Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

Un outil d'aide à la rédaction technique en français avec conformité aux langues contrôlées

Hreshvik Sewraj[1*] Emilia Sadaoui[1*] Petra Dadić[1,2] Liana Ermakova[1]
(1) HCTI, Université de Bretagne Occidentale, 29200 Brest, France
(2) LabSTICC, IMT Atlantique, 29200 Brest, France
* Contribution égale
hreshvik.sewraj@univ-brest.fr, emilia.sadaoui@etudiant.univ-brest.fr,
petra.dadic@imt-atlantique.fr, liana.ermakova@univ-brest.fr

RÉSUMÉ

Un outil d'aide à la rédaction technique en français avec conformité aux langues contrôlées

Les documents techniques et scientifiques en français sont souvent difficiles d'accès en raison d'une terminologie complexe et de structures syntaxiques élaborées, tandis que les outils soutenant la rédaction en langue contrôlée pour le français restent limités. Cela crée des défis tant pour l'accessibilité que pour la documentation technique professionnelle. Nous présentons un outil de réécriture hybride qui combine un grand modèle de langage Qwen2.5 7B (12) avec des conseils basés sur des règles pour simplifier et reformuler des textes techniques français. Le système propose quatre modes : Light, Intermediate, FALC (Facile à Lire et à Comprendre) et rédaction technique, alignés sur les principes de l'ISO 82079-1. Évalué sur un corpus de 70 textes scientifiques et 9 documents techniques, le système produit une simplification progressive tout en maintenant la fidélité sémantique. En mode rédaction technique, il évalue la clarté, la cohérence terminologique et la qualité structurelle, avec une conformité à l'ISO 82079-1. Ces résultats suggèrent que les approches hybrides combinant LLM et règles peuvent soutenir à la fois la simplification orientée accessibilité et la rédaction technique contrôlée pour la documentation française.

ABSTRACT

A French Technical Writing Support Tool with Controlled-Language Compliance

French technical and scientific documents are often difficult to access due to complex terminology and syntactic structures, while tools supporting controlled-language writing in French remain limited. This creates challenges for both accessibility and professional technical documentation. We present a hybrid rewriting tool that combines a large language model Qwen2.5 7B (12) with rule-based guidance to simplify and reformulate French technical texts. The system provides four modes: Light, Intermediate, FALC (*Facile à Lire et à Comprendre*), and Technical Writing aligned with ISO 82079-1 principles. Evaluated on a corpus of 70 scientific and 9 technical documents, the system produces progressive simplification while maintaining semantic fidelity. In Technical Writing mode, it assesses clarity, terminology consistency, and structural quality, with partial ISO 82079-1 compliance. These results suggest that hybrid LLM and rule-based approaches can support both accessibility-oriented simplification and controlled technical writing for French documentation.

MOTS-CLÉS : rédaction technique, simplification de texte, langue contrôlée, LLMs, accessibilité.

KEYWORDS: technical writing, text simplification, controlled language, LLMs, accessibility.

Actes de la 21e Conférence en Recherche d'Information et Applications.
Volume 1 : articles scientifiques originaux, pages 197–208.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

1 Introduction

La rédaction technique vise à rendre les informations complexes claires, cohérentes et accessibles grâce à l'utilisation d'une grammaire contrôlée et de vocabulaires restreints. Cependant, les outils soutenant la simplification et la réécriture contrôlée des textes techniques et scientifiques français restent rares.

L'accès à l'information scientifique et technique est essentiel pour la recherche, l'éducation et la prise de décision industrielle. Pourtant, ces textes sont souvent difficiles à comprendre en raison d'une terminologie dense, de structures syntaxiques longues et de conventions spécifiques au domaine. Ces caractéristiques affectent également les tâches de recherche d'information (RI), car une formulation complexe et une variabilité lexicale peuvent réduire l'indexabilité des documents et entraver l'appariement requête-document. Améliorer la clarté textuelle et la cohérence structurelle profite donc à la fois à la lisibilité humaine et à l'efficacité de la recherche.

Les progrès récents des grands modèles de langage (LLMs) ont démontré de fortes capacités en matière de réécriture et de simplification de textes. Cependant, la plupart des approches existantes se concentrent sur l'anglais ou sur des corpus de domaine général et intègrent rarement des contraintes de langues contrôlées ou des normes de documentation professionnelle telle que l'ISO 82079-1. De plus, les métriques d'évaluation couramment utilisées en simplification de texte, comme SARI (17) et BLEU (10), ne sont pas entièrement calibrées pour le français, ce qui motive l'inclusion d'une évaluation centrée sur l'humain.

Pour répondre à ces limitations, nous présentons un outil de réécriture de textes techniques français combinant la génération neuronale à la base de Qwen (12) avec des conseils explicites basés sur des règles. Le système prend en charge une simplification à plusieurs niveaux pour les publics non experts (Light, Intermediate, et FALC – *Facile à Lire et à Comprendre*) ainsi qu'un mode dédié *Rédaction Technique* pour la documentation professionnelle. Dans le mode de la rédaction technique, l'outil soit reformule le texte selon les principes des langues contrôlées, favorisant la clarté, la cohérence terminologique et la réduction de l'ambiguïté dans la documentation critique pour la sécurité, les manuels techniques et les supports pédagogiques, soit vérifie si le texte donné est conforme à ces normes.

L'outil fournit un pipeline complet : les utilisateurs peuvent saisir du texte directement ou télécharger des fichiers `.txt` ou `.pdf`, qui sont automatiquement extraits, concaténés et rendus éditables avant la réécriture. Le système intègre une vérification grammaticale et des fonctionnalités d'exportation, permettant son intégration dans des flux de travail de rédaction technique.

Les contributions de ce travail sont :

- une architecture hybride combinant LLM et règles pour la réécriture et l'évaluation de textes techniques français selon les principes des langues contrôlées (ISO 82079-1);
- une simplification multi-niveaux alignée sur les directives d'accessibilité, y compris une réécriture orientée FALC ;
- une évaluation expérimentale sur des corpus scientifiques et techniques évaluant la qualité de la simplification, la fidélité sémantique et la conformité aux langues contrôlées.

En combinant la simplification orientée vers l'accessibilité avec le soutien à la rédaction technique professionnelle, ce système répond au manque d'outils dédiés pour le français et explore comment les approches hybrides LLM–règles peuvent améliorer à la fois la lisibilité et la qualité structurelle

des documents techniques.

2 Contexte et travaux connexes

La simplification de texte s'est traditionnellement appuyée sur des règles artisanales ou la traduction automatique statistique, les premiers systèmes se concentrant sur des caractéristiques de surface telles que la division des phrases et la substitution lexicale (14). Des modèles neuronaux comme T5 (13) et BART (9) ont permis une réécriture monolingue plus flexible, mais la plupart des travaux restent centrés sur l'anglais, avec une adaptation limitée au français. Les ressources françaises, comme le FALC (Facile à Lire et à Comprendre), fournissent des lignes directrices pour une écriture accessible mais manquent de corpus volumineux pour l'entraînement ou l'évaluation. Des ensembles de données anglais comme WikiLarge (18) et Newsela (16) existent, mais leur utilisation en few-shot ou zero-shot avec des modèles multilingues est encore limitée. Les métriques d'évaluation standard telles que SARI (17) et BLEU (11) ne sont pas calibrées pour le français, soulignant la nécessité d'une analyse d'erreurs basée sur l'humain.

De nombreuses organisations ciblant des publics internationaux adoptent un langage clair, par exemple « plain English » ou « lenguaje claro », mais pour la documentation critique pour la sécurité, la norme ISO 82079-1 est préférée. L'ISO 82079-1 (iso) est une norme internationale qui encadre la rédaction de la documentation d'utilisation des produits (*information for use*), en visant la clarté, la cohérence terminologique, la lisibilité et l'objectivité, via des modalités normatives (*shall/should/may*). Dans ce travail, nous nous limitons à six critères opérationnels dérivés de cette norme (Tableau 1) ; l'élargissement à d'autres dimensions constitue une perspective de travaux futurs. Les langues contrôlées restreignent la grammaire et le vocabulaire pour réduire l'ambiguité et la complexité (5), et incluent des règles d'écriture et des exigences de cohérence terminologique. Des outils existants comme le Boeing Simplified English Checker (BSEC) (boe), HyperSTE (6) et le parseur espagnol simplifié (5) soutiennent l'anglais et l'espagnol, mais des analyseurs validés pour le français font défaut. Inspirés par TechScribe (4), nous explorons une approche hybride utilisant des conseils basés sur des règles combinés à une vérification grammaticale via LanguageTool (3) pour les textes techniques français.

3 Présentation du système

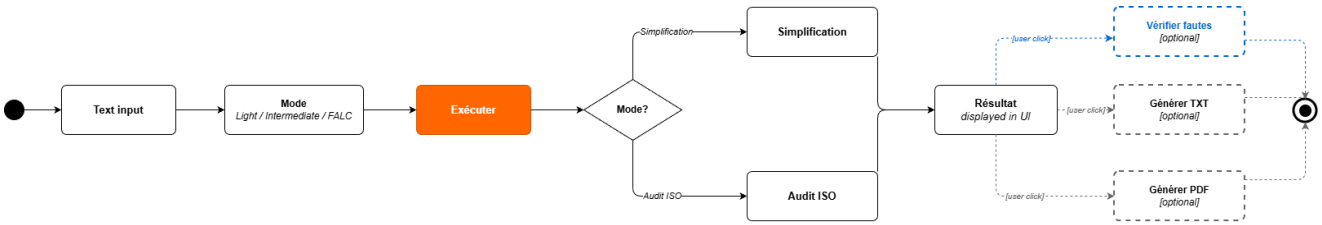

Figure 1: Flux de travail du système. Les flèches pleines indiquent le chemin de traitement principal ; les flèches en pointillés indiquent des actions facultatives déclenchées par l'utilisateur.

Notre outil fournit un pipeline complet pour simplifier, réécrire et évaluer des textes techniques français tout en préservant le sens et en améliorant la lisibilité et la qualité professionnelle (Fig. 1).

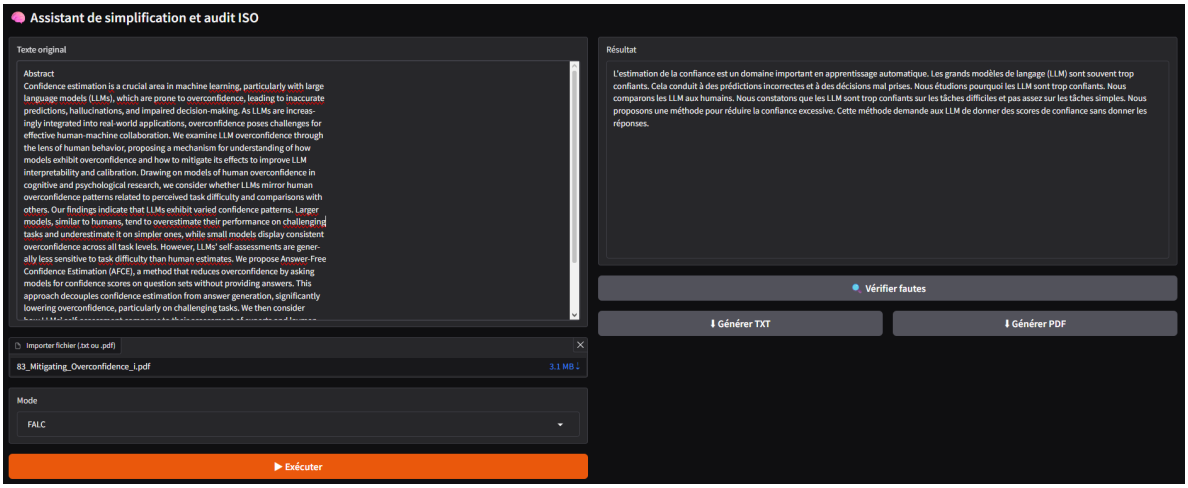

Figure 2: Interface de l'application. À gauche : saisie de texte et sélection du mode. À droite : panneau de sortie avec actions de post-traitement optionnelles.

Les utilisateurs peuvent saisir du texte directement ou télécharger des fichiers `.txt` ou `.pdf`. Le système extrait et concatène le contenu, fournissant une interface entièrement éditable avant la simplification.

3.1 Détails d'implémentation

3.1.1 Mode Simplification

Réécriture à niveaux contrôlés. Le système implémente trois modes de réécriture distincts (Light, Intermediate, FALC), chacun guidé par un prompt et une configuration de règles dédiées. Les trois premiers modes appliquent des stratégies de simplification graduées ciblant les publics non experts :

- **Light :** Ajustements mineurs pour améliorer la lisibilité tout en préservant la structure et la terminologie d'origine.
- **Intermediate :** Simplifie le vocabulaire et restructure les phrases en 10 à 15 mots avec une idée par phrase.
- **FALC :** Produit des phrases très courtes et explicites pour une accessibilité maximale.

Exemple (FALC) — Entrée : « Les interactions entre les variables phénotypiques et génotypiques conditionnent l'expression différentielle des protéines membranaires. » *Sortie :* « Les gènes influencent les protéines. Les protéines se trouvent sur la membrane cellulaire. Elles ne sont pas toutes activées de la même façon. »

3.1.2 Mode Rédaction Technique

Le mode Rédaction Technique cible les documents professionnels. Il opère en deux sous-modes distincts : (1) l'**audit ISO**, qui évalue la conformité du texte soumis aux six critères de l'ISO 82079-1

et produit un rapport structuré identifiant les manquements ; (2) la **réécriture contrôlée**, qui reformule les passages non conformes selon les mêmes critères. Ces deux sous-modes correspondent aux deux prompts reproduits en Annexe B : le prompt d'audit évalue un critère à la fois (six passes séparées) ; le prompt de reformulation reçoit le rapport d'audit en entrée et produit la version réécrite. L'outil :

- Évalue la conformité à la norme ISO 82079-1 selon six critères : Clarté & précision, Structure normative, Formulation des exigences, Objectivité, Cohérence et Lisibilité.
- Produit un rapport de conformité structuré identifiant les passages non conformes et les critères associés.
- Reformule les passages non conformes selon les principes du langage contrôlé, norme ISO 82079-1.
- Applique une terminologie cohérente, la voix active et un formatage structuré des instructions.

Exemple (Rédaction Technique) — Entrée : « Il convient de s'assurer que le dispositif est correctement éteint avant toute intervention de maintenance. » *Sortie (ISO 82079-1) :* « Éteindre le dispositif avant toute opération de maintenance. Contrôler que le voyant d'alimentation est éteint. »

Le Tableau 1 détaille les six critères d'évaluation et leur définition opérationnelle, tels qu'ils sont instanciés dans les prompts du pipeline d'audit.

Table 1: Critères ISO 82079-1 et définitions opérationnelles utilisés dans le pipeline d'audit.

Critère	**Définition opérationnelle**
Clarté & précision	Absence de formulations ambiguës, de termes imprécis et de doubles sens.
Structure normative	Présence d'une hiérarchie explicite, de sections identifiables et d'un ordre adapté au contexte normatif.
Formulation des exigences	Usage correct des modalités normatives (*shall / should / may*) ; exigences vérifiables et non vagues.
Objectivité	Absence de formulations subjectives, de jugements de valeur et de biais rédactionnels.
Cohérence	Absence de contradictions internes, de terminologie incohérente et de répétitions contradictoires.
Lisibilité	Phrases de longueur maîtrisée, vocabulaire accessible et mise en forme adaptée au public cible.

Ce mode met en œuvre un pipeline en deux étapes où l'évaluation de la conformité précède la reformulation : le système identifie d'abord les passages ne satisfaisant pas aux critères ISO 82079-1, puis procède à une réécriture en langage contrôlé afin de corriger les manquements détectés. Cette architecture unifiée établit que la simplification et la conformité normative reposent sur les mêmes principes fondamentaux de clarté, de précision et de cohérence terminologique.

L'outil est implémenté en Python utilisant Qwen2.5-7B (12) accessible via Hugging Face Transformers pour la génération basée sur des prompts. Le choix de Qwen2.5-7B est motivé par sa déployabilité locale, qui garantit la confidentialité des documents industriels, ses performances documentées sur le français parmi les modèles de taille comparable (12), et les contraintes matérielles du cadre

académique. Une comparaison avec un modèle de plus grande taille est laissée en perspective. De plus, les modèles de petite taille tendent à réduire la variance des hallucinations dans les tâches de réécriture contrainte.

Les prompts et les règles de simplification sont stockés dans un fichier de configuration dédié `rules.txt`. Ce fichier contient 14 règles textuelles inspirées de l'ISO 82079-1, couvrant notamment : l'usage de mots simples et courants, la limitation des phrases à une idée, la voix active, la suppression des mots inutiles, la cohérence terminologique (un terme par concept), et l'évitement des nominalisations (voir Annexe A). Ces règles sont injectées directement dans le prompt de chaque mode sous forme d'instructions textuelles.

Le pipeline de simplification repose sur du *few-shot prompting* : le prompt inclut une instruction système, les 14 règles, deux exemples travaillés par mode — rédigés manuellement par les auteurs à partir de textes hors corpus et validés conjointement pour s'assurer qu'ils illustrent fidèlement le niveau de simplification visé — puis le texte à traiter. Le prompt varie selon le mode sélectionné (Light, Intermediate, FALC) ; son contenu complet est reproduit en Annexe A. Le mode Rédaction Technique repose sur un pipeline en deux étapes : (i) un prompt d'audit évalue le texte selon six critères ISO 82079-1 dans des passes d'inférence séparées (une passe par critère, afin de limiter les interférences entre dimensions) ; (ii) un second prompt distinct reçoit le rapport d'audit et produit la version reformulée. Les six prompts d'audit sont reproduits en Annexe B.

L'extraction de texte prend en charge les fichiers `.txt` via le décodage UTF-8 et les fichiers `.pdf` via PyPDF2. Les sorties du modèle sont post-traitées pour supprimer les lignes artefactuelles (répétitions partielles du prompt, tokens parasites en fin de séquence et phrases incomplètes produites à la limite du budget de tokens) et tronquées à un token `<END>` désigné.

Le pipeline de traitement, illustré dans la Figure 1, suit une séquence linéaire avec une étape de post-traitement optionnelle. L'utilisateur fournit une entrée textuelle par saisie directe ou en téléchargeant un fichier `.txt` ou `.pdf`, puis sélectionne un mode de fonctionnement. Lors de l'exécution, le système achemine la requête vers l'une des deux branches : Simplification (Light, Intermediate ou FALC) ou Audit ISO. Les deux branches produisent un résultat affiché dans l'interface (Figure 2). Trois actions optionnelles sont ensuite disponibles à la demande de l'utilisateur : vérification grammaticale et orthographique via LanguageTool (3), et export au format `.txt` ou `.pdf`.

L'interface basée sur Gradio expose ces étapes directement. Le panneau gauche fournit une zone de texte éditable pré-remplie avec le contenu du document extrait et un sélecteur de mode ; le bouton *Exécuter* déclenche le traitement. Le panneau droit affiche la sortie réécrite ou auditée à côté des actions de post-traitement optionnelles, visuellement distinctes du flux de travail principal pour refléter leur caractère non obligatoire.

4 Protocole expérimental

4.1 Simplification

Nous avons soumis 70 textes scientifiques français couvrant la biotechnologie, les neurosciences, l'écologie et la science des matériaux au système dans chacun des trois modes (Light, Intermediate, FALC), produisant 210 sorties au total. Chaque texte a été traité via du few-shot prompting : le prompt incluait des définitions explicites du mode et deux exemples travaillés par mode, rédigés

manuellement par les auteurs à partir de textes hors corpus et validés conjointement pour s'assurer qu'ils illustrent fidèlement le niveau de simplification visé. Le modèle a été interrogé une fois par mode et par texte sans aucun filtrage a posteriori.

Les sorties ont été évaluées selon quatre dimensions à l'aide d'un pipeline d'évaluation automatique. Les métriques de fidélité lexicale (BLEU, ROUGE, Jaccard) mesurent la préservation du contenu informatif par rapport à la source, distincte de la qualité de simplification lexicale ; elles ne capturent pas la fluidité de la réécriture et doivent être interprétées comme indicateurs de conservation sémantique uniquement. Les métriques **structurelles** comprennent le taux de division (ratio du nombre de phrases en sortie par rapport à la source, où les phrases sont délimitées par la ponctuation) et le nombre moyen de mots par phrase (W/S). La **lisibilité** est mesurée par l'indice Flesch-Kincaid Grade Level (FKG), largement utilisé en simplification pour son interprétabilité et son applicabilité sans corpus de référence :

$$\text{FKG (8)} = 0.39\,\frac{W}{S} + 11.8\,\frac{\text{Syl}}{W} - 15.59$$

et le score Flesch Reading Ease dans l'adaptation Kandel-Moles pour le français :

$$\text{FRE (7)} = 207 - 1.015\,\frac{W}{S} - 73.6\,\frac{\text{Syl}}{W}$$

où W est le nombre de tokens, S le nombre de phrases, et Syl le nombre total de syllabes estimé via une heuristique de groupes de voyelles, en utilisant l'implémentation fournie par la bibliothèque `textstat` ; cette approximation est une limite connue, et les valeurs absolues doivent être interprétées comme des tendances relatives plutôt que comme des mesures certifiées.

4.2 Rédaction Technique

Un sous-ensemble distinct de 9 textes instructifs et procéduraux, issus de manuels de procédures industrielles, de spécifications de produits et d'extraits normatifs ISO, a été réservé exclusivement pour l'évaluation du mode Rédaction Technique. Ce corpus a été sélectionné car la conformité aux langues contrôlées est principalement significative dans des contextes instructifs et normatifs ; les résumés scientifiques, bien que syntaxiquement complexes, ne constituent pas le champ d'application naturel de l'ISO 82079-1. Tous les documents ont été vérifiés comme étant exempts de contenu personnellement identifiable ou commercialement sensible avant utilisation.

Chaque texte a été soumis au pipeline d'audit ISO, chaque critère étant évalué dans une passe d'inférence séparée au moyen d'un prompt dédié, afin de limiter les interférences entre dimensions lors de la notation, produisant 54 évaluations au niveau des critères (9 textes $\times$ 6 critères). Pour évaluer la fiabilité de la notation, une étudiante en master Rédaction Technique a annoté les 9 textes selon les mêmes six critères de l'ISO 82079-1 en utilisant la même échelle à trois niveaux (*High*, *Partial*, *Low*), servant de référence experte humaine. La notation humaine a été réalisée de manière indépendante des scores du modèle, en s'appuyant sur les définitions officielles des six critères telles que formulées dans les prompts d'évaluation ISO 82079-1 utilisés pour le pipeline automatique, constituant ainsi le guide de cotation commun aux deux annotateurs. L'accord modèle-humain a été mesuré via le taux d'accord exact, l'exactitude (accuracy), et l'erreur absolue moyenne (MAE), ainsi que le biais moyen (différence signée entre le score LLM et le score humain). Ces métriques sont définies comme suit :

- **Accord exact** : nombre d'évaluations pour lesquelles le score attribué par le modèle est identique à l'annotation humaine, rapporté au nombre total d'évaluations.
- **Exactitude** : proportion d'évaluations en accord exact, exprimée en pourcentage.
- **Biais moyen** : différence moyenne signée $\bar{s}^{\text{LLM}} - \bar{s}^{\text{human}}$; une valeur négative indique une sous-estimation systématique par le modèle.

L'erreur absolue moyenne (MAE) est définie comme suit :

$$\text{MAE (15)} = \frac{1}{N} \sum_{i=1}^{N} \left| s_i^{\text{LLM}} - s_i^{\text{human}} \right| \tag{1}$$

où N est le nombre total d'évaluations au niveau des critères, s_i^{LLM} est le score attribué par le modèle, et s_i^{human} est l'annotation humaine correspondante, tous deux sur l'échelle à trois niveaux $\{1, 2, 3\}$. Étant donné l'absence de multiples annotateurs, la fiabilité inter-annotateurs n'a pas pu être estimée ; ces résultats doivent donc être interprétés comme une analyse descriptive du comportement de notation du système par rapport à une seule référence experte plutôt que comme une évaluation de conformité pleinement validée.

5 Résultats

5.1 Simplification

Les Tableaux 2 et 3 rapportent les résultats de l'évaluation pour les trois niveaux de simplification. Toutes les métriques de surface suivent le gradient progressif attendu. Le nombre moyen de mots par phrase a diminué de 19,5 (source) à 11,4 (Light) et 8,5 (FALC), avec des taux de division supérieurs à 1,0 à tous les niveaux, confirmant que le modèle segmente plutôt que de tronquer les phrases sources. La lisibilité s'est améliorée de manière monotone de FRE 17,15 à 39,27, bien que tous les niveaux soient restés dans la bande *Difficile* (FRE < 50), reflétant la densité technique du corpus.

Table 2: Métriques de simplification consolidées par niveau.

Niveau	W/S	Split	FKG	FRE	LWR (%)
Original	19.50	—	19.28	17.15	51.8
Light	11.36	1.151	15.55	28.91	49.0
Intermediate	9.13	1.332	14.27	33.72	47.1
FALC	8.49	1.308	13.23	39.27	44.6

W/S : moy. mots/phrase ; Split : taux de division ; FKG : Flesch-Kincaid Grade Level (plus bas = plus facile) ; FRE : Flesch Reading Ease (Kandel-Moles) ; LWR : ratio mots longs ($\geq$7 car.)

Les métriques de fidélité (Tableau 3) diminuent monotoniquement de Light à FALC, conformément à la divergence structurelle croissante par rapport à la source. Des scores ROUGE-1 modérés (0,44–0,59) indiquent que les mots de contenu clés sont conservés tandis que la syntaxe de surface est activement reformulée, une propriété souhaitable pour un système de simplification graduée.

Table 3: Métriques de fidélité (BLEU, ROUGE, Jaccard) calculées par rapport au texte source.

Niveau	BLEU	ROUGE-1	ROUGE-2	ROUGE-L	Jaccard
Light	0.3802	0.5927	0.3480	0.5527	0.4289
Intermediate	0.2887	0.5023	0.2505	0.4588	0.3439
FALC	0.2224	0.4396	0.1837	0.3920	0.2939

5.2 Rédaction Technique

Table 4: Scores de conformité ISO 82079-1 par critère ($n = 9$ textes uniques). H = High (3), P = Partial (2), L = Low (1).

Critère	H n (%)	P n (%)	L n (%)	Moyenne (/3)
Clarté & précision	2 (22%)	7 (78%)	0 (0%)	2.22
Structure normative	0 (0%)	9 (100%)	0 (0%)	2.00
Formulation des exigences	0 (0%)	8 (89%)	1 (11%)	1.89
Objectivité	4 (44%)	5 (56%)	0 (0%)	2.44
Cohérence	8 (89%)	1 (11%)	0 (0%)	2.89
Lisibilité	0 (0%)	9 (100%)	0 (0%)	2.00
Global	**14 (26%)**	**39 (72%)**	**1 (2%)**	**2.24**

Le Tableau 4 présente la distribution des scores de conformité produits par Qwen2.5-7B (12) sur 54 évaluations (9 textes $\times$ 6 critères). La distribution est inégale : le critère Cohérence reçoit majoritairement des notes *High*, tandis que les cinq autres critères reçoivent presque exclusivement des notes *Partial*. Ce contraste indique que le modèle ne se comporte pas de la même façon selon la dimension évaluée.

Cette asymétrie s'explique en partie par la nature de l'entraînement : les modèles autorégressifs sont optimisés pour produire des textes cohérents, ce qui les rend naturellement plus performants sur ce critère. En revanche, la quasi-absence de variation sur la Structure Normative, la Lisibilité et la Formulation des Exigences (aucun texte noté *High* ou *Low* sur les deux premiers critères) suggère que le modèle applique une politique de notation fixe sur ces dimensions, sans distinguer les textes entre eux. L'origine de ce comportement (invariance réelle du corpus, prudence face à l'incertitude, ou manque de connaissances sur les normes du domaine) ne peut être déterminée à partir des seules distributions observées.

Le Tableau 5 compare les notes du modèle à celles d'un expert humain sur cinq métriques d'accord. Trois résultats ressortent : (*i*) le modèle attribue systématiquement des notes inférieures à l'expert sur l'ensemble des critères ; (*ii*) cet écart est plus prononcé sur la Formulation des Exigences que sur la Cohérence, ce qui indique que la sous-estimation est concentrée sur les critères les plus complexes sur le plan normatif ; (*iii*) l'accord exact atteint 89% sur la Cohérence contre 0 à 22% sur les autres critères, confirmant que la qualité de l'alignement modèle-humain dépend fortement du critère considéré.

Ces résultats peuvent s'interpréter selon deux hypothèses concurrentes. L'*hypothèse de calibrage* suppose que le modèle applique des standards plus stricts que l'expert : là où ce dernier accepte une conformité partielle, le modèle la pénalise. L'*hypothèse de connaissance* suppose plutôt que le modèle manque des repères normatifs nécessaires pour attribuer une note *High* sur les critères prescriptifs, et se replie par défaut sur *Partial*. La seconde hypothèse est mieux soutenue par les données, car l'écart varie selon les critères, ce que la première hypothèse ne prédit pas. La taille

Table 5: Comparaison entre les annotations humaines et Qwen2.5-7B (12) selon les critères ISO 82079-1 ($n = 9$ textes). Les annotations humaines sont uniformément notées High (3).

Critère	Humain	LLM	Accord exact	Préc.	MAE (pts)	Biais
Clarté & précision	3.00	2.22	2/9	22%	0.78	−0.78
Structure normative	3.00	2.00	0/9	0%	1.00	−1.00
Formulation des exigences	3.00	1.89	0/9	0%	1.11	−1.11
Objectivité	3.00	2.44	4/9	44%	0.56	−0.56
Cohérence	3.00	2.89	8/9	89%	0.11	−0.11
Lisibilité	3.00	2.00	0/9	0%	1.00	−1.00
Global	**3.00**	**2.24**	**14/54**	**26%**	**0.76**	**−0.76**

Accord exact = nombre de textes où le score LLM est égal au score humain. Préc. = précision (exactitude). MAE = erreur absolue moyenne. Biais = moyenne LLM − moyenne humain ; les valeurs négatives indiquent une sous-estimation systématique.

réduite du corpus (n=9) ne permet toutefois pas de conclusion définitive ; trancher entre ces deux explications nécessiterait un corpus plus large, plusieurs annotateurs, et des expériences ciblées sur le rôle des instructions.

6 Conclusion

Cette étude a présenté une architecture hybride combinant LLM et règles pour la réécriture graduée de textes techniques français selon quatre modes, répondant à un manque documenté d'outils en langue contrôlée pour le français. Évalué sur 70 documents scientifiques, le système produit une simplification progressive systématique (longueur moyenne de phrase : 19,5 à 8,5 mots ; FRE : 17,15 à 39,27), bien que l'ensemble des niveaux reste dans la bande de lisibilité Difficile (FRE < 50). La principale limitation inclut l'absence de jugements humains de fluidité. Les travaux futurs devraient prioriser le décodage contraint pour cibler des plages de lisibilité bornées et l'évaluation humaine de l'adéquation de la simplification.

Sur les 9 documents techniques, le système atteint une conformité partielle à l'ISO 82079-1, avec un accord modèle-humain élevé sur la Cohérence (89% d'accord exact) mais une sous-estimation systématique sur les critères normativement complexes, notamment la Formulation des Exigences. Ce profil est cohérent avec un déficit de connaissance normative plutôt qu'avec un biais de calibrage global, dans la mesure où l'écart varie selon les critères plutôt que de rester uniforme. Les résultats du tableau 5 ont une valeur descriptive, reposant sur une seule annotatrice, et ne permettent pas de généralisation ; l'annotation multi-annotateurs avec estimation de la fiabilité inter-annotateurs constitue la priorité principale des travaux futurs.

Remerciements

Cette recherche a été financée en tout ou partie, par l'Agence Nationale de la Recherche (ANR) au titre du projet ANR-22-CE23-0019-01.

References

[iso] IEC/IEEE CD 82079-1.

[boe] Simplified English Checker.

[3] (2025). Languagetool – grammar, style, and spell checker. https://languagetool.org/. Open source tool for multilingual grammar checking.

[4] (2025). Techscribe project. https://www.techscribe.co.uk/. Information about controlled language and technical writing.

[5] CASCALES R. & SUTCLIFFE R. (2020). A specification and validating parser for simplified technical spanish. In *Proceedings of the Conference on Technical Writing*.

[6] COMMUNICATIONS S. (2023). Hyperste – simplified technical english checker. https://www.smartcommunications.com/hyperste. Accessed: 2025-10-18.

[7] KANDEL L. & MOLES A. (1958). Application de l'indice de Flesch à la langue française. *Cahiers Études de Radio-Télévision*, **19**, 253–274.

[8] KINCAID J. P., FISHBURNE R. P., ROGERS R. L. & CHISSOM B. S. (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel*. Rapport interne Research Branch Report 8-75, Naval Air Station Memphis.

[9] LEWIS M., LIU Y., GOYAL N., GHAZVININEJAD M., MOHAMED A., LEVY O., STOYANOV V. & ZETTLEMOYER L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.

[10] PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2002a). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, p. 311–318: Association for Computational Linguistics.

[11] PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2002b). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, p. 311–318, USA: Association for Computational Linguistics. DOI : 10.3115/1073083.1073135.

[12] QWEN, :, YANG A., YANG B., ZHANG B., HUI B., ZHENG B., YU B., LI C., LIU D., HUANG F., WEI H., LIN H., YANG J., TU J., ZHANG J., YANG J., YANG J., ZHOU J., LIN J., DANG K., LU K., BAO K., YANG K., YU L., LI M., XUE M., ZHANG P., ZHU Q., MEN R., LIN R., LI T., TANG T., XIA T., REN X., REN X., FAN Y., SU Y., ZHANG Y., WAN Y., LIU Y., CUI Z., ZHANG Z. & QIU Z. (2025). Qwen2.5 technical report.

[13] RAFFEL C., SHAZEER N., ROBERTS A., LEE K., NARANG S., MATENA M., ZHOU Y., LI W. & LIU P. J. (2023). Exploring the limits of transfer learning with a unified text-to-text transformer.

[14] SIDDHARTHAN A. (2006). Syntactic simplification and text cohesion. *Research on Language and Computation*, **4**(1), 77–109. DOI : 10.1007/s11168-006-9011-1.

[15] WILLMOTT C. & MATSUURA K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, **30**, 79–82.

[16] XU W., CALLISON-BURCH C. & NAPOLES C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, **3**, 283–297. DOI : 10.1162/tacl_a_00139.

[17] XU W., CALLISON-BURCH C. & NAPOLES C. (2016). Optimizing statistical machine translation for text simplification. In *Proceedings of the Transactions of the Association for Computational Linguistics*, volume 4, p. 401–415.

[18] ZHANG X. & LAPATA M. (2017). Sentence simplification with deep reinforcement learning. In M. PALMER, R. HWA & S. RIEDEL, Éds., *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, p. 584–594, Copenhagen, Denmark: Association for Computational Linguistics. DOI : 10.18653/v1/D17-1062.