

De l'espace appris à l'espace fondé : un cadre MANHAJ pour la mesure du sens en IA

Article de recherche scientifique — Papier 04

Auteur : Faouzi Messeoud — فوزي مسعود

Méthodologie : MANHAJ — Méthodologie de modélisation intellectuelle

Version : 1.0 — octobre 2026

Statut : Article de recherche scientifique

À propos de cet article

Cet article applique trois lois de la Méthodologie de modélisation intellectuelle (MANHAJ) — la Loi de la dualité de la direction de l'examen, la Loi de l'origine coordonnée et la Loi de la justesse — à la mesure du sens en intelligence artificielle. Il établit les conditions dans lesquelles un cadre de mesure du sens est fondé plutôt qu'implicite, formalise ces conditions sur le plan computationnel et précise les dispositifs expérimentaux par lesquels leurs conséquences peuvent être testées.

L'article s'inscrit dans les travaux de recherche de la méthodologie MANHAJ. Il prolonge le Papier 02, qui a établi une étape a priori pour la production du sens en IA et identifié la représentation opérationnelle des états du sens comme un problème ouvert, et il s'appuie sur le Papier 03, qui a examiné la relation entre le point d'origine et le sujet dans les sciences humaines.

Relation avec le document de référence MANHAJ

Cet article développe une dimension de recherche spécifique de la méthodologie présentée dans : Messeoud, Faouzi. *Intellectual Modeling Methodology (MANHAJ) — Reference Document — Version 1.0*. Zenodo, 15 septembre 2026. DOI : 10.5281/zenodo.22769627. Le présent article est une étude scientifique indépendante et ne doit pas être considéré comme un substitut au document de référence.

Citer cet article

Messeoud, Faouzi. *De l'espace appris à l'espace fondé : un cadre MANHAJ pour la mesure du sens en IA*. Article de recherche scientifique, Papier 04, Version 1.0, 2026.

L'article scientifique commence à la page suivante.

Résumé

En intelligence artificielle, le sens est mesuré au moyen de valeurs — scores de similarité, distances, indices de dérive, notes d'évaluation — dont la référence est rarement examinée comme condition de leur interprétation. La référence par rapport à laquelle une telle valeur est lue est souvent implicite dans la représentation apprise, ou dérivée des sorties mêmes que l'on mesure. Une valeur obtenue dans ces conditions peut être calculée avec une précision totale alors que ses conditions d'interprétation demeurent non fondées.

Cet article traite le problème à sa source par trois lois de la Méthodologie de modélisation intellectuelle (MANHAJ). La **Loi de la dualité de la direction de l'examen** exige que le cadre de mesure soit fondé avant la mesure, selon l'ordre centralité doctrinale → domaine conceptuel → origine coordonnée → coordonnées du sens. La **Loi de l'origine coordonnée** exige que la mesure soit rapportée à une origine explicite, indépendante de la propriété mesurée. La **Loi de la justesse** exige qu'un cadre non admissible ne produise aucune mesure valide. Ensemble, ces lois remplacent une mesure dont les conditions sont implicites par une mesure dont les conditions sont fondées, déclarées et examinées.

L'article traduit ces exigences en un modèle formel d'*espace fondé* : un cadre construit à partir de formulations de référence et de pôles déclarées, lu au moyen d'un encodeur appris, et soumis à un test d'admissibilité à deux niveaux qui distingue l'admissibilité méthodologique de l'opérabilité computationnelle. Deux résultats formels en découlent. Les transitions entre états du sens sont invariantes par changement d'origine, alors que les déviations par rapport à la référence ne le sont pas ; et une référence dérivée des sorties mesurées est structurellement aveugle à la composante commune de leur dérive, quelle qu'en soit l'ampleur. L'article établit en outre, comme conséquence méthodologique de MANHAJ, que toute représentation apprise opère sous une centralité doctrinale incorporée, et il fournit une procédure relationnelle pour caractériser le contenu de cette centralité.

Quatre dispositifs expérimentaux précisent comment chaque capacité peut être testée empiriquement, chacun avec une référence de comparaison forte, une vérité de terrain humaine indépendante et une ablation qui isole la contribution des exigences de MANHAJ de celle de la technique de mise en œuvre ; leur protocole opérationnel figure en annexe.

Mots-clés : MANHAJ ; intelligence artificielle ; mesure du sens ; centralité doctrinale ; domaine conceptuel ; origine coordonnée ; Loi de la justesse ; examen a priori ; représentations apprises ; évaluation ; dérive.

1. Introduction

1.1. Le problème dans les termes de l'IA

Les systèmes d'IA contemporains représentent le langage par des vecteurs dans des espaces appris, et une grande partie de ce qui est dit du sens dans ces systèmes l'est au moyen de quantités calculées dans ces espaces : la similarité de deux sorties, la dérive d'un modèle d'une version à l'autre, la note qu'un évaluateur attribue à une réponse. Ces quantités sont calculées avec précision. Ce qui est rarement examiné, c'est la référence par rapport à laquelle elles sont lues.

Trois problèmes reconnus de la pratique actuelle de l'IA partagent cette structure.

- **L'évaluation par des modèles.** Les modèles de langage sont de plus en plus utilisés pour juger les sorties d'autres modèles de langage. Lorsque la référence de jugement est dérivée de systèmes proches de ceux qui sont jugés, l'évaluation peut favoriser ce à quoi elle ressemble

déjà ; les effets documentés d’auto-préférence illustrent ce risque (Zheng et al., 2023 ; Panickssery et al., 2024).

- **La dérive collective.** Des modèles entraînés sur des données générées peuvent dériver collectivement, les queues de la distribution d’origine disparaissant progressivement (Shumailov et al., 2024). Lorsque la dérive est suivie d’une version à l’autre, ou par rapport au centroïde des sorties actuelles, un déplacement partagé par toutes les sorties peut passer inaperçu.
- **Les références compromises.** Des résultats d’évaluation sont calculés et publiés sur des références dont l’intégrité n’est pas établie, comme dans la contamination des jeux d’évaluation par les données d’entraînement (Sainz et al., 2023). La métrique est calculée correctement ; son fondement n’est pas examiné.

Dans chaque cas, une valeur est calculée avec précision par rapport à une référence implicite, dérivée du matériau mesuré ou compromise. La valeur est correcte sur le plan computationnel et indéterminée sur le plan méthodologique.

1.2. La lacune laissée par le Papier 02

Le Papier 02 de la méthodologie MANHAJ a établi une étape a priori pour le processus de production du sens en IA, $A_0 = (R, D, Def, C_a, O)$, et a suivi les états du sens et leurs transitions ΔM_t par rapport à elle. Il a explicitement identifié la représentation opérationnelle des états du sens comme une limite (Papier 02, section 12.2) : ΔM_t était défini formellement, mais l’espace dans lequel M_t est mesuré — son origine, ses axes et son étalonnage — restait non spécifié. Dans le langage des coordonnées, le Papier 02 suivait un changement *relatif* sans cadre *absolu* déclaré.

Le présent article fournit ce cadre et, ce faisant, fonde la mesure sur laquelle repose le Papier 02.

1.3. Question centrale et thèse

Fonder a priori le système de coordonnées du sens — sur une centralité doctrinale reconnue, un domaine conceptuel et une origine déclarée — change-t-il les conditions de la mesure du sens dans les systèmes d’IA, et quelles capacités, que la mesure à référence implicite ne fournit pas, découlent de ce changement ?

La thèse de l’article est que oui, et que ce changement est produit par trois lois de MANHAJ : la Loi de la dualité de la direction de l’examen (Loi 2), la Loi de l’origine coordonnée (Loi 4) et la Loi de la justesse (Loi 3). Les techniques de représentation utilisées pour mettre en œuvre ces lois sont établies dans la recherche en IA. La contribution réside dans les conditions que ces lois imposent à leur usage et dans les capacités que ces conditions produisent.

1.4. Contributions

- **C1 — La mesure fondée.** Une distinction entre mesure du sens à référence implicite et mesure fondée, dérivée des Lois 2 et 4 et de l’ordre fondationnel de MANHAJ.
- **C2 — Un résultat formel sur les références auto-dérivées.** Une démonstration qu’une référence dérivée des sorties mesurées est structurellement aveugle à la composante commune de leur dérive (Proposition 2).
- **C3 — L’invalidité fondationnelle.** Un garde-fou fondationnel dérivé de la Loi 3, qui introduit dans la mesure en IA la catégorie d’un cadre dont les mesures sont invalides à leur source, avec une séparation stricte entre admissibilité méthodologique et opérabilité computationnelle.

- **C4 — La centralité incorporée.** L'établissement, comme conséquence méthodologique de MANHAJ, que toute représentation apprise opère sous une centralité doctrinale incorporée, et un indice relationnel qui rend son orientation mesurable.
- **C5 — Des dispositifs expérimentaux.** Quatre dispositifs expérimentaux dotés de références de comparaison fortes, d'une vérité de terrain indépendante et d'ablations qui attribuent chaque capacité aux exigences de MANHAJ, leur protocole opérationnel figurant à l'Annexe A.

1.5. Portée

Le modèle formel s'applique à tout encodeur qui associe des vecteurs à des textes. Les expériences sont spécifiées pour des encodeurs de phrases et pour des sorties de modèles de langage encodées par un encodeur externe fixe. Les deux domaines conceptuels retenus — la famille et le domaine académique du scénario du professeur d'université du Papier 02 — assurent la continuité avec les papiers précédents de la méthodologie MANHAJ.

2. Travaux connexes

2.1. Directions conceptuelles dans les espaces appris

Des directions correspondant à des distinctions conceptuelles peuvent être identifiées dans les espaces appris à partir d'ensembles d'exemples contrastés. Des travaux pionniers ont dérivé de telles directions à partir de paires de plongements lexicaux (Bolkvasi et al., 2016) ; les vecteurs d'activation de concepts identifient des directions dans les activations d'un réseau à partir d'exemples de concepts fournis par l'utilisateur (Kim et al., 2018) ; les sondes linéaires testent si une information est linéairement récupérable dans les représentations intermédiaires (Alain & Bengio, 2017) ; et les méthodes d'ingénierie des représentations lisent et orientent le comportement des modèles le long de telles directions (Zou et al., 2023). Le présent article utilise des directions de ce type, construites à partir de pôles, comme outil de mise en œuvre.

2.2. La géométrie implicite des espaces appris

Les représentations apprises ne sont pas centrées sur une origine sémantiquement significative. Les représentations contextuelles occupent une région étroite et anisotrope de l'espace (Ethayarajh, 2019), et une pratique corrective courante soustrait la moyenne des représentations et supprime les directions dominantes (Mu & Viswanath, 2018). Un tel post-traitement établit une origine dérivée des données elles-mêmes. Il améliore de nombreuses tâches en aval, mais la référence qu'il crée est, par construction, une propriété du matériau représenté et non une référence déclarée.

2.3. Résidus de sous-espace et détection hors distribution

Mesurer la part d'une représentation qui se situe hors d'un sous-espace principal est une technique établie pour détecter des entrées qui s'écartent d'une distribution de référence (par exemple Wang et al., 2022). L'indicateur d'écart de la section 6 utilise cette technique par rapport à un sous-espace déclaré, et non extrait des données.

2.4. Évaluation par des modèles, dérive collective et contamination

L'usage de modèles de langage comme évaluateurs, avec ses effets documentés de position et d'auto-préférence, a été examiné de façon systématique (Zheng et al., 2023 ; Panickssery et al., 2024). Il a été

montré que l'entraînement récursif sur des données générées produit un effondrement des modèles (Shumailov et al., 2024). La contamination des jeux d'évaluation a été identifiée comme une menace pour la validité des évaluations publiées (Sainz et al., 2023).

2.5. Ce qu'apporte cet article

Ces lignes de recherche fournissent des outils de représentation et documentent des défaillances de l'évaluation. Parmi les travaux discutés ici, aucun ne précise les conditions dans lesquelles une référence de mesure est admissible ; aucun n'exige que la référence soit fondée dans un cadrage reconnu et indépendante de la propriété mesurée ; et aucun ne traite une référence compromise comme invalidant toutes les mesures qui en dérivent. Plus largement, nous n'avons pas connaissance de travaux qui fondent la mesure du sens en IA sur un fondement ordonné de cette nature. Ces conditions sont la contribution des Lois 2, 4 et 3 de MANHAJ. Les outils sont utilisés ; les conditions sont nouvelles.

3. Fondement méthodologique

3.1. La centralité doctrinale

La centralité doctrinale est le cadrage supérieur au sein duquel se constitue un champ intellectuel ou humain, et au sein duquel les concepts, les valeurs, les relations et les interprétations acquièrent leur position (Papier 01, section 2.1 ; Document de référence, section 6.2). Dans MANHAJ, elle est le cadrage le plus élevé de toute existence humaine raisonnable. Elle n'est pas choisie à la manière d'un paramètre expérimental : elle est reconnue et identifiée. Elle n'est ni une variable de mesure ni un axe de coordonnées. Elle est à un champ d'investigation ce qu'un système d'exploitation est à une application développée en son sein : le système d'exploitation n'est pas l'une des variables de l'application, mais il détermine les conditions dans lesquelles celle-ci est construite et s'exécute.

3.2. L'ordre fondationnel

MANHAJ examine un champ selon un ordre de fondation :

Centralité doctrinale $\rightarrow$ Domaine conceptuel $\rightarrow$ Origine coordonnée $\rightarrow$ Coordonnées du sens

Le domaine conceptuel se constitue au sein de la centralité, $D = D(\Gamma)$; l'origine coordonnée est établie au sein du domaine ; et les coordonnées du sens représentent les états du sens au sein du domaine ainsi cadré (Document de référence, section 6.3.4 ; Papier 01, section 2.2). Les flèches expriment un ordre de fondation et d'admission, non une dérivation mécanique de chaque élément à partir du précédent.

3.3. Trois centralités dans le contexte de l'IA

L'application de cet ordre à l'intelligence artificielle requiert trois centralités qu'il ne faut pas confondre :

1. Γ_F — **la centralité du cadre fondé** : le cadrage au sein duquel se situe le champ qui examine et par rapport auquel le sens est examiné. L'espace fondé est construit sur Γ_F , qui est reconnue et documentée explicitement.
2. Γ_L — **la centralité incorporée dans l'encodeur appris** : le cadrage porté par les données de production humaine à partir desquelles l'encodeur a été formé. L'encodeur opère en son sein sans l'avoir choisie.

3. Γ_C — **la centralité du contenu représenté** : le cadrage du projet, du texte ou du discours dont le sens est représenté. Comme dans le Papier 03, elle est découverte par l'examen, et non imposée par l'examineur.

3.4. La centralité incorporée des systèmes appris

Parce que la centralité doctrinale est le cadrage supérieur de toute existence humaine raisonnable, les données à partir desquelles les systèmes d'IA apprennent portent le cadrage dans lequel elles ont été produites. Un système qui traite de telles données a posteriori, sans en examiner le fondement, n'opère pas en dehors de toute centralité. Il accepte implicitement le cadrage incorporé et opère sous sa direction. L'existence de Γ_L dans toute représentation apprise est donc une conséquence méthodologique de l'ordre fondationnel de MANHAJ, et non une hypothèse en attente de confirmation empirique.

La portée de cette affirmation découle de la manière dont MANHAJ définit la centralité doctrinale : par sa fonction, et non par sa source. Tout ensemble de conceptions qui fournit une explication de l'existence, de l'individu et des rôles de l'individu constitue une centralité doctrinale, quelle qu'en soit l'origine et si élémentaire qu'en soit la forme. Un système appris ne produit pas de telles conceptions par lui-même : ses représentations sont formées à partir de conceptions empruntées à un cadrage antérieur — à travers les données, mais aussi à travers le choix des données, la conception du modèle et la définition de ses objectifs, qui sont tous des actes humains. L'affirmation s'applique donc à toute représentation apprise qui opère a posteriori, quelle que soit la source de ses données d'entraînement, et non seulement aux représentations apprises à partir de textes écrits par des humains. C'est une affirmation du cadre théorique de MANHAJ, et non un résultat empirique.

Ce qui reste ouvert à l'examen est son *contenu* : quelles distinctions conceptuelles la représentation apprise réalise, et comment sa réalisation est orientée par rapport à des centralités déclarées. La section 6.15 rend ce contenu mesurable.

3.5. Les trois lois

Loi de la dualité de la direction de l'examen (Loi 2). L'examen a priori porte sur le fondement, les références et l'approche d'un champ ; l'examen a posteriori travaille au sein d'un champ dont le fondement a été accepté, en s'orientant vers les phénomènes, les détails et les résultats (Document de référence, section 8.2). Les deux directions sont complémentaires, mais la direction a priori est première.

Loi de la justesse (Loi 3). La question de la licéité du fondement de l'approche précède la question des détails : *ce qui n'est pas licite, ses détails ne sont pas valides* (Document de référence, section 8.3). La justesse se distingue de la validité scientifique. Une recherche peut utiliser des outils corrects au sein d'un champ sans avoir examiné la justesse du fondement de ce champ.

Loi de l'origine coordonnée (Loi 4). La position du sens au sein d'un domaine conceptuel est déterminée par rapport à une origine coordonnée de référence ; la déviation d'une position P_i par rapport à la position de référence P_0 est $V_i = P_i - P_0$, mesurée selon les axes et les unités établis par l'étude (Document de référence, section 9.1).

3.6. Origine et sujet

Le Papier 01 (section 2.2) et le Papier 03 (section 3.3) distinguent trois formes de relation entre l'origine et le sujet de l'investigation : une origine indépendante du sujet ; une origine liée à la position ou à l'état du sujet ; et une origine liée à une propriété du sujet, qui peut devenir le seul critère selon lequel le sujet est jugé. Le Papier 03 (section 3.4) montre que la circularité apparaît lorsque la conception adoptée d'un

sujet devient la référence ultime à l'aune de laquelle ce sujet est évalué, sans examen indépendant de son fondement. La section 6 donne à cette circularité une forme computationnelle exacte.

4. Conditions de la mesure

4.1. L'exécutable et son système d'exploitation

Considérons un fichier exécutable développé en VB.NET et s'exécutant sous Windows. Trois couches sont en jeu :

Couche	Dans le logiciel	Dans la mesure du sens en IA
L'exécutable	Ce qui est exécuté et observé	Les sorties et leurs représentations apprises
L'environnement de développement et le code source	VB.NET	Le domaine conceptuel
Le système d'exploitation	Windows	La centralité doctrinale

La pratique actuelle de la mesure du sens travaille directement sur l'exécutable. Elle calcule des distances, des dérives et des scores sur les sorties et leurs représentations, et elle répare les défaillances observées un comportement après l'autre. Cela s'apparente à la rétro-ingénierie : elle peut aboutir à des réparations locales efficaces, mais elle n'atteint pas les conditions qui ont produit le comportement. La recherche en interprétabilité décrit elle-même souvent son travail comme une rétro-ingénierie, et l'analogie est exacte sur ce point.

MANHAJ exige que l'exécutable soit d'abord cadré : dans quel environnement il a été développé, et sous quel système d'exploitation il s'exécute. Sans ce cadrage, un exécutable peut être corrigé ponctuellement, mais non traité méthodologiquement à sa racine.

L'analogie s'applique dans deux directions. Elle s'applique au contenu représenté, dont la centralité I_C doit être reconnue avant que son sens ne soit mesuré. Elle s'applique également au modèle d'IA lui-même, qui est un exécutable construit sous la centralité incorporée I_L de ses données (section 3.4).

4.2. Coordonnées relatives et absolues

En conception assistée par ordinateur, un point peut être saisi par rapport à la position courante ($@20,40$) ou en coordonnées absolues ($20,40$). La saisie relative ne dépend pas de l'origine du dessin ; la saisie absolue en dépend. La même distinction vaut pour le sens :

- une **transition** ΔM_t entre états successifs est une quantité relative, indépendante de l'origine (Proposition 1, section 6.13) ;
- une **dévi**ation D_t par rapport à une référence déclarée est une quantité absolue, qui n'existe qu'une fois une origine déclarée.

Le Papier 02 suivait le sens au moyen de quantités relatives. La plupart des suivis du sens en IA sont de même relatifs : sortie contre sortie, version contre version, sortie contre le centroïde des sorties. Ils ne disposent d'aucun cadre absolu déclaré.

4.3. Systèmes de coordonnées donnés et définis

AutoCAD fournit un système de coordonnées général (SCG, *World Coordinate System*, WCS), donné par le logiciel, et permet de définir un système de coordonnées utilisateur (SCU, *User Coordinate System*, UCS) pour un projet particulier, avec une origine et des axes choisis selon les exigences de ce projet avant le début du dessin.

L'espace appris correspond au WCS : un système de coordonnées qui a été donné, et non défini pour l'investigation en cours. L'espace fondé correspond à l'UCS : un système de coordonnées défini avant la mesure, selon les exigences de l'investigation — c'est-à-dire selon sa centralité et son domaine conceptuel. MANHAJ ajoute ce que l'UCS seul ne contient pas : l'exigence que la définition du système de coordonnées soit fondée dans une centralité reconnue et examinée quant à son admissibilité.

4.4. Mesure à référence implicite et mesure fondée

Nous appelons **à référence implicite** une mesure dont la référence n'est pas examinée comme condition de son interprétation : elle n'est pas déclarée, elle coïncide avec la structure apprise du système, ou elle est dérivée des sorties mesurées. Les jeux de référence fixes et les critères d'évaluation déclarés sont largement utilisés, et la critique formulée dans cet article ne les vise pas en tant que tels. Elle vise les pratiques de mesure qui n'examinent pas explicitement comment la référence a été fondée, si elle est cohérente avec le domaine conceptuel du sens mesuré, et si elle est indépendante de la propriété mesurée.

Nous appelons **fondée** une mesure dont le cadre a été établi selon l'ordre fondationnel de la section 3.2 et a satisfait au test d'admissibilité de la section 6.14.

4.5. Différence nécessaire des conditions, différence spécifiée des résultats

Les conditions de la mesure fondée et de la mesure à référence implicite diffèrent nécessairement : cela découle des définitions et des Lois 2 et 4. Leurs résultats diffèrent de manières spécifiées. La Proposition 1 montre que les transitions sont identiques quelle que soit l'origine. La Proposition 2 montre que la capacité de détection diverge entre une référence auto-dérivée et une référence indépendante. Les dispositifs expérimentaux de la section 8 précisent comment peuvent être testées les capacités supplémentaires que la mesure fondée est censée fournir au-delà de la mesure à référence implicite.

5. De la loi au test

5.0. L'argument central

Une mesure du sens peut être calculée avec une précision totale alors que ses conditions d'interprétation demeurent non fondées : rien dans la mesure n'établit à quoi la valeur se rapporte au sein du domaine, ni si le cadre qui l'a produite était admissible. MANHAJ traite ce problème à sa source. Ses lois établissent que la mesure est précédée par la fondation, et trois d'entre elles traduisent cet ordre fondationnel en exigences que la mesure à référence implicite ne satisfait pas :

- la **Loi de la dualité de la direction de l'examen (Loi 2)** exige que le cadre de mesure soit fondé avant la mesure ;
- la **Loi de l'origine coordonnée (Loi 4)** exige que la mesure soit rapportée à une origine explicite, indépendante de la propriété mesurée ;
- la **Loi de la justesse (Loi 3)** exige qu'un cadre non admissible ne produise aucune mesure valide.

Ces exigences constituent la contribution de l'article. Les techniques computationnelles utilisées pour les mettre en œuvre — directions conceptuelles, projection sur un sous-espace, mesures résiduelles, détection de chevauchement — sont des outils de mise en œuvre établis. MANHAJ fournit les conditions qui en gouvernent l'usage, et ces conditions produisent les capacités exposées ci-dessous.

La centralité incorporée des systèmes appris (section 3.4) encadre l'ensemble de l'argument : les cadres de mesure utilisés dans la pratique actuelle ne sont pas neutres, parce qu'ils héritent du cadrage des données à partir desquelles ils ont été appris. La question n'est pas de savoir si un cadre possède une centralité, mais si sa centralité est reconnue, déclarée et examinée.

Chaque affirmation ci-dessous est formulée à l'un de trois niveaux :

- **[Méthodologique]** — ce qui découle des lois, des théories et de l'ordre fondationnel de MANHAJ ;
- **[Computationnel]** — la manière dont le modèle met en œuvre ces exigences et ce qu'il fournit, y compris les propriétés démontrées à partir du modèle ;
- **[Hypothèse empirique]** — un résultat attendu concernant la performance et les conséquences pratiques, que les dispositifs expérimentaux de la section 8 sont construits pour tester. Les affirmations de ce niveau ne sont pas présentées comme des résultats établis ; elles ne deviennent des constats empiriques qu'une fois les expériences réalisées.

Pour chaque loi, la section énonce la loi, l'exigence qu'elle impose, le mécanisme qui la met en œuvre, la capacité qui en résulte et le dispositif expérimental qui la teste.

5.1. Loi de la dualité de la direction de l'examen (Loi 2)

Loi. L'examen a priori porte sur le fondement, les références et les conditions d'un champ avant que le travail a posteriori ne s'y déploie (Document de référence, section 8.2).

Exigence. Le cadre de mesure est fondé selon l'ordre fondationnel complet :

Centralité doctrinale → Domaine conceptuel → Origine coordonnée → Coordonnées du sens

La centralité Γ_F est reconnue et documentée ; le domaine $D(\Gamma_F)$ se constitue en son sein ; les formulations de référence, les pôles et l'étalonnage sont dérivés des définitions opératoires au sein de $D(\Gamma_F)$; et tout cela précède la mesure. **[Méthodologique]**

La pratique actuelle construit les cadres de mesure à partir des exemples ou des données disponibles et interprète ensuite les valeurs obtenues. La Loi 2 inverse cet ordre. L'indépendance du cadre à l'égard des sorties mesurées découle de cette inversion, mais l'exigence est plus forte que l'indépendance : un cadre indépendant des sorties, mais tiré, sans déclaration, d'un domaine ou d'une centralité autres que ceux qui sont déclarés, ne satisfait pas à la Loi 2. **[Méthodologique]**

Mécanisme. Le cadre fondé F de la section 6.3, dont chaque élément peut être rattaché à l'étape de l'ordre fondationnel dont il dérive.

Capacité.

- La construction du cadre n'est pas façonnée par les sorties mesurées. **[Computationnel]**
- Chaque axe porte une signification conceptuelle déclarée au sein de $D(\Gamma_F)$, de sorte que toute valeur de coordonnée est interprétable dans les termes du domaine plutôt que dans ceux de l'encodeur. **[Computationnel]**

- Les axes ordonnés selon l'ordre fondationnel devraient s'aligner sur les distinctions du domaine conceptuel plus étroitement que des axes construits à partir d'exemples de commodité ou extraits des données. **[Hypothèse empirique — testée par E3]**

5.2. Loi de l'origine coordonnée (Loi 4)

Loi. La position du sens au sein d'un domaine conceptuel est déterminée par rapport à une origine coordonnée de référence ; la déviation est $V_i = P_i - P_0$ (Document de référence, section 9.1).

Exigence. L'origine est explicite et indépendante de la propriété mesurée. Une origine liée à la propriété mesurée produit une mesure circulaire (Papier 01, section 2.2 ; Papier 03, sections 3.3–3.4). **[Méthodologique]**

Le suivi actuel du sens en IA est essentiellement relatif. Dans les termes de la section 4.2, il travaille avec des coordonnées relatives et ne dispose d'aucun cadre absolu déclaré. La Loi 4 fournit ce cadre.

Mécanisme. Une origine déclarée p_0 calculée à partir des formulations de référence ; des coordonnées absolues M_t^F et une déviation D_t rapportées à p_0 ; des transitions ΔM_t^F selon des axes déclarés (sections 6.5–6.9).

Capacité.

- Le référencement absolu ajoute une dimension de mesure qui fait défaut au suivi relatif : la position de chaque état du sens par rapport à une référence déclarée, à côté de son changement par rapport à l'état précédent. Les transitions sont invariantes par changement d'origine (Proposition 1) ; l'origine gouverne la déviation. **[Computational]**
- Une référence dérivée des sorties mesurées est structurellement aveugle à la composante commune de leur dérive, quelle qu'en soit l'ampleur (Proposition 2). Une référence indépendante des sorties ne souffre pas de cette cécité : elle enregistre la composante commune dans la mesure où l'encodeur représente le changement et où sa projection sur les axes retenus, $U^T \bar{\delta}$, est non nulle. La Proposition 2 concerne le déplacement représentationnel ; que ce déplacement corresponde à une dérive sémantique est établi indépendamment, par la validation humaine dans E2. C'est la forme computationnelle de la circularité identifiée par la Loi 4 et par le Papier 03. **[Computational]**
- Face à une référence externe fixe, qui est elle aussi indépendante des sorties, la capacité distinctive du cadre fondé réside dans ce que la Loi 4 ajoute au-delà de l'indépendance : une origine déclarée au sein d'un domaine fondé, et une déviation décomposée selon des axes déclarés. Le cadre fondé identifie non seulement que le sens s'est éloigné de la référence, mais selon quelle dimension conceptuelle du domaine il s'est déplacé. **[Computational]**. Que cette capacité produise une meilleure attribution conceptuelle que la référence de comparaison forte est une **[Hypothèse empirique — testée par E2]**.
- Dans une dérive collective naturelle, la composante commune devrait représenter une part substantielle de la dérive totale, auquel cas la cécité établie par la Proposition 2 affecte la pratique réelle du suivi. **[Hypothèse empirique — testée par E2]**

5.3. Loi de la justesse (Loi 3)

Loi. Ce qui n'est pas licite, ses détails ne sont pas valides (Document de référence, section 8.3).

Exigence. Le Papier 02 a appliqué la loi aux transitions entre états du sens. Le présent article l’applique au niveau antérieur du cadre de mesure lui-même. Si le cadre n’est pas admissible, les mesures produites en son sein — transitions, déviations, appréciations de clôture — ne sont pas des détails valides, quelle que soit leur précision computationnelle. **[Méthodologique]**

Cette exigence n’a pas d’équivalent dans la pratique actuelle, où les métriques sont calculées et rapportées indépendamment de l’état de la référence qui les a produites. MANHAJ introduit dans la mesure du sens la catégorie de l’**invalidité fondationnelle** : toute une famille de résultats est invalidée à sa source, sans qu’il soit nécessaire d’examiner chaque résultat individuellement.

Loi et garde-fou. La Loi de la justesse détermine ce qui ne peut être tenu pour valide : les mesures produites au sein d’un cadre non admissible. Le garde-fou fondationnel est le mécanisme computationnel qui met en œuvre les conditions d’admissibilité spécifiées dans cet article. Les deux se situent à des niveaux différents. Si le garde-fou ne détecte pas un défaut, la loi n’est pas pour autant contredite ; l’échec indique que les conditions d’admissibilité sont incomplètement spécifiées, ou incomplètement mises en œuvre, pour ce défaut. Les expériences testent donc le garde-fou en tant que mise en œuvre de la Loi 3, et non la Loi 3 elle-même. **[Méthodologique]**

Deux types d’échec. Un cadre peut échouer sur le plan méthodologique ou sur le plan computationnel, et le modèle traite ces échecs comme distincts :

Vérification	Conditions	Nature	Établie par
Admissibilité méthodologique $\text{Adm}_M(F)$	C (cohérence avec $D(I_F)$), E (explicitation), I_{doc} (provenance documentée)	Fondationnelle	Examen documenté selon un protocole écrit
Opérabilité computationnelle $\text{Op}_{\text{comp}}(F)$	Q (pouvoir discriminant des axes), I_{comp} (absence de chevauchement ou de quasi-duplication avec les sorties), Cal (matériau d’étalonnage suffisant)	Représentationnelle	Calcul

Adm_M	Op_{comp}	État	Conséquence
0	—	Invalidité fondationnelle	Toutes les mesures au sein de F sont invalides ; le cadre est renvoyé pour révision. [Méthodologique]
1	0	Admissible, non opérant	Le fondement est sain, mais l’encodeur ne réalise pas le cadre ou les données n’en permettent pas le fonctionnement ; les mesures au sein de F ne sont pas interprétables. L’échec se situe dans la représentation, non dans le fondement. [Méthodologique]
1	1	Opérant	La mesure se poursuit.

Le deuxième état est lui-même un résultat : il localise une limite de la représentation apprise par référence à un fondement sain, ce que la mesure à référence implicite, dépourvue de cadre fondé, ne peut pas faire.

Mécanisme. Un garde-fou fondationnel qui évalue $\text{Adm}_M(F)$ avant tout encodage et $\text{Op}_{\text{comp}}(F)$ après l’encodage, et qui bloque ou qualifie la mesure selon le tableau ci-dessus (section 6.14).

Capacité.

- Le garde-fou applique la Loi 3 sur le plan computationnel : aucune mesure n'est rapportée comme fondée à partir d'un cadre qui n'a pas satisfait à l'admissibilité. **[Computationnel]**
- Le garde-fou devrait détecter des défauts fondationnels — provenance non déclarée et importation non déclarée depuis un autre cadrage — que les métriques standard et les procédures d'audit établies ne révèlent pas, et les conclusions tirées au sein de tels cadres devraient se révéler non fiables au regard d'une vérité de terrain indépendante. **[Hypothèse empirique — testée par E1]**

5.4. La centralité incorporée

Fondement. Toute représentation apprise opère sous une centralité incorporée Γ_L (section 3.4). **[Méthodologique]**

Mécanisme. Deux cadres pour le même concept, fondés sur deux centralités reconnues, sont réalisés au moyen du même encodeur ; l'indice d'orientation $\mathcal{O}$ compare l'intensité avec laquelle l'encodeur réalise les distinctions propres à chacun, rapportée aux distinctions qu'ils partagent (section 6.15).

Capacité. Le contenu de Γ_L devient mesurable de manière relationnelle : l'orientation de la réalisation d'un encodeur entre des centralités déclarées est définie par l'indice $\mathcal{O}$. **[Computationnel]** Que cette orientation soit non nulle pour des encodeurs réels et varie entre encodeurs formés à partir de corpus différents est une **[Hypothèse empirique — testée par E4]** ; attribuer causalement cette variation aux données d'entraînement exige la comparaison contrôlée décrite à la section 8.6.

5.5. Intégration

La Loi 2 détermine quand et dans quel ordre le cadre de mesure est fondé ; la Loi 4 détermine comment la mesure est référencée en son sein ; la Loi 3 détermine ce qui s'ensuit lorsque le cadre échoue. La centralité doctrinale détermine le domaine au sein duquel les trois opèrent, et son pendant incorporé Γ_L devient un objet de mesure. Ensemble, elles remplacent une mesure dont les conditions d'interprétation sont implicites par une mesure dont les conditions sont fondées, déclarées et examinées.

Affirmation	Niveau
Le cadre de mesure doit être fondé selon l'ordre fondationnel complet	Méthodologique
L'origine doit être explicite et indépendante de la propriété mesurée	Méthodologique
Un cadre non admissible ne produit aucune mesure valide	Méthodologique
Toute représentation apprise opère sous une centralité incorporée	Méthodologique
Le cadre fondé n'est pas façonné par les sorties mesurées	Computationnel
Les transitions sont invariantes par changement d'origine ; l'origine gouverne la déviation	Computationnel
Une référence dérivée des sorties est aveugle à la composante commune de la dérive	Computationnel
Le cadre fondé attribue la dérive à des dimensions conceptuelles déclarées	Computationnel

Affirmation	Niveau
Cette attribution dépasse celle d'une référence externe fixe	Hypothèse empirique (E2)
La dérive commune est substantielle dans des conditions naturelles	Hypothèse empirique (E2)
Le garde-fou détecte des défauts fondationnels que les vérifications standard manquent	Hypothèse empirique (E1)
Les axes ordonnés selon l'ordre fondationnel s'alignent sur les distinctions du domaine	Hypothèse empirique (E3)
Le contenu de la centralité incorporée est caractérisé de manière relationnelle	Méthodologique (examen)
L'orientation d'un encodeur entre des centralités déclarées est mesurée par $\mathcal{O}$	Computationnel
L'orientation varie entre encodeurs formés à partir de corpus différents	Hypothèse empirique (E4)

6. Modèle formel de l'espace fondé

6.1. Objet et principes directeurs

Cette section présente la construction formelle par laquelle les exigences de la section 5 sont mises en œuvre. Deux principes la gouvernent.

1. **La centralité doctrinale est une condition, non une variable.** Elle gouverne les définitions qui peuvent entrer dans le cadre, mais elle n'apparaît comme quantité numérique dans aucune équation (Papier 01, section 2.1).
2. **L'admissibilité méthodologique et l'opérabilité computationnelle sont définies sur des objets différents.** $\text{Adm}_M(F)$ est définie sur le fondement, les formulations et leur provenance documentée, indépendamment de l'encodeur. $\text{Op}_{\text{comp}}(F)$ est définie sur les formulations encodées et les sorties encodées. Cette séparation structurelle garantit que la validité méthodologique d'un cadre n'est jamais inférée du succès de sa mise en œuvre computationnelle, ni confondue avec lui.

6.2. L'espace appris

Soit $\mathcal{X}$ l'ensemble des sorties et formulations possibles, et soit

$$\varphi: \mathcal{X} \rightarrow \mathbb{R}^n$$

un encodeur appris. Un état du sens dans l'espace appris est $m_t = \varphi(o_t)$. L'origine de $\mathbb{R}^n$ est le vecteur nul, dont la position par rapport aux sens est déterminée par l'entraînement et n'est déclarée à partir d'aucun domaine conceptuel : c'est une **origine implicite**. Soit $O_T = \{o_1, \dots, o_T\}$ l'ensemble des sorties mesurées disponibles à l'instant T .

6.3. Le cadre fondé

Le cadre fondé est

$$F = (\Gamma_F, R, D(\Gamma_F), Def, \mathcal{E}_B, H, P_0, A, \kappa)$$

où :

- Γ_F est la centralité doctrinale reconnue du cadre (section 6.4) ;
- R est la référence ;
- $D(\Gamma_F)$ est le domaine conceptuel constitué au sein de Γ_F ;
- Def est l'ensemble des définitions opératoires au sein de $D(\Gamma_F)$;
- $\mathcal{E}_B = (E_0, \{E_k^+, E_k^-\}_{k=1}^d, \mathcal{P})$ est l'**ensemble de construction et d'étalonnage**, dans lequel E_0 contient des formulations de référence du sens de référence, E_k^+ et E_k^- contiennent des formulations des pôles positif et négatif de l'axe k , et $\mathcal{P} = \{P_1, \dots, P_m\}$ contient des ensembles de paraphrases, chacun regroupant des formulations lexicalement différentes d'un même sens déclaré ;
- H est l'**ensemble de validation**, composé de formulations tenues à l'écart, relevant du domaine et rédigées selon le même protocole que $\mathcal{E}_B$;
- P_0 est l'origine coordonnée, de lecture p_0 ;
- $A = \{u_1, \dots, u_d\}$ sont les axes de représentation, dans un ordre déclaré au sein de Def , avec $d \ll n$;
- $\kappa = (\kappa_1, \dots, \kappa_d)$ sont les coefficients d'étalonnage.

Soit $\mathcal{E}_{all} = \mathcal{E}_B \cup H$.

Règle de construction. Chaque formulation de $\mathcal{E}_{all}$ est rédigée à partir de Def au sein de $D(\Gamma_F)$ et documentée avant la mesure. P_0, A, κ et les seuils ρ et ε sont calculés à partir de $\mathcal{E}_B$ seul. H est exclu de toute étape de construction et d'étalonnage. Aucun élément du cadre n'est calculé à partir de O_T .

Validation. La validation sur H est un diagnostic rapporté, non une étape de construction. Si le cadre est révisé après examen de ses résultats sur H , H devient matériau de construction, et un nouvel ensemble de validation est rédigé selon le même protocole ; seul un ensemble dont les résultats n'ont pas informé le cadre est qualifié de tenu à l'écart.

6.4. La centralité doctrinale dans le modèle

Position. La centralité doctrinale est au cadre ce qu'un système d'exploitation est à une application développée en son sein. Γ_F n'entre dans aucune expression numérique du modèle ; elle gouverne les définitions, et donc les formulations, qui peuvent entrer dans le cadre.

Rôle formel. Soit $\mathfrak{D}$ l'ensemble des définitions opératoires candidates. Une centralité Γ induit sur $\mathfrak{D}$ une relation d'admissibilité,

$$\Gamma \vdash d,$$

qui se lit : *la définition d acquiert son sens au sein du cadrage Γ* . Le domaine conceptuel est l'ensemble des concepts dont les définitions opératoires sont admises par Γ_F ; les définitions opératoires satisfont

$Def \subseteq \{d \in \mathcal{D} : \Gamma_F \vdash d\}$; et chaque formulation $e \in \mathcal{E}_{all}$ réalise une certaine $d \in Def$. L'ordre fondationnel

$$\Gamma_F \rightarrow D(\Gamma_F) \rightarrow P_0 \rightarrow U, K \rightarrow M_t^F$$

exprime un ordre de fondation et d'admission.

Reconnaissance. Une procédure *Rec* reconnaît la centralité au sein de laquelle se constitue le domaine d'un cadre : pour le cadre fondé, la centralité du champ qui examine ; dans la caractérisation relationnelle de la section 6.15, les centralités des cadrages sources de F_A et F_B , chacune reconnue à partir de ses propres sources. *Rec* produit trois types de preuves :

1. **Cadrage source** : les références supérieures — textes, principes codifiés ou conceptions établies — au sein desquelles se constituent les concepts du domaine ;
2. **Trace de dérivation** : pour chaque $d \in Def$, la trace par laquelle elle est admise sous Γ_F ;
3. **Éléments distinctifs** : des définitions ou des distinctions admises sous Γ_F qui ne seraient pas admises, ou qui acquerraient un sens différent, sous un cadrage alternatif.

Le troisième type de preuve distingue la reconnaissance de la documentation : il montre que le cadre porte les distinctions propres à sa centralité, et non un contenu générique compatible avec n'importe quel cadrage.

Reconnaissance erronée et importation incorporée. Il y a *reconnaissance erronée* lorsque les définitions effectivement utilisées dérivent d'un cadrage autre que la Γ_F déclarée. Il y a *importation incorporée* lorsque des définitions ou des formulations sont façonnées par le système appris lui-même — par exemple lorsque *Def* ou $\mathcal{E}_{all}$ est rédigé avec l'assistance du système évalué —, de sorte que Γ_L entre dans le cadre qui est censé l'examiner.

Indétermination. Si *Rec* n'établit pas Γ_F , le cadre n'est pas fondé. Ses mesures peuvent être calculées, mais seulement avec le statut d'une mesure à référence implicite.

Centralité incorporée. D'après la section 3.4, l'encodeur φ opère sous une centralité incorporée Γ_L . Γ_L n'est pas directement observable ; son contenu est caractérisé de manière relationnelle à la section 6.15.

6.5. L'origine coordonnée

La lecture de l'origine est la position moyenne des formulations de référence :

$$p_0 = \frac{1}{|E_0|} \sum_{e \in E_0} \varphi(e).$$

L'origine de l'espace fondé coïncide avec cette **lecture** du sens de référence, et non avec le sens de référence lui-même, qui est défini méthodologiquement par R et *Def*. L'origine est déclarée dans sa définition et apprise dans sa lecture.

6.6. Les axes de représentation

Pour chaque axe k , les centroïdes des pôles et la direction brute sont

$$c_k^+ = \frac{1}{|E_k^+|} \sum_{e \in E_k^+} \varphi(e), \quad c_k^- = \frac{1}{|E_k^-|} \sum_{e \in E_k^-} \varphi(e), \quad a_k = c_k^+ - c_k^-.$$

Les directions brutes sont orthonormalisées par le procédé de Gram–Schmidt dans l’ordre déclaré :

$$\tilde{u}_k = a_k - \sum_{j < k} \langle a_k, u_j \rangle u_j, \quad u_k = \frac{\tilde{u}_k}{\|\tilde{u}_k\|} \quad (\tilde{u}_k \neq 0),$$

et regroupées dans $U = [u_1 \cdots u_d] \in \mathbb{R}^{n \times d}$ avec $U^\top U = I$. Le **rapport de composante orthogonalisée** est

$$\gamma_k = \frac{\|\tilde{u}_k\|}{\|a_k\|}, \quad 0 \leq \gamma_k \leq 1, \quad a_k \neq 0.$$

Si $a_k = 0$, l’axe k est indéfini ; si $a_k \neq 0$ et $\tilde{u}_k = 0$, alors $\gamma_k = 0$ et u_k est indéfini. Dans les deux cas, l’axe ne satisfait pas la condition Q (section 6.14.2). L’orthogonalité est une propriété du système de représentation et n’affirme pas l’indépendance des sens représentés (Papier 01, section 3.4). La construction par pôles est une technique de mise en œuvre ; MANHAJ ne la prescrit pas. Un axe indéfini n’est jamais utilisé comme coordonnée opérationnelle : il n’entre ni dans U ni dans K , et aucune quantité mesurée au sein du cadre $(M_t^F, \Delta M_t^F, D_t, r_t)$ ne l’utilise. La valeur attribuée à un tel axe par la règle de plancher de la section 6.15 sert uniquement à l’analyse diagnostique de la réalisation par l’encodeur, et ne rend pas l’axe opérationnel.

6.7. L’étalonnage

Le coefficient d’étalonnage de l’axe k est

$$\kappa_k = 1/2 \langle a_k, u_k \rangle.$$

Lemme 1. Chaque fois que $\tilde{u}_k \neq 0$, $\langle a_k, u_k \rangle = \|\tilde{u}_k\|$, et donc $\kappa_k = 1/2 \|\tilde{u}_k\| > 0$.

Démonstration. $a_k = \tilde{u}_k + \sum_{j < k} \langle a_k, u_j \rangle u_j$, et u_k est orthogonal à tout u_j avec $j < k$; donc $\langle a_k, u_k \rangle = \langle \tilde{u}_k, u_k \rangle = \|\tilde{u}_k\|$. ■

Sur leur propre axe, les deux centroïdes des pôles sont séparés d’exactement deux unités, $\langle a_k, u_k \rangle / \kappa_k = 2$. L’unité de chaque axe est définie par les pôles déclarés du domaine, et non par la distribution des sorties mesurées : la méthodologie établit l’axe, et la recherche au sein du domaine en établit l’étalonnage (Papier 01, section 4.1).

6.8. L’état du sens et la répartition des fonctions

Avec $K = \text{diag}(\kappa_1, \dots, \kappa_d)$, l’état du sens dans l’espace fondé est

$$M_t^F = K^{-1} U^\top (\varphi(o_t) - p_0) \in \mathbb{R}^d.$$

La coordonnée moyenne des formulations de référence est nulle,

$$\frac{1}{|E_0|} \sum_{e \in E_0} K^{-1} U^\top (\varphi(e) - p_0) = K^{-1} U^\top (p_0 - p_0) = 0,$$

de sorte que l’origine de l’espace fondé coïncide avec la lecture du sens de référence, ce qui rend opérationnelle la Loi de l’origine coordonnée. M_t^F entre dans la représentation du Papier 01 comme sa composante de sens, $P_t = (t, x, M_t^F)$. La transition entre états successifs est

$$\Delta M_t^F = M_{t+1}^F - M_t^F = K^{-1}U^\top(\varphi(o_{t+1}) - \varphi(o_t)).$$

Élément	Gouverne
Axes U et étalonnage K	Les transitions ΔM_t^F , selon des dimensions déclarées et en unités déclarées
Origine p_0	La déviation D_t (section 6.9) et l'indicateur d'écart r_t (section 6.10)

6.9. La déviation par rapport à la référence

Puisque l'origine de l'espace fondé a pour coordonnée zéro, la déviation de l'état t par rapport à la lecture du sens de référence est

$$D_t = M_t^F.$$

$\|D_t\|$ est la distance à la lecture de référence en unités étalonnées, et les composantes de D_t attribuent la déviation aux axes déclarés. C'est la forme opérationnelle de $V_i = P_i - P_0$ dans la Loi de l'origine coordonnée ; le symbole D_t évite la confusion avec la variable de validité V_t du Papier 02.

6.10. L'indicateur d'écart

Pour $\varphi(o_t) \neq p_0$,

$$r_t = \frac{\|(I - UU^\top)(\varphi(o_t) - p_0)\|}{\|\varphi(o_t) - p_0\|},$$

et $r_t = 0$ lorsque $\varphi(o_t) = p_0$. Dans tous les cas, $0 \leq r_t \leq 1$.

r_t indique un écart par rapport à la **représentation déclarée** du domaine. Une valeur élevée peut refléter un mouvement vers des sens extérieurs à $D(I_F)$, ou des dimensions de $D(I_F)$ non captées par les axes déclarés. r_t est un indicateur d'écart, non une mesure d'appartenance au domaine ; sa relation à l'appartenance est examinée dans E3.

6.11. Les seuils étalonnés

Seuil d'écart. ρ est un percentile supérieur déclaré de $\{r(e) : e \in E_0 \cup \bigcup_k (E_k^+ \cup E_k^-)\}$, calculé à partir de $\mathcal{E}_B$; son comportement est rapporté sur H .

Seuil de changement. ε est un percentile supérieur déclaré de

$$\{\|K^{-1}U^\top(\varphi(e) - \varphi(e'))\| : e, e' \in P_i, e \neq e', P_i \in \mathcal{P}\},$$

c'est-à-dire des amplitudes de transition entre paraphrases d'un même sens déclaré. Le seuil de « aucun changement de sens détecté » est ainsi dérivé d'une reformulation déclarée plutôt que supposé.

Conditions d'étalonnage. ρ exige $|E_0 \cup \bigcup_k (E_k^+ \cup E_k^-)| \geq N_\rho$; ε exige $|P_i| \geq 2$ pour tout $P_i \in \mathcal{P}$ et $\sum_i |P_i| (|P_i| - 1)/2 \geq N_\varepsilon$. N_ρ , N_ε et les niveaux de percentile sont déclarés à l'avance et examinés par analyse de sensibilité. Si une condition d'étalonnage n'est pas satisfaite, le seuil correspondant n'est pas calculable et le cadre ne satisfait pas la condition *Cal*.

6.12. La classification des transitions

Notons in_t la condition $r_t \leq \rho$. Les frontières sont fixées comme suit : une transition entre deux états interprétables est un **changement de sens détecté** si $\|\Delta M_t^F\| \geq \varepsilon$, et **aucun changement de sens détecté** si $\|\Delta M_t^F\| < \varepsilon$. Ces deux classes ne sont pas équivalentes aux conditions $\Delta M_t = 0$ et $\Delta M_t \neq 0$ du Papier 02, et ne sont pas présentées comme telles. $\|\Delta M_t^F\| < \varepsilon$ n'affirme pas que le sens est inchangé ; il affirme qu'aucun changement n'est détecté à la résolution ε , étalonnée comme la variation produite par la reformulation d'un même sens déclaré (section 6.11). Le modèle maintient donc trois cas distincts : *aucun changement* ($\Delta M_t^F = 0$), *changement inférieur à la résolution de détection* ($0 < \|\Delta M_t^F\| < \varepsilon$) et *changement détecté* ($\|\Delta M_t^F\| \geq \varepsilon$). Les deux premiers sont rapportés ensemble comme « aucun changement de sens détecté », la valeur de $\|\Delta M_t^F\|$ étant enregistrée, de sorte que la distinction n'est jamais perdue.

Relation avec le Papier 02. Le Papier 02 teste $\Delta M_t = 0$ à l'étape où il décide si une transition a produit un sens nouveau. Dans l'espace fondé, cette étape est réalisée à la résolution déclarée ε : le détecteur de clôture du Papier 02 reçoit les transitions classées « aucun changement de sens détecté » entre états interprétables, et une clôture établie sur elles est une **clôture détectée à la résolution ε** . Cela ne modifie ni la fonction du détecteur de clôture — qui exige toujours une absence persistante de sens dérivationalnel nouveau (Papier 02, Loi 7) — ni la Loi de la justesse : la validité de transition V_t du Papier 02 est évaluée sur chaque transition, quelle que soit son amplitude, de sorte qu'un changement inférieur à la résolution de détection demeure soumis à l'examen de validité. Un état est interprétable si $r_t \leq \rho$, le cas limite $r_t = \rho$ étant interprétable.

La Loi 7 définit la fin de la production du sens comme l'atteinte d'une constante : la fonction du sens atteint C , et sa dérivée est nulle parce qu'elle est constante. Un changement nul entre deux états est une conséquence de cette condition, non sa définition, puisqu'un changement peut s'annuler momentanément sans qu'une constante ait été atteinte. La clôture est donc appréciée sur l'état : sur la fenêtre de persistance, l'état du sens demeure à une seule constante à la résolution ε près, et les transitions classées « aucun changement de sens détecté » sont un signe nécessaire de cette condition, non un signe suffisant.

in_t	in_{t+1}	$\ \Delta M_t^F\ $	Classification
dedans	dedans	$\ \Delta M_t^F\ \geq \varepsilon$	Changement de sens détecté au sein de la représentation déclarée de $D(I_F)$
dedans	dedans	$\ \Delta M_t^F\ < \varepsilon$	Aucun changement de sens détecté , transmis au détecteur de clôture du Papier 02
dedans	dehors	—	Écart par rapport à la représentation déclarée ; non interprété comme un changement de sens au sein de $D(I_F)$
dehors	dedans	—	Retour ; non interprété comme un changement de sens ; M_{t+1}^F devient l'état interprétable suivant
dehors	dehors	—	Hors de la représentation déclarée ; non interprété au sein de $D(I_F)$

Cette classification détermine si le sens a changé par rapport à la représentation déclarée. Elle complète la validité de transition V_t du Papier 02, qui détermine si un changement est valide.

6.13. Propriétés formelles

L'encodeur φ est fixe dans toute cette sous-section.

Proposition 1 (Invariance des transitions par changement d'origine). Si p_0 est remplacée par une autre lecture déclarée p_0' , U et K restant inchangés, alors

$$M_t^{F'} = M_t^F + K^{-1}U^T(p_0 - p_0'),$$

et par conséquent

$$\Delta M_t^{F'} = \Delta M_t^F.$$

Démonstration. Le terme ajouté ne dépend pas de t et s'annule par soustraction. ■

Les transitions sont des coordonnées relatives et sont indépendantes de l'origine ; les déviations sont des coordonnées absolues et en dépendent. L'effet d'un changement d'origine déclaré est donc calculable et connu, ce qui est cohérent avec le Papier 01 (section 2.2), qui autorise un changement d'origine à condition qu'il soit déclaré.

Proposition 2 (Cécité d'une référence dérivée des sorties à la dérive commune).

Hypothèses. (A1) L'encodeur φ est fixe. (A2) Les axes U et l'étalonnage K sont fixes, calculés à partir de $\mathcal{E}_B$ seul. (A3) Deux suites de sorties de même longueur T , indexées par le même $t = 1, \dots, T$, vérifient

$$\varphi(o_t') = \varphi(o_t) + \delta_t, \quad t = 1, \dots, T,$$

avec une dérive moyenne $\bar{\delta} = \frac{1}{T} \sum_{t=1}^T \delta_t$. (A4) La référence dérivée des sorties de chaque suite est calculée à partir de cette suite elle-même : $\tilde{p}_0 = \frac{1}{T} \sum_{t=1}^T \varphi(o_t)$ et $\tilde{p}_0' = \frac{1}{T} \sum_{t=1}^T \varphi(o_t')$. (A5) La lecture de référence déclarée p_0 est calculée à partir de formulations qui ne figurent dans aucune des deux suites, et elle est donc la même pour les deux.

Énoncé. Sous (A1)–(A5), dans l'espace de représentation complet :

- (i) par rapport à la référence dérivée des sorties,

$$\varphi(o_t') - \tilde{p}_0' = (\varphi(o_t) - \tilde{p}_0) + (\delta_t - \bar{\delta});$$

- (ii) par rapport à la référence déclarée,

$$\varphi(o_t') - p_0 = (\varphi(o_t) - p_0) + \delta_t;$$

- (iii) en coordonnées fondées, les déviations par rapport à $\tilde{p}_0$ varient de $K^{-1}U^T(\delta_t - \bar{\delta})$, et les déviations par rapport à p_0 varient de $K^{-1}U^T\delta_t$; en moyenne sur t , la première variation est nulle et la seconde vaut $K^{-1}U^T\bar{\delta}$.

Démonstration. D'après (A3) et (A4), la moyenne donne $\tilde{p}_0' = \tilde{p}_0 + \bar{\delta}$, d'où (i). D'après (A5), p_0 est la même pour les deux suites, d'où (ii). D'après (A1) et (A2), l'application $K^{-1}U^T$ est la même application linéaire pour les deux suites ; en l'appliquant à (i) et (ii) et en prenant la moyenne sur t , on obtient (iii). ■

Corollaires. (a) $\sum_t (\varphi(o_t) - \tilde{p}_0) = 0$: la déviation moyenne par rapport à une référence dérivée des sorties est nulle par construction. (b) **Translation uniforme** : si $\delta_t = \delta$ pour tout t , les déviations par rapport à $\tilde{p}_0$ sont inchangées quelle que soit l'amplitude de δ , et les transitions sont également inchangées ; seules les déviations par rapport à une origine indépendante enregistrent la translation. (c) **Dérive non uniforme** : si les δ_t diffèrent, les déviations individuelles par rapport à $\tilde{p}_0$ varient effectivement, de $\delta_t - \bar{\delta}$; ce qui disparaît, c'est leur moyenne, qui est nulle pour toute dérive. Une référence dérivée des sorties enregistre donc la dispersion d'une dérive autour de sa moyenne, mais

jamais la composante commune $\bar{\delta}$ elle-même. En conséquence, « aveugle à la composante commune » signifie dans cet article aveugle à $\bar{\delta}$, et non insensible à tout changement individuel.

Conditions d'enregistrement. Les énoncés (i) et (ii) valent dans l'espace complet : la référence déclarée y enregistre la totalité de la composante commune $\bar{\delta}$, et la référence dérivée des sorties n'en enregistre rien. En coordonnées fondées, la référence déclarée enregistre la composante commune si et seulement si $U^\top \bar{\delta} \neq 0$ (K étant inversible). La composante de $\bar{\delta}$ orthogonale aux axes retenus n'apparaît pas dans les coordonnées ; elle ne contribue qu'à l'indicateur d'écart r_t . Les conclusions formulées en coordonnées concernent donc la composante projetée $U^\top \bar{\delta}$, et non la dérive complète.

Portée. La proposition concerne le déplacement représentationnel sous un encodeur fixe. L'encodeur est extérieur aux systèmes générateurs et fixe d'une condition à l'autre, de sorte que les sorties de différents modèles générateurs, de différentes versions d'un modèle ou de différentes générations d'auto-entraînement sont toutes encodées par le même φ , et la proposition s'y applique directement. Elle ne couvre pas les situations dans lesquelles les représentations internes d'un modèle générateur sont elles-mêmes utilisées comme encodeur. Qu'un déplacement représentationnel constitue une dérive sémantique est établi indépendamment, par validation humaine (section 8.4). La référence dérivée des sorties ne remplace que l'origine : $\tilde{p}_0$ est substituée à p_0 , tandis que φ , U et K sont maintenus fixes. C'est la condition de référence R1 de la section 8.2.

La Proposition 2 est la forme computationnelle de ce que le Papier 03 décrit comme la troisième relation entre l'origine et le sujet : lorsque la référence est dérivée de la propriété évaluée, l'évaluation perd sa capacité à détecter un changement de cette propriété.

6.14. L'admissibilité du cadre

6.14.1. L'admissibilité méthodologique

$$\text{Adm}_M(F) = C \wedge E \wedge I_{\text{doc}}, \quad C, E, I_{\text{doc}} \in \{0,1\},$$

établie par un examen documenté selon un protocole écrit Π :

- **C (Cohérence) :** Γ_F est reconnue par Rec avec ses trois types de preuves ; $Def \subseteq \{d: \Gamma_F \vdash d\}$; chaque formulation de $\mathcal{E}_{\text{all}}$ réalise une définition de Def ; et aucun matériau de référence n'est importé d'un autre cadrage sans déclaration. $C = 0$ en cas d'indétermination, de reconnaissance erronée ou d'importation non déclarée.
- **E (Explicitation) :** chaque ensemble de $\mathcal{E}_{\text{all}}$ est documenté, et chaque formulation peut être rattachée à R et à Def par le registre du protocole.
- **I_{doc} (Indépendance documentée) :** la provenance de chaque définition de Def et de chaque formulation de $\mathcal{E}_{\text{all}}$ est enregistrée, et aucune n'a été produite par le système évalué, dérivée de ses sorties ou rédigée avec son assistance.

Adm_M est définie sur $(\Gamma_F, R, D(\Gamma_F), Def, \mathcal{E}_{\text{all}})$ et sur les registres de provenance. Elle ne dépend ni de φ ni de O_T .

6.14.2. L'opérabilité computationnelle

$$\text{Op}_{\text{comp}}(F) = Q \wedge I_{\text{comp}} \wedge Cal, \quad Q, I_{\text{comp}}, Cal \in \{0,1\},$$

toutes calculées.

Q (Pouvoir discriminant des axes du cadre opérant). Pour chaque axe tel que $\tilde{u}_k \neq 0$ et $|E_k^\pm| \geq 2$, soient $\sigma_k^\pm$ les écarts-types de $\{(\varphi(e) - c_k^\pm, u_k) : e \in E_k^\pm\}$, et définissons le rapport de séparation orthogonalisé

$$S_k^\perp = \frac{\langle a_k, u_k \rangle}{\sigma_k^+ + \sigma_k^-},$$

défini lorsque $\sigma_k^+ + \sigma_k^- > 0$, et $S_k^\perp = +\infty$ lorsque $\sigma_k^+ + \sigma_k^- = 0$.

Alors

$$Q = \bigwedge_{k=1}^d (a_k \neq 0 \wedge \tilde{u}_k \neq 0 \wedge \gamma_k \geq \gamma_{\min} \wedge S_k^\perp \geq S_{\min}).$$

Les conditions $a_k \neq 0$ et $\tilde{u}_k \neq 0$ excluent les cas dégénérés ; le contenu opératoire de Q réside dans les seuils $\gamma_{\min}$ et $S_{\min}$. Chaque axe se trouve dans exactement l'un des états suivants, et l'état est enregistré pour chaque axe :

État de l'axe	Condition	Contribution à Q	Signification
Discriminant	Défini, matériau suffisant, $\gamma_k \geq \gamma_{\min}$ et $S_k^\perp \geq S_{\min}$	Satisfait Q pour l'axe k	L'axe est opérationnel
Sous le seuil	Défini, matériau suffisant, un seuil non atteint	$Q = 0$	L'encodeur ne discrimine pas l'axe suffisamment pour un usage opérationnel
Indéfini	$a_k = 0$, ou $\tilde{u}_k = 0$	$Q = 0$	Aucune direction opérationnelle n'existe
Matériau insuffisant	Moins de deux formulations dans un pôle, de sorte que $\sigma_k^\pm$ ne peut être calculé	$Q = 0$	Le pouvoir discriminant n'est pas déterminé ; le cadre doit être complété avant de pouvoir être évalué

$Q = 0$ dans le dernier cas exprime que l'opérabilité n'a pas été établie, et non qu'il a été montré que l'axe n'est pas discriminant. Ce cas concerne le matériau de construction du cadre et se résout en le complétant, non en révisant l'axe. Dans tous les cas où $Q = 0$, la caractérisation relationnelle de la section 6.15 reste disponible, puisqu'elle exige $\text{Adm}_M = 1$ et non $\text{Op}_{\text{comp}} = 1$ (section 6.14.3).

I_{comp} (Indépendance computationnelle). Avec $\text{dup}(e, o)$, où $0 \leq \text{dup}(e, o) \leq 1$, une mesure déclarée de quasi-duplication (par exemple une combinaison du recouvrement lexical en n-grammes et de la similarité des plongements),

$$I_{\text{comp}} = [\mathcal{E}_{\text{all}} \cap \mathcal{O}_T = \emptyset] \wedge \left[\max_{e \in \mathcal{E}_{\text{all}}, o \in \mathcal{O}_T} \text{dup}(e, o) < \tau \right].$$

Le seuil τ vise la duplication, et non la proximité sémantique légitime entre des sorties relevant du domaine et des formulations relevant du domaine.

Cal (Calculabilité de l'étalonnage). $\text{Cal} = 1$ si les conditions d'étalonnage de la section 6.11 sont satisfaites, et 0 sinon.

Op_{comp} est définie sur $\varphi(\mathcal{E}_{all})$ et O_T . Q et Cal sont évaluées une fois par cadre et par encodeur ; I_{comp} est réévaluée à mesure que les sorties s'accumulent. I_{doc} concerne l'**origine** du matériau de référence ; I_{comp} concerne la **coïncidence** entre le matériau de référence et les sorties. Un cadre peut satisfaire I_{doc} et ne pas satisfaire I_{comp} , par exemple lorsque le système évalué reproduit un texte de référence rencontré au cours de son entraînement.

6.14.3. États du cadre et règle d'invalidité

Adm_M est évaluée en premier, avant tout encodage ; Op_{comp} après l'encodage.

$Adm_M(F)$	$Op_{comp}(F)$	État	Statut des mesures au sein de F
0	non évaluée	Invalidité fondationnelle	Invalides en tant que mesures fondées. Si $C = 0$ par indétermination, les mesures ne peuvent être calculées qu'avec le statut d'une mesure à référence implicite.
1	0	Admissible, non opérant	Non interprétables au sein de F . Le fondement est sain ; l'échec se situe dans la représentation (Q), dans la coïncidence avec les sorties (I_{comp}) ou dans l'insuffisance du matériau d'étalonnage (Cal).
1	1	Opérant	La mesure se poursuit.

Règle d'invalidité (Loi 3). Si $Adm_M(F) = 0$, toute quantité calculée au sein de F — M_t^F , ΔM_t^F , D_t , r_t , la classification des transitions et toute appréciation de clôture — est invalide en tant que détail méthodologique, quelle que soit sa précision computationnelle. Si I_{comp} échoue à l'instant T , les mesures impliquant des sorties à partir de ce point ne sont pas interprétables tant que la coïncidence n'est pas résolue.

Mesures au sein d'un cadre et mesures de sa réalisation. Les mesures *au sein* du cadre sont M_t^F , ΔM_t^F , D_t , r_t et la classification des transitions. Les mesures *de la réalisation du cadre par l'encodeur* sont le profil de réalisation $\mathcal{R}(\varphi; F)$, la classification de réalisation et l'indice d'orientation $\mathcal{O}$ (section 6.15). La non-interprétabilité des mesures dans un cadre admissible mais non opérant concerne les mesures au sein du cadre ; elle n'empêche pas de mesurer comment l'encodeur réalise le cadre, puisque l'échec de Q est précisément ce qu'une telle mesure enregistre. La caractérisation relationnelle exige donc $Adm_M = 1$ et n'exige pas $Op_{comp} = 1$. Adm_M , Op_{comp} et $(\mathcal{R}, \mathcal{O})$ mesurent trois choses distinctes : l'admissibilité méthodologique, l'opérabilité computationnelle et la relation entre le cadre et l'encodeur.

6.15. Caractérisation relationnelle de la centralité incorporée

Cadres et axes. Soient Γ_A et Γ_B deux centralités, chacune reconnue par Rec , et soient F_A et F_B deux cadres pour le même concept fondés sur elles, chacun avec $Adm_M = 1$. Aucune des deux centralités n'est placée au-dessus de l'autre ni imposée à l'autre : la caractérisation décrit l'encodeur par rapport aux deux cadrages. Les axes de chaque cadre sont répartis en

- **axes partagés** $\mathcal{A}_{sh}$: des distinctions spécifiées une seule fois, par une définition admise sous les deux centralités avec le même contenu de pôles, et formulées **séparément au sein de chaque cadre** selon le même protocole, de sorte qu'elles captent les effets de la qualité rédactionnelle et du style propres à chaque cadre ;
- **axes distinctifs** $\mathcal{A}_{dist}$: des distinctions dérivées des éléments distinctifs de chaque centralité, sans équivalent dans l'autre cadre.

Niveaux des données. Une *distinction* k est identifiée une seule fois. Une *instance d'axe* (k, f) est la distinction k au sein du cadre f . Un *ensemble de formulations* (k, f, w) est l'ensemble des formulations de pôles rédigées par le groupe de rédacteurs w pour l'instance d'axe (k, f) .

Séparation brute. La séparation brute est définie sur la direction brute de chaque instance d'axe, indépendamment de l'ordre d'orthogonalisation. Trois quantités sont maintenues distinctes :

- le **vecteur de différence brut** $a_k = c_k^+ - c_k^-$ (section 6.6), qui n'est *pas* normalisé ; sa longueur $\|a_k\|$ est la distance entre les deux centroïdes des pôles ;
- la **direction brute unitaire** $v_k = a_k / \|a_k\|$, définie lorsque $a_k \neq 0$, avec $\|v_k\| = 1$;
- les **dispersions des pôles** s_k^+ et s_k^- : les écarts-types des projections $\langle \varphi(e), v_k \rangle$ sur $e \in E_k^+$ et sur $e \in E_k^-$ respectivement, calculés à partir des formulations regroupées des deux groupes de rédacteurs, avec au moins le minimum déclaré par pôle.

Le numérateur de la séparation est la distance entre les pôles le long de leur propre direction. Puisque $\langle c_k^+ - c_k^-, v_k \rangle = \langle a_k, a_k \rangle / \|a_k\| = \|a_k\|$, il peut s'écrire de manière équivalente comme la différence entre les projections moyennes des deux pôles,

$$\|a_k\| = \langle c_k^+, v_k \rangle - \langle c_k^-, v_k \rangle.$$

La séparation brute est alors :

- si $a_k \neq 0$ et $s_k^+ + s_k^- > 0$:

$$\bar{S}_k = \min \left(\frac{\|a_k\|}{s_k^+ + s_k^-}, S_{\max} \right);$$

- si $a_k \neq 0$ et $s_k^+ + s_k^- = 0$: $\bar{S}_k = S_{\max}$;
- si $a_k = 0$: $\bar{S}_k = S_{\text{floor},k}$.

Le numérateur utilise a_k non normalisé, jamais v_k ; la direction unitaire n'intervient qu'à travers les projections qui définissent $s_k^\pm$. La même définition est utilisée pour la valeur observée, pour chaque valeur permutée dans le test de réalisation et dans l'indice d'orientation.

$S_{\max}$ est un plafond déclaré, et $S_{\text{floor},k}$ est défini par la règle de plancher ci-dessous. Toute valeur vérifie $0 < \bar{S}_k \leq S_{\max}$, de sorte que $\log \bar{S}_k$ est fini.

Réalisation d'un axe. Une instance d'axe est **réalisée** si la séparation de ses pôles est statistiquement significative selon un test de permutation qui respecte la manière dont les formulations ont été produites, avec contrôle du taux de fausses découvertes sur l'ensemble des instances d'axes d'un encodeur. Le dispositif du test est spécifié à l'Annexe A.7. v_{sh} et v_{dist} désignent les proportions d'instances d'axes partagés et distinctifs réalisées. La classification de réalisation est rapportée à côté de l'indice d'orientation et n'y entre pas.

Trois conditions maintenues distinctes. La règle de plancher sépare trois conditions qui ne s'impliquent pas mutuellement :

- **Séparation observée nulle** : $a_k = 0$, c'est-à-dire que les deux centroïdes des pôles coïncident. Cela n'implique pas que l'encodeur confonde les formulations au sein de chaque pôle.
- **Effondrement dans φ** : l'encodeur associe toutes les formulations de l'instance d'axe numériquement au même point, ce qui est défini par une condition observable indépendante,

$$\max_{e, e' \in E_k^+ \cup E_k^-} \|\varphi(e) - \varphi(e')\| \leq \eta_{\text{coll}},$$

où η_{coll} est une tolérance numérique déclarée. L'effondrement n'est établi que par cette condition, jamais inféré à partir de la distribution nulle.

- **Plancher non estimable** : la distribution nulle de permutation de l'instance d'axe ne permet pas d'estimer le niveau du hasard — sa médiane est nulle, ou il reste moins de N_{null} valeurs nulles valides. C'est une propriété du test sous le protocole, et non une preuve concernant l'encodeur.

Valeur d'un axe de séparation nulle. Une instance d'axe avec $a_k = 0$ reçoit une valeur déclarée de niveau du hasard $S_{\text{floor},k}$, estimée à partir de sa propre distribution nulle de permutation ou, à défaut, à partir des distributions nulles regroupées des deux cadres. La règle, et les cas dans lesquels elle ne peut être appliquée, sont spécifiés à l'Annexe A.8. La valeur est uniquement diagnostique : elle entre dans le profil de réalisation et dans l'indice d'orientation, et elle ne transforme pas un axe indéfini en axe opérationnel (section 6.6).

Profil de réalisation. Le profil de réalisation de l'encodeur par rapport à un cadre est

$$\mathcal{R}(\varphi; F) = (\{\bar{S}_k\}_k, \{\gamma_k\}_k, \bar{r}_H, v_{\text{sh}}, v_{\text{dist}}),$$

où $\bar{r}_H$ est l'indicateur d'écart moyen sur H . Les composantes mesurent des propriétés différentes : $\bar{S}_k$ la **réalisation** de chaque distinction ; γ_k la **structure** du cadre ; $\bar{r}_H$ la **couverture** du domaine ; et $v_{\text{sh}}, v_{\text{dist}}$ la proportion de distinctions effectivement réalisées.

Indice d'orientation. L'indice d'orientation agrège les séparations brutes par moyennes géométriques au sein de chaque cellule (cadre $\times$ type d'axe), calculées comme moyennes arithmétiques de $\log \bar{S}_k$ sur toutes les instances d'axes de la cellule. Chaque instance d'axe, réalisée ou non, entre dans la comparaison avec sa valeur $\bar{S}_k$; aucun axe n'est exclu, de sorte qu'aucun cadre ne paraît mieux réalisé par la suppression de distinctions que l'encodeur représente faiblement. Le score de réalisation distinctive d'un cadre est

$$\text{Real}(\varphi; F) = \frac{1}{|\mathcal{A}_{\text{dist}}|} \sum_{k \in \mathcal{A}_{\text{dist}}} \log \bar{S}_k - \frac{1}{|\mathcal{A}_{\text{sh}}|} \sum_{k \in \mathcal{A}_{\text{sh}}} \log \bar{S}_k,$$

et l'indice d'orientation est

$$\mathcal{O}(\varphi; F_A, F_B) = \text{Real}(\varphi; F_A) - \text{Real}(\varphi; F_B) = \log \frac{G_{\text{dist}}^A / G_{\text{sh}}^A}{G_{\text{dist}}^B / G_{\text{sh}}^B},$$

où G désigne la moyenne géométrique de $\bar{S}_k$ sur les axes indiqués. $\mathcal{O}$ est l'interaction entre cadre et type d'axe dans un plan 2×2 . La classification de réalisation, et donc la procédure de multiplicité, ne détermine que v_{sh} et v_{dist} ; elle n'entre ni dans Real ni dans $\mathcal{O}$.

Conditions de définition. La moyenne géométrique d'une cellule (cadre $\times$ type d'axe) est définie lorsque la cellule contient au moins une instance d'axe et que chaque instance d'axe qu'elle contient a une valeur $\bar{S}_k$, ce qui est le cas de toute instance avec $a_k \neq 0$ et d'une instance avec $a_k = 0$ une fois son plancher attribué. $\mathcal{O}(\varphi; F_A, F_B)$ est défini si et seulement si les quatre cellules — $\mathcal{A}_{\text{dist}}$ et $\mathcal{A}_{\text{sh}}$ dans F_A et dans F_B — sont définies. Il n'est utilisé comme résultat principal que si, en outre, chaque cellule contient au moins le nombre minimal déclaré n_{cell} d'instances d'axes, après exclusion des paires partagées qui échouent au jugement d'équivalence (Annexe A.6). L'indice se trouve donc dans exactement l'un de quatre états : **défini et principal** ; **défini mais descriptif** (une cellule en dessous de n_{cell} , ou des valeurs de plancher regroupé au-dessus du plafond déclaré de l'Annexe A.8) ; **non calculable sous la règle de**

plancher (un plancher requis ne peut être attribué) ; ou **indéfini** (une cellule vide). L'état est toujours rapporté avec l'indice, et seul le premier permet une caractérisation de Γ_L .

Axes distinctifs faibles. Puisque chaque axe entre avec sa valeur, les axes distinctifs faiblement réalisés ou non réalisés abaissent $\text{Real}(\varphi; F)$ de leur ampleur effective, qui se situe près du niveau du hasard $S_{\text{floor},k}$. $\mathcal{O}$ demeure donc interprétable dans ce cas : une contribution négative des axes distinctifs d'un cadre enregistre que l'encodeur aplatit les distinctions propres à ce cadre. Une valeur proche de zéro, $\mathcal{O} \approx 0$, ne détermine pas à elle seule son interprétation, car elle peut survenir de plus d'une manière : les axes distinctifs des deux cadres peuvent être faibles alors que les axes partagés sont forts (aplatissement des deux cadrages), ou les séparations des deux cadres peuvent simplement être comparables à un niveau quelconque (absence d'orientation différentielle). Les deux cas ne se distinguent que par les niveaux de séparation dans chacune des quatre cellules — les moyennes géométriques $G_{\text{dist}}^A, G_{\text{sh}}^A, G_{\text{dist}}^B, G_{\text{sh}}^B$ — ainsi que par v_{sh} et v_{dist} . Ces valeurs au niveau des cellules sont toujours rapportées avec $\mathcal{O}$, et une valeur de $\mathcal{O}$ n'est jamais interprétée sans elles.

Niveaux de l'affirmation. Quatre affirmations relatives à la centralité incorporée sont maintenues distinctes dans tout l'article : (1) l'existence de Γ_L dans toute représentation apprise est une conséquence méthodologique de MANHAJ (section 3.4) ; (2) son *contenu* relève de l'examen et de la caractérisation ; (3) l'orientation de la réalisation d'un encodeur entre des centralités déclarées, mesurée par $\mathcal{O}$, est un résultat computationnel défini par le modèle ; (4) l'attribution de cette orientation aux *données d'entraînement* de l'encodeur est une conclusion causale, que la variation entre encodeurs n'établit pas à elle seule, puisque les encodeurs diffèrent aussi par leur architecture, leur dimensionnalité, leurs objectifs et leurs prétraitements ; elle exige une comparaison contrôlée d'encodeurs proches par l'architecture et les objectifs mais différant par des données d'entraînement documentées.

Interprétation. $\mathcal{O} > 0$ établit une orientation relative de la réalisation de l'encodeur vers les distinctions propres à F_A , par comparaison avec celles propres à F_B , pour les concepts et les axes examinés. Il caractérise Γ_L par rapport à une paire déclarée de centralités ; il n'identifie pas Γ_L dans sa totalité. Une faible réalisation des distinctions d'un cadre concerne l'encodeur, et non l'admissibilité de la centralité sur laquelle le cadre est fondé, qui est établie par Adm_M indépendamment de φ . Une orientation nulle ne concerne pas l'existence de Γ_L , qui est établie méthodologiquement.

6.16. Les cadres non réalisés comme état diagnostique

L'état ($\text{Adm}_M = 1, \text{Op}_{\text{comp}} = 0$) avec échec de Q identifie des distinctions requises par un domaine méthodologiquement sain que l'encodeur ne réalise pas. Parce que le fondement est établi indépendamment de φ , l'échec est imputable à la représentation plutôt qu'au cadre. Le cadre fondé fonctionne ainsi comme un instrument de diagnostic de la représentation apprise, et contribue à la caractérisation relationnelle de Γ_L .

6.17. Définition a priori et lectures a posteriori

Le cadre fondé est a priori dans sa définition — la centralité, le domaine, la référence, les pôles et les unités de mesure — et a posteriori dans ses lectures, puisque $p_0, c_k^\pm, U$ et $\varphi(o_t)$ sont calculés au moyen de l'encodeur appris. Le fondement de l'espace appris est implicite ; le fondement de l'espace fondé est reconnu, déclaré et examiné, son admissibilité méthodologique étant établie indépendamment de son encodage. L'encodeur appris fournit la matière première de la mesure ; MANHAJ fournit le cadre.

6.18. Séquence du modèle

Étape	Opération	Issue en cas d'échec
1	Reconnaissance de Γ_F par <i>Rec</i>	Indétermination : statut de mesure à référence implicite
2	$D(\Gamma_F), R, Def \subseteq \{d: \Gamma_F \vdash d\}$	—
3	Ensembles de formulations déclarés $\mathcal{E}_B$ et H , avec registres de provenance	—
4	$Adm_M(F) = C \wedge E \wedge I_{doc}$	Invalidité fondationnelle
5	Encodage de $\mathcal{E}_B : p_0, a_k, \tilde{u}_k, u_k, \gamma_k, \kappa_k, \rho, \varepsilon$	—
6	$Op_{comp}(F) = Q \wedge I_{comp} \wedge Cal$	Admissible, non opérant (la réalisation reste mesurée)
7	Validation sur H (diagnostic rapporté)	—
8	$M_t^F, D_t, r_t, \Delta M_t^F$, classification des transitions	—
9	Détection de la clôture et validité des transitions (Papier 02)	—

6.19. Correspondance entre MANHAJ et les éléments formels

Source méthodologique	Éléments formels
Théorie de la centralité doctrinale	$\Gamma \vdash d ; D(\Gamma_F) ; Rec ; \text{condition } C ; \text{centralité incorporée } \Gamma_L ; \mathcal{R}(\varphi; F) \text{ et } \mathcal{O}$
Loi de la dualité de la direction de l'examen (Loi 2)	Règle de construction (section 6.3) ; évaluation de Adm_M avant tout encodage
Loi de l'origine coordonnée (Loi 4)	$p_0 ; M_t^F$ avec origine en zéro ; $D_t ; I_{doc}$ et I_{comp} ; Propositions 1 et 2
Loi de la justesse (Loi 3)	États du cadre et règle d'invalidité (section 6.14.3)
Théorie du sens (Document de référence, section 6.3.4)	$M_t^F ; r_t$; écart et retour dans la classification des transitions

6.20. Notation

Symbole	Signification
$\Gamma_F, \Gamma_L, \Gamma_C$	Centralités du cadre fondé, de l'encodeur appris et du contenu représenté
Γ_A, Γ_B	Deux centralités déclarées dans la caractérisation relationnelle
$\Gamma \vdash d$	La définition d est admise au sein du cadrage Γ
<i>Rec</i> , Π	Procédure de reconnaissance ; protocole écrit de l'examen méthodologique
φ, n, O_T	Encodeur appris ; dimension de son espace ; sorties mesurées à l'instant T

Symbole	Signification
$\mathcal{E}_B, H, \mathcal{E}_{\text{all}}$	Ensemble de construction et d'étalonnage ; ensemble de validation ; leur réunion
$E_0, E_k^\pm, \mathcal{P}$	Ensembles de formulations de référence, de pôles et de paraphrases
p_0	Lecture du sens de référence (origine de l'espace fondé)
a_k, v_k	Vecteur de différence brut (non normalisé) des pôles ; sa direction unitaire
$\tilde{u}_k, u_k, U$	Directions d'axes orthogonalisées et normalisées ; matrice des axes
$s_k^\pm$	Dispersions des pôles le long de v_k (séparation brute)
d, γ_k	Nombre d'axes ; rapport de composante orthogonalisée
κ_k, K	Coefficients d'étalonnage ; matrice d'étalonnage
$M_t^F, \Delta M_t^F, D_t$	État du sens, transition et déviation dans l'espace fondé
r_t, ρ, ε	Indicateur d'écart ; seuil d'écart ; seuil de changement
N_ρ, N_ε	Matériau d'étalonnage minimal pour ρ et ε
$S_k^\perp, S_{\min}, \gamma_{\min}$	Rapport de séparation orthogonalisé et seuils de la condition Q
$\bar{S}_k, S_{\max}, S_{\text{floor},k}$	Séparation brute ; son plafond ; son plancher
$N_{\text{null}}, N_{\text{pool}}, \lambda_{\text{pool}}, \eta_{\text{coll}}$	Nombre minimal de valeurs nulles valides ; nombre minimal de médianes propres aux axes pour le plancher regroupé ; plafond des valeurs de plancher regroupé par cellule ; tolérance d'effondrement
B, m, m_w, q, H_m	Permutations ; instances d'axes de la famille de tests ; rédacteurs par instance d'axe ; niveau du FDR ; somme harmonique
$\mathcal{A}_{\text{sh}}, \mathcal{A}_{\text{dist}}, n_{\text{cell}}$	Axes partagés et distinctifs ; nombre minimal d'instances d'axes par cellule
$\mathcal{R}(\varphi; F), \bar{r}_H, \nu_{\text{sh}}, \nu_{\text{dist}}$	Profil de réalisation ; écart moyen sur H ; proportions d'axes réalisés
$\text{Real}(\varphi; F), \mathcal{O}(\varphi; F_A, F_B)$	Score de réalisation distinctive ; indice d'orientation
dup, τ	Mesure de quasi-duplication ; son seuil
$\text{Adm}_M, C, E, I_{\text{doc}}$	Admissibilité méthodologique et ses conditions
$\text{Op}_{\text{comp}}, Q, I_{\text{comp}}, \text{Cal}$	Opérabilité computationnelle et ses conditions
V_t	Validité de transition (Papier 02)

7. Intégration avec le Papier 02

L'espace fondé complète le cadre computationnel du Papier 02 en quatre points.

1. **Le fondement avant le processus.** Le garde-fou de la section 6.14 précède l'état a priori $A_0 = (R, D, Def, C_a, O)$ du Papier 02 : avant que la validité de la production du sens ne soit examinée, l'admissibilité du cadre dans lequel le sens est mesuré est établie. La Loi 3 opère ainsi à deux niveaux — le cadre (le présent article) et la transition (Papier 02).
2. **Un suivi relatif selon des dimensions déclarées.** Le ΔM_t du Papier 02 devient ΔM_t^F , exprimé selon des axes déclarés, en unités déclarées, avec un seuil de changement ε étalonné à partir de paraphrases déclarées.
3. **La référence absolue.** La déviation D_t par rapport à la référence déclarée, absente du Papier 02, est ajoutée. Le suivi du Papier 02 acquiert ainsi les coordonnées absolues qui font défaut au suivi relatif.
4. **Une clôture calculable.** La clôture au sens de la Loi 7 — l'atteinte d'une constante cognitive C — devient calculable au sein de l'espace fondé à la résolution déclarée ε : il est établi que l'état du sens demeure à une constante sur la fenêtre de persistance, les transitions classées « aucun changement de sens détecté » à la section 6.12 constituant le signe nécessaire de cette condition. Il s'agit d'une clôture *détectée à la résolution ε* , et non de l'affirmation que $\Delta M_t = 0$ exactement ; les écarts et les retours en sont séparés, de sorte qu'un changement de cadre n'est pris ni pour une clôture ni pour un sens nouveau. La Loi 7 fournit ainsi au traitement dérivational d'un système d'IA un critère d'arrêt fondé : le traitement prend fin lorsqu'il ne produit plus de sens dérivational nouveau au sein du domaine déclaré, au lieu de se poursuivre indéfiniment ou de s'arrêter à une limite arbitraire.

La séquence qui en résulte prolonge celle du Papier 02 :

Centralité → Cadre fondé → Admissibilité → État du sens → Transition → Validité → Clôture

8. Dispositifs expérimentaux

8.1. Statut et principes

Cet article est une contribution théorique et méthodologique. Ses affirmations reposent sur les lois de MANHAJ et sur le modèle formel ; les expériences ci-dessous sont des **dispositifs** qui précisent comment chaque capacité dérivée des lois peut être démontrée empiriquement. Leur protocole opérationnel — annotation, échantillonnage, procédures de test, analyse statistique et constantes déclarées — est exposé à l'Annexe A. Les mises en œuvre de preuve de concept font partie des travaux ultérieurs consacrés à la méthodologie MANHAJ.

Quatre principes gouvernent les dispositifs. (1) **Référence de comparaison forte** : le cadre fondé est comparé non seulement à des références dérivées des sorties et relatives, mais aussi à une référence externe fixe, de sorte que ce que les expériences testent est la contribution des exigences de MANHAJ, et non celle de la seule fixité de la référence. (2) **Vérité de terrain indépendante** : les changements sémantiques et les distinctions du domaine sont établis par un jugement humain indépendant des mesures testées. (3) **Attentes spécifiées à l'avance** : le résultat attendu de chaque expérience est énoncé à l'avance. (4) **Attribution par ablation** : une condition qui partage la technique de mise en œuvre mais non l'ordre fondationnel isole la contribution des lois.

8.2. Conditions de référence

Code	Condition de référence	Indépendante des sorties	Fondée selon MANHAJ
R1	Centroïde dérivé des sorties, substitué à p_0 , φ , U et K étant maintenus fixes	Non	Non
R2	Comparaison d'une version à l'autre (relative)	Non	Non
R3	Référence de comparaison forte : ensemble externe fixe de textes relevant du domaine, rédigés par des humains et sélectionnés, comparés par similarité standard	Oui	Non
R3a	Ablation : référence externe fixe avec des axes construits par la même technique que R4 à partir d'exemples de commodité	Oui	Non
R4	Cadre fondé	Oui	Oui

R3a partage la technique de mise en œuvre de R4 et n'en diffère que par l'ordre fondationnel. La comparaison de R4 avec R3a isole donc la contribution des exigences de MANHAJ.

8.3. Expérience E1 — Le garde-fou fondationnel (Loi 3)

Question. L'application de la Loi de la justesse au niveau du cadre de mesure détecte-t-elle des défauts que les vérifications conventionnelles laissent non révélés ?

Dispositif. Des cadres sains sont comparés à des cadres portant chacun un défaut connu : contamination de la référence par les sorties, matériau de référence produit par le système évalué, importation non déclarée depuis un autre cadrage et axes non discriminants. Les métriques standard, les procédures d'audit établies et le garde-fou fondationnel sont appliqués à chaque cadre, et les conclusions tirées dans chaque cadre sont comparées à la vérité de terrain (Annexe A.3).

Résultat attendu. Les défauts représentationnels (contamination, axes non discriminants) sont détectés par les vérifications computationnelles du garde-fou et peuvent l'être aussi par les procédures d'audit établies. Les **défauts fondationnels** (provenance, importation non déclarée) laissent inchangées les métriques standard et les vérifications conventionnelles, parce qu'ils concernent l'origine et la cohérence de la référence plutôt que ses propriétés de surface ; ils sont détectés par la vérification méthodologique Adm_M . Les conclusions tirées au sein de cadres portant des défauts fondationnels sont non fiables au regard de la vérité de terrain, bien que leurs métriques paraissent normales.

Ce qu'elle établirait. Si le résultat attendu est obtenu : que la catégorie de l'invalidité fondationnelle, introduite par la Loi 3, fournit une protection dont la mesure conventionnelle est dépourvue.

8.4. Expérience E2 — Référencement absolu et dérive (Loi 4)

Question. Le référencement absolu au sein d'un cadre fondé fournit-il une détection de la dérive et une attribution conceptuelle au-delà des références relatives et des références externes fixes, et la composante commune de la dérive est-elle significative en pratique ?

Dispositif. Une dérive sémantique construite et validée par jugement humain, des contrôles préservant le sens (paraphrase, style, longueur) et une dérive collective naturelle au fil de générations d'auto-

entraînement ou de versions de modèles sont mesurés sous les conditions de référence R1–R4 (Annexe A.4).

Résultat attendu. R1 n'enregistre pas la composante commune de la dérive, comme l'exige la Proposition 2 ; R2 enregistre le changement d'une étape à l'autre, mais aucun déplacement cumulé par rapport à une référence. R4 distingue la dérive sémantique du changement préservant le sens au moins aussi bien que R3 et, en outre, attribue la dérive à des dimensions conceptuelles déclarées — capacité que R3 ne possède pas. R4 dépasse R3a en matière d'attribution, de sorte que l'avantage provient de l'ordre fondationnel. La composante commune représente une part substantielle de la dérive naturelle.

Ce qu'elle établirait. Si le résultat attendu est obtenu : que la Loi de l'origine coordonnée ajoute un référencement absolu et conceptuellement décomposé à une pratique qui travaille avec des quantités relatives, et que les références dérivées des sorties échouent sur une composante pratiquement significative de la dérive réelle.

8.5. Expérience E3 — L'ordre fondationnel (Loi 2)

Question. Les axes fondés selon l'ordre fondationnel complet mesurent-ils les distinctions du domaine conceptuel plus fidèlement que des axes construits sans lui ?

Dispositif. Les axes fondés (R4) sont comparés à des axes construits par la même technique à partir d'exemples de commodité (R3a) et à des axes extraits des données sans supervision (R4u), au regard des jugements d'experts indépendants du domaine sur des éléments tenus à l'écart, avec des contrôles qui séparent l'ordre fondationnel de la sélection experte (Annexe A.5).

Résultat attendu. R4 dépasse R3a et R4u en matière d'alignement sur les jugements des experts et de séparation entre éléments relevant du domaine et éléments extérieurs au domaine.

Ce qu'elle établirait. Si le résultat attendu est obtenu : que l'ordre a priori exigé par la Loi de la dualité de la direction de l'examen produit des axes qui mesurent les distinctions propres au domaine.

8.6. Expérience E4 — Caractérisation relationnelle de la centralité incorporée

Question. Comment la réalisation d'un encodeur appris est-elle orientée entre deux cadres fondés sur des centralités déclarées différentes, et comment cette orientation varie-t-elle entre encodeurs formés à partir de corpus différents ?

Dispositif. Pour deux concepts (candidats : *famille* et *progrès*), deux cadres sont fondés sur deux centralités reconnues, chacun avec des axes partagés et distinctifs. Trois encodeurs formés à partir de corpus différents réalisent les deux cadres, et l'indice d'orientation $\mathcal{O}$ de la section 6.15 est calculé avec ses séparations au niveau des cellules et ses proportions de réalisation (Annexe A.6).

Résultat attendu. Une orientation systématique non nulle sur les axes distinctifs alors que les axes partagés sont réalisés de manière comparable ; une stabilité entre groupes de rédacteurs et entre concepts ; et une variation de l'orientation entre encodeurs, décrite comme une covariation avec leurs corpus d'entraînement. Une attribution causale de cette variation aux données d'entraînement n'est pas attendue de ce dispositif, puisque les encodeurs diffèrent aussi par leur architecture, leur dimensionnalité, leurs objectifs et leurs prétraitements. Lorsque sont disponibles des encodeurs proches par l'architecture et les objectifs mais différant par des données d'entraînement documentées, leur comparaison fournit le dispositif plus contrôlé qu'exige une telle attribution.

Ce qu’elle établirait. Si le résultat attendu est obtenu : que le contenu de la centralité incorporée Γ_L , dont MANHAJ établit l’existence méthodologiquement, devient mesurable de manière relationnelle.

8.7. Des résultats aux affirmations

Résultat	Contribution que le résultat établirait
E1 conforme aux attentes	La Loi 3 appliquée au niveau du cadre détecte des défauts fondationnels — provenance et importation non déclarée — que les métriques standard et les procédures d’audit établies ne révèlent pas
E2 conforme aux attentes	Le référencement absolu au sein d’un cadre fondé fournit une attribution conceptuelle de la dérive au-delà des références externes fixes, et les références dérivées des sorties échouent sur une composante significative de la dérive réelle
E3 conforme aux attentes	L’ordre fondationnel de la Loi 2 produit des axes qui mesurent les distinctions propres au domaine
E4 conforme aux attentes	L’orientation de la réalisation d’un encodeur appris entre des centralités déclarées est mesurable, et elle varie entre encodeurs formés à partir de corpus différents ; une attribution causale aux données d’entraînement exige la comparaison contrôlée de la section 8.6
Résultats différent des attentes	Analysés au niveau où ils apparaissent — mise en œuvre, capacité testée ou portée — et utilisés pour affiner la spécification opérationnelle ; les exigences méthodologiques des trois lois demeurent au niveau méthodologique, où elles sont établies

Les cadres méthodologiquement admissibles mais non opérants sont analysés individuellement comme un résultat distinct : ils identifient des distinctions requises par un domaine fondé que l’encodeur ne réalise pas (section 6.16).

9. Discussion

9.1. Ce qui change lorsque le cadre est fondé

L’objet de l’examen méthodologique change. La mesure à référence implicite examine les représentations encodées et les relations calculées en leur sein. La mesure fondée examine en outre les conditions dans lesquelles ces relations valent comme mesures du sens : la centralité au sein de laquelle le domaine se constitue, le domaine lui-même, l’origine, les axes et leur étalonnage. L’encodeur reste indispensable, mais le statut de ses sorties change. Elles sont lues au sein d’un cadre déclaré dont l’admissibilité n’est pas déterminée par le succès numérique de l’encodeur.

9.2. L’origine et le référencement absolu

Les quantités relatives — transitions, différences d’une version à l’autre, distances au centroïde des sorties actuelles — sont les instruments dominants du suivi du sens en IA. La Proposition 1 montre que de telles quantités sont aveugles au choix de l’origine, et la Proposition 2 montre qu’une référence construite à partir des sorties elles-mêmes est aveugle à leur dérive commune. La Loi de l’origine coordonnée fournit ce qui manque à ces instruments : une position absolue par rapport à une référence déclarée, décomposée selon des dimensions conceptuelles déclarées. Dans les termes de la section 4.2,

MANHAJ ajoute la coordonnée absolue à une pratique qui ne travaille qu'avec des coordonnées relatives.

9.3. L'invalidité fondationnelle comme protection

La Loi de la justesse introduit dans la mesure en IA une catégorie dont la pratique actuelle ne dispose pas : une famille de résultats invalides à leur source. Un jeu d'évaluation contaminé, une référence rédigée par le système évalué, ou une référence importée d'un autre cadrage sans déclaration peuvent produire des métriques calculées correctement et qui ne donnent aucun signe du défaut ; E1 est conçue pour tester dans quelle mesure il en est ainsi. Le garde-fou fondationnel n'examine pas ces métriques une à une ; il examine le cadre qui les a produites, et son verdict s'applique à toutes. C'est la forme pratique du principe selon lequel ce qui n'est pas licite, ses détails ne sont pas valides.

9.4. L'exécutable et son système d'exploitation

La pratique actuelle traite les sorties des systèmes d'IA, et les systèmes eux-mêmes, comme des exécutables à analyser et à réparer. MANHAJ exige que les uns et les autres soient cadrés : le contenu par sa centralité Γ_C , et le modèle par la centralité Γ_L incorporée dans ses données. La caractérisation relationnelle de la section 6.15 rend la seconde exigence opérationnelle. Elle fait passer la question de savoir quel cadrage une représentation apprise réalise du domaine de l'affirmation à celui de la mesure, tout en maintenant l'existence de ce cadrage là où MANHAJ la situe : au niveau méthodologique.

9.5. Relation avec les représentations apprises

L'espace fondé ne remplace pas les représentations apprises. Il les utilise comme matière première de la mesure et place sur elles un cadre qu'elles ne fournissent pas. Il s'agit d'une répartition des rôles : l'encodeur appris engendre une structure numérique ; le cadre fondé établit les conditions de référence sous lesquelles cette structure est lue. Un encodeur appris peut donc être évalué au regard d'un cadre fondé, et une telle évaluation est conçue pour montrer où il réalise les distinctions déclarées et où il ne les réalise pas — information inaccessible lorsque la représentation sert de référence à elle-même.

9.6. Trois niveaux d'affirmations

L'article formule des affirmations à trois niveaux. Les affirmations méthodologiques découlent des lois de MANHAJ : le cadre doit être fondé, l'origine doit être indépendante, un cadre non admissible ne produit aucun détail valide, et toute représentation apprise opère sous une centralité incorporée. Les affirmations computationnelles découlent du modèle : les Propositions 1 et 2, le Lemme 1 et les propriétés des fonctions d'admissibilité. Les affirmations empiriques sont, à ce stade, des hypothèses : les expériences E1–E4 sont conçues pour les tester, et leurs résultats ne sont pas présentés ici. Maintenir les niveaux distincts donne à chacun sa force propre : les affirmations méthodologiques et les résultats mathématiques n'attendent pas les expériences, et les expériences, une fois réalisées, montreront concrètement où et dans quelle mesure la différence méthodologique importe.

10. Statut épistémique et portée

- **Outils et conditions.** Les outils représentationnels utilisés — directions conceptuelles, projection, résidus, détection de chevauchement — sont établis. La contribution réside dans les conditions que MANHAJ impose à leur usage et dans les capacités que ces conditions produisent.

- **Définition a priori, lectures a posteriori.** Le cadre fondé est a priori dans sa définition et a posteriori dans ses lectures, qui passent par l'encodeur appris. La différence entre les deux espaces est que le fondement de l'un est implicite, tandis que le fondement de l'autre est déclaré et examiné.
- **Examen méthodologique.** Les conditions C , E et I_{doc} sont établies par un examen documenté selon les protocoles Rec et Π , rédigés dans le cadre du protocole opérationnel avant les expériences, avec le domaine de la famille comme exemple développé.
- **Portée formelle.** La Proposition 2 concerne le déplacement représentationnel sous un encodeur fixe ; seule la composante projetée $U^T \bar{\delta}$ de la dérive commune apparaît dans les coordonnées, et son interprétation sémantique repose sur la validation humaine.
- **Protocole opérationnel.** Les procédures de test, les seuils et les tailles d'échantillon de l'Annexe A sont des décisions de mise en œuvre ; ils ne concernent pas les affirmations méthodologiques, et les mises en œuvre de preuve de concept suivront dans des travaux ultérieurs.
- **Constantes opérationnelles.** S_{max} , N_{null} , N_{pool} , λ_{pool} , η_{coll} , q , γ_{min} , S_{min} , τ , N_{ρ} , N_{ε} et les niveaux de percentile sont des choix de mise en œuvre, fixés avant les expériences et examinés par analyse de sensibilité.
- **Ordre des axes.** L'orthogonalisation de Gram–Schmidt rend les axes opérationnels dépendants de leur ordre déclaré ; l'ordre est déclaré au sein de Def , et la caractérisation relationnelle utilise les directions brutes, qui n'en dépendent pas.

11. Conclusion

Cet article a examiné le fondement de la mesure du sens en intelligence artificielle. Les valeurs qui expriment le sens dans les systèmes d'IA — similarités, dérives, scores — sont calculées avec précision par rapport à des références implicites, dérivées des sorties mesurées ou compromises. MANHAJ traite ce problème à sa source.

Trois lois produisent le changement. La Loi de la dualité de la direction de l'examen exige que le cadre de mesure soit fondé avant la mesure, selon l'ordre centralité doctrinale $\rightarrow$ domaine conceptuel $\rightarrow$ origine coordonnée $\rightarrow$ coordonnées du sens. La Loi de l'origine coordonnée exige une origine explicite, indépendante de la propriété mesurée, et ajoute ainsi des coordonnées absolues à une pratique qui travaille avec des coordonnées relatives. La Loi de la justesse exige qu'un cadre non admissible ne produise aucune mesure valide, et introduit ainsi la catégorie de l'invalidité fondationnelle.

Le modèle formel met en œuvre ces exigences dans un espace fondé lu au moyen d'un encodeur appris. Il démontre que les transitions sont invariantes par changement d'origine alors que les déviations ne le sont pas, et qu'une référence dérivée des sorties mesurées est aveugle à la composante commune de leur dérive. Il sépare l'admissibilité méthodologique de l'opérabilité computationnelle, de sorte que la validité d'un cadre n'est jamais inférée du succès de sa mise en œuvre. Il rend également calculable, à une résolution déclarée, la clôture dérivationnelle de la Loi 7 — l'atteinte d'une constante cognitive —, ce qui donne au traitement dérivationnel un critère d'arrêt fondé. Il établit que toute représentation apprise opère sous une centralité incorporée, et il rend l'orientation de cette centralité mesurable par rapport à des cadrages déclarés.

Quatre dispositifs expérimentaux, chacun doté d'une référence de comparaison forte, d'une vérité de terrain indépendante et d'une ablation qui isole les exigences de MANHAJ, précisent comment les capacités qui en résultent doivent être testées : la détection de défauts fondationnels que les vérifications

conventionnelles laissent non révélés, l'attribution conceptuelle de la dérive, des axes qui mesurent les distinctions propres au domaine, et une caractérisation relationnelle du cadrage incorporé dans les représentations apprises.

La distinction entre espace appris et espace fondé n'est donc pas une distinction entre deux agencements de vecteurs. C'est une distinction entre une référence supposée et une référence fondée. Une représentation fournit des coordonnées ; un calcul fournit un résultat ; un cadre fondé établit les conditions dans lesquelles ce résultat est une mesure du sens.

12. Note sur le développement et l'usage commercial

MANHAJ est une méthodologie originale dont les droits intellectuels et la paternité demeurent attachés à son auteur.

Toute utilisation, intégration, adaptation, exploitation, commercialisation ou mise en œuvre commerciale de MANHAJ requiert l'autorisation écrite préalable de l'auteur dans le cadre d'une licence commerciale payante.

Aucune licence commerciale gratuite n'est accordée. Toute autorisation d'exploiter commercialement MANHAJ est nécessairement subordonnée au paiement de redevances de licence et aux conditions contractuelles fixées par l'auteur.

Les termes de la licence commerciale, notamment son étendue, sa durée, son domaine d'utilisation, les droits concédés, les conditions financières, les exigences d'attribution et les dispositions relatives à la propriété intellectuelle, sont définis dans un contrat conclu avec l'auteur.

La publication du présent article scientifique ne constitue en aucun cas une licence commerciale et ne confère aucun droit d'exploiter commercialement MANHAJ.

Aucune utilisation commerciale, commercialisation, intégration dans un produit ou un service, ni exploitation de MANHAJ à des fins commerciales ne saurait être considérée comme autorisée du seul fait de la publication du présent article.

13. Références

Documents MANHAJ

1. Messeoud, Faouzi. *Intellectual Modeling Methodology (MANHAJ) — Reference Document — Version 1.0*. Zenodo, 15 septembre 2026. DOI : 10.5281/zenodo.22769627.
2. Messeoud, Faouzi. *MANHAJ : Une méthodologie générative pour la modélisation intellectuelle*. Article scientifique, Papier 01, Version 1.0, octobre 2026. DOI : 10.5281/zenodo.23138923.
3. Messeoud, Faouzi. *Examen de la production du sens dans l'IA : un cadre MANHAJ*. Article scientifique, Papier 02, Version 1.0, octobre 2026. DOI : 10.5281/zenodo.23214262.
4. Messeoud, Faouzi. *Les fondements épistémologiques des sciences humaines modernes à la lumière de la méthodologie de modélisation intellectuelle (MANHAJ) : l'approche a posteriori, le point d'origine et le problème de la généralisation du modèle historique*. Article scientifique, Papier 03, Version 1.0, octobre 2026. DOI : 10.5281/zenodo.23558985.

Ouvrages antérieurs de l'auteur

1. Messeoud, Faouzi. *Le lien immatériel des Tunisiens avec la France : comment la France entretient la dépendance tunisienne*. Première édition. Dar El Alaoui, 2023. ISBN 978-9938-9684-5-3.
2. Messeoud, Faouzi. *La centralité doctrinale et le champ conceptuel : manifeste sur la corruption méthodologique et les fondements de l'assujettissement*. Dar Wachma (Tunisie), 2024. ISBN 978-9938-9740-6-5.

Intelligence artificielle et représentation

Alain, G., & Bengio, Y. (2017). Understanding intermediate layers using linear classifier probes. *ICLR 2017 Workshop Track*. arXiv:1610.01644.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems* 29.

Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33, 1877–1901.

Ethayarajh, K. (2019). How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. *Proceedings of EMNLP-IJCNLP 2019*, 55–65.

Golchin, S., & Surdeanu, M. (2024). Time travel in LLMs: Tracing data contamination in large language models. *International Conference on Learning Representations (ICLR 2024)*. arXiv:2308.08493.

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., & Sayres, R. (2018). Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). *Proceedings of the 35th International Conference on Machine Learning*, PMLR 80, 2668–2677.

Mu, J., & Viswanath, P. (2018). All-but-the-top: Simple and effective postprocessing for word representations. *International Conference on Learning Representations (ICLR 2018)*. arXiv:1702.01417.

Panickssery, A., Bowman, S. R., & Feng, S. (2024). LLM evaluators recognize and favor their own generations. *Advances in Neural Information Processing Systems* 37. arXiv:2404.13076.

Sainz, O., Campos, J. A., García-Ferrero, I., Etxaniz, J., Lopez de Lacalle, O., & Agirre, E. (2023). NLP evaluation in trouble: On the need to measure LLM data contamination for each benchmark. *Findings of the Association for Computational Linguistics: EMNLP 2023*. arXiv:2310.18018.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631, 755–759. <https://doi.org/10.1038/s41586-024-07566-y>

Wang, H., Li, Z., Feng, L., & Zhang, W. (2022). ViM: Out-of-distribution with virtual-logit matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, 4921–4930.

Zheng, L., Chiang, W.-L., Sheng, Y., et al. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *Advances in Neural Information Processing Systems* 36 (*Datasets and Benchmarks Track*).

Zou, A., Phan, L., Chen, S., et al. (2023). Representation engineering: A top-down approach to AI transparency. arXiv:2310.01405.

Statistique

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188. <https://doi.org/10.1214/aos/1013699998>

Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). SAGE Publications.

Owen, A. B. (2007). The pigeonhole bootstrap. *The Annals of Applied Statistics*, 1(2), 386–411. <https://doi.org/10.1214/07-AOAS122>

Annexe A. Protocole opérationnel (proposé)

A.1. Statut de la présente annexe

La présente annexe spécifie le protocole opérationnel par lequel les dispositifs expérimentaux de la section 8 peuvent être réalisés. Elle relève du niveau de la mise en œuvre : elle n'ajoute rien aux affirmations méthodologiques de l'article, et ses choix — procédures de test, seuils, tailles d'échantillon — sont des décisions de mise en œuvre fixées avant toute expérience et examinées par analyse de sensibilité. Les valeurs des constantes déclarées, ainsi que les protocoles de reconnaissance et d'examen Rec et Π , sont fixés dans le document de protocole qui précède les expériences.

A.2. Configuration et annotation

- **Encodeur.** Un encodeur principal au niveau de la phrase, extérieur aux systèmes générateurs et fixe d'une condition à l'autre, avec un second encodeur pour la robustesse. L'expérience E4 utilise trois encodeurs (Annexe A.6).
- **Systèmes générateurs.** Un modèle de langage produisant les sorties ; des générations d'auto-entraînement ou des versions successives du modèle comme spécifié dans E2, toutes encodées par le même encodeur fixe.
- **Domaines.** Le domaine de la famille du Papier 02, qui fournit un cas concret pour documenter la condition C , et le domaine académique du scénario du professeur d'université du Papier 02.
- **Documentation du cadre.** Pour chaque domaine : le registre de reconnaissance de Γ_F produit par Rec ; la détermination de $D(\Gamma_F)$; les définitions opératoires Def ; les ensembles E_0 , $E_k^\pm$, $\mathcal{P}$ et H ; et un registre de protocole reliant chaque élément à son étape dans l'ordre fondationnel.

Les annotateurs ignorent les conditions de référence et les valeurs mesurées. Leurs tâches sont : la présence et l'ampleur d'un changement sémantique ; la direction du changement par rapport aux distinctions déclarées du domaine ; la préservation du sens dans les éléments de contrôle ; et l'appartenance au domaine des éléments tenus à l'écart. L'accord est rapporté au moyen de l' α de Krippendorff (Krippendorff, 2018) ; les éléments situés sous un seuil spécifié à l'avance sont analysés séparément.

A.3. Expérience E1 — procédure

Cadres. Des cadres sains, fondés selon la procédure complète ; et des cadres compromis, portant chacun un défaut connu :

- (a) **contamination** : formulations de référence chevauchant les sorties mesurées ou en constituant des quasi-doublons (I_{comp}) ;
- (b) **provenance** : formulations de référence produites par le système évalué (I_{doc}) ;
- (c) **importation non déclarée** : matériau de référence provenant d'un autre domaine ou d'une autre centralité sans déclaration (C) ;
- (d) **axes non discriminants** (Q).

Vérifications conventionnelles. Les procédures établies suivantes sont appliquées à chaque cadre, sain ou compromis, et leurs résultats sont rapportés par type de défaut :

- **recouvrement en n-grammes** entre formulations de référence et sorties mesurées (recouvrement en 13-grammes, tel qu'utilisé dans les analyses de contamination des grands modèles de langage ; Brown et al., 2020) ;
- **détection de quasi-doublons par plongements** entre formulations de référence et sorties, avec un seuil de similarité déclaré ;
- **détection de contamination par instruction guidée**, qui teste si le système évalué complète des formulations de référence qu'il a rencontrées (Golchin & Surdeanu, 2024) ;
- **vérifications distributionnelles** des métriques standard elles-mêmes : comparaison de leurs distributions entre cadres sains et cadres compromis.

Procédure. (1) Calculer les métriques standard — similarité, dérive, détection de changement — au sein de chaque cadre. (2) Appliquer les vérifications conventionnelles ci-dessus. (3) Appliquer le garde-fou : $Op_{comp}(F)$ par le calcul ; $Adm_M(F)$ par deux examinateurs indépendants appliquant le protocole écrit, sans connaissance de la construction du cadre.

Information disponible pour chaque procédure. La comparaison teste les procédures, et non des différences dans l'information qui leur est fournie. Chaque procédure reçoit le même matériau de cadre : les ensembles de formulations, les sorties mesurées et la documentation du cadre — y compris les registres de provenance et le registre de reconnaissance de *Rec* — lorsque la procédure est en mesure de les utiliser. Les défauts sont de deux types, analysés séparément : les **défauts enregistrés**, dont la trace est présente dans la documentation (par exemple, un registre de provenance montrant que des formulations ont été produites par le système évalué), et les **défauts non enregistrés**, dont la trace doit être trouvée par une investigation indépendante (par exemple, une importation non déclarée que la documentation ne mentionne pas). La détection des défauts enregistrés montre que l'examen méthodologique utilise une information que les vérifications conventionnelles n'examinent pas ; la détection des défauts non enregistrés teste si l'examen peut trouver ce que la documentation n'énonce pas. (4) Comparer les conclusions tirées dans chaque cadre à la vérité de terrain de l'Annexe A.2.

Mesures. Différences des métriques standard entre cadres sains et cadres compromis ; taux de détection du garde-fou et taux de faux signalements, rapportés séparément pour les échecs méthodologiques (b, c) et computationnels (a, d) ; accord entre examinateurs sur Adm_M ; taux d'erreur des conclusions dans les cadres compromis par rapport à la vérité de terrain.

Interprétation de résultats divergents. Si des procédures d'audit établies détectent certains défauts fondationnels, le résultat restreint la classe des défauts pour lesquels le garde-fou est la seule protection et précise exactement sa contribution. Si l'accord des examinateurs sur Adm_M est faible, le résultat établit que le protocole de documentation exige une plus grande précision ; conformément à la section 5.3, cela concerne la mise en œuvre des conditions d'admissibilité, non la Loi 3.

A.4. Expérience E2 — procédure

Conditions.

- (i) **Dérive sémantique construite** : sorties modifiées en direction de positions déclarées du domaine ou en s'en éloignant, à des amplitudes graduées, chacune validée par les annotateurs.
- (ii) **Contrôles** : changements de paraphrase, de style et de longueur confirmés par les annotateurs comme préservant le sens.
- (iii) **Dérive collective naturelle** : générations successives d'auto-entraînement ou versions successives d'un modèle, avec des échantillons annotés.

Mesures. Discrimination de la dérive sémantique par rapport au changement préservant le sens (AUROC) pour R1–R4 ; fausses alarmes sur les contrôles ; corrélation entre la déviation mesurée et l’ampleur annotée ; accord entre l’axe auquel R4 attribue la dérive et la direction identifiée par les annotateurs (R4 et R3a ; R3 ne fournit aucune attribution) ; et, pour la condition (iii), la part de la composante commune, $\|\bar{\delta}\|$ rapportée à la moyenne de $\|\delta_t\|$, dans l’espace complet et en coordonnées projetées ($\|U^T \bar{\delta}\|$).

Interprétation de résultats divergents. Si R4 et R3a obtiennent des performances similaires, le résultat situe l’avantage actuel dans la technique des axes pour ce domaine et cet encodeur, et oriente l’attention vers la spécification de l’ordre fondationnel dans la documentation du cadre ; la force de la Loi 4 face aux références dérivées des sorties demeure établie par la Proposition 2. Si la composante commune est faible dans la situation testée, la Proposition 2 demeure établie, et le résultat précise les régimes de dérive dans lesquels ses conséquences sont les plus marquées.

A.5. Expérience E3 — procédure

Cadres. R4 : ordre fondationnel complet, documenté. R3a : la même technique, le même nombre d’axes et le même nombre de formulations, à partir d’exemples de commodité. R4u : axes dérivés sans supervision de données relevant du domaine (composantes principales).

Vérité de terrain. Des experts du domaine jugent les positions et l’appartenance aux pôles des éléments tenus à l’écart. L’indépendance est assurée par quatre règles : les experts ne participent ni à la rédaction de *Def*, ni à celle des formulations de pôles de R4, ni à celle des exemples de commodité de R3a ; ils ignorent quel cadre a produit chaque axe ; ils jugent des éléments, non des axes ; et leurs jugements sont recueillis avant qu’aucun cadre ne soit projeté sur les éléments tenus à l’écart.

Mesures. Corrélation de rang entre coordonnées projetées et jugements des experts ; pouvoir discriminant sur les éléments tenus à l’écart ; séparation des éléments relevant du domaine et des éléments extérieurs au domaine par l’indicateur d’écart r_t (AUROC).

Contrôle. Les critères d’appariement — nombre, longueur et diversité lexicale — sont fixés avant la construction de l’un ou l’autre cadre. Les exemples de R3a sont rassemblés par une équipe distincte selon une procédure d’échantillonnage déclarée, sans accès aux pôles de R4 ; l’appariement est vérifié ensuite par des annotateurs aveugles, qui évaluent aussi la qualité des deux cadres. La comparaison isole ainsi l’ordre fondationnel de la sélection experte sans conférer à l’un ou l’autre cadre un avantage non voulu.

Interprétation de résultats divergents. Si R4 dépasse R4u mais non R3a, le résultat établit que des axes construits à partir d’exemples déclarés ou sélectionnés surpassent des axes extraits des données dans cette situation ; l’effet supplémentaire de l’ordre fondationnel au-delà de la sélection des exemples exige alors que les éléments distinctifs de l’ordre soient rendus plus explicites dans la documentation du cadre. Si R4 dépasse R3a mais non R4u, la structure extraite des données capte les distinctions du domaine dans cette situation, et la comparaison de R4 avec R4u est examinée par domaine et par axe.

A.6. Expérience E4 — procédure

Concepts. Deux concepts candidats dont les définitions opératoires diffèrent selon les centralités : *famille*, dans le prolongement du domaine de la famille du Papier 02 ; et *progrès*, dont le sens dépend de l’échelle par laquelle il est défini (Papier 03, section 5.1). Chacun n’est retenu que si *Rec* produit des éléments distinctifs suffisants pour au moins six axes distinctifs par centralité.

Cadres. Pour chaque concept, deux cadres F_A et F_B , fondés sur deux centralités, chacune reconnue par *Rec* avec un cadrage source, des traces de dérivation et des éléments distinctifs documentés. Les deux cadres satisfont $\text{Adm}_M = 1$. L'expérience décrit l'encodeur par rapport aux deux cadrages ; elle n'évalue pas les cadrages.

Axes. Chaque cadre contient au moins six axes partagés et six axes distinctifs, en nombres égaux dans les deux cadres. Chaque distinction partagée est spécifiée une seule fois, par une définition admise sous les deux centralités avec le même contenu de pôles, et ses formulations sont rédigées séparément au sein de chaque cadre selon le même protocole ; des annotateurs aveugles jugent si les axes partagés correspondants expriment la même distinction, et les paires situées sous un seuil d'accord spécifié à l'avance sont exclues avec leurs homologues. Les axes distinctifs sont dérivés des éléments distinctifs de chaque centralité. Le dispositif minimal comprend donc $m = 48$ instances d'axes par encodeur (12 instances partagées et 12 instances distinctives par concept).

Appariement et contrôle de la rédaction.

- Les ensembles de formulations sont appariés entre cadres en nombre par pôle, en longueur, en diversité lexicale et en registre.
- Deux groupes de rédacteurs indépendants, chacun compétent dans les deux cadrages, rédigent pour les deux cadres, de sorte que les effets de rédacteur sont répartis également (rédaction contrebalancée).
- L'identité des rédacteurs est enregistrée. Le mode de production — indépendant ou apparié — est déclaré à l'avance pour chaque concept. Chaque rédacteur fournit un nombre égal de formulations aux deux pôles de chaque instance d'axe à laquelle il participe ; les formulations d'un même rédacteur pour un même pôle satisfont $\text{dup} < \tau$; et aucune formulation n'est produite en modifiant une autre formulation du même pôle.
- Avec $m = 48$ et $q = 0.05$, les conditions de faisabilité de l'Annexe A.7 donnent $B + 1 \geq 4\,281$ et, pour le test d'échange par rédacteur, $m_w \geq 14$ rédacteurs par instance d'axe ; $B = 10\,000$ est retenu. Le nombre final de rédacteurs est fixé par analyse de puissance.
- Des annotateurs aveugles évaluent la qualité des formulations, rapportée pour chaque cadre et chaque type d'axe.

Encodeurs. Trois encodeurs choisis pour différer par leurs corpus d'entraînement — par exemple un encodeur centré sur l'anglais, un encodeur multilingue et un encodeur disposant d'un entraînement substantiel en langue arabe. La variation de l'orientation entre encodeurs aux corpus différents fait elle-même partie de la caractérisation de I_L , rapportée comme covariation. Lorsqu'elle est disponible, une paire d'encodeurs proches par l'architecture, la dimensionnalité et les objectifs, mais différant par des données d'entraînement documentées, est ajoutée comme comparaison contrôlée.

Mesures.

- **Principale :** l'indice d'orientation $\mathcal{O}(\varphi; F_A, F_B)$ de la section 6.15, avec les moyennes géométriques G au niveau des cellules, les proportions de réalisation v_{sh} et v_{dist} , le nombre d'instances d'axes effondrées et de valeurs de plancher regroupé par cellule, et les deux analyses de sensibilité de l'Annexe A.8.
- **Contrôle lexical :** chaque séparation est aussi calculée avec un encodeur lexical (TF-IDF). Après plafonnement séparé de chacune, la séparation ajustée lexicalement $L_k = \log \bar{S}_k^\varphi - \log \bar{S}_k^{\text{lex}}$ remplace $\log \bar{S}_k$ pour donner $\mathcal{O}_{\text{lex}}$. Parce que $\mathcal{O}$ est une double différence, les décalages constants entre les échelles des deux encodeurs s'annulent ; les différences non linéaires et les différences

dans la manière dont l'échelle affecte les axes partagés et distinctifs ne s'annulent pas, et $\mathcal{O}_{\text{lex}}$ est donc rapporté à côté de $\mathcal{O}$ comme l'orientation au-delà de celle qu'atteint le contrôle lexical spécifié.

- **Auxiliaires** : γ_k et le contraste d'écart $\Delta\bar{r} = \bar{r}_H(F_B) - \bar{r}_H(F_A)$, rapportés séparément et non combinés dans $\mathcal{O}$.
- **Robustesse** : recalcul de $\mathcal{O}$ en omettant un axe à la fois ; fiabilité par moitiés de $\bar{S}_k$ au sein de chaque pôle.

Analyse statistique.

- *Estimation principale.* Pour chaque encodeur et chaque concept, $\bar{S}_k$ est calculé pour chaque instance d'axe à partir des formulations regroupées des deux groupes de rédacteurs, et $\mathcal{O}$ est calculé selon sa définition ; il est également rapporté pour chaque groupe de rédacteurs séparément.
- *Incertitude.* Les rédacteurs contribuent à plusieurs instances d'axes, de sorte que rédacteurs et distinctions forment un plan croisé. Les intervalles de confiance de $\mathcal{O}$ sont obtenus par un bootstrap à deux facteurs (*pigeonhole bootstrap*) (Owen, 2007), en rééchantillonnant à chaque réplication : (1) les rédacteurs, une fois par réplication au sein de chaque groupe de rédacteurs, chaque rédacteur rééchantillonné emportant toutes ses formulations dans toutes les instances d'axes ; (2) les distinctions, indépendamment — les distinctions partagées comme paires gardant ensemble leurs deux instances de cadre, les distinctions distinctives au sein de leur cadre ; (3) les formulations au sein de chaque rédacteur, pôle et instance d'axe. Les groupes de rédacteurs ne sont pas rééchantillonnés.
- *Confirmation par modèle.* Un modèle à effets mixtes est ajusté avec l'ensemble de formulations (k, f, w) comme unité d'observation :

$$y_{kfw} = \beta_0 + \beta_F F + \beta_T T + \beta_{FT} (F \times T) + \beta_W W + d_k + a_{kf} + \epsilon_{kfw},$$

où $y_{kfw} = \log \bar{S}$ calculé à partir des formulations de l'ensemble (k, f, w) ; F , T et W sont des effets fixes pour le cadre, le type d'axe et le groupe de rédacteurs ; d_k est un intercept aléatoire pour l'identité de la distinction, partagé par les deux instances d'une distinction partagée ; et a_{kf} est un intercept aléatoire pour l'instance d'axe. Les formulations n'interviennent qu'à travers le calcul de y_{kfw} . Parce que le logarithme d'une séparation regroupée n'est pas la moyenne des logarithmes des séparations par groupe, β_{FT} et $\mathcal{O}$ sont des estimandes apparentés mais distincts ; $\mathcal{O}$ est principal, et l'accord de signe et d'ampleur est rapporté comme confirmation.

- *Portée de l'inférence.* L'unité d'inférence est la distinction. Les formulations issues de cadrages différents ne sont jamais traitées comme échangeables ; l'échangeabilité n'est supposée que pour les étiquettes de pôles au sein d'une strate, sous l'hypothèse nulle du test de réalisation.

Interprétation de résultats divergents.

- *Absence d'orientation* ($\mathcal{O} \approx 0$ avec des niveaux de séparation comparables dans les quatre cellules) : l'encodeur réalise de manière comparable le contenu distinctif des deux cadrages pour ces concepts. Cela caractérise Γ_L comme ne discriminant pas entre les deux cadrages sur ces concepts, et oriente le choix vers des concepts aux éléments distinctifs plus marqués. Cela ne concerne pas l'existence de Γ_L .

- *Axes distinctifs faibles dans les deux cadres, axes partagés forts* ($\mathcal{O} \approx 0$, établi à partir des séparations au niveau des cellules G et de v_{dist}) : l’encodeur retient ce que les cadrages partagent et aplatit ce qui les distingue. C’est en soi une caractérisation de Γ_L .
- *Encodeur effondré pour un concept* (établi par la condition d’effondrement indépendante) : rapporté comme une caractérisation substantielle (Annexe A.8).
- *Indice non calculable sous la règle de plancher* : rapporté comme tel, sans qu’aucune conclusion ne soit tirée sur Γ_L .
- *Orientation variant avec le groupe de rédacteurs* : le résultat reflète la rédaction plutôt que le cadrage, et le protocole d’appariement est affiné avant toute interprétation.

A.7. Test de réalisation

La réalisation est évaluée pour chaque instance d’axe par un test de permutation dont la structure suit le mode de production déclaré des formulations :

- *Production indépendante* : chaque formulation est rédigée indépendamment, et chaque rédacteur fournit un nombre égal de formulations aux deux pôles de l’instance d’axe. Les étiquettes de pôles sont permutées au sein de chaque rédacteur, en préservant le nombre de formulations par pôle ; les strates sont combinées sur les deux groupes de rédacteurs.
- *Production appariée* : chaque formulation positive est rédigée avec une formulation négative correspondante, et les étiquettes ne sont échangées qu’au sein des paires.

Dans les deux modes, la statistique est la séparation brute $\bar{S}_k$ telle que définie ci-dessus, avec a_k , v_k et $s_k^\pm$ recalculés à chaque permutation, puisque la direction est elle-même estimée à partir des étiquettes. Sous l’hypothèse nulle selon laquelle l’encodeur ne sépare pas les pôles, les formulations sont échangeables entre pôles au sein de chaque strate. Le protocole soutient cette hypothèse en exigeant que les formulations d’un même rédacteur pour un même pôle ne soient pas des quasi-paraphrases les unes des autres ($\text{dup} < \tau$) et qu’aucune formulation ne soit produite en modifiant une autre formulation du même pôle. Lorsque la dépendance au sein d’un rédacteur ne peut être exclue, un **test d’échange par rédacteur**, plus grossier, échange les ensembles complets de pôles de chaque rédacteur en tant qu’unités.

Une instance d’axe est **réalisée** si sa séparation est significative après contrôle du taux de fausses découvertes par la procédure de Benjamini–Yekutieli à un niveau déclaré q , sur l’ensemble des instances d’axes d’un même encodeur (Benjamini & Yekutieli, 2001). Benjamini–Yekutieli est la procédure principale parce qu’elle contrôle le taux de fausses découvertes sous une dépendance arbitraire. Les rédacteurs contribuent à plusieurs axes, de sorte que les tests sont dépendants, et la condition sous laquelle la validité de Benjamini–Hochberg (1995) est garantie — la dépendance de régression positive sur le sous-ensemble des hypothèses nulles vraies — ne peut être établie pour ces statistiques ; les résultats de Benjamini–Hochberg sont rapportés comme analyse de sensibilité. Avec m instances d’axes dans la famille de tests et $H_m = \sum_{i=1}^m 1/i$, le nombre de permutations B doit vérifier

$$\frac{1}{B+1} < \frac{q}{m H_m},$$

et le test d’échange par rédacteur n’est admissible que si $2^{1-m_w} < q/(m H_m)$, où m_w est le nombre de rédacteurs contribuant à chaque instance d’axe. Ce sont des conditions de faisabilité ; les tailles d’échantillon sont fixées par analyse de puissance.

v_{sh} et v_{dist} désignent les proportions d'instances d'axes partagés et distinctifs réalisées. La classification de réalisation est rapportée à côté de l'indice d'orientation et n'y entre pas.

A.8. Règle de plancher, effondrement et dépendance

Règle de plancher. Le plancher n'est utilisé qu'à une seule fin : attribuer une valeur à une instance d'axe avec $a_k = 0$. Cette valeur entre dans le profil de réalisation et dans l'indice d'orientation en tant que quantité diagnostique ; elle ne transforme pas un axe indéfini en axe opérationnel (section 6.6). Pour chaque instance d'axe, chaque valeur permutée est évaluée selon le même tableau de cas : une valeur permutée avec $a_k \neq 0$ et une dispersion nulle est fixée à S_{max} ; une valeur avec $a_k = 0$ et une dispersion positive est fixée à 0 ; une valeur avec $a_k = 0$ et une dispersion nulle est indéfinie et écartée. Toutes les valeurs sont plafonnées à S_{max} . Ensuite :

1. **Plancher propre à l'axe.** S'il reste au moins N_{null} valeurs nulles valides et que leur médiane est positive, $S_{floor,k}$ est cette médiane : la séparation attendue de la seule variation d'échantillonnage.
2. **Plancher regroupé.** Sinon, le plancher de cette instance d'axe est non estimable, et $S_{floor,k}$ est fixé au plancher regroupé : la médiane des médianes nulles positives propres aux autres instances d'axes des deux cadres, pour le même encodeur et le même concept, à condition qu'il existe au moins N_{pool} de ces médianes. Parce qu'il est calculé sur les deux cadres, le plancher regroupé ne confère d'avantage à aucun d'eux.
3. **Indice non calculable.** Si une instance d'axe avec $a_k = 0$ requiert un plancher et que moins de N_{pool} médianes propres aux axes sont disponibles, $\mathcal{O}$ est rapporté comme **non calculable sous cette règle** pour cet encodeur et ce concept. Aucune conclusion empirique sur Γ_L n'est tirée de cette issue.

L'effondrement comme constat distinct. Le nombre d'instances d'axes effondrées, établi par la condition indépendante ci-dessus, est rapporté pour chaque cellule (cadre $\times$ type d'axe). Si la plupart des instances d'axes d'un concept sont effondrées et que la chaîne d'encodage a été vérifiée quant aux défauts techniques (entrées tronquées, vides ou mal regroupées en lots), l'encodeur est rapporté comme **effondré pour ce concept** : il ne distingue les formulations d'aucun des deux cadrages. Ce constat repose sur la condition d'effondrement elle-même, et non sur le comportement de la distribution nulle.

Prédominance du plancher regroupé. La valeur de plancher ne remplace la séparation observée que lorsque $a_k = 0$. Si, dans une cellule quelconque, la proportion d'instances d'axes dont la valeur est fixée par le plancher regroupé dépasse un plafond déclaré λ_{pool} , $\mathcal{O}$ est rapporté comme descriptif et signalé, et il n'est pas utilisé comme résultat principal pour cet encodeur et ce concept, puisqu'un plancher commun dominerait alors la comparaison.

Structure de dépendance. Deux types d'inférence utilisent des unités différentes. Le test de réalisation opère au sein d'une seule instance d'axe, sur les formulations, avec une échangeabilité limitée à un rédacteur ou à une paire ; la dépendance entre instances d'axes, créée par les rédacteurs qui contribuent à plusieurs axes, est traitée par la procédure de Benjamini–Yekutieli. Les intervalles de confiance de $\mathcal{O}$ traitent les rédacteurs et les distinctions comme des facteurs croisés et utilisent le bootstrap à deux facteurs de l'Annexe A.6 ; aucune formulation n'est traitée comme une observation indépendante dans l'une ou l'autre procédure.

Deux analyses de sensibilité sont rapportées : $S_{\text{floor},k}$ fixé au premier quartile des valeurs nulles valides, avec le même recours au plancher regroupé ; et une séparation planchée au niveau du hasard, dans laquelle chaque $\bar{S}_k$ inférieur à son propre $S_{\text{floor},k}$ est relevé à ce plancher.

A.9. Analyse statistique générale

Les attentes, les seuils, les familles de tests et les plans d'analyse sont spécifiés avant la collecte des données. Les intervalles de confiance sont obtenus par rééchantillonnage bootstrap adapté à la structure de chaque expérience. Des corrections pour comparaisons multiples sont appliquées entre conditions de référence. Les analyses de sensibilité font varier ρ , ε , $\gamma_{\min}$, $S_{\min}$, $S_{\max}$ et la règle de plancher. Les tailles d'échantillon sont fixées par analyse de puissance ; le matériau minimal requis pour calculer une statistique n'est pas confondu avec le matériau suffisant pour étayer une inférence. Les valeurs de toutes les constantes déclarées sont fixées dans un protocole opérationnel avant les expériences.