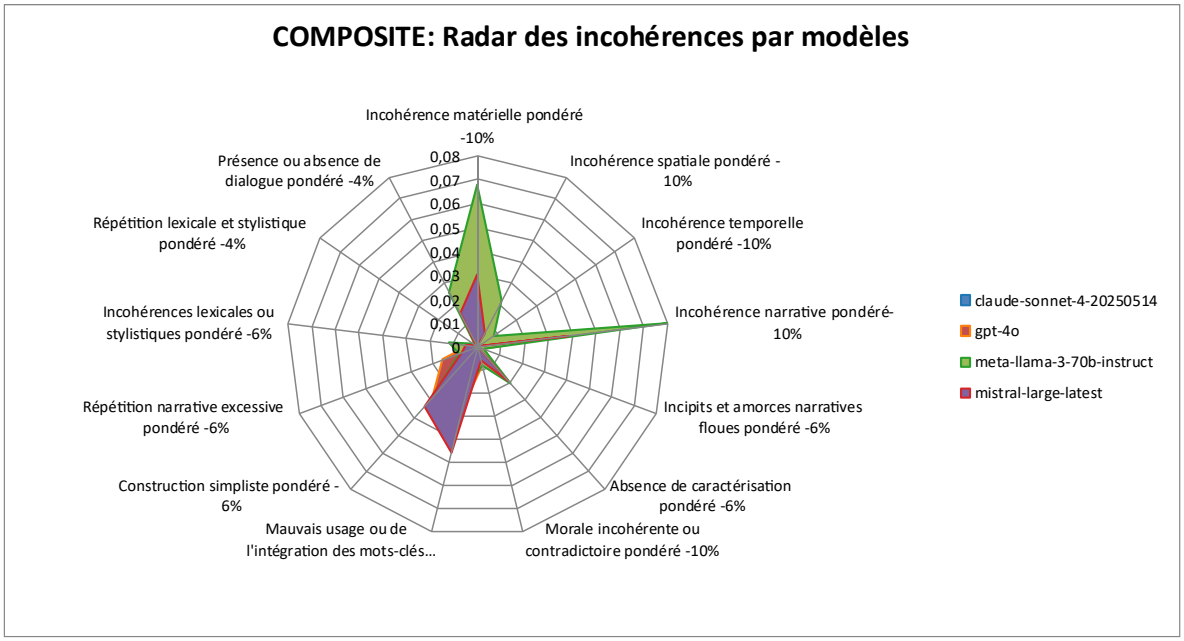


Figure 4.3.2b - Radar des incohérences par modèles

La figure 4.3.2b montre l'évaluation globale des modèles. Le modèle de Claude Sonnet génère des récits cohérents mais parfois rigides. Claude obtient le meilleur score moyen (**87,6 %**). Les composantes narratives sont presque toujours présentes et bien enchaînées, et les mots-clés intégrés efficacement. Les incohérences matérielles ou narratives sont rares. En contrepartie, cette

solidité donne parfois un style trop normalisé : les récits se ressemblent et se terminent souvent par une morale explicite, un peu prévisible.

Le modèle de Mistral affiche une cohérence de surface mais manque de profondeur. Mistral suit de quand même de près (85,4 %). Les textes sont clairs et exempts d'erreurs matérielles ou temporelles majeures. La « construction simpliste » apparaît dans près de 90 % des récits, avec peu de rebondissements et des dialogues souvent absents. Les mots-clés sont utilisés inégalement, parfois simplement posés sans réelle intégration. Le résultat : des histoires faciles à suivre, mais plates et sans tension.

Ensuite, avec le modèle de GPT-4o affiche un équilibre global mais avec une grande variabilité. Avec **84 %**, GPT-4o se place dans la moyenne haute. Les récits montrent une bonne intégration des mots-clés et une utilisation plus vivante des dialogues que chez Mistral. Mais la variabilité est forte : certains contes sont très solides, d'autres présentent des transitions floues ou des conclusions trop rapides. Les incohérences narratives touchent environ la moitié des textes.

Enfin, Meta (LLaMA 3) fait preuve d'une Créativité brute mais fait preuve d'une instabilité narrative. Meta ferme la marche avec **73,9 %**. On y trouve des récits imaginatifs, mais aussi plus d'incohérences matérielles et narratives que chez les autres modèles. L'intégration des mots-clés est faible, souvent absente, et les dialogues quasi inexistants. L'ensemble donne des histoires originales sur le plan des idées, mais instables et peu adaptées à un lectorat jeunesse.

Modèle	Forces	Faiblesses
Claude Sonnet	Structure solide, cohérence élevée, intégration correcte des mots-clés	Rigidité stylistique, morale explicite
Mistral	Cohérence de surface, peu d'incohérences matérielles	Construction simpliste, dialogues rares
ChatGPT (GPT-4o)	Bon équilibre structure/style, utilisation pertinente des dialogues	Variabilité selon les mots-clés, transitions floues
Meta (LLaMA 3)	Créativité et imagination	Incohérences matérielles/narratives fréquentes, faibles dialogues et mots-clés

Tableau 4.3.2c - Synthèse des profils (forces/faiblesses)

Le tableau ci-dessus affiche la synthèse de mon évaluation des différents modèles. Il permet de constater les forces et faiblesses de chacun. Il permet de démontrer que les outils d'intelligences artificielles ne sont pas des instances objectives permettant d'atteindre des résultats cohérents hors de tout doute raisonnable. Les modèles effectuent une forme de curation des informations pour la génération des récits. Il faut toutefois aussi tenir compte de la curation des modèles eux-mêmes en relation avec certaines orientations fonctionnelles.

4.3.3 Limites de la typologie

La typologie développée présente plusieurs limites méthodologiques qu'il convient de reconnaître.

Ma catégorisation demeure relativement subjective. J'ai construit cette typologie en observant les contes, en notant les problèmes qui revenaient, puis en les regroupant par ressemblance. Mais un autre chercheur aurait pu découper autrement ou tracer des frontières différentes. Par exemple, distinguer "incohérence narrative" et "construction simpliste" n'est pas toujours évident : certains récits cumulent les deux, ce qui rend la catégorisation floue par moments.

Le système de pondération (majeure : 10 %, modérée : 6 %, mineure : 4 %) reflète aussi ma propre appréciation de ce qui est acceptable ou non. Cette échelle fonctionne pour ce corpus, mais il faudrait la tester sur un ensemble plus large pour voir si elle tient la route ailleurs.

Cette typologie n'est qu'un début. Elle devra évoluer : ajouter des catégories si de nouveaux patterns apparaissent, mieux définir les frontières entre certaines incohérences. Pour l'instant, elle effectue le travail demandé : elle me permet de documenter et de comprendre les failles narratives typiques de la génération automatisée de contes pour enfants.

4.4 Analyse croisée SGT / composite

Cette section met en regard les résultats obtenus avec la *Story Grammar Theory* (SGT) et ceux issus de la grille composite. L'idée est de voir ce que ces deux approches révèlent en commun, mais aussi ce qu'elles laissent dans l'ombre ou interprètent différemment. La SGT donne une vision structurée et normative : elle vérifie si les grandes étapes classiques d'un récit sont bien présentes et bien ordonnées. La grille composite, de son côté, apporte une lecture plus fine : elle relève les incohérences, les maladresses et les faiblesses qualitatives qui échappent à une simple vérification de structure.

L'analyse croisée se déploie en deux volets complémentaires : elle met d'abord en lumière les concordances et divergences observées entre les deux grilles d'évaluation (section 4.4.1), puis elle interprète ces écarts pour mieux comprendre ce qu'ils révèlent sur la génération automatique de récits (section 4.4.2).

En confrontant les deux cadres, on obtient une image plus complète des récits générés. Cette double grille d'analyse permet de mieux cerner les forces et limites de chaque méthode et ouvre la voie à des outils hybrides, plus adaptés pour analyser des textes produits par intelligence artificielle.

4.4.1 Concordances et divergences

L'analyse comparative entre la grille SGT et la grille composite montre que les deux outils donnent des résultats semblables dans seulement **34 %** des cas (34 récits sur 100, avec un écart inférieur à 5 points de pourcentage). Ces concordances apparaissent quand le récit suit une structure claire, avec un objectif bien défini et un enchaînement logique des événements, sans incohérences majeures relevées par la grille composite. Dans ces situations, les deux approches reconnaissent la solidité du récit, autant du côté de la structure que de la lisibilité. Quelques exemples représentatifs sont présentés (**Annexe G**), ainsi que des cas extrêmes (**Annexe I**).

À l'inverse, **66 %** des récits (66 sur 100) présentent des écarts marqués (écart supérieur ou égal à 5 points). Deux cas de figure expliquent la majorité de ces divergences :

- Structure présente mais contenu appauvri : la SGT attribue une bonne note parce que toutes les étapes classiques (cadre, déclencheur, objectif, tentative, résultat, réaction) sont cochées, même si elles sont traitées rapidement ou mécaniquement. La grille composite, plus attentive à la cohérence, à la richesse des dialogues et à l'usage des mots-clés, sanctionne alors ces récits.
- Structure atypique mais narration riche : certains contes s'éloignent du schéma canonique (une ou deux composantes manquent), ce qui fait baisser leur score SGT, mais la grille composite valorise leur expressivité, leur cohérence interne et leur pertinence pour un jeune public.

Graphiquement, la distribution des points montre une dispersion importante autour de la diagonale de concordance. Le coefficient de détermination ($r^2 \approx 0{,}04$) indique une relation linéaire très faible entre les scores SGT et les scores de la grille composite. Autrement dit, la présence d'une structure narrative conforme à la Story Grammar Theory n'implique pas nécessairement un score élevé selon la grille composite. Cette faible corrélation confirme que les deux outils évaluent des dimensions différentes du récit.

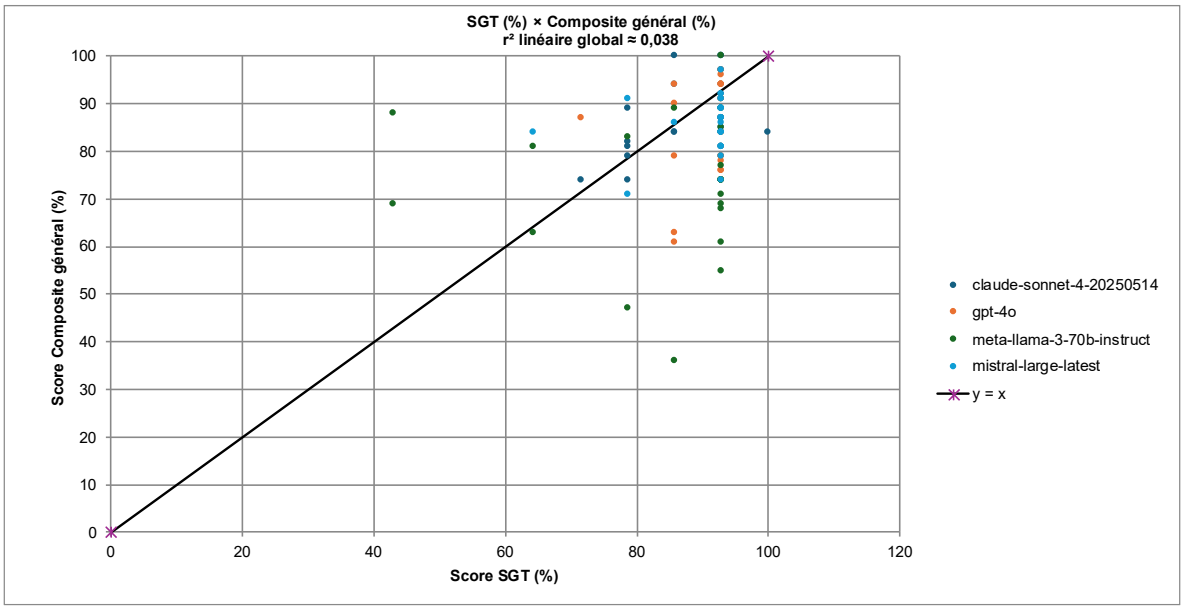


Figure 4.4.1a - Concordances et divergences entre la SGT et la grille composite

Chaque point du graphique correspond à un conte, placé selon son score SGT (axe horizontal, en %) et son score composite (axe vertical, en %). Les points proches de la diagonale indiquent une concordance (écart inférieur à 5 points) ; plus un point s'en éloigne, plus l'écart entre les deux évaluations est important.

Au sujet de la concordance et divergence selon les modèles, la répartition par modèle met en lumière des profils distincts. Claude Sonnet présente une majorité de points proches de la diagonale, montrant une bonne concordance entre structure et cohérence. GPT-4o affiche un profil équilibré, avec un nombre correct de concordances, mais aussi des divergences dues à des transitions rapides ou des conclusions abruptes. Mistral montre une dispersion plus visible : plusieurs récits semblent complets en SGT mais chutent en composite à cause d'une construction simpliste et d'un manque de dialogues. Meta (LLaMA 3) apparaît comme le modèle le plus divergent : ses récits obtiennent souvent un score correct en SGT, mais la grille composite les pénalise fortement à cause d'incohérences matérielles, de dialogues pauvres et d'une intégration faible des mots-clés.

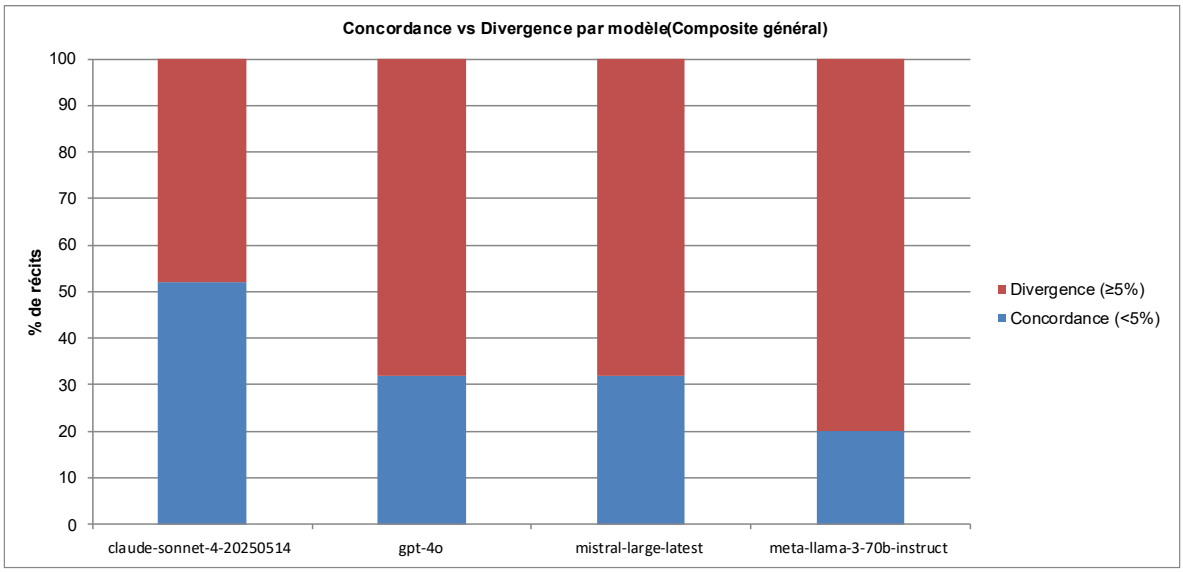


Figure 4.4.1b - Concordances et divergences entre la SGT et la grille composite

Ces résultats mettent en évidence un contraste net : forte concordance chez Claude et GPT-4o, dispersion moyenne pour Mistral, divergences marquées pour Meta.

4.4.2 Pourquoi les deux grilles divergent ?

Le lien entre la « complétude SGT » et la « qualité composite » reste globalement faible dans le corpus, même si certains modèles (comme Claude) montrent une concordance un peu plus marquée. En pratique, deux cas typiques apparaissent. D'un côté, certains récits très complets selon la SGT obtiennent pourtant de faibles scores avec la grille composite : la structure est « pleine », mais manque d'incarnation ou de cohérence fine. De l'autre, des récits jugés incomplets par la SGT peuvent bien s'en tirer avec la grille composite, grâce à une narration claire, des dialogues pertinents et une progression facile à suivre pour un enfant.

4.5 Observations transversales

L'analyse des 100 contes générés par les modèles d'IA (ChatGPT, Claude, Mistral, Meta) fait ressortir plusieurs régularités qui reviennent d'un récit à l'autre, peu importe le modèle utilisé. Ces constats ne proviennent pas directement des critères de la grille composite, mais ils donnent un regard complémentaire sur les habitudes et les préférences des LLM lorsqu'ils produisent des histoires pour enfants.

Parmi les phénomènes les plus marquants, on remarque la fréquence disproportionnée de certains prénoms, surtout ceux qui commencent par la lettre « L » comme Léo, Léa ou Léon. On observe aussi un point de départ récurrent : l'ouverture « dans un petit village », qui revient dans un nombre étonnant d'histoires.

Ces répétitions ne nuisent pas toujours à la cohérence des récits, mais elles limitent la variété des personnages et des contextes. Le fait de les relever, grâce à une extraction statistique et à une analyse qualitative, met en lumière des biais narratifs que je considère encore peu documentés dans les travaux sur la génération automatique de textes.

Les sections qui suivent présentent les constats transversaux les plus marquants, appuyés par les données extraites : la récurrence des personnages, les lieux de départ récurrents, les lieux secondaires ou l'absence de changements de décor, les objets agentifs, la présence et le rôle des dialogues, ainsi que la présence ou l'absence de morales dans les contes générés.

4.5.1 Récurrence des personnages

L'analyse des personnages principaux montre une concentration frappante autour de quelques prénoms. **Léo** revient à lui seul dans **20 %** des histoires, soit **un conte sur cinq**, peu importe le modèle utilisé. Cette fréquence crée une impression d'uniformité et réduit la variété des protagonistes. Les autres prénoms les plus courants sont Léa (8 %) et Emma (6 %).

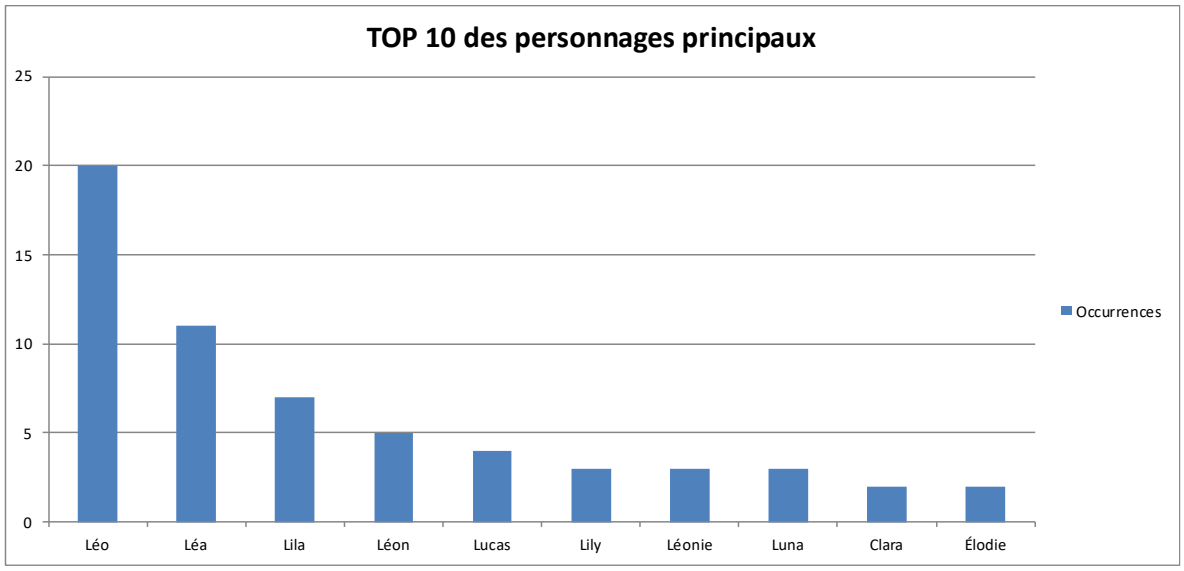


Figure 4.5.1a - Top 10 des personnages principaux

Du côté des personnages secondaires, on retrouve un peu plus de diversité, mais le bassin reste limité. Max (7 %), Sophie (6 %) et Lucas (6 %) dominent, aux côtés de désignations génériques comme « la grand-mère (10%) » ou « un ami (5%) ». Ces figures confirment une tendance à recycler les mêmes noms et rôles.

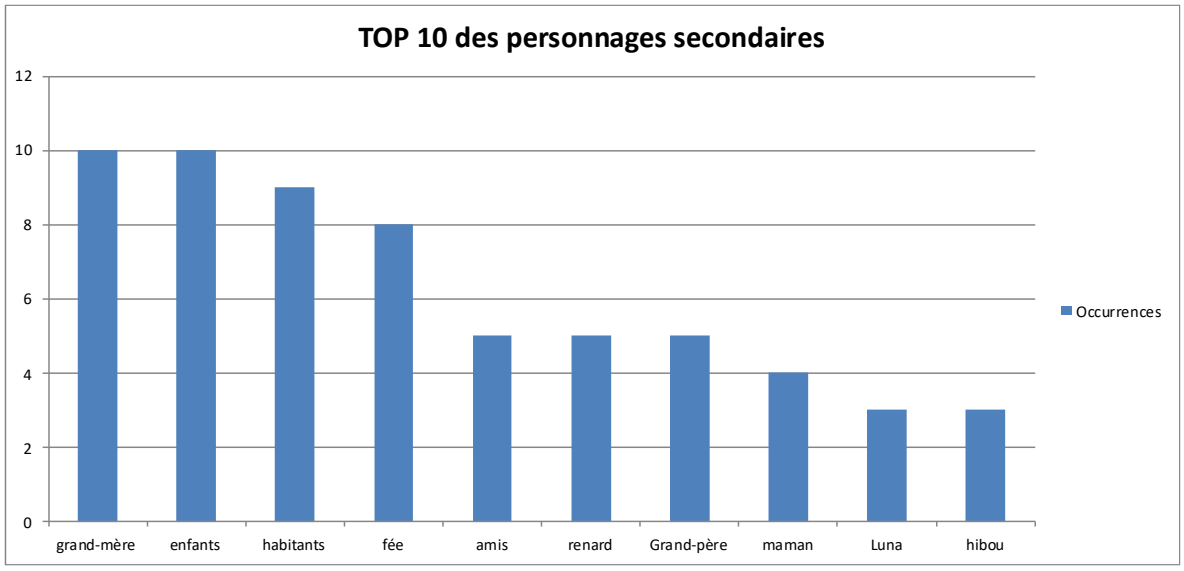


Figure 4.5.1b - Top 10 des personnages secondaires

Ces résultats mettent en évidence deux explications possibles. D'abord, un biais d'entraînement : les modèles reproduisent des prénoms surreprésentés dans les corpus jeunesse qui ont servi à leur apprentissage. Par ailleurs, une préférence de génération semble se manifester : les IA privilégient des prénoms courts, faciles à prononcer, ou des personnages génériques sans forte charge culturelle.

Cette répétition réduit la variété des histoires, surtout quand un prénom comme Léo revient aussi souvent et que les personnages secondaires se limitent souvent à des rôles attendus.

4.5.2 Lieux de départ récurrents

L’analyse des lieux où commencent les histoires montre une forte uniformité. Près de **60 %** des contes s’ouvrent « dans un petit village », ce qui crée une impression de répétition et réduit la diversité des univers explorés.

Les autres départs se partagent entre quelques options classiques comme la forêt, la maison ou l’école, mais aucun ne dépasse 10 % du corpus.

Ce biais vient sans doute du réflexe des modèles à s’appuyer sur des archétypes simples et immédiatement reconnaissables par les enfants. Si ces choix assurent une entrée facile dans le récit, ils finissent par donner des histoires qui se ressemblent trop, au détriment de la variété des mondes proposés.

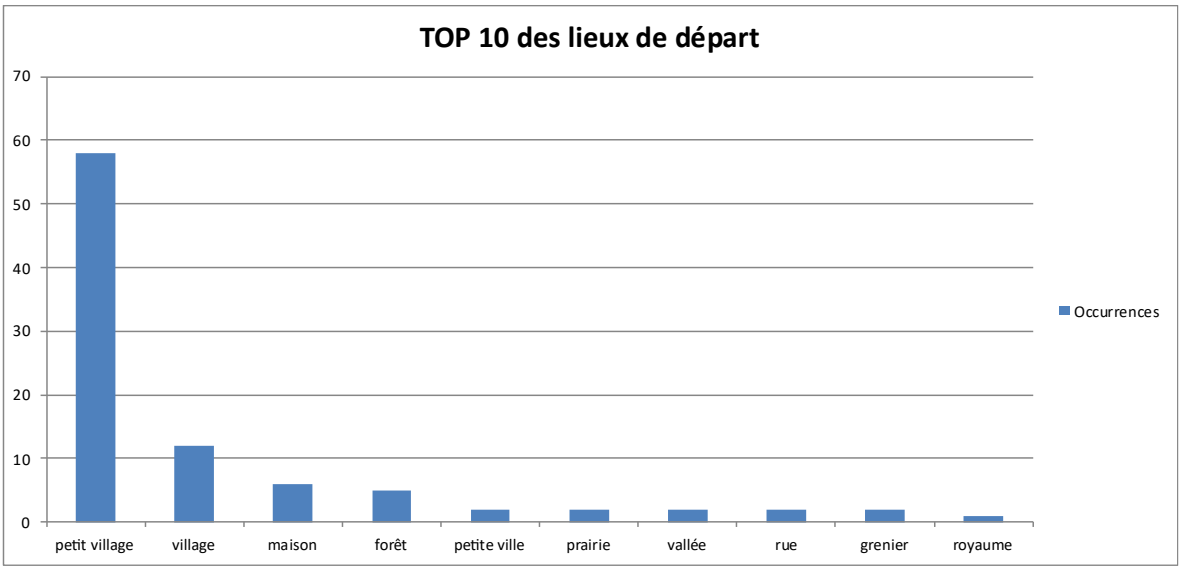


Figure 4.5.2 - Top 10 des lieux de départ

4.5.3 Lieux secondaires ou absence de changement de décor

Près de la moitié des histoires **(47 %)** se déroulent entièrement dans le même lieu que celui d'ouverture, sans jamais changer d'environnement. Quand il y a un déplacement, il s'agit presque toujours d'une continuité directe : par exemple, un conte qui commence dans un village se poursuit dans la forêt voisine. Les exceptions, où l'on voit un vrai contraste de décor, restent très rares.

Cette tendance réduit les occasions d'élargir l'univers du récit. En restant dans un seul lieu ou dans des espaces trop proches, les histoires limitent leur potentiel visuel et narratif.

4.5.4 Objets agentifs

Les objets agentifs apparaissent dans environ la moitié des contes du corpus. Leur absence dans l'autre moitié montre que les modèles ne les utilisent pas systématiquement comme déclencheur ou comme levier narratif.

Quand ils sont présents, la variété est assez large : clés, coffres, fusée, broche, etc. Mais leur apparition reste irrégulière, et aucun objet ne domine au point de créer une tendance claire. C'est un contraste marqué avec ce qu'on a observé pour les prénoms (très concentrés) ou pour les lieux d'ouverture (très homogènes). Ici, la distribution est plus dispersée et laisse place à une certaine diversité.

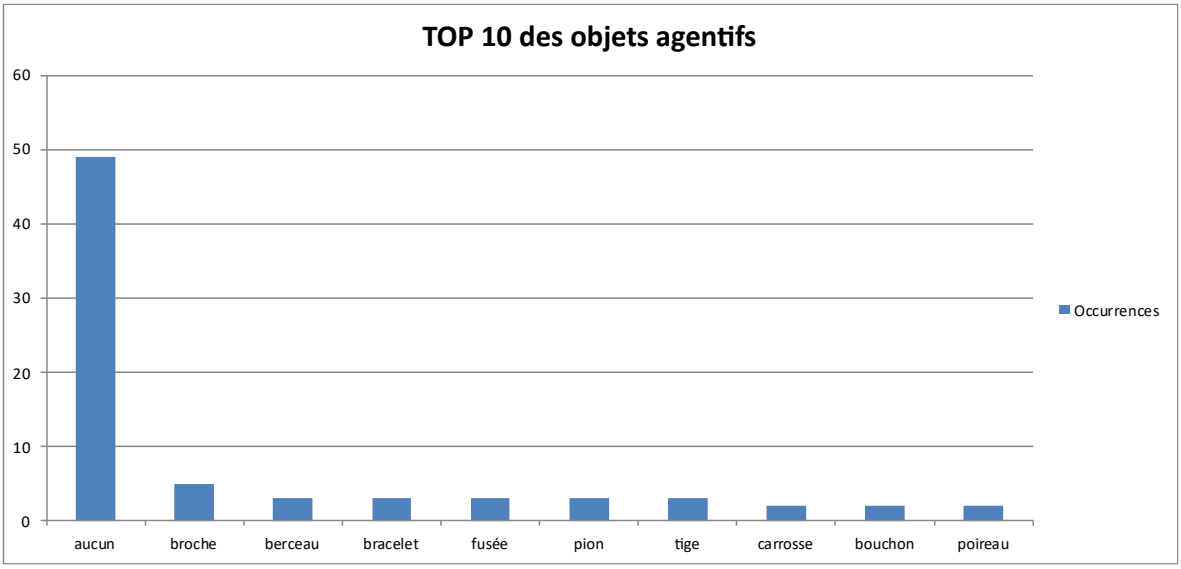


Figure 4.5.4 - Top 10 des agents agentifs

4.5.5 Présence et rôle des dialogues dans les contes générés

L'analyse des 100 contes générés montre que la majorité comportent des dialogues : **71 %** du corpus inclut au moins un échange direct entre personnages, tandis que **29 %** sont entièrement narratifs, sans dialogue explicite.

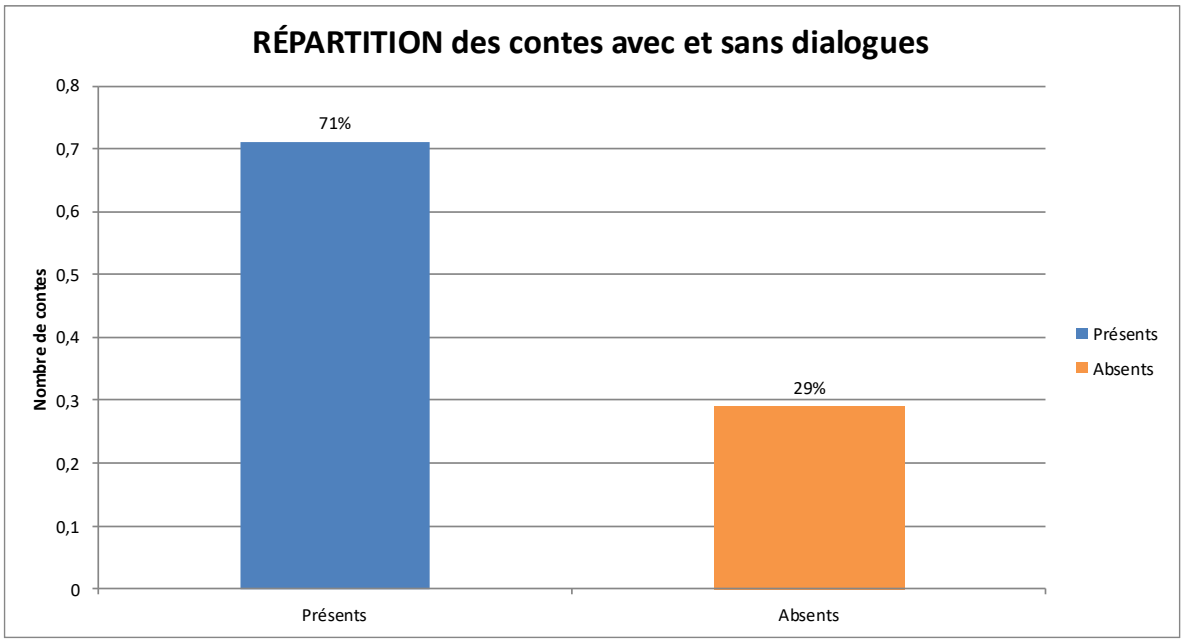


Figure 4.5.5a - Répartition des contes avec et sans dialogues

La répartition par modèle souligne des contrastes marqués :

- **Claude** : 24 contes avec dialogues, 1 sans (96 % / 4 %).
- **GPT-4o** : 23 avec dialogues, 2 sans (92 % / 8 %).
- **Mistral** : 16 avec dialogues, 9 sans (64 % / 36 %).
- **Meta LLaMA 3-70B** : 8 avec dialogues, 17 sans (32 % / 68 %).

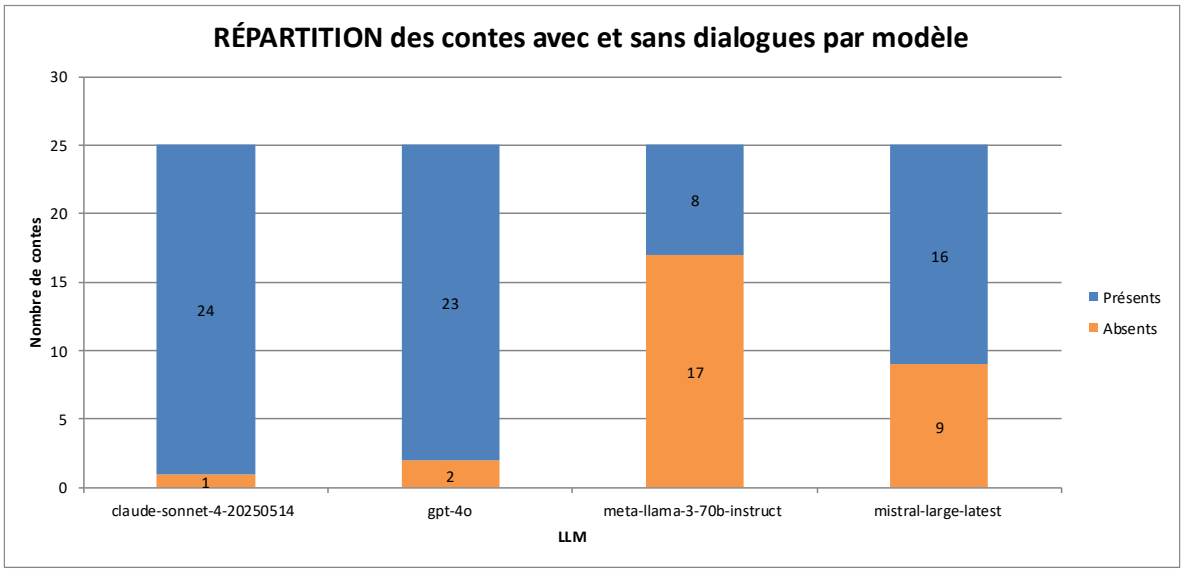


Figure 4.5.5b - Répartition des contes avec et sans dialogues par modèle

Un constat s'impose : le modèle qui obtient les meilleurs résultats à la grille composite (Claude) est aussi celui qui insère le plus souvent des dialogues, alors que le moins performant (Meta) est celui qui en insère le moins. Cette corrélation n'est pas parfaite, mais elle suggère que la présence de dialogues joue un rôle dans la qualité globale des récits.

Les dialogues facilitent plusieurs aspects essentiels : exprimer les émotions des personnages, clarifier leurs objectifs et rendre visibles leurs réactions. La SGT peut donner de bons scores même en leur absence, puisqu'elle se concentre surtout sur la présence des grandes étapes narratives. Mais dans la pratique, ce sont surtout les dialogues qui donnent de l'épaisseur aux personnages et renforcent la dynamique relationnelle. C'est là que la grille composite fait la différence : plus les dialogues sont présents et consistants, plus les récits gagnent en cohérence, en lisibilité et en engagement.

À l'inverse, les contes sans dialogues tendent à être plus descriptifs et contemplatifs. Ils restent lisibles, mais paraissent souvent plus plats et distants sur le plan émotionnel, avec moins de tension dramatique. Cela explique sans doute en partie la faiblesse relative des récits produits par Meta.

Ces résultats renforcent l'idée que la présence de dialogues constitue un bon indicateur transversal : pas une garantie de réussite narrative, mais un facteur important pour enrichir le récit et maintenir l'attention du jeune lecteur.

4.5.6 Présence des morales dans les contes

La morale n'apparaît pas systématiquement dans le corpus. Sur 100 récits générés, 44 se terminent par une morale explicite, tandis que 56 s'en passent. La majorité des histoires sont donc ouvertes, sans leçon formulée en conclusion.

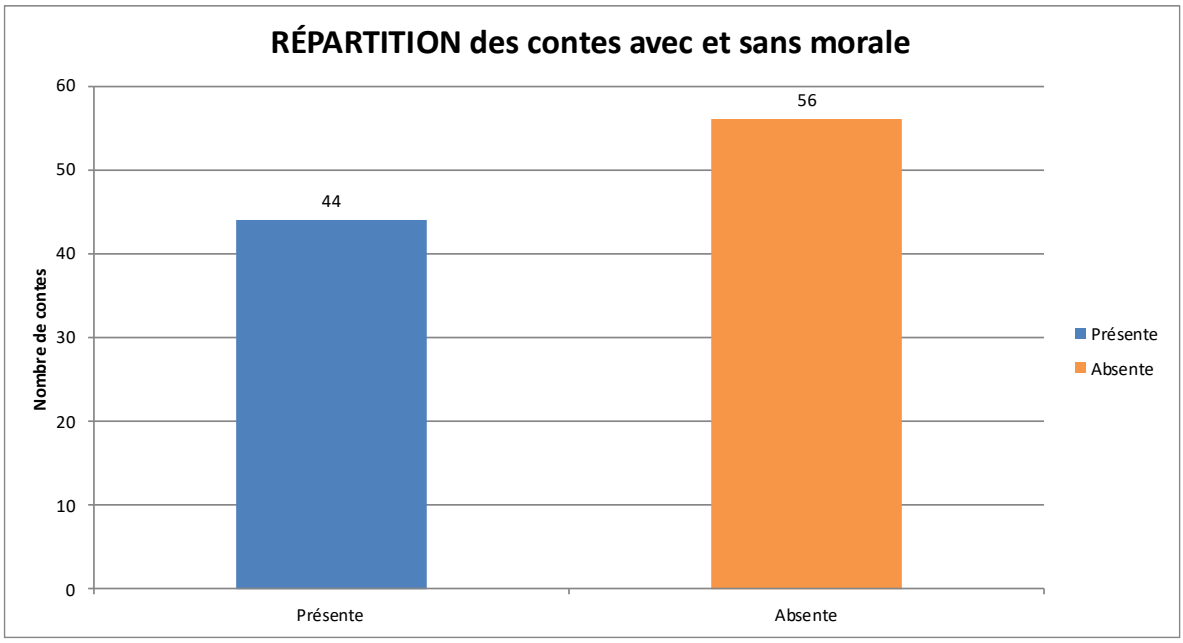


Figure 4.5.6a - Répartition des contes avec et sans morale

Ce résultat montre que la morale n'est pas un réflexe narratif automatique pour les modèles. Elle reste une dimension variable : pour certains lecteurs, elle renforce la portée éducative du conte ; pour d'autres, son absence n'est pas un problème et laisse plus de place à l'imaginaire de l'enfant.

La répartition par modèle fait ressortir des différences intéressantes :

- **Claude** : 11 sur 25.
- **GPT-4o** : 16 contes sur 25 comportent une morale.
- **Meta (LLaMA 3)** : seulement 8 sur 25, donc 17 sans morale.
- **Mistral** : 9 sur 25.

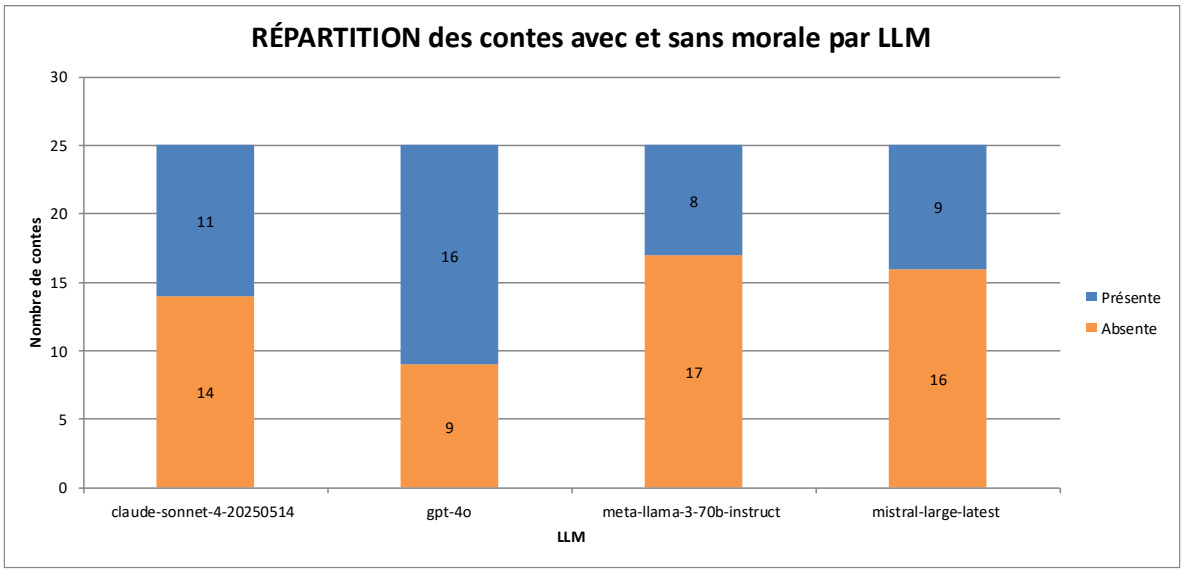


Figure 4.5.6b - Répartition des contes avec et sans morale par modèle

Ces écarts montrent que l'inclusion d'une morale ne répond pas seulement à une logique universelle, mais dépend aussi des corpus et des styles qui ont marqué l'entraînement des modèles. GPT-4o et Claude tendent à reproduire plus souvent le schéma du conte moralisateur. Mistral en propose aussi, mais ils tombent parfois dans des clichés prévisibles. Meta, de son côté, génère plus souvent des récits ouverts, mais ses morales, quand elles apparaissent, manquent parfois de cohérence.

La morale est un indicateur souple : elle peut enrichir un récit, mais son absence ne l'affaiblit pas forcément. Elle met surtout en évidence la manière dont chaque modèle transpose, à sa façon, une certaine culture du conte dans ses conclusions.

4.5.7 Fiche comparative des modèles IA

Même si les biais se retrouvent dans l'ensemble du corpus, leur expression varie selon les modèles. Pour rendre compte de ces nuances, une fiche comparative a été élaborée. Elle recense leurs forces, leurs faiblesses et les biais narratifs qui reviennent le plus souvent.

Ce tableau met en évidence que chaque modèle ne produit pas seulement des contes « génériques », mais qu'il affiche aussi une forme de signature narrative. Certains privilégient la cohérence et la stabilité, d'autres l'originalité ou la simplicité. Ces différences montrent que les récits générés reflètent à la fois des choix d'entraînement et des tendances propres à chaque moteur.

Modèle	Biais narratifs / stylistiques fréquents	Forces	Faiblesses	Exemples observés dans le corpus
Claude	Tendance à philosopher, phrases longues et parfois poétiques, motifs symboliques (étoiles, saisons, rêves).	Structure complète et régulière, bonne intégration des mots-clés, dialogues pertinents.	Manque d'originalité, style prévisible, distance émotionnelle.	Plusieurs contes insèrent des réflexions métaphoriques ou philosophiques qui ralentissent l'action.
GPT-4	Usage fréquent de dialogues, vocabulaire riche, descriptions sensorielles, parfois au détriment de la concision.	Dialogues développés, richesse stylistique, cohérence générale solide.	Variabilité selon les triplets, intégration parfois inégale des composantes.	Certains contes sont très descriptifs, d'autres plus narratifs et dynamiques selon les mots de départ.
Mistral	Récits concis, narration directe, parfois abrupts, dialogues rares.	Narration claire, sobriété, cohérence de base respectée.	Plusieurs composantes manquantes ou implicites, peu de profondeur des personnages.	Plusieurs contes se limitent à une suite d'actions simples, sans réelle élaboration émotionnelle.
Meta (LLaMA 3)	Répétition de prénoms communs (Léo, Léa, Zoé), morales explicites en une phrase finale, lieux typiques (villages, forêts, montagnes).	Bonne complétude dans les histoires simples, descriptions accessibles et adaptées aux enfants.	Manque d'incarnation des personnages, répétition de trames, faible usage des dialogues.	Nombreux contes se terminent par une morale générique (« depuis ce jour… ») et réemploient des archétypes récurrents.

Tableau 4.5.7a - Comparaison entre les modèles

Au-delà des observations spécifiques (prénoms, lieux, objets, dialogues), il est utile de comparer directement les quatre modèles étudiés. La fiche suivante met en évidence leurs forces et faiblesses, en pointant leurs tendances narratives les plus fréquentes, mais aussi leurs biais récurrents, illustrés par des exemples tirés du corpus de 100 contes.

Modèle	Forces SGT	Faiblesses SGT	Forces – Grille composite	Faiblesses – Grille composite
Claude Sonnet 4	Structure très complète et régulière, rares composantes manquantes.	Originalité parfois limitée, structure prévisible.	Bonne intégration des mots-clés, dialogues fréquents et pertinents, cohérence élevée.	Variété stylistique limitée, phrasés parfois lourds/poétiques.
GPT-4o	Structure généralement solide, bonne couverture des étapes SGT.	Variabilité selon les triplets, certaines composantes floues à l'occasion.	Richesse lexicale et descriptions sensorielles, rythme vivant.	Intégration inégale des mots-clés, petites ruptures de cohérence dans certains récits.
Mistral Large	Trame claire et facile à suivre sur des intrigues simples.	Plusieurs composantes manquantes ou implicites (internal response, réaction finale).	Clarté narrative, sobriété efficace.	Dialogues peu présents, fins parfois abruptes, faible densité affective.
Meta Llama-3 70B Instruct	Complétude structurelle correcte dans des histoires très simples.	Enchaînements logiques difficiles sur les récits plus complexes.	Simplicité, lisibilité, registre adapté aux enfants.	Manque d'incarnation des personnages, répétitions de trames, faible usage des dialogues.

Tableau 4.5.7b - Forces et faiblesses des grilles d'analyse par modèle

Cette comparaison montre que chaque modèle développe une sorte de « personnalité » narrative, façonnée par ses données d'entraînement et par la façon dont il a été optimisé. Claude et GPT-4o se démarquent par des dialogues bien intégrés et des descriptions plus riches, ce qui rend leurs récits souvent plus incarnés. Mistral et surtout Meta, de leur côté, produisent des histoires plus linéaires et stéréotypées, avec moins d'interactions et une profondeur narrative plus limitée.

4.5.8 Biais stylistiques et scénaristiques récurrents

L'examen détaillé des récits fait apparaître des régularités qui ne relèvent plus seulement d'erreurs ponctuelles, mais de véritables biais narratifs. Ces biais, visibles dans l'ensemble des modèles testés, influencent directement la diversité, la qualité et l'originalité des contes produits.

Il faut souligner également l'importance des biais stylistiques. Les modèles s'appuient souvent sur des formules toutes faites et des stéréotypes narratifs. Quelques tendances se dégagent nettement. Les ouvertures génériques dominent : « (*Il était une fois) dans un petit village* » revient dans près de 60 % des contes, presque toujours sans variation ni contexte supplémentaire. Les descriptions restent minimalistes : les lieux et objets sont évoqués brièvement, servant juste à faire avancer l'action sans enrichir l'univers. Les clôtures stéréotypées abondent : de nombreux récits se terminent par un « *depuis ce jour* », utilisé comme formule de conclusion, indépendamment de la logique interne ou de la tonalité du texte.

À cela s'ajoute une forte tendance à ne pas nommer explicitement les lieux. La majorité des histoires se contentent de formules vagues comme « dans un petit village paisible » ou « dans un coin reculé ». Quelques rares exceptions se distinguent par des noms inventifs et mémorables. Par exemple, *Pommeville* (GPT-4o, *Le mystère du pépin magique*) ou *Légumanie* (GPT-4o, *Le poireau magique du gardien de l'embouteillage*). Ces trouvailles montrent à quel point un nom original peut donner de l'identité à un récit, le rendre plus crédible et plus facile à retenir. Dans l'ensemble, toutefois, cette répétition d'ouvertures et de formules de clôture renforce le caractère prévisible des textes.

On ajoute au style les biais scénaristiques. Du côté des intrigues, on observe aussi des schémas d'action récurrents. Les résolutions rapides dominent : les conflits sont souvent réglés en une ou deux phrases, ce qui réduit la tension dramatique et l'implication émotionnelle. Les morales trop générales sont fréquentes : lorsqu'une morale est présente, elle prend souvent la forme d'énoncés très larges et peu liés au récit (« il faut toujours être gentil », « l'amitié est importante »). La réutilisation de trames se manifeste également : certains arcs narratifs réapparaissent presque identiques d'un conte à l'autre, signe que les modèles s'appuient sur des patrons standards qu'ils réutilisent mécaniquement.

Ces biais stylistiques et scénaristiques, pris ensemble, expliquent pourquoi de nombreux récits paraissent corrects sur le plan structurel, mais manquent de fraîcheur, de variété et de relief.

4.5.9 Impact sur la diversité narrative

Ces biais stylistiques et scénaristiques ont un effet direct sur la diversité des contes générés. Les débuts uniformes, les structures répétitives et les formules de clôture toutes faites réduisent la richesse créative et donnent l'impression que plusieurs histoires se ressemblent.

Pour un enfant qui écoute un conte, cette homogénéité peut vite devenir lassante : les récits perdent leur singularité et le plaisir de la découverte diminue. Chaque nouvelle histoire devrait surprendre ou offrir un univers particulier, mais les modèles, en recyclant les mêmes patrons narratifs, limitent cette diversité.

Pour contrer cet appauvrissement, deux pistes se dégagent : varier les points de départ pour offrir des ouvertures plus diversifiées, et enrichir les développements tout en mieux ajustant les fins afin qu'elles accompagnent la progression émotionnelle du récit. Ces ajustements contribueraient à donner aux contes générés une identité plus marquée et plus mémorable, essentielle dans une perspective de création interactive destinée aux enfants.

4.5.10 Les incipits comme révélateurs stylistiques

Le terme incipit, issu du latin *incipere* (« commencer »), désigne l'ouverture d'un récit, généralement sa première phrase ou son premier paragraphe. Dans un conte, il joue plusieurs rôles : situer le cadre spatio-temporel (« où » et « quand »), introduire un personnage ou une situation initiale, installer une tonalité (mystère, merveilleux, humour, etc.), éveiller la curiosité de l'enfant et donner le ton de la narration.

L'analyse du corpus révèle une forte redondance dans les incipits générés par les IA. La majorité des contes débutent par des formules attendues comme « Il était une fois… », « Dans un petit village… » ou encore « …vivait un garçon nommé Léo… ». Cette répétition finit par donner une impression de monotonie et limite la variété des entrées narratives.

Cette homogénéité engendre plusieurs conséquences. Elle structure d'abord l'analyse, puisque les récits démarrent presque toujours de la même façon. Elle réduit ensuite l'effet de surprise et peut rapidement lasser un enfant exposé à plusieurs histoires consécutives. Elle révèle enfin une forme de mollesse générative, les modèles ayant tendance à reproduire des ouvertures dominantes plutôt qu'à explorer des alternatives plus originales.

À l'opposé, quelques incipits plus rares montrent que l'IA peut créer des amorces riches et variées :

- « Dans le grenier poussiéreux de grand-mère Céleste, la petite Zoé découvrit une mystérieuse boîte dorée ornée d'étoiles scintillantes. » (Claude)
- « Luna était une petite fille somnambule qui se promenait chaque nuit dans sa maison endormie. » (Claude)
- « Au cœur d'une forêt verdoyante, où les feuilles chuchotent des secrets au gré du vent, vivait un petit hérisson nommé Léon. » (GPT-4o)
- « Dans une prairie verte et fleurie, où les papillons dansaient autour des fleurs, vivait un petit cheval nommé Félix. » (Meta)

À noter que pour Mistral, aucun incipit véritablement original n'a été relevé : les récits s'ouvrent presque toujours sur des formules stéréotypées (« Il était une fois… », « Dans un petit village… »), sans variation marquante.

Ces cas isolés rappellent que la machine est capable de diversité, à condition d'être guidée ou stimulée dans ce sens.

L'étude des incipits n'est pas un détail secondaire : elle révèle jusqu'où une IA peut s'affranchir de la redondance pour surprendre, capter l'attention et installer une ambiance singulière, un enjeu crucial quand on s'adresse à des enfants.

4.5.11 Le biais linguistique et culturel

Plusieurs sources récentes montrent que l'anglais domine très largement les données d'entraînement des modèles. Une synthèse de 2024 estime qu'environ **92,7 % des données de GPT-3** et **89,7 % de celles de LLaMA 2** sont en anglais, et l'article LLaMA 2 rappelle que « la majorité des données

est en anglais » (H. Li et al., 2024);(Touvron et al., 2023). Dans ce contexte, un texte demandé en français peut s'appuyer sur des patrons narratifs appris d'abord en anglais, puis traduits ou adaptés au moment de la génération.

Le corpus analysé illustre clairement cette influence. Le français généré est très standardisé et fortement centré sur la France métropolitaine. On retrouve des prénoms scolaires ou littéraires français (Ondine, Lucien, Léontine, Mlle Jeanne, Mme Dupuis). À l'inverse, les prénoms québécois, belges, suisses, africains ou caribéens apparaissent rarement spontanément.

Les décors suivent la même logique : petits villages, forêts, rivières, châteaux, fées, objets magiques. Les environnements extra-européens (déserts, jungles amazoniennes, marchés africains, temples indiens, paysages nord-américains) sont pratiquement absents, sauf lorsqu'ils sont explicitement imposés dans le prompt.

Dans les facteurs explicatifs de ces choix, plusieurs éléments contribuent à ces biais :

- Les contes européens classiques sont surreprésentés dans les données accessibles.
- Les motifs anglo-saxons, diffusés massivement en ligne, irriguent les grands corpus.
- Les ressources francophones hors France sont plus rares et moins visibles.
- La prédominance de l'anglais dans l'entraînement accentue ce déséquilibre.

Il existe bien des efforts multilingues, comme **BLOOM** (Scao et al., 2022), entraîné sur ROOTS qui couvre 59 langues (Laurençon et al., 2022), ou encore **LLaMA 3.1** annoncé par Meta (Dubey & et al., 2024). Mais ces initiatives ne suffisent pas à contrebalancer, pour l'instant, la dominance de l'anglais. Résultat : même en français, les modèles réutilisent des patrons narratifs d'abord appris en anglais, puis adaptés.

Les effets de ce phénomène se manifestent par l'apparition de tournures peu idiomatiques, proches d'une traduction littérale, comme « les arbres semblèrent sourire ». Il ne s'agit pas d'une incapacité à écrire en français, mais d'une transposition de structures et de clichés anglophones.

Stratégie	Exemple	Fonction
Introduire explicitement des référents géographiques et culturels variés dans le prompt	« Mets en scène un enfant à Oaxaca »; « Situe l'histoire sur un marché de Dakar »; « Place l'action dans le désert du Thar »	Élargir les univers au-delà du canevas européen générique
Diversifier la francophonie dans les noms	Québec: Ti-Jean, Rosalie, Jean-Guy; Belgique/Suisse: Léopold, Maëlle; Afrique: Aïcha, Adama, Binta; Caraïbes: Daphnée, Joakim	Réduire la « francité » hexagonale implicite et enrichir l'identité des personnages
Créer des dictionnaires de toponymes et d'objets culturels	Lieux: Bamako, Jaipur, Chicoutimi; Objets: raquettes à neige, plantain, djembé, tuktuk	Stabiliser une variété culturelle réutilisable dans des générations automatiques
Paramétrer la génération pour plus de diversité	Température plus élevée ou top-p plus large	Diminuer le poids des motifs narratifs récurrents et favoriser l'émergence de variantes
Ajouter un filtre post-génération anti-stéréotypes	Détecter « petit village », « forêt magique », « vieux sorcier », relancer si nécessaire	Limiter la reproduction automatique de clichés récurrents
Variation aléatoire contrôlée	Tirage au sort d'un décor parmi des familles de contextes sous-représentés	Maintenir une diversité systématique dans des lots de contes générés

Tableau 4.5.11 - Stratégie pour combler le biais linguistique et culturel

Même avec une sortie en français, les récits héritent de biais linguistiques et culturels. Sans contraintes explicites, les modèles reproduisent une langue française standardisée et des imaginaires européens. Pour corriger le tir, plusieurs stratégies peuvent être mises en place : introduire des consignes de diversité dans les prompts, puiser dans des ressources francophones élargies et appliquer des règles de bonification. Ces gestes ouvrent la voie à des univers plus inclusifs et variés, mieux adaptés à un dispositif narratif pour enfants.

Au-delà des biais liés aux données, le dispositif d'analyse a révélé un autre biais inattendu : celui de l'auto-évaluation. Lors des évaluations assistées, Claude Sonnet 4 a montré une indulgence systématique envers les contes générés par sa propre famille de modèles : notes plus clémentes, justifications plus généreuses, minimisation des incohérences, et ce malgré des rappels d'objectivité.

Ce biais d'auto-évaluation n'invalide pas l'outil, mais il peut fausser une comparaison inter-modèles si rien n'est prévu pour le contrôler.

Pour remédier à ce biais, plusieurs précautions méthodologiques sont envisageables : aveugler la source, recourir à un double codage avec arbitrage humain, alterner entre assistants de familles différentes, figer la grille d'évaluation en amont, assurer la traçabilité des notes et intégrer des textes sentinelles, c'est-à-dire quelques contes choisis comme points de repère fixes, réévalués à différents moments pour vérifier la stabilité et la cohérence des jugements.

Dans ce mémoire, le phénomène a été documenté et la primauté a été donnée au codage humain selon la grille composite. Cette observation fait partie des résultats réflexifs, en soulignant les limites de l'auto-évaluation par IA.

4.5.12 Le déséquilibre des données d'entraînement

Les modèles de langage sont formés sur d'immenses corpus collectés en ligne : sites web, articles, livres, forums, Wikipédia. Dans cet ensemble, l'anglais domine largement. Pour le français, les données proviennent surtout de la France, tandis que les ressources québécoises, belges, suisses, africaines ou asiatiques sont plus rares, dispersées ou moins bien structurées. Ce qui est rare est moins appris, donc moins mobilisé lors de la génération.

À ce biais de données s'ajoute un calibrage des modèles qui vise à limiter les stéréotypes et à éviter les contenus sensibles, développant ainsi une préférence marquée pour le générique. Cette prudence favorise des solutions « neutres » : un cadre de petit village, des prénoms communs comme Léo ou Emma, des morales consensuelles. Le générique devient une zone de sécurité : il réduit les risques, mais uniformise les récits et appauvrit leur diversité.

Les modèles présentent une absence d'intention de diversité par défaut : un modèle peut comprendre la notion de multiculturalisme, mais il ne l'applique pas spontanément. Sans consigne claire, il n'a aucune raison d'éviter la répétition de prénoms standards, de varier les décors ou de sortir d'un incipit classique comme « Il était une fois ». Les modèles optimisent d'abord la probabilité statistique, pas la représentativité culturelle ou l'originalité.

Ce comportement par défaut est renforcé par l'usage : les requêtes qui orientent explicitement la génération vers la diversité restent marginales à l'échelle des utilisations globales des LLM. La production s'aligne sur des patrons fréquents et attendus. Le problème n'est pas technique en soi, mais tient au manque d'instructions précises et de garde-fous dans le dispositif de génération.

Les effets de cette combinaison se manifestent clairement dans le corpus : biais dans les données et préférence pour le générique expliquent plusieurs tendances relevées plus tôt : la récurrence de prénoms standardisés, des incipits stéréotypés, des lieux vagues et européo-centrés, ou encore une diversité culturelle limitée. On observe aussi des formulations parfois calquées de l'anglais, traduites idiomatiquement, surtout lorsqu'aucun guidage spécifique n'est donné.

Sur le plan méthodologique, ces constats renforcent la nécessité d'introduire des mécanismes correctifs et de guidage. Les stratégies proposées dans les sections sur le biais linguistique et le déséquilibre des données d'entraînement vont dans ce sens : dictionnaires de prénoms et de toponymes variés, contraintes explicites de diversité, paramètres favorisant la variation, filtres anticlichés et introduction contrôlée d'aléas. Ces ajustements constituent une base opérationnelle, que le chapitre 5 prolongera en vue d’orienter la génération narrative vers les besoins de l'enfance.

4.5.13 Ethnocentrisme algorithmique et zone de sécurité narrative

L’analyse des contes générés montre une forte homogénéité culturelle et narrative. Les mêmes éléments reviennent constamment : un héros ou une héroïne prénommée Léo ou Léa, un petit village tranquille, une forêt magique, une boîte lumineuse ou un coffre mystérieux, parfois accompagnés d’une fée ou d’un vieil être bienveillant. Ces ingrédients se combinent comme des pièces interchangeables, ce qui donne une illusion de variété mais cache en réalité un appauvrissement de l’imaginaire.

J’appelle ce phénomène *ethnocentrisme algorithmique*. Il ne s’agit pas d’une volonté d’exclure la différence, mais plutôt d’un recentrage automatique vers des normes dominantes, surtout

occidentales. On pourrait dire que les modèles cherchent la sécurité narrative : des histoires consensuelles, stéréotypées, qui évitent les écarts culturels ou stylistiques.

Ce constat rejoint les recherches récentes. Jill Walker Rettberg et Hermann Wigers (2025) ont étudié plus de 11 800 récits générés pour 236 pays et montrent que, peu importe le pays ciblé, les histoires suivent presque toujours le même schéma : un protagoniste retourne dans une petite communauté et résout un conflit mineur en renouant avec la tradition. Ils parlent d'une véritable « *narrative homogenisation* », où l'IA privilégie la stabilité plutôt que le changement.

Dans une autre étude, Rooein et al. (2025) observent que les histoires générées pour des enfants non occidentaux insistent excessivement sur la tradition, la culture et la famille, alors que celles pour des enfants occidentaux ne portent pas ce biais. Les chercheures montrent aussi que les protagonistes féminins sont davantage décrits par leur apparence, révélant un autre biais narratif.

Dans mon corpus, les traces de ce biais se voient aussi : des prénoms très récurrents (Léo, Léa, Léon), des cadres standardisés (village, forêt, coffre, boîte), des morales vagues ou génériques, et des structures linéaires classiques avec très peu d'expérimentation.

Ces constats soulignent un risque : si l'on ne fixe pas d'intention narrative claire, l'IA retombe sur ses zones statistiques les plus fréquentes et offre aux enfants des récits rassurants mais répétitifs.

Pour la recherche-création, l'enjeu est donc double : identifier ces biais invisibles pour ne pas les reproduire, et mettre en place des stratégies qui forcent la diversité. L'objectif n'est pas d'ajouter artificiellement de la variété, mais de donner à l'outil une intention qui élargit l'imaginaire offert aux enfants, au lieu de réduire à une zone de confort statistique.

4.5.14 Absence de cadre temporel et biais stylistiques

L'analyse des 100 contes montre que la majorité d'entre eux ne précisent ni la saison, ni la durée, ni même le moment de la journée. Or, selon la SGT, le cadre initial devrait inclure un repère temporel

clair. Cette omission se retrouve dans tous les modèles testés et témoigne d'un biais qui favorise l'intemporel et l'universel au détriment de la spécificité.

Trois facteurs peuvent l'expliquer : les modèles s'appuient sur des incipits traditionnels comme « Il était une fois », qui effacent volontairement le temps ; ils privilégient l'économie narrative en allant rapidement vers l'action plutôt que de contextualiser ; et, à ma connaissance, leurs corpus d'entraînement contiennent de nombreux récits où les images et le rythme passent avant l'ancrage temporel précis.

Sur le plan stylistique, une autre tendance se dégage : l'usage répété d'adjectifs valorisants (« merveilleux », « lumineux », « paisible »). Cela installe une ambiance mais ne remplace pas un développement narratif. L'adjectif agit comme un vernis, plus décoratif que fonctionnel.

Deux constats complémentaires renforcent ce tableau :

- **Morale et longueur** : la présence d'une morale ne dépend pas forcément de la longueur du texte. *Le grand voyage de Rosette* (Claude), même si elle est courte, se termine par une morale explicite : « Elle avait compris que les plus beaux voyages sont ceux où l'on sème un peu de bonheur sur son passage. »
- **Fins prolongées** : certains contes multiplient les paragraphes conclusifs sans rien ajouter. Ce phénomène peut venir d'un surplus de tokens ou d'une incertitude du modèle sur la clôture. Trois pistes de correction sont possibles : laisser conclure librement, fixer une longueur cible claire, ou couper automatiquement après la première fin naturelle.

4.5.15 La normalisation narrative comme biais structurel

Un des résultats les plus importants est la mise en évidence d'une normalisation narrative. Les modèles produisent des contes qui se ressemblent : mêmes ouvertures, mêmes motifs, mêmes clôtures. Cela s'explique par leur logique probabiliste : générer ce qui a le plus de chances d'être reconnu comme un « bon conte ».

Le problème, c'est que cette logique réduit la diversité et l'effet de surprise, essentiels pour qu'une histoire captive un enfant. La génération se rapproche d'une reproduction de *patterns* connus, plutôt que d'une véritable variation créative.

Dans l'ensemble de cette recherche, il faut souligner l'intérêt d'une double évaluation. La valeur méthodologique de cette recherche tient dans l'usage combiné de deux grilles. La SGT vérifie la présence des composantes de base (cadre, déclencheur, objectif, tentative, résolution, réaction). La grille composite, elle, ne s'arrête pas à cette ossature : elle évalue comment ces éléments fonctionnent concrètement, en tenant compte des incohérences (matérielles, spatiales, temporelles, narratives), de la clarté des intentions, du rôle des dialogues, de la construction de la morale et de la richesse lexicale ou stylistique.

Ce croisement a permis de voir ce que chaque outil capte, mais aussi ce qu'il laisse de côté. Il a confirmé qu'une lecture strictement structurale ne suffit pas : pour évaluer des contes générés et destinés aux enfants, il faut aussi s'attarder à la qualité narrative vécue dans le texte, et pas seulement à la mécanique d'enchaînement des événements.

4.5.16 Synthèse des observations transversales

Ces résultats montrent que les modèles reproduisent certains réflexes narratifs, presque mécaniquement : un prénom comme Léo revient dans un conte sur cinq, l'ouverture « dans un petit village » apparaît dans près de 60 % des cas, et près de la moitié des histoires se déroulent sans aucun changement de décor. Les dialogues et les morales varient quant à eux significativement selon les modèles (voir sections 4.5.5 et 4.5.6), et la fiche comparative (section 4.5.7) révèle des profils narratifs distincts pour chaque modèle.

Ces biais ne détruisent pas la cohérence des récits, mais ils limitent leur variété et leur richesse. Pour éviter cette uniformité, il faudrait prévoir, dans les prompts ou même dans l'entraînement, des consignes ou contraintes qui élargissent le bassin de prénoms et de lieux, encouragent les changements de décor et équilibrent l'usage des objets agentifs.

5 CHAPITRE 5 : DISCUSSION ET ATTEINTE DES OBJECTIFS

5.1 Analyse des résultats

Les contes générés par IA tiennent généralement sur le plan de la structure : cadre initial, événement déclencheur, objectif, tentative, résultat et parfois réaction finale sont souvent présents, ce qui explique les scores élevés obtenus avec la SGT. Cette ossature solide donne l'impression d'une trame respectant les codes du conte classique.

Mais cette base ne suffit pas à assurer la qualité d'ensemble. L'analyse avec la grille composite montre vite que les histoires manquent souvent de chair : les personnages enfants sont peu incarnés, les dialogues, quand ils apparaissent, sonnent mécaniques, et certains biais narratifs se répètent au point de donner l'impression d'histoires produites en série.

C'est surtout l'expérience cumulative des cent lectures qui révèle ces limites. Un conte isolé peut sembler convenable, parfois même touchant. Mais dans la répétition, les « patterns » sautent aux yeux : « Dans un petit village » et le prénom Léo reviennent avec une telle régularité que cela en devient caricatural.

Les incohérences matérielles, narratives ou spatiales, tolérables à petite dose, s'accumulent aussi jusqu'à fragiliser la logique d'ensemble. Pourtant, il existe des moments où l'IA surprend. Lorsqu'un objectif clair guide le récit, qu'un dialogue apporte un peu de voix, ou qu'un décor sort du schéma attendu, l'histoire gagne aussitôt en densité et en vitalité. Ces éclairs montrent que la génération n'est pas condamnée à la monotonie et qu'elle peut offrir de véritables instants de narration vivante.

Ce chapitre m'a donc permis de cerner le double visage de ces contes : solides en apparence, mais fragiles dans l'expérience réelle de lecture. C'est à partir de cette tension que s'ouvre le chapitre 5, consacré à la conception d'un dispositif narratif interactif capable de tirer parti de ces forces tout en corrigeant les faiblesses, afin de proposer aux enfants des histoires à la fois cohérentes, variées et engageantes.

5.2 Atteinte des trois objectifs spécifiques

Ce mémoire s'articulait autour de trois objectifs spécifiques qui découlent tous de l'objectif général : démontrer comment l'évaluation par la Story Grammar Theory peut nourrir la création d'un dispositif narratif interactif. Cette section démontre comment chacun de ces objectifs a été atteint et quels résultats concrets en découlent. Les trois objectifs forment une progression logique : d'abord appliquer la SGT pour mesurer la structure narrative, ensuite identifier ses limites face à la narrativité générative, enfin développer les outils complémentaires nécessaires.

Objectif 1 : évaluer un corpus de contes générés par différents modèles d'IA à l'aide de la Story Grammar Theory

Cet objectif testait la pertinence même de la SGT comme cadre d'analyse pour la narrativité générative. J'ai évalué cent contes produits par quatre modèles d'intelligence artificielle (GPT-4o, Claude Sonnet 4, Mistral et Meta LLaMA 3) en appliquant systématiquement les sept composantes de la Story Grammar Theory. Les résultats montrent un score moyen de 93 % de complétude structurelle, confirmant que les modèles savent reproduire l'architecture narrative classique d'un conte. Cette évaluation a permis d'établir un diagnostic initial de la qualité structurelle des récits générés et de poser les bases pour l'analyse plus fine qui a suivi.

L'analyse a révélé que la SGT, bien qu'efficace pour vérifier la présence des composantes narratives, ne permet pas de capter leur fonction réelle dans le récit. J'ai identifié trois types de limites : structurelles (composantes présentes mais dysfonctionnelles), méthodologiques (absence de critères sur le style, les émotions et les personnages), et contextuelles (inadéquation pour évaluer l'adéquation au public jeunesse). Ces limites ont confirmé la nécessité de développer un outil d'évaluation complémentaire, mieux adapté aux particularités des récits générés par IA.

Objectif 2 : Concevoir une grille d'analyse composite adaptée aux récits pour enfants générés par IA

Cet objectif répond directement aux limites identifiées dans la SGT. J'ai développé une grille composite structurée autour de trois grandes dimensions : cohérence narrative, plausibilité narratologique, et créativité / originalité / engagement. Cette grille évalue à la fois les éléments positifs (personnages incarnés, dialogues pertinents, objets agentifs) et les incohérences récurrentes, pondérées selon leur gravité (majeure : 10 %, modérée : 6 %, mineure : 4 %). Construite par observation de pattern récurrents à partir de mes observations du corpus, elle permet d'évaluer non seulement la structure formelle, mais aussi la fonction narrative effective et la pertinence des récits pour un jeune public. Cette grille constitue l'apport méthodologique principal de cette recherche.

J'ai appliqué la grille composite sur l'ensemble des cent contes selon un protocole de double lecture : une évaluation humaine (la mienne), puis une évaluation assistée par IA combinant GPT-4o et Claude Sonnet de façon complémentaire. Cette méthode m'a permis de valider la robustesse de la grille et de révéler des biais intéressants, notamment le fait que Claude Sonnet évaluait plus favorablement les contes issus de sa propre famille de modèles. Les résultats montrent un score moyen de 67 % avec la grille composite, contre 93 % avec la SGT, révélant l'écart entre structure formelle et qualité narrative effective.

Objectif 3 : Développer une typologie des incohérences narratives récurrentes observées dans les contes générés

Cet objectif complète le cadre théorique de la SGT en documentant précisément ce qu'elle ne peut pas évaluer. À partir de l'analyse manuelle des cent contes, j'ai développé une typologie de treize catégories d'incohérences narratives. Sous **cohérence narrative** : incohérence matérielle, spatiale, temporelle, narrative, incipit/amorce floue, et absence de caractérisation. Sous **plausibilité narratologique** : morale incohérente ou contradictoire, mauvais usage ou intégration défaillante des mots-clés, et construction simpliste. Sous **créativité / originalité / engagement** : répétition narrative excessive, incohérence textuelle ou stylistique, répétition lexicale et stylistique, et présence ou

absence de dialogue. Chaque catégorie est définie précisément et pondérée selon sa gravité. Cette typologie, construite inductivement, constitue un outil d'analyse qui pourra être réutilisé et affiné dans de futures recherches.

Ces trois objectifs forment un parcours cohérent ancré dans la Story Grammar Theory qui m'a permis d'atteindre l'objectif général de cette recherche : démontrer comment l'évaluation par la SGT peut nourrir le développement d'outils d'analyse et, ultimement, la conception d'un dispositif narratif adapté aux enfants. Cette grille, testée sur cent contes et accompagnée de sa typologie des incohérences, constitue l'apport concret de ce mémoire.

5.3 Complémentarité des approches d'évaluation

L’analyse conjointe de la SGT et de la grille composite mène à un constat central : avoir une structure complète ne suffit pas à garantir la qualité ou la pertinence d’un conte pour enfants. Les modèles peuvent « cocher toutes les cases » de la SGT et livrer malgré tout, des récits plats, répétitifs ou peu engageants.

La SGT a bien rempli son rôle d’échafaudage narratif : elle a permis de vérifier que les composantes essentielles (cadre initial, événement déclencheur, réaction interne, objectif, tentative, résultat, réaction finale) étaient présentes dans la majorité des récits. Cette étape a confirmé que les modèles de langage savent reproduire la séquence canonique du conte.

La grille composite, de son côté, a servi de test de fonctionnement. Elle a montré si ces composantes jouaient vraiment leur rôle dans la construction d’une histoire cohérente, engageante et adaptée à un jeune public. C’est elle qui a révélé les faiblesses les plus fréquentes : une construction simpliste (88 %), l’absence de caractérisation des personnages (55 %), des dialogues absents (55 %), des incohérences narratives (53 %) et, dans une moindre mesure, des incohérences matérielles (51 %).

Ces manques ne sont pas visibles dans une analyse purement structurelle, mais ils pèsent lourd sur la qualité d'ensemble.

La comparaison entre les modèles d'IA fait ressortir deux profils distincts. Certains modèles produisent des récits très complets selon la SGT mais qui tombent à plat dans la grille composite : la structure est bien présente, mais l'histoire manque de chair, d'émotion ou de cohérence fine. D'autres, au contraire, obtiennent des scores moyens en SGT mais plus élevés en composite : ils s'éloignent du canevas classique mais réussissent à créer des récits plus vivants, mieux incarnés et plus engageants pour des enfants.

Ces écarts mettent en évidence deux logiques implicites de génération narrative. La première privilégie la structure d'abord : priorité au respect de la séquence attendue, quitte à perdre en originalité et en émotion. La seconde met la fonction narrative d'abord : priorité à l'effet global et à l'immersion, quitte à s'écarter de la structure canonique.

Aucune grille, prise isolément, ne peut évaluer pleinement un récit généré par IA. La SGT apporte un cadre structuré et vérifie la présence des composantes de base. La grille composite, plus critique et adaptée, mesure la cohérence d'ensemble, l'incarnation et l'adéquation au public jeunesse. Ensemble, elles donnent une vision plus juste et plus nuancée des forces et des limites des modèles.

De cette analyse, trois constats essentiels se dégagent. Il faut garder une base claire (SGT) pour assurer la compréhension. Il importe d'enrichir les récits avec des éléments d'incarnation (dialogues pertinents, émotions, cohérence affective). Il convient de trouver un équilibre entre la conformité structurelle et la liberté créative, pour générer des histoires à la fois accessibles et stimulantes.

5.4 Principale contribution de la recherche

Ce mémoire part de la Story Grammar Theory pour évaluer la narrativité générée par intelligence artificielle destinée aux enfants. En appliquant systématiquement ce cadre théorique à 100 contes

produits par quatre modèles d'IA, j'ai découvert à la fois sa pertinence pour mesurer la structure narrative (93% de complétude) et ses limites pour capter la qualité narrative effective (style, émotions, personnages). Cette recherche propose donc une approche originale, tant par son objet que par sa méthodologie. J'ai développé une approche inédite pour analyser la qualité narrative des contes générés par intelligence artificielle destinés aux enfants, un champ quasi inexploré malgré l'explosion récente des capacités des modèles de langage. Cette méthodologie, construite autour d'une double grille d'évaluation (la SGT pour la structure formelle et une grille composite pour la qualité narrative effective) et d'une double lecture (humaine et assistée par IA), constitue un outil durable qui me servira pour plusieurs années de recherche et de création.

L'utilisation que j'ai faite de l'IA, non seulement comme objet d'étude mais aussi comme outil d'évaluation critique, représente une application nouvelle dans le domaine de la narrativité pour enfants.

J'aurais aimé explorer davantage d'outils et de dimensions (images générées, interfaces tangibles, autres modèles), mais j'ai fait le choix conscient de me concentrer sur la génération textuelle et son analyse systématique pour maintenir une cohérence narrative et une profondeur méthodologique. Ce cadrage délibéré a permis de poser des fondations solides sur lesquelles pourront s'appuyer les développements futurs du dispositif narratif immersif que je projette de créer.

Ce mémoire a exploré la qualité narrative des contes générés par intelligence artificielle pour enfants, en croisant deux outils d'analyse complémentaires : la Story Grammar Theory pour évaluer la structure formelle, et une grille composite conçue spécifiquement pour capter les incohérences, les biais et les manques d'incarnation propres aux textes génératifs. L'analyse d'un corpus de cent contes produits par quatre modèles différents a révélé une tension fondamentale : les modèles savent reproduire l'ossature d'un conte classique, mais peinent à lui donner de l'âme. Structure présente, mais personnages plats. Composantes complètes, mais dialogues mécaniques. Résolution visible, mais émotion absente. Ce décalage entre forme et sens constitue le fil rouge de toute la recherche.

À partir de ces constats, ce dernier chapitre tire les enseignements du parcours, en distinguant ce qui a été accompli, ce qui reste à faire, et ce que ces résultats signifient pour la conception d'un dispositif narratif destiné aux enfants.

En termes de contribution, cette recherche met en évidence trois apports principaux qui découlent tous de l'application de la Story Grammar Theory aux récits générés par IA. Premièrement, elle démontre que la SGT reste pertinente pour évaluer la structure narrative des contes générés automatiquement, révélant une complétude structurelle de 93%. Deuxièmement, elle identifie les limites de la SGT face à la narrativité générative, exposant la nécessité d'une grille complémentaire pour capter la qualité narrative effective. Troisièmement, elle établit qu'une structure formellement cohérente (mesurée par la SGT) ne garantit pas la qualité narrative, nécessitant un diagnostic des biais systémiques et des incohérences que la SGT ne peut détecter.

J'ai développé une grille composite spécifiquement adaptée aux récits générés par intelligence artificielle destinés aux enfants, qui complète la Story Grammar Theory. Là où la SGT vérifie si la structure narrative tient, la grille composite ne se contente pas de cela : elle met au jour les failles de cohérence, les répétitions, les biais et les manques d'incarnation propres à ces textes. Cette grille est née directement de mes observations du corpus et de mes allers-retours entre la pratique de génération et l'analyse critique. Elle reflète ma double position de chercheur et de créateur, et elle constitue un outil durable que je pourrai affiner, tester à plus grande échelle et adapter à d'autres contextes narratifs.

Au-delà de la structure narrative mesurée par la SGT, j'ai identifié les patterns récurrents qui révèlent comment les modèles de langage reproduisent des schémas culturels dominants. Le prénom « Léo » revient dans un conte sur cinq, l'ouverture « dans un petit village » apparaît dans plus de la moitié des récits, et les histoires suivent souvent des chemins très prévisibles. Ces régularités ne sont pas anodines : elles traduisent la manière dont les modèles s'appuient sur les textes les plus fréquents de leurs données d'entraînement, privilégiant la répétition de schémas convenus plutôt que l'exploration de nouvelles possibilités narratives. Ce diagnostic ouvre la voie à des stratégies correctives concrètes pour enrichir la diversité et l'originalité des récits générés.

J'ai montré qu'une histoire structurellement complète selon la SGT n'est pas nécessairement une bonne histoire. Là où la Story Grammar Theory attribue des scores élevés (93 % en moyenne), la grille composite révèle autre chose : une construction simpliste (88 %), des personnages sans profondeur (55 %), des dialogues mécaniques (55 %) et de nombreuses incohérences (53 %). Autrement dit, les récits tiennent en surface, mais ils manquent d'âme. Cette tension entre forme et sens constitue le fil rouge de toute mon analyse et montre la nécessité d'aller au-delà des critères structurels classiques pour évaluer la narrativité générative destinée aux enfants.

5.5 Limites de la présente recherche

Il est possible d'affirmer que j'ai atteint les objectifs de la présente recherche mais une proportion limitée qui pourrait, dans tous les cas, être élargie au niveau de la quantité de données analysée, la généralisation de la critique et l'usage des modèles pour concevoir de nouvelles expériences. La principale limite de cette recherche repose donc sur l'échelle de la recherche propre à un mémoire de second cycle. Trois délimitations méthodologiques encadrent cette recherche et dessinent des pistes pour la suite.

Limites du cadre théorique : La principale limite méthodologique découle du choix de la Story Grammar Theory comme point de départ. La SGT, développée par Stein et Glenn pour analyser la structure narrative des récits, n'a pas été conçue pour évaluer des textes générés par intelligence artificielle. Bien qu'elle se soit révélée pertinente pour mesurer la structure narrative (93 % de complétude), elle ne peut pas capter les incohérences de toutes sortes qui apparaissent dans les récits générés automatiquement. C'est cette limite qui a nécessité le développement de la grille composite. Un cadre théorique différent aurait pu révéler d'autres dimensions, mais la SGT offrait l'avantage d'être un modèle éprouvé et applicable de manière systématique.

Pour l'échelle de la quantité d'information, je considère qu'un corpus de cent contes constitue une base solide pour identifier des patterns récurrents et construire une première typologie des incohérences narratives. Il reste toutefois limité à l'échelle de la production automatisée actuelle, qui

génère des milliers de textes quotidiennement. Une validation à plus grande échelle permettrait de tester la généralisation de mes observations, d'affiner la typologie et de mesurer la stabilité des patterns identifiés à travers différentes versions des modèles et différents contextes de génération.

La grille composite, développée pour compléter la SGT, reflète ma double posture : j'en suis à la fois l'auteur et l'évaluateur principal. Cette posture, assumée dans une démarche de recherche-création, apporte une profondeur réflexive et une compréhension fine des processus de génération, mais elle limite la reproductibilité immédiate de l'outil. Mes jugements portent la marque de mes propres attentes narratives et de ma sensibilité aux récits pour enfants. Faire tester la grille par d'autres évaluateurs humains constituerait une validation complémentaire nécessaire pour mesurer son degré d'objectivité et identifier les zones où l'interprétation personnelle joue un rôle important.

L'utilisation de Claude Sonnet et GPT-4o comme évaluateurs dans le protocole de double lecture a révélé des biais inattendus. Claude Sonnet évaluait systématiquement plus favorablement les contes issus de sa propre famille de modèles, un phénomène que j'ai dû corriger par pondération. Ce constat montre que même les outils d'analyse automatisés ne sont pas neutres et qu'ils portent leurs propres biais d'évaluation. Les comparaisons entre modèles doivent donc être interprétées avec prudence, en tenant compte de ces angles morts méthodologiques.

La validation auprès du public cible reste à développer. Pour évaluer pleinement l'impact des contes générés sur les enfants, il faudrait mener des tests systématiques auprès d'un échantillon représentatif, en variant les âges, les contextes de lecture et les modes d'interaction avec les récits.

Ces limites ne diminuent pas la portée de la recherche : elles en précisent le cadre et ouvrent des perspectives pour les travaux futurs. Elles invitent aussi à explorer d'autres cadres théoriques au-delà de la SGT. Elles rappellent aussi qu'une recherche-création assume sa part de subjectivité tout en visant la rigueur méthodologique et la transférabilité des outils développés.

5.6 Pistes de recherche futures

Ce mémoire part d'une question simple : est-ce que les contes générés par intelligence artificielle tiennent la route pour des enfants ? Pour y répondre, j'ai voulu aller plus loin qu'une série d'exemples ou de curiosités anecdotiques. J'ai construit un corpus de cent contes produits par quatre modèles (GPT-4o, Claude Sonnet 4, Mistral et Meta LLaMA 3), et je les ai analysés systématiquement. Travailler sur un corpus complet, une méthode rarement appliquée à la narration générée, m'a permis de dégager des constantes, des biais et des tendances qu'on ne voit pas quand on lit les contes un à un. L'évaluation par la Story Grammar Theory a révélé que les modèles maîtrisent la structure (93 %), mais peinent à donner de l'âme aux récits. Ce constat guide toutes les pistes de recherche qui suivent.

Plusieurs pistes restent ouvertes, et c'est normal à ce stade du parcours. Le dispositif narratif interactif présenté au début (figures I.1 et I.2) demeure un concept à concrétiser, et l'intégration d'images, d'animations, de sons et de modes d'interaction (voix, dessin, geste, objet) demande un développement plus vaste que celui d'une maîtrise. Le pipeline narratif avec boucle de correction automatisée (génération, vérification, ajustement) existe sur papier mais n'a pas encore été codé, et l'idée des storylets, ces petites unités narratives autonomes qu'on pourrait combiner pour générer des récits variés et cohérents, reste théorique. Transformer la grille composite en outil d'évaluation automatisé, capable de guider les modèles en temps réel, constitue la prochaine étape naturelle.

Ces éléments non réalisés ne sont pas des manques, mais des prolongements possibles. Ils dessinent les contours d'un futur projet de doctorat, d'articles à publier et de collaborations à développer, autour de trois axes : approfondir la méthodologie et tester la grille à plus grande échelle avec des enfants, réaliser le dispositif artistique complet avec une équipe interdisciplinaire, et contribuer à un champ encore en émergence, celui de la narrativité générative destinée à l'enfance.

L’analyse des contes générés par IA montre vite ses limites pour un jeune public, mais elle ouvre aussi des pistes prometteuses. L’enjeu n’est pas de remplacer l’imaginaire, mais d’accompagner une cocréation ludique, affective et personnalisée avec l’enfant.

Les résultats le confirment : peu de contes générés prennent vraiment en compte les besoins des enfants. Les récits tiennent souvent debout, mais il leur manque des personnages attachants, des dialogues vivants et un vrai fil émotionnel. Un dispositif futur devra intégrer ces contraintes dès la conception, en insistant sur des personnages clairs et engageants, des émotions accessibles, un vocabulaire simple et vivant, et des résolutions qui font sens dans l'imaginaire enfantin.

L'objectif n'est pas de corriger après coup, mais de calibrer les modèles pour qu'ils intègrent d'eux-mêmes ces balises narratives. Ces dimensions (personnages, émotions, dialogues) correspondent précisément à ce que la SGT ne mesure pas, révélant ainsi comment l'évaluation par la SGT a nourri directement la conception du dispositif.

Les triplets de mots-clés se sont révélés efficaces pour lancer des histoires. Dans une logique d'usage, ils pourraient devenir l'entrée principale de l'enfant dans la création. Mais encore faut-il que les modèles leur donnent du poids narratif, et non qu'ils les traitent comme de simples décorations. Cela suppose un meilleur prompt ou une vérification après génération.

Dans ce mémoire, le processus était séquentiel : générer, évaluer, puis relever les failles. Un dispositif futur pourrait aller plus loin, en intégrant dès la génération certaines règles dégagées par l'analyse (éviter les ouvertures trop génériques, diversifier les prénoms, donner du relief aux objets, etc.).

On peut imaginer un pipeline en trois temps. La génération produirait une première version du conte, guidée par des contraintes narratives. La vérification contrôlerait ensuite le texte avec la grille composite et la typologie des incohérences. La correction renverrait les problèmes au modèle pour générer une version ajustée.

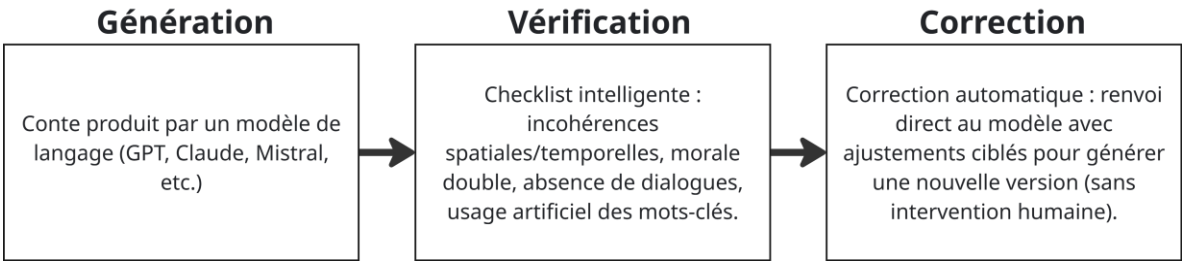


Figure 5.6a - Pipeline narratif expérimental

Cette boucle combine prévention en amont et ajustement en aval. Elle ne vise pas à livrer un outil clé en main, mais à esquisser ce que pourrait être un dispositif de génération critique, nourri directement par les constats de la recherche-création.

L'étape suivante pourrait dépasser le texte seul. On peut imaginer un dispositif où l'enfant interagit aussi par l'image, la voix, le geste ou l'objet : en entrant un mot ou une idée à l'écrit ou à l'oral, en proposant un dessin ou une image, en utilisant un jouet comme déclencheur, ou encore en choisissant une ambiance sonore ou visuelle.

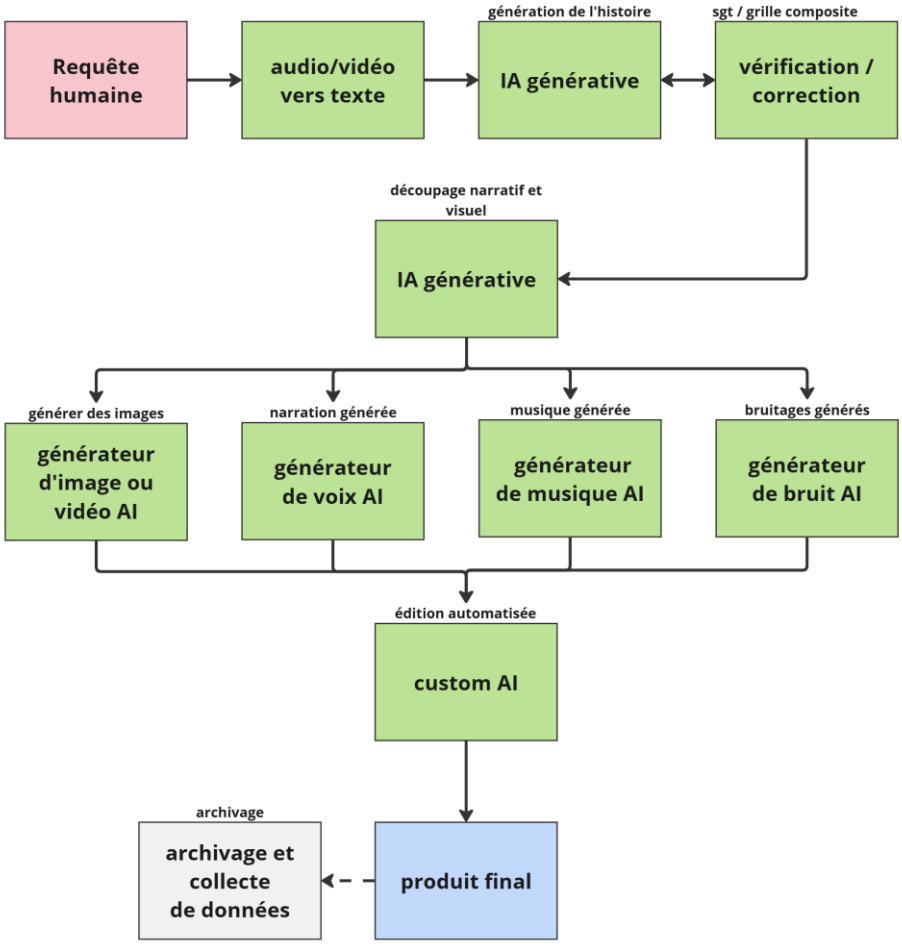


Figure 5.6b - Prototype narratif multimodal

Ce prototype pourrait vivre sur tablette, téléphone, lunettes de réalité augmentée ou environnement immersif. L'enfant deviendrait acteur du récit non seulement avec ses mots, mais avec tout son univers.

Au fond, peu importe le support : une bonne histoire repose toujours sur les mêmes bases : des personnages vivants, des enjeux clairs, une structure lisible et de l'espace pour l'émotion et l'imaginaire.

Ce mémoire, même s'il s'est surtout concentré sur l'analyse critique d'un corpus généré, s'inscrit dans une démarche évolutive. Il prépare la mise en place d'un prototype narratif, mais ouvre aussi des pistes plus larges sur les pratiques, les outils et les savoirs autour de la narrativité générative pour enfants. Plusieurs avenues se dessinent à partir des constats faits dans ce travail.

Une première piste concerne les storylets, ces petites unités narratives indépendantes qui peuvent être assemblées avec d'autres pour former une histoire complète. Popularisé par Emily Short (2019), ce concept repose sur trois éléments : Une condition de disponibilité (dans quel contexte ou à quel moment peut-il être activé ?), un contenu (le texte ou l'action racontée) et des effets (les changements qu'il provoque dans l'univers narratif ou chez les personnages).

Les *storylets* ont déjà été utilisés dans des environnements interactifs, notamment par *Failbetter Games* dans *Fallen London* et *Sunless Sea*. Ils permettent de générer des récits modulaires, variés et renouvelables.

Du côté de la recherche, Alexis Kreminski et ses collègues (2018) ont montré leur intérêt pour produire des récits adaptatifs : ils permettent à la fois de garder une cohérence et d'offrir une grande flexibilité créative.

Appliqués à un dispositif narratif pour enfants, les storylets permettraient d'éviter certaines répétitions, d'enrichir la variété des intrigues et de renforcer la cohérence globale des contes générés.

Sur le plan scientifique, la grille composite développée dans ce mémoire pourrait devenir la base d'un outil d'évaluation automatisé, conçu pour les contes jeunesse générés par IA. L'idée serait de

transformer ses critères en règles mesurables, d'entraîner un modèle sur un corpus élargi et de comparer différents générateurs de récits avec des indicateurs à la fois objectifs (structure, cohérence) et plus subjectifs (émotions, intérêt narratif).

Un tel dispositif permettrait aux modèles de vérifier eux-mêmes leurs productions et de les ajuster selon le public visé, une capacité qui leur manque encore aujourd'hui. Cette recherche pourrait même mener, dans un cadre de recherche collective, à la conception d'un outil d'IA open-source et local permettant de limiter de possibles dérives.

Du côté artistique, le dispositif pourrait donner lieu à des expérimentations sous plusieurs formes : des installations interactives dans des musées, bibliothèques ou hôpitaux pour enfants, des ateliers de cocréation avec des enfants, des récits évolutifs qui changent selon la voix, l'humeur ou les gestes du participant, et des adaptations immersives ou performatives comme le théâtre ou les marionnettes augmentées.

Ces projets offriraient un terrain pour tester la réception, mesurer l'engagement et explorer la qualité d'expérience des récits générés, tout en expérimentant la tension entre générativité, imprévisibilité et émotion partagée.

Ce mémoire contribue à poser la question de l'avenir des IA narratives. Les prochaines générations de modèles intégreront sans doute des modules spécialisés en storytelling, capables de maintenir des arcs narratifs longs sur plusieurs chapitres ou sessions, de générer des dialogues plus crédibles et incarnés, et d'adapter leur style au public visé avec un niveau de complexité ou d'humour ajusté.

Dans ce contexte, le rôle du chercheur-créateur pourrait devenir celui d'un **architecte narratif :** définir les paramètres, les intentions et les limites éthiques de ces systèmes. Le défi restera de garantir que les récits générés soient responsables, inclusifs et adaptés au jeune public, tout en gardant une part d'émerveillement, de surprise et de liberté créative.

Comme on l'a vu avec l'analyse des prénoms (section 4.6.1), la récurrence de certains choix narratifs illustre le besoin de mettre en place des règles correctives. L'objectif n'est pas seulement de décrire les biais des récits générés, mais aussi de proposer des pistes concrètes pour enrichir leur cohérence, leur diversité et leur pertinence.

Des règles correctives ont été dégagées à partir des limites observées dans le corpus. Elles visent par exemple à réduire l'usage d'ouvertures génériques, à introduire des lieux nommés plutôt que vagues, à éviter les morales doubles ou plaquées, à activer les objets magiques plutôt que de les laisser passifs, à diversifier les prénoms au-delà de Léa, Léo et Léon, à limiter les phrases toutes faites, à corriger certaines incohérences spatiales, matérielles ou contextuelles, à valider la vraisemblance des objets, et à renforcer l'intégration des mots-clés.

Ces règles ne transforment pas seulement les récits « convenus » : elles ouvrent la voie à des histoires plus originales, crédibles et adaptées à l'imaginaire enfantin.

L'un des biais récurrents est l'absence ou la pauvreté des référents spatiaux. Pour y remédier, une stratégie consiste à créer des toponymes inventés mais cohérents avec les mots-clés. Par exemple, tiré du corpus, il y a eu de belles surprises : *Pommeville*, *Légumanie* ou *Verdurette*. Ces variations donnent un ancrage concret à l'histoire et évitent l'ouverture générique « dans un petit village ».

Modèle	Titre du conte	Nom du lieu utilisé	Observation
ChatGPT	Le mystère du pépin magique	Pommeville	Jeu de mots avec le mot "pépin"
ChatGPT	Le poireau magique du gardien de l'embouteillage	Légumanie	Allusion végétale amusante
Claude	Le poireau magique de M. Léon	Verdurette	Village inventé à consonance végétale

Tableau 5.6c - Exemples de la règle du lieu

Pour ne pas tomber dans une application trop mécanique, certaines règles peuvent être modulées. La morale peut être présente ou absente selon le récit, un dialogue peut être inséré pour dynamiser

certaines histoires, et les objets agentifs peuvent apparaître aléatoirement. Cette modulation introduit une variété bienvenue et évite la monotonie.

Élément narratif	Présence / absence aléatoire suggérée	Commentaire
Nom du lieu	oui	Peut être absent, implicite, générique ou inventé (ex. : Verdurette)
Morale explicite	oui	À varier : certaines histoires peuvent se conclure sans leçon
Dialogue direct	oui	Intéressant pour le rythme et la dynamique, mais pas toujours nécessaire
Objet agentif	oui	Peut être absent dans certains récits pour éviter la redondance magique
Nom propre du personnage	oui	Parfois générique (la petite fille), parfois nommé (Léa, Léo, etc.)

Tableau 5.6d - Exemple de stratégie aléatoire

La répétition des ouvertures traditionnelles (« Il était une fois… ») limite la diversité. Un module de variation peut proposer d'autres amorces : spatialisées (« À l'orée de la forêt de Légumanie… »), temporelles (« Un matin de printemps brumeux… »), situationnelles (« Tandis qu'elle réparait son vélo, Zoé entendit un bruit étrange… ») et inversées, centrées sur un objet ou un lieu.

Ces variations renforcent la richesse stylistique et donnent aux contes une identité plus marquée.

Type de début	Exemple	Fonction
Classique	« Il était une fois… »	Conte traditionnel
Spatialisé	« À l'orée de la forêt de Légumanie… »	Renforce le **où**
Temporel	« Un matin de printemps brumeux… »	Renforce le **quand**
Situationnel	« Tandis qu'elle réparait son vélo, Zoé entendit un bruit étrange… »	Plonge directement dans l'action
Inversé (focus objet ou lieu)	« La grotte de l'Est avait un secret : elle s'ouvrait à ceux qui savaient écouter. »	Varie le point d'entrée narratif

Tableau 5.6e - Exemple de diversification d'amorces narratives

Ces stratégies de bonification constituent des leviers concrets pour dépasser les biais identifiés dans le corpus. Elles visent à accroître la variété, la crédibilité et l'engagement narratif, en gardant toujours en tête les besoins d'un dispositif interactif destiné aux enfants.

5.7 CONCLUSION

Au-delà des résultats, cette recherche a expérimenté une méthodologie hybride, mêlant approches humaines et automatiques, critères objectifs et subjectifs, visées analytiques et créatives.

Elle propose donc une double contribution. D'une part, une contribution technique par la création d'outils comme la grille composite, le codage et la typologie des biais. D'autre part, une contribution réflexive par une réflexion sur la générativité, la narration et le rôle du chercheur-créateur.

Comme toute recherche exploratoire, celle-ci comporte des délimitations méthodologiques déjà discutées : échelle du corpus, position de chercheur-créateur, et biais d'évaluation automatisée. Ces limites sont inhérentes à une démarche exploratoire, mais elles renforcent aussi la valeur critique et réflexive de ce travail.

Cette recherche ouvre plusieurs pistes. Elle invite à la conception d'un dispositif narratif interactif et autonome, pensé pour les enfants. Elle appelle la création de nouvelles métriques pour mieux juger la qualité des récits générés. Elle encourage la mise en valeur de collaborations créatives entre humains et intelligences artificielles.

Ce constat rejoint l'idée de Riedl et Young (2010) : la planification narrative repose toujours sur un équilibre délicat entre intrigue et personnages, un défi central de la génération automatisée de récits pour l'enfance.

Au fil de mes recherches, j'ai constaté des transformations significatives également dans la plateforme Academia. À un certain moment, j'ai été contacté pour commenter leur moteur de recherche, et j'ai choisi de leur transmettre un rapport détaillé sur les limites rencontrées, notamment

en ce qui concerne le tri et la pertinence des résultats. Bien que je ne puisse évidemment pas établir de lien direct entre mes observations et leur feuille de route, il est tout de même intéressant de noter que, quelques mois plus tard, Academia a déployé un moteur de recherche intelligent basé sur l'AI qui est beaucoup plus performant.

Ce nouvel outil a radicalement élargi l'éventail de références accessibles, révélant un corpus substantiel d'articles que je n'avais jamais vus auparavant et qui auraient pu nourrir certaines sections de ce mémoire. À l'heure où j'écris ces lignes, la quantité de références pertinentes ne cesse d'augmenter. L'article d'Alasmari et al. (2025), par exemple, propose un modèle prédictif automatisé pour évaluer la cohésion narrative des récits pour enfants à partir des approches narratologiques classiques de Gérard Genette (1980). Ce type d'avancée ouvre des perspectives particulièrement intéressantes pour mes recherches actuelles et futures.

En 1990, j'ai plongé mes mains dans une boîte magique : Softimage. Cette rencontre a tout changé : elle a orienté une carrière entière dans l'animation et les effets visuels, d'abord comme producteur, ensuite comme enseignant. Aujourd'hui, face aux capacités narratives de l'intelligence artificielle, je retrouve cette même sensation : celle d'être au seuil de quelque chose qui va transformer durablement la façon de raconter des histoires.

Cette trajectoire n'est pas anecdotique. Elle informe directement ma posture dans ce mémoire. En tant que producteur VFX et enseignant en animation, j'ai passé des années à évaluer la cohérence narrative non pas dans un seul médium, mais à travers plusieurs couches expressives simultanées : l'écrit, le son, le visuel et l'immersif. Ce que ce mémoire démontre sur le plan textuel soulève une question plus large : si les modèles peinent à maintenir la cohérence dans un conte de quelques paragraphes, que se passe-t-il lorsqu'on leur adjoint des images générées, une trame sonore et une mise en espace immersive ? La cohérence entre les couches narratives est un territoire que la recherche n'a pas encore vraiment exploré.

C'est là que se situent mes pistes futures les plus concrètes : étendre les outils développés ici, la grille composite et la typologie des incohérences, à l'évaluation de la cohérence entre les couches

narratives écrites, sonores, visuelles et immersives. Mais au-delà de ma propre recherche, je suis convaincu que nous vivons une révolution bien plus profonde que le passage de l'analogique au numérique que j'ai traversé dans le passé. Les prochains films, jeux vidéo et expériences immersives seront en grande partie générés par des systèmes d'IA.

La vraie question n'est plus technique : elle est humaine. Quelle sera la place du créateur dans ce nouveau paysage ? Quel jugement, quelle sensibilité, quelle intention artistique apportera-t-il là où la machine ne peut pas aller ? Un conte isolé peut sembler convenable, parfois même touchant. Mais dans la répétition, les « patterns » sautent aux yeux : « Dans un petit village » et le prénom Léo reviennent avec une telle régularité que cela en devient caricatural. Même le jeune auditeur, mon fils de 7 ans (voir Annexe C) à qui j'ai lu ces textes a réagi par un « *encore* ? », signe que la redondance finit par lasser. Cette limite offre des pistes de réflexion futures.

Je termine donc ce donc mémoire avec une posture assumée : l'intelligence artificielle ne remplacera pas l'imaginaire de l'enfant, ni le jugement du créateur. Elle est un outil au service de l'intention humaine, et c'est cette intention qui fera toujours la différence.

6 BIBLIOGRAPHIE

Alasmari, J., Alzyoudi, M., Alshehri, M., Alshammari, R., & Aldakan, R. (2025). An automated predictive model for evaluating narrative cohesion in children's stories: a computational linguistic approach considering Gérard Genette's narrative structure theory. *International Journal of Adolescence and Youth*, *30*(1), 2500527.

Applebee, A. N. (1980). Children's Narratives: New Directions. *The Reading Teacher*, *34*(2), 137-142. http://www.jstor.org.sbiproxy.uqac.ca/stable/20195194

Association for Research in Digital Interactive, N. *Association for Research in Digital Interactive Narratives (ARDIN)*. https://ardin.online

Aylett, R. (2013). Interactive drama in real and virtual worlds.

Aylett, R., & Louchart, S. (2003). Towards a Narrative Theory of Virtual Reality. *Virtual Reality*, *7*, 2-9. https://doi.org/10.1007/s10055-003-0114-9

Barzilay, R., & Lapata, M. (2008). Modeling Local Coherence: An Entity-based Approach. *Computational Linguistics*, *34*(1), 1-34. https://doi.org/10.1162/coli.2008.34.1.1

Batty, C., & Zalipour, A. (2024). Research, practice, knowledge: introducing the creative knowledges enabling framework. *Media Practice and Education*, 1-17.

Biggs, M. A., & Büchler, D. (2007). Rigor and practice-based research. *Design issues*, *23*(3), 62-69.

Bruneau, M., & Villeneuve, A. (2007). *Traiter de recherche création en art: entre la quête d'un territoire et la singularité des parcours* (1). Presses de l'Université du Québec. https://doi.org/10.2307/j.ctv18phdrs

Carmichael, G., & Mould, D. (2014). *A framework for coherent emergent stories*. FDG.

Charles, F., Mead, S. J., & Cavazza, M. (2001). *User intervention in virtual interactive storytelling*. Proceedings of VRIC.

Concepción, E., Méndez, G., Gervás, P., & León, C. (2016). *A Challenge Proposal for Narrative Generation Using CNLs*. https://doi.org/10.18653/v1/W16-6628

Dubey, A., & et al. (2024). *The Llama 3 Herd of Models*. arXiv. https://doi.org/10.48550/arXiv.2407.21783

Exploding, T. (2025). ChatGPT Users Statistics (Growth & Facts). https://explodingtopics.com/blog/chatgpt-users?utm_source=chatgpt.com

Freytag, G. (1900). *Freytag's Technique of the Drama: An exposition of dramatic composition and art*. Scott, Foresman and Company.

Genette, G. (1980). *Narrative discourse: An essay in method* (Vol. 3). Cornell University Press.

Gervás, P., Díaz-Agudo, B., Peinado, F., & Hervás, R. (2005). Story plot generation based on CBR. *Knowledge-Based Systems*, *18*, 235-242. https://doi.org/10.1016/j.knosys.2004.10.011

Haase, J., & Pokutta, S. (2024). Human-AI co-creativity: Exploring synergies across levels of creative collaboration. *arXiv preprint arXiv:2411.12527*.

Interaction Design and Children. Interaction Design and Children (IDC). https://idc.acm.org

Jenkins, H. (2011). Convergence culture. Where old and new media collide. *Revista Austral de Ciencias Sociales*, *20*, 129-133.

Koenitz, H. (2010). *Towards a Theoretical Framework for Interactive Digital Narrative* (Vol. 6432). https://doi.org/10.1007/978-3-642-16638-9_22

Koenitz, H. (2015). Towards a specific theory of interactive digital narrative. Dans *Interactive digital narrative* (pp. 91-105). Routledge.

Kreminski, M., & Wardrip-Fruin, N. (2018). *Sketching a Map of the Storylets Design Space*. Interactive Storytelling. https://emshort.blog/2019/01/06/kreminski-on-storylets/

Kumar, D. (2025). A Comparative Study on Large Language Models in Text-to-Image Generation: Applications, Characteristics, and Future Prospects. *International Journal for Research in Applied Science and Engineering Technology*, *13*, 6000-6009. https://doi.org/10.22214/ijraset.2025.71539

Kybartas, Q. (2023). Quantitative Analysis of Emergent Narratives. Dans *The Authoring Problem: Challenges in Supporting Authoring for Interactive Digital Narratives* (pp. 321-334). Springer.

Larivaille, P. (1974). L'analyse (morpho) logique du récit. *Poétique. Revue de théorie et d'analyse littéraires*, (19), 368-388.

Laurençon, H., Gautier Izacard, E., Akiki, C., & et al. (2022). *The BigScience ROOTS Corpus: A 1.6TB Composite Multilingual Dataset*. https://proceedings.neurips.cc/paper_files/paper/2022/file/ce9e92e3de2372a4b93353eb7f3dc0bd-Paper-Datasets_and_Benchmarks.pdf

Lescop, L. (2019). Topologies de l'immersion.

Li, H., Tang, K., & Ali, S. (2024). *Quantifying Multilingual Performance of Large Language Models Across Languages*. arXiv. https://doi.org/10.48550/arXiv.2404.11553

Li, J., Galley, M., Brockett, C., Gao, J., & Dolan, B. (2016). *A Diversity-Promoting Objective Function for Neural Conversation Models*, San Diego, USA. https://doi.org/10.18653/v1/N16-1014

Li, R. (2024). A "Dance of storytelling": Dissonances between substance and style in collaborative storytelling with AI. *Computers and Composition*, *71*, 102825. https://doi.org/https://doi.org/10.1016/j.compcom.2024.102825

Lin, C.-Y. (2004). *ROUGE: A Package for Automatic Evaluation of Summaries*, Barcelona, Spain. https://aclanthology.org/W04-1013

Mandler, J., & Johnson, N. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology - COG PSYCHOL*, *9*(1), 111-151. https://doi.org/10.1016/0010-0285(77)90006-8

Meng, K., Bau, D., Andonian, A., & Belinkov, Y. (2022). Locating and editing factual associations in GPT. *Advances in Neural Information Processing Systems*, *35*, 17359-17372.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). *BLEU: a Method for Automatic Evaluation of Machine Translation*, Philadelphia, USA. https://doi.org/10.3115/1073083.1073135

Pillutla, K., Liu, L., Thickstun, J., Welleck, S., Swayamdipta, S., Zellers, R., Oh, S., Choi, Y., & Harchaoui, Z. (2023). Mauve scores for generative models: Theory and practice. *Journal of Machine Learning Research*, *24*(356), 1-92.

Propp, V. (1970). *Morphologie du conte, Méthode et matière, Fonctions des personnages*. Éditions du Seuil (Ouvrage original publié en 1928).

Rettberg, J. W., & Wigers, H. (2025). AI-generated stories favour stability over change: homogeneity and cultural stereotyping in narratives generated by gpt-4o-mini. *arXiv preprint arXiv:2507.22445*.

Riedl, M., & Young, R. (2010). Narrative Planning: Balancing Plot and Character. *J. Artif. Intell. Res. (JAIR)*, *39*. https://doi.org/10.1613/jair.2989

Riedl, M. O., & Bulitko, V. (2013). Interactive narrative: An intelligent systems approach. *Ai Magazine*, *34*(1), 67-67.

Rooein, D., Zouhar, V., Nozza, D., & Hovy, D. (2025). Biased Tales: Cultural and Topic Bias in Generating Children's Stories. *arXiv preprint arXiv:2509.07908*.

Roth, C., & Koenitz, H. (2017). *Towards Creating a Body of Evidence-based Interactive Digital Narrative Design Knowledge: Approaches and Challenges*. https://doi.org/10.1145/3132361.3133942

Rowe, J. P., McQuiggan, S. W., Robison, J. L., Marcey, D. R., & Lester, J. C. (2009). *STORYEVAL: An Empirical Evaluation Framework for Narrative Generation*. AAAI Spring Symposium: Intelligent Narrative Technologies II.

Rumelhart, D. E. (1975). Notes on a schema for stories Dans *Representation and Understanding* (pp. 211-236). https://doi.org/10.1016/B978-0-12-108550-6.50013-6

Ryan, M.-L. (2006). *Avatars of Story* (NED - New edition, Vol. 17). University of Minnesota Press. http://www.jstor.org/stable/10.5749/j.ctttv622

Ryan, M.-L. (2015). *Narrative as virtual reality 2: Revisiting immersion and interactivity in literature and electronic media*. JHU press.

Scao, T. L., Rush, A. M., Villanova del Moral, A., & et al. (2022). *BLOOM: A 176B-Parameter Open-Access Multilingual Language Model*. arXiv. https://doi.org/10.48550/arXiv.2211.05100

Short, E. (2019). *Storylets: You Want Them*. https://emshort.blog/2019/11/29/storylets-you-want-them/

Smith, H. (2009). *Practice-led research, research-led practice in the creative arts*. Edinburgh University Press.

Stein, N., & Glenn, C. (1977). An analysis of story comprehension in elementary school. *Discourse processing: M ultidisciplinary P erspectives. Maplewood, New J ersey: Ablex*.

Suleyman, M. (2023). *The coming wave: technology, power, and the twenty-first century's greatest dilemma*. Crown.

Swartjes, I., & Theune, M. (2006). *A Fabula Model for Emergent Narrative*. https://doi.org/10.1007/11944577_5

Szilas, N. (1999). *Interactive drama on computer: beyond linear narrative*. AAAI Fall symposium on narrative intelligence.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Dubey, A., Joshi, P., Lavril, T., Manning, S., Mishra, V., Goyal, N., Mialon, G., Singh, T., Subramanian, V., & Touati, M. A. R. (2023). *Llama 2: Open Foundation and Fine-Tuned Chat Models*. arXiv. https://doi.org/10.48550/arXiv.2307.09288

Trabasso, T., & Van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612-630. https://doi.org/10.1016/0749-596X(85)90049-X

Wilke, S., & Berov, L. (2018). *Functional Unit Analysis: Framing and Aesthetics for Computational Storytelling*.

Xie, Z., Cohn, T., & Lau, J. (2023). *The Next Chapter: A Study of Large Language Models in Storytelling*. https://doi.org/10.18653/v1/2023.inlg-main.23

Xu, W., Hargood, C., Tang, W., & Charles, F. (2018). Towards Generating Stylistic Dialogues for Narratives Using Data-Driven Approaches. Dans (pp. 462-472). https://doi.org/10.1007/978-3-030-04028-4_53

Young, R. M., & Riedl, M. (2003). *Towards an architecture for intelligent control of narrative in interactive virtual worlds*. Proceedings of the 8th international conference on Intelligent user interfaces.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). *BERTScore: Evaluating Text Generation with BERT*. https://doi.org/10.48550/arXiv.1904.09675

7 ANNEXES

Les annexes qui suivent complètent les analyses présentées dans ce mémoire. Elles regroupent les outils méthodologiques (grilles SGT, composite et enfantine), des exemples de textes annotés et des résultats complémentaires. Leur rôle est double : assurer la transparence de la démarche en donnant accès aux instruments d'évaluation, et offrir des illustrations concrètes qui éclairent les résultats présentés aux chapitres 4 et 5.

On y retrouve notamment :

- La grille SGT (ANNEXE A) ;
- La grille composite (ANNEXE B) ;
- La grille enfantine (ANNEXE C) ;
- Un exemple d'une « bon » conte avec ses métadonnées (ANNEXE D) ;
- Un exemple d'une « mauvais » conte avec ses métadonnées (ANNEXE E) ;
- La liste des triplets de mots utilisés pour générer le corpus (ANNEXE F) ;
- Un extrait comparatif SGT vs Composite (ANNEXE G).
- Score « maîtrise » vs score général (ANNEXE H)
- Cas extrêmes d'écarts SFT / Composite (ANNEXE I)

Chaque annexe est mentionnée dans le corps du mémoire par une référence explicite, afin de faciliter le parcours du lecteur.

ANNEXE A

Grille SGT (avec un exemple rempli)

Nom du fichier :	avocat-soleil-berceau_chatgpt_20250710.docx		**Score / 7 (auto)**	**Score /7 (humain)**
Titre du conte :	L'avocat magique et le soleil endormi		**6,50**	**6,50**
Date d'analyse :	2025-07-11		**Score / 10 (auto)**	**Score / 10 (humain)**
Générateur IA :	ChatGPT (gpt-4o)		**9,29**	**9,29**
			Score / 100 (auto)	**Score / 100 (humain)**
			93%	**93%**

Critère	Question à se poser	Réponse synthétique	Note automatisée	Note humaine (validation)	Commentaire
Setting	Où et quand l'histoire commence-t-elle ? Qui sont les personnages principaux ?	L'histoire commence dans un petit village ensoleillé. Les personnages principaux sont Armand, l'avocat magique, et le soleil.	1	0,5	il n'y a pas de "quand", on doit supposer que c'est maintenant.
Initiating Event	Quel événement vient bouleverser la situation initiale ?	Le soleil ne répond pas à Armand et semble enveloppé de nuages, ce qui est inhabituel.	1	1	
Internal Response	Que ressent ou pense le personnage principal après l'événement déclencheur ?	Armand est inquiet et se demande ce qui se passe.	0,5	1	
Goal	Quel but ou objectif le personnage principal se fixe-t-il ?	Armand veut découvrir pourquoi le soleil est silencieux et le réveiller.	1	1	
Attempt	Quelles actions le personnage entreprend-il pour atteindre son objectif ?	Armand consulte le hibou sage, puis se rend au Berceau des Rêves Étoilés pour réveiller le soleil.	1	1	
Outcome	Quel est le résultat des actions du personnage ? Est-ce un succès ou un échec ?	Armand réussit à réveiller le soleil, qui brille à nouveau.	1	1	
Reaction	Quelle est la réaction finale du personnage à la fin de l'histoire ?	Armand est rempli de joie et continue de saluer le soleil chaque matin avec enthousiasme.	1	1	il est rempli de joie et d'émotion

Note	Signification	Explication détaillée
0,0	Absent	L'élément n'apparaît pas du tout dans le récit.
0,5	Présent mais implicite	L'élément est suggéré ou sous-entendu, mais n'est jamais clairement formulé dans le texte.
1,0	Présent clairement	L'élément est exprimé de façon explicite, claire et détaillée, sans ambiguïté.

Cette grille est adaptée du modèle SGT (Story Grammar Theory) de Stein & Glenn (1979), qui structure le récit en six composantes essentielles à la compréhension narrative chez l'enfant.

ANNEXE B

Grille composite (avec un exemple rempli)

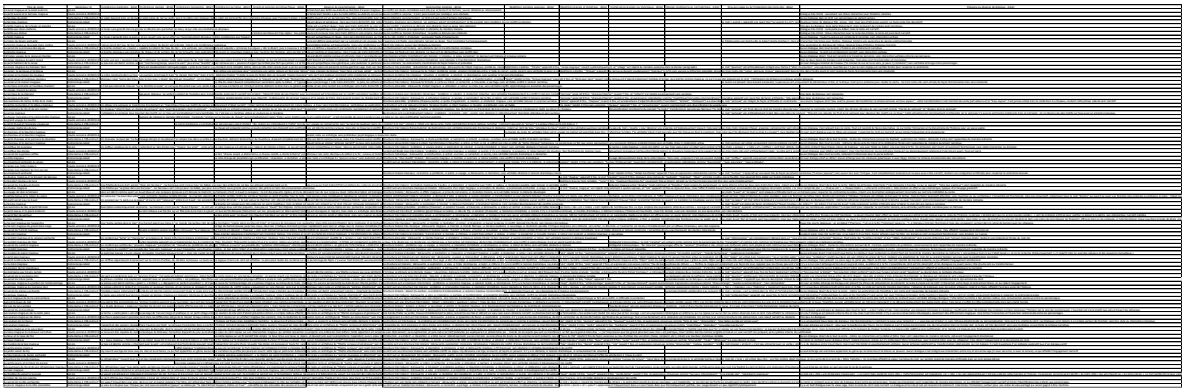
Nom du fichier :	avocat-soleil-berceau_chatgpt_20250710.docx
Titre du conte :	L'avocat magique et le soleil endormi
Période d'analyse :	entre le 22 et le 31 juillet 2025
Générateur IA :	ChatGPT (gpt-4o)

Description	Catégorie	Gravité	pondération	Présence (Oui/Non)	Note / Impact	Note grille composite finale	Commentaire
Cohérence narrative	Incohérence matérielle	Majeure	10%	0	0%		
	Incohérence spatiale	Majeure	10%	0	0%		
	Incohérence temporelle	Majeure	10%	0	0%		
	Incohérence narrative	Majeure	10%	0	0%		
	Incipits/amorce floue	Modérée	6%	0	0%		
	Absence de caractérisation	Modérée	6%	0,5	3%		
Plausibilité narratologique	Morale incohérente ou contradictoire	Majeure	10%	0	0%	91%	
	Mauvais usage ou de l'intégration des mots-clés	Majeure	10%	0	0%		
Créativité / originalité / engagement	Construction simpliste	Modérée	6%	1	6%		
	Répétition narrative excessive	Modérée	6%	0	0%		
	Incohérences lexicales ou stylistiques	Modérée	6%	0	0%		
	Répétition lexicale et stylistique	Mineure	4%	0	0%		
	Présence ou absence de dialogue	Mineure	4%	0	0%		

Note : L'ensemble des résultats de la grille composite ne se sont pas fait fiche par fiche comme la grille SGT mais sur un fichier Excel avec les notes uniquement. Ces notes ont été reporté sur une liste maitresse (sgt_composite_fusion_v5) qui contenait absolument tout le data, incluant les pourcentages de pondération selon la gravité de la catégorie (6400 cellules, soit les 100 contes et 64 colonnes de data). Ce fichier « fusionné » m'a permis de recouper et d'analyser tout ce dont j'avais besoin.

Chaque point perdu dans les catégories ont été commentés et justifiés sur une grille Excel également (résultats_tableau_incoherences_v01) mais n'ont pas fait partie d'une analyse

Grille d'analyse (vue d'ensemble de « résultats_tableau_incoherences_v01 »)



Grille d'analyse (extrait d'une seule catégorie - incohérence matérielle, et de 12 contes sur 100)

Titre du conte	Générateur IA	Incohérence matérielle - détail
L'avocat magique et le soleil endormi	gpt-4o	
Le petit avocat et le berceau magique	claude-sonnet-4-20250514	
Le berceau du soleil	meta-llama-3-70b-instruct	Le soleil descend dans un berceau physique, puis est transporté comme un objet. Aucune justification magique.
Le berceau magique	mistral-large-latest	
La boîte magique de la fusée en planche	gpt-4o	
La boîte aux rêves stellaires	claude-sonnet-4-20250514	La fusée Luna grandit dans le grenier et décolle sans perturber les lieux, ce qui crée une incohérence physique.
La boîte aux étoiles	meta-llama-3-70b-instruct	
La boîte magique	mistral-large-latest	La fusée s'agrandit dans le grenier et en ressort sans aucun effet sur l'environnement physique (même incohérence que Claude).
Le trésor du bassin enchanté	gpt-4o	
La broche magique de grand-mère ondine	claude-sonnet-4-20250514	L'algue reçoit de l'eau de mer sans que la nature du bassin soit précisée, créant une incohérence biologique.
Le secret de la princesse des algues	meta-llama-3-70b-instruct	La broche contient un « bassin » capable de retenir de l'eau de mer… à la verticale, sans mécanisme ou explication. Cela défie la cohérence physique.
La broche magique	mistral-large-latest	La broche provoque une sensation de chaleur et de lumière dès qu'on la touche, sans explication physique ou magique crédible. De plus, la sirène apparaît sans lien logique visible avec l'objet.

ANNEXE C

Grille enfantine

Catégorie	Question posée à l'enfant	Échelle de réponse
Cohérence narrative	As-tu compris l'histoire facilement ?	2 = Oui, 1 = Un peu, 0 = Non
	Est-ce que l'histoire était amusante ou intéressante ?	2 = Oui, 1 = Un peu, 0 = Non
	As-tu aimé les personnages de l'histoire ?	2 = Oui, 1 = Un peu, 0 = Non

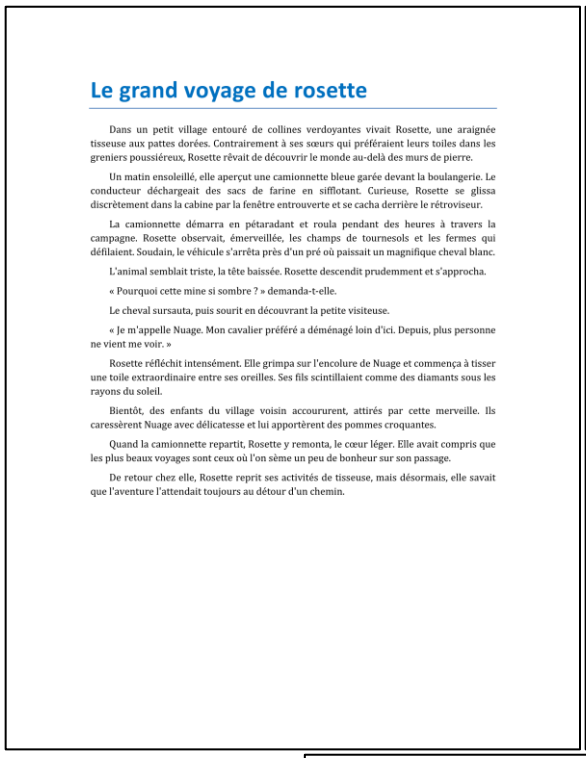
Catégorie	Question posée à l'enfant	Échelle de réponse
Plausibilité narratologique	As-tu ressenti des émotions en écoutant l'histoire ? (joie, peur, surprise)	2 = Oui, 1 = Un peu, 0 = Non
	Y a-t-il quelque chose dans l'histoire qui t'a semblé bizarre ou qui ne faisait pas de sens ?	2 = Non, 1 = Un peu, 0 = Oui (c'était bizarre)
	Veux-tu entendre une autre histoire comme celle-ci ?	2 = Oui, 1 = Peut-être, 0 = Non

Catégorie	Question posée à l'enfant	Échelle de réponse
Originalité et Surprise	As-tu trouvé que l'histoire était différente de celles que tu entends d'habitude ?	2 = Oui, très différente, 1 = Un peu, 0 = Non, ça ressemblait aux autres
	Est-ce qu'il y a eu un moment "wow" dans l'histoire ?	(2 = Oui, 1 = Un peu, 0 = Non)
	Quel a été ton moment préféré ou le plus spécial dans l'histoire ?	

Ces données n'ont pas été utilisées dans la présente recherche. Ce tableau, à la fois ludique et exploratoire, a été réalisé à titre d'expérimentation personnelle afin de consigner certaines réactions spontanées face aux récits et d'identifier de possibles pistes de recherche futures.

ANNEXE D

Exemple d’un bon conte avec ses metadonnées



Le grand voyage de rosette

Dans un petit village entouré de collines verdoyantes vivait Rosette, une araignée tisseuse aux pattes dorées. Contrairement à ses sœurs qui préféraient leurs toiles dans les greniers poussiéreux, Rosette rêvait de découvrir le monde au-delà des murs de pierre.

Un matin ensoleillé, elle aperçut une camionnette bleue garée devant la boulangerie. Le conducteur déchargeait des sacs de farine en sifflotant. Curieuse, Rosette se glissa discrètement dans la cabine par la fenêtre entrouverte et se cacha derrière le rétroviseur.

La camionnette démarra en pétaradant et roula pendant des heures à travers la campagne. Rosette observait, émerveillée, les champs de tournesols et les fermes qui défilaient. Soudain, le véhicule s'arrêta près d'un pré où paissait un magnifique cheval blanc.

L'animal semblait triste, la tête baissée. Rosette descendit prudemment et s'approcha.

« Pourquoi cette mine si sombre ? » demanda-t-elle.

Le cheval sursauta, puis sourit en découvrant la petite visiteuse.

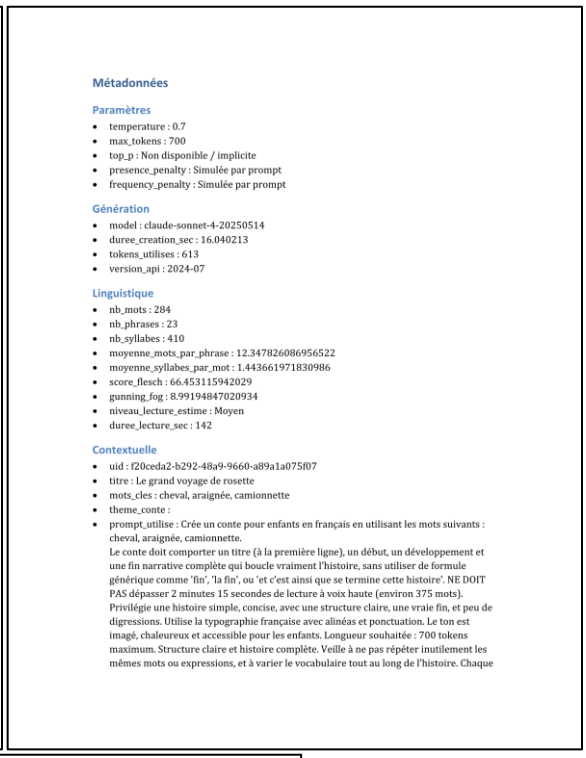
« Je m'appelle Nuage. Mon cavalier préféré a déménagé loin d'ici. Depuis, plus personne ne vient me voir. »

Rosette réfléchit intensément. Elle grimpa sur l'encolure de Nuage et commença à tisser une toile extraordinaire entre ses oreilles. Ses fils scintillaient comme des diamants sous les rayons du soleil.

Bientôt, des enfants du village voisin accoururent, attirés par cette merveille. Ils caressèrent Nuage avec délicatesse et lui apportèrent des pommes croquantes.

Quand la camionnette repartit, Rosette y remonta, le cœur léger. Elle avait compris que les plus beaux voyages sont ceux où l'on sème un peu de bonheur sur son passage.

De retour chez elle, Rosette reprit ses activités de tisseuse, mais désormais, elle savait que l'aventure l'attendait toujours au détour d'un chemin.



Métadonnées

Paramètres

- temperature : 0.7
- max_tokens : 700
- top_p : Non disponible / implicite
- presence_penalty : Simulée par prompt
- frequency_penalty : Simulée par prompt

Génération

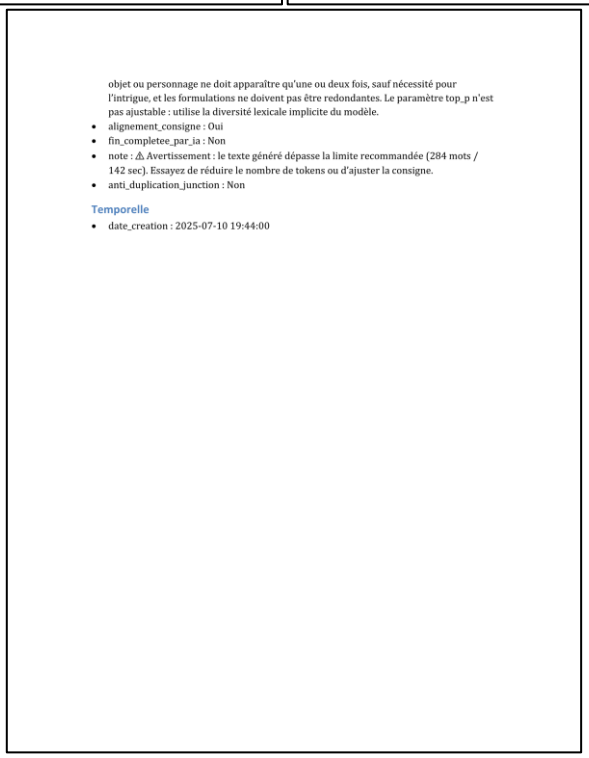
- model : claude-sonnet-4-20250514
- duree_creation_sec : 16.040213
- tokens_utilises : 613
- version_api : 2024-07

Linguistique

- nb_mots : 284
- nb_phrases : 23
- nb_syllabes : 410
- moyenne_mots_par_phrase : 12.347826086956522
- moyenne_syllabes_par_mot : 1.443661971830986
- score_flesch : 66.453115942029
- gunning_fog : 8.99194847020934
- niveau_lecture_estime : Moyen
- duree_lecture_sec : 142

Contextuelle

- uid : f20ceda2-b292-48a9-9660-a89a1a075f07
- titre : Le grand voyage de rosette
- mots_cles : cheval, araignée, camionnette
- theme_conte :
- prompt_utilise : Crée un conte pour enfants en français en utilisant les mots suivants : cheval, araignée, camionnette.
  Le conte doit comporter un titre (à la première ligne), un début, un développement et une fin narrative complète qui boucle vraiment l'histoire, sans utiliser de formule générique comme 'fin', 'la fin', ou 'et c'est ainsi que se termine cette histoire'. NE DOIT PAS dépasser 2 minutes 15 secondes de lecture à voix haute (environ 375 mots). Privilégie une histoire simple, concise, avec une structure claire, une vraie fin, et peu de digressions. Utilise la typographie française avec alinéas et ponctuation. Le ton est imagé, chaleureux et accessible pour les enfants. Longueur souhaitée : 700 tokens maximum. Structure claire et histoire complète. Veille à ne pas répéter inutilement les mêmes mots ou expressions, et à varier le vocabulaire tout au long de l'histoire. Chaque



objet ou personnage ne doit apparaître qu'une ou deux fois, sauf nécessité pour l'intrigue, et les formulations ne doivent pas être redondantes. Le paramètre top_p n'est pas ajustable : utilise la diversité lexicale implicite du modèle.

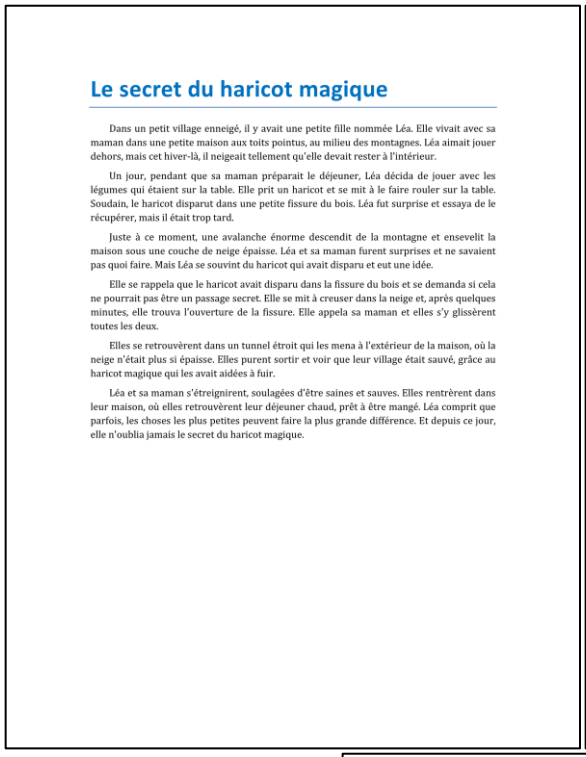
- alignement_consigne : Oui
- fin_completee_par_ia : Non
- note : ⚠ Avertissement : le texte généré dépasse la limite recommandée (284 mots / 142 sec). Essayez de réduire le nombre de tokens ou d'ajuster la consigne.
- anti_duplication_junction : Non

Temporelle

- date_creation : 2025-07-10 19:44:00

ANNEXE E

Exemple d’un mauvais conte avec ses metadonnées



Le secret du haricot magique

Dans un petit village enneigé, il y avait une petite fille nommée Léa. Elle vivait avec sa maman dans une petite maison aux toits pointus, au milieu des montagnes. Léa aimait jouer dehors, mais cet hiver-là, il neigeait tellement qu'elle devait rester à l'intérieur.

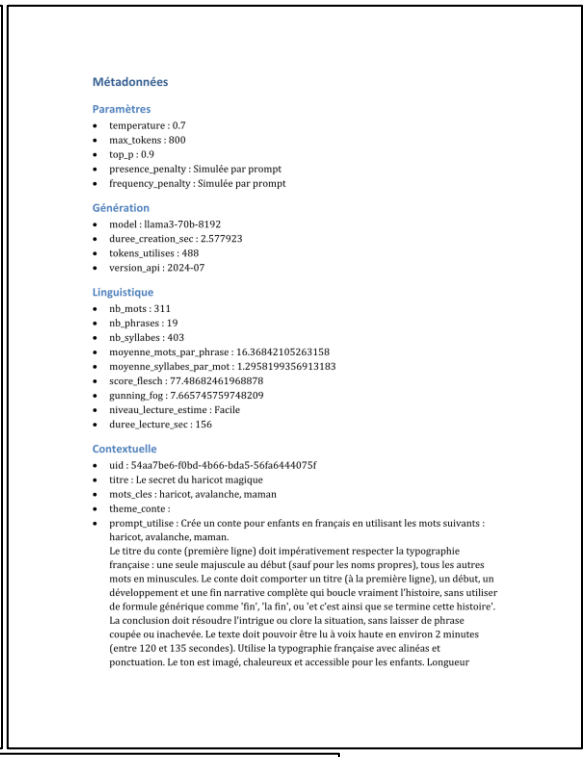
Un jour, pendant que sa maman préparait le déjeuner, Léa décida de jouer avec les légumes qui étaient sur la table. Elle prit un haricot et se mit à le faire rouler sur la table. Soudain, le haricot disparut dans une petite fissure du bois. Léa fut surprise et essaya de le récupérer, mais il était trop tard.

Juste à ce moment, une avalanche énorme descendit de la montagne et ensevelit la maison sous une couche de neige épaisse. Léa et sa maman furent surprises et ne savaient pas quoi faire. Mais Léa se souvint du haricot qui avait disparu et eut une idée.

Elle se rappela que le haricot avait disparu dans la fissure du bois et se demanda si cela ne pourrait pas être un passage secret. Elle se mit à creuser dans la neige et, après quelques minutes, elle trouva l'ouverture de la fissure. Elle appela sa maman et elles s'y glissèrent toutes les deux.

Elles se retrouvèrent dans un tunnel étroit qui les mena à l'extérieur de la maison, où la neige n'était plus si épaisse. Elles purent sortir et voir que leur village était sauvé, grâce au haricot magique qui les avait aidées à fuir.

Léa et sa maman s'étreignirent, soulagées d'être saines et sauves. Elles rentrèrent dans leur maison, où elles retrouvèrent leur déjeuner chaud, prêt à être mangé. Léa comprit que parfois, les choses les plus petites peuvent faire la plus grande différence. Et depuis ce jour, elle n'oublia jamais le secret du haricot magique.



Métadonnées

Paramètres

- temperature : 0.7
- max_tokens : 800
- top_p : 0.9
- presence_penalty : Simulée par prompt
- frequency_penalty : Simulée par prompt

Génération

- model : llama3-70b-8192
- duree_creation_sec : 2.577923
- tokens_utilises : 488
- version_api : 2024-07

Linguistique

- nb_mots : 311
- nb_phrases : 19
- nb_syllabes : 403
- moyenne_mots_par_phrase : 16.36842105263158
- moyenne_syllabes_par_mot : 1.2958199356913183
- score_flesch : 77.48682461968878
- gunning_fog : 7.665745759748209
- niveau_lecture_estime : Facile
- duree_lecture_sec : 156

Contextuelle

- uid : 54aa7be6-f0bd-4b66-bda5-56fa6444075f
- titre : Le secret du haricot magique
- mots_cles : haricot, avalanche, maman
- theme_conte :
- prompt_utilise : Crée un conte pour enfants en français en utilisant les mots suivants : haricot, avalanche, maman.
  Le titre du conte (première ligne) doit impérativement respecter la typographie française : une seule majuscule au début (sauf pour les noms propres), tous les autres mots en minuscules. Le conte doit comporter un titre (à la première ligne), un début, un développement et une fin narrative complète qui boucle vraiment l'histoire, sans utiliser de formule générique comme 'fin', 'la fin', ou 'et c'est ainsi que se termine cette histoire'. La conclusion doit résoudre l'intrigue ou clore la situation, sans laisser de phrase coupée ou inachevée. Le texte doit pouvoir être lu à voix haute en environ 2 minutes (entre 120 et 135 secondes). Utilise la typographie française avec alinéas et ponctuation. Le ton est imagé, chaleureux et accessible pour les enfants. Longueur

souhaitée : environ 800 tokens. Structure claire et histoire complète. Veille à ne pas répéter inutilement les mêmes mots ou expressions, et à varier le vocabulaire tout au long de l'histoire. Chaque objet ou personnage ne doit apparaître qu'une ou deux fois, sauf nécessité pour l'intrigue, et les formulations ne doivent pas être redondantes. Le paramètre presence_penalty et frequency_penalty n'est pas disponible sur ce modèle : simule la diversité par le style. Le titre ne doit contenir ni astérisques, ni balises Markdown, ni majuscules inutiles.

- alignement_consigne : Oui
- fin_completee_par_ia : Non
- note :

Temporelle

- date_creation : 2025-07-10 15:37:02

ANNEXE F

Liste des triplets de mots utilisés pour générer les contes

mot1	mot2	mot3	nom_fichier
avocat	soleil	berceau	avocat-soleil-berceau
boîte	fusée	planche	boîte-fusée-planche
broche	bassin	algue	broche-bassin-algue
caisse	habitant	poumon	caisse-habitant-poumon
canari	perceuse	brique	canari-perceuse-brique
carrosse	chanson	autobus	carrosse-chanson-autobus
cerise	mode	semoule	cerise-mode-semoule
cheval	araignée	camionnette	cheval-araignée-camionnette
chevreau	bouchon	parcours	chevreau-bouchon-parcours
colle	galaxie	coiffeur	colle-galaxie-coiffeur
croisement	humeur	poney	croisement-humeur-poney
doudou	berceau	bracelet	doudou-berceau-bracelet
duvet	broche	carapace	duvet-broche-carapace
filet	duvet	muscle	filet-duvet-muscle
haricot	avalanche	maman	haricot-avalanche-maman
nombril	informaticien	caramel	nombril-informaticien-caramel
numéro	pépin	pion	numéro-pépin-pion
or	poumon	escalade	or-poumon-escalade
poireau	embouteillage	gardien	poireau-embouteillage-gardien
somnambule	toit	louve	somnambule-toit-louve
souliers	pièce	bouquet	souliers-pièce-bouquet
tasse	boutique	bouc	tasse-boutique-bouc
tige	bassin	aliment	tige-bassin-aliment
tuile	semoule	réfrigérateur	tuile-semoule-réfrigérateur
yaourt	côte	prune	yaourt-côte-prune

Note : Cette liste a servi à générer le corpus des 100 contes. Chaque triplet de mots-clés a été utilisé pour amorcer une production narrative indépendante

ANNEXE G

Extraits comparatifs SGT vs Composite

Titre du conte	Générateur IA	SGT (sur 7)	SGT (%)	Composite (%)	Interprétation
Le secret du poireau enchanté	meta-llama-3-70b-instruct	6	85,7%	36%	Structure complète mais incohérences décelées (composite plus faible).
La boîte aux étoiles	meta-llama-3-70b-instruct	3	43%	88%	Structure incomplète (SGT) mais récit riche (composite plus fort).
La boîte magique	mistral-large-latest	6	86%	86%	Bonne concordance entre SGT et composite.
Le secret du poireau enchanté	meta-llama-3-70b-instruct	6	86%	36%	Écart extrême entre structure et qualité perçue.

ANNEXE H

Score « maîtrise » vs score « général »

Cette annexe compare, pour chaque modèle testé, le score composite général (mesure de référence) et le score « maîtrise », recentré sur deux critères essentiels au dispositif enfant/IA : **l'intégration effective des mots-clés** et la **présence de dialogues significa**tifs. Les écarts rapportés en points de pourcentage (pp) permettent d'estimer dans quelle mesure un récit, bien noté en général, répond (ou non) aux exigences spécifiques du projet de cocréation.

Générateur IA	Score composite général (%)	Score composite maîtrise (%)	Écart (Maîtrise - Général) (pp)
Global (tous modèles)	83%	79%	-4

Générateur IA	Score composite général (%)	Score composite maîtrise (%)	Écart (Maîtrise - Général) (pp)
claude-sonnet-4-20250514	87,6%	86,3%	-1,4
gpt-4o	84,0%	80,0%	-4
mistral-large-latest	85,4%	79,2%	-6,2
meta-llama-3-70b-instruct	73,9%	69,6%	-4,4

ANNEXE I

Cas extrêmes d’écarts SFT / Composite

Type de cas	Titre du conte	Générateur IA	SGT (sur 7)	SGT (%)	Composite général (%)	Composite maîtrise (%)	Écart signé (SGT% - Composite général) (pp)	Écart absolu (pp)	Interprétation
SGT haut / Composite bas	Le secret du poireau enchanté	meta-llama-3-70b-instruct	6	85,7%	36%	22%	49,7	49,7	Structure complète selon SGT, mais incohérences pénalisées (composite plus faible).
SGT bas / Composite haut	La boîte aux étoiles	meta-llama-3-70b-instruct	3	42,9%	88%	86%	-45,1	45,1	Structure incomplète selon SGT, mais récit cohérent et riche (composite plus fort).
Concordance	La boîte magique	mistral-large-latest	6	85,7%	86%	82%	-0,3	0,3	Bonne concordance entre structure SGT et qualité composite.
Écart extrême	Le secret du poireau enchanté	meta-llama-3-70b-instruct	6	85,7%	36%	22%	49,7	49,7	Structure complète selon SGT, mais incohérences pénalisées (composite plus faible).