

Quantifier sans objectiver : protocole pilote pour une pondération interprétative des théories des relations internationales, 1815–2026

Sylvain Geffroy

Mai 2026

Résumé

Cet article propose et éprouve un protocole de pondération interprétative des grandes familles théoriques des relations internationales. Sa visée n'est pas de naturaliser les théories en variables, ni de prétendre mesurer objectivement la part « réelle » du réalisme, du libéralisme ou du constructivisme dans l'histoire internationale ; elle est réflexive. Lorsqu'un analyste qualifie une période de réaliste, libérale, impériale, dépendantiste ou constructiviste, il accomplit déjà une hiérarchisation tacite des cadres : un acte de classement, indissociable de la position des théories dans le champ savant et de l'histoire d'une discipline longtemps pensée comme « science sociale américaine » (Hoffmann, 1977; Bourdieu, 1976). Le protocole de *Strates RI* vise à rendre cette opération explicite, auditable, révisable et comparable, dans la lignée d'une sociologie des relations internationales attentive aux conditions de production du savoir (Badie and Smouts, 1992; Devin, 2002). L'article clarifie le statut métrique faible des scores, distingue scores bruts et scores normalisés, formalise trois couches d'application – réalité stratégique, doctrine politique et influence académique –, propose une matrice de non-confusion taxonomique adossée à la cartographie reçue des paradigmes (Battistella, 2012), justifie une périodisation longue de 1815 à 2026, établit un codebook de scoring 0–100, et applique le protocole à un pilote dérivé sur trois périodes : 1815–1848, 1945–1962 et 1991–2001. Le pilote code sept familles théoriques sur trois couches, documente les éléments probants et contradictoires, produit une analyse de sensibilité aux pondérations, une simulation explicitement marquée de second codage, et un audit de cohérence méthodologique aboutissant à un verdict explicite. Les trois périodes éprouvent trois difficultés distinctes : l'anachronisme et l'ordre pré-disciplinaire (1815–1848, où la méthode va jusqu'à ne calculer aucune synthèse), un cas robuste (1945–1962) et un cas controversé d'ordre libéral hégémonique (1991–2001). La position défendue est délibérément prudente : les scores ne sont pas des mesures cardinales fortes, mais des quasi-indices interprétatifs contrôlés, dont la valeur tient moins à la précision numérique affichée qu'à la transparence du codage, à la discipline anti-circularité, à la traçabilité bibliographique et à la révisabilité. C'est en ce sens que l'on peut *quantifier sans objectiver*.

Mots-clés : relations internationales ; théorie des relations internationales ; méthode qualitative ; scoring ; validité ; fiabilité inter-codeurs ; anachronisme ; périodisation ; réalisme ; libéralisme ; constructivisme ; École anglaise ; économie politique internationale ; postcolonialisme.

1 Introduction

Les relations internationales raisonnent volontiers par diagnostics synthétiques. Une période est dite « réaliste » lorsqu'elle semble dominée par la compétition de puissance ; « libérale » lorsqu'elle paraît structurée par les institutions, le droit et l'interdépendance ; « constructiviste » lorsque les normes, les identités et les processus de légitimation semblent décisifs ; « marxiste » ou « dépendantiste » lorsque les hiérarchies de l'économie mondiale rendent mieux compte des dynamiques observées ; « postcoloniale » lorsque les traces impériales, les hiérarchies raciales ou les asymétries de savoir deviennent centrales. Ces diagnostics sont commodes ; ils sont aussi méthodologiquement opaques. Cet article prend au sérieux deux exigences trop souvent

disjointes : celle d'une discipline réellement globale, qui ne se réduise pas à l'extension des catégories dominantes à des objets non occidentaux (Acharya, 2014; Badie, 1992), et celle d'une lecture praxéologique des configurations stratégiques, attentive à ce que les acteurs font autant qu'à ce qu'ils proclament (Aron, 1962).

Dire que 1945–1962 est une période réaliste n'est pas une simple description. C'est une opération de pondération : l'analyste décide que la bipolarité, les alliances, le dilemme de sécurité, la dissuasion nucléaire, la guerre de Corée et les crises de Berlin expliquent mieux la structure du système que d'autres dimensions pourtant présentes, comme l'institutionnalisation onusienne, Bretton Woods, la décolonisation ou les normes de souveraineté (Morgenthau, 1948; Waltz, 1979; Gaddis, 2005; Westad, 2005). Dire que 1991–2001 est un moment libéral effectue aussi une pondération : l'expansion institutionnelle et normative est mise au premier plan, tandis que la condition hégémonique américaine de cet ordre est parfois reléguée au second plan (Krauthammer, 1990; Ikenberry, 2001, 2011; Cox, 1981).

Le projet *Strates RI* part de cette observation. Il ne cherche pas à substituer un calcul au jugement historique, mais à rendre visibles – et donc discutables – les opérations interprétatives déjà à l'œuvre dans les diagnostics théoriques. Caractériser une période, c'est hiérarchiser des cadres concurrents selon leur capacité à rendre raison du réel : un geste de classement qui engage la position des théories dans le champ savant autant que les faits invoqués (Bourdieu, 1976). Le score n'est donc pas une mesure empirique directe ; c'est un indice argumenté – la formalisation explicite d'un jugement comparatif, adossée à des critères, à des éléments probants, à des contradictions documentées, à des références et à des niveaux de confiance.

La contribution de cet article est triple. Premièrement, il propose un protocole méthodologique complet pour coder le poids relatif de familles théoriques sur plusieurs périodes historiques. Deuxièmement, il applique ce protocole à un cas pilote – le système international de 1945 à 1962, étendu à 1815–1848 et 1991–2001 – afin de montrer concrètement ce que produit la méthode, ce qu'elle stabilise et ce qu'elle laisse contestable. Troisièmement, et c'est le point souvent escamoté des travaux méthodologiques, il montre que le dispositif ne se contente pas de formaliser des intuitions connues : il engendre des propositions substantielles et falsifiables sur les relations internationales, que la seule lecture qualitative laisse implicites. Trois d'entre elles sont énoncées plus bas (écart doctrine/réalité comme signature hégémonique ; montée de l'influence académique comme marqueur datable d'émergence disciplinaire ; sensibilité aux pondérations comme propriété de la période). Ce passage du manifeste méthodologique au cas pilote est décisif : une architecture de scoring n'est pas encore une méthode ; elle ne devient méthode que lorsqu'elle est testée sur un cas, exposée à la sensibilité des pondérations et soumise à un protocole de fiabilité.

L'argument central est volontairement restrictif : une quantification interprétative est défendable seulement si elle renonce à l'illusion de précision forte. Les scores 0–100 ne sont pas des mesures cardinales du réel. Ils sont des quasi-intervalles pragmatiques, utiles pour comparer, visualiser et auditer des jugements qualitatifs. Leur légitimité dépend d'un codebook explicite, d'une discipline anti-anachronisme, d'une distinction entre scores bruts et scores normalisés, d'une justification de la périodisation et d'une procédure de révision. Sans ces garde-fous, *Strates RI* ne vaudrait pas mieux qu'une intuition chiffrée. On lira donc ce texte pour ce qu'il est : une contribution méthodologique adossée à un pilote de preuve de concept, destinée à la discussion sur les manières de coder le savoir des relations internationales – ni une étude empirique du poids réel des théories, ni un manifeste programmatique. Les données mobilisées sont une amorce non vérifiée ; ce sont le protocole et sa discipline qui sont soumis à l'examen.

2 Contribution et positionnement dans la littérature

La littérature des relations internationales est structurée par des traditions théoriques concurrentes. Le réalisme classique et structurel place au centre l'anarchie, la puissance, la sécurité, l'intérêt national et les effets de polarité (Morgenthau, 1948; Waltz, 1979; Mearsheimer, 2001).

Le libéralisme et l'institutionnalisme libéral analysent le rôle des régimes, des institutions, de l'interdépendance, des préférences domestiques et de la paix démocratique (Keohane and Nye, 1977; Keohane, 1984; Doyle, 1983; Moravcsik, 1997). Le constructivisme insiste sur les normes, les règles, les identités, la socialisation et la constitution sociale des intérêts (Onuf, 1989; Kratochwil, 1989; Wendt, 1999; Finnemore and Sikkink, 1998). L'École anglaise traite l'ordre international comme société internationale faite d'institutions primaires, de souveraineté, de diplomatie, de droit et d'équilibre (Wight, 1977; Bull, 1977; Watson, 1992; Buzan, 2014). Les traditions marxistes, dépendantistes et d'économie politique internationale déplacent l'attention vers capitalisme mondial, impérialisme, dépendance, hégémonie matérielle, production et finance (Hobson, 1902; Lenin, 1917; Frank, 1966; Cardoso and Faletto, 1979; Wallerstein, 1974; Cox, 1981; Gilpin, 1981; Strange, 1988, 1996). Les approches critiques, poststructuralistes, féministes, postcoloniales, décoloniales et raciales interrogent les conditions de production du savoir, les hiérarchies de genre, de race et d'empire, ainsi que les exclusions constitutives de la discipline (Ashley, 1984; Walker, 1993; Tickner, 1992; Enloe, 1989; Said, 1978; Spivak, 1988; Fanon, 1961; Vitalis, 2015; Acharya and Buzan, 2019).

La méthode proposée ne prétend pas arbitrer définitivement entre ces traditions. Elle vise à rendre plus explicite leur combinaison historique. L'originalité n'est pas d'ajouter une nouvelle théorie des RI. Elle est de proposer un dispositif de codage qui oblige l'analyste à dire : quelle famille théorique s'applique à quelle couche, selon quels éléments probants, avec quelle capacité discriminante, quels contre-éléments et quel niveau de confiance.

Ce projet relève de la méthodologie qualitative structurée plutôt que de la mesure statistique classique. Il emprunte à la construction conceptuelle et à la validité de mesure l'exigence de relier concept, indicateurs et procédures de codage (Adcock and Collier, 2001; Goertz, 2006; Collier and Levitsky, 1997). Il emprunte aux méthodes comparatives l'idée qu'une catégorie doit être stabilisée avant d'être appliquée à plusieurs cas (Ragin, 1987). Il emprunte aux études de cas et au process tracing l'exigence de documenter les mécanismes, les éléments probants et les inférences, au lieu de se contenter d'une attribution intuitive (George and Bennett, 2005; Gerring, 2007; Brady and Collier, 2010; Bennett and Checkel, 2015). Enfin, il emprunte à la sociologie historique la prudence face aux périodisations longues et aux continuités artificielles (Tilly, 1992; Hobsbawm, 1994; Kennedy, 1987).

Le dispositif gagne toutefois à être lu à la croisée de deux traditions méthodologiques rarement mises en dialogue. D'un côté, le souci anglophone de la validité de mesure et de la transparence des inférences, rappelé ci-dessus. De l'autre, la *sociologie des relations internationales* de tradition française, qui tient l'objet « système international » pour une construction historiquement située et place au centre la réflexivité de l'observateur (Merle, 1988; Badie and Smouts, 1992; Devin, 2002). C'est de cette seconde tradition que *Strates RI* hérite son geste fondateur : non pas mesurer le monde, mais objectiver les schèmes par lesquels la discipline le découpe en « moments » théoriques. La taxinomie mobilisée n'est pas inédite – elle reprend la cartographie reçue des paradigmes (Battistella, 2012) – mais elle est ici soumise à une discipline de codage explicite, couche par couche. L'apport propre est donc de doter cette sociologie réflexive d'un instrument reproductible : un codage daté, sourcé, sensible aux pondérations et révisable, là où la critique réflexive reste le plus souvent au registre de l'essai.

La contribution doit être formulée négativement autant que positivement. Positivement, l'article fournit un protocole de pondération interprétative. Négativement, il refuse quatre raccourcis : réduire une période à une théorie unique ; normaliser les scores comme s'ils étaient des parts objectives du réel ; appliquer rétrospectivement des théories modernes comme si elles avaient existé historiquement ; agréger des traditions critiques hétérogènes comme si elles formaient un bloc théorique homogène.

2.1 Trois propositions substantielles et falsifiables

Un reproche prévisible adressé à ce genre de dispositif est la trivialité : montrer que « le réalisme est élevé pendant la guerre froide » ne ferait que reformuler en chiffres un consensus déjà acquis. Cette objection vaut tant que le modèle ne sert qu'à confirmer la famille dominante. Mais la valeur analytique du dispositif est ailleurs : dans les seconds rôles, les écarts entre couches et les renversements de classement. La structure à trois couches et la sensibilité aux pondérations rendent énonçables trois propositions que la lecture qualitative laisse habituellement implicites. Chacune est formulée de manière à pouvoir être réfutée par un codage sourcé et indépendant.

P1 — L'écart entre doctrine politique et réalité stratégique mesure une condition hégémonique, non la nature de l'ordre. Lorsqu'une famille domine la couche *doctrine politique* très au-delà de son score de *réalité stratégique*, l'écart ne dit pas que l'ordre est de ce type ; il dit qu'un acteur dominant a réussi à imposer sa lecture comme doctrine commune. Sur le pilote, 1991–2001 affiche une doctrine libérale à 93.25 pour une réalité stratégique libérale nettement moindre et un réalisme stratégique maintenu à 74.25 : l'écart est la trace de l'unipolarité américaine, pas la preuve d'un monde devenu libéral. *Condition de réfutation* : si un codage sourcé de périodes sans acteur prépondérant — ainsi 1815–1848, multipolaire, où la doctrine réaliste (77.25) est *inférieure* à la réalité stratégique (81.25), sans inflation doctrinale — révélait des écarts doctrine/réalité aussi grands que sous unipolarité, la proposition tomberait. P1 est donc testable par comparaison systématique des configurations de polarité. Une réserve s'impose néanmoins : avec un codeur unique, l'écart pourrait refléter le cadre interprétatif de l'analyste autant que la condition hégémonique elle-même – un codeur déjà convaincu de l'unipolarité a pu coder haut la doctrine libérale. Seul un codage indépendant (*infra*) peut écarter cette circularité ; à ce stade, P1 est une hypothèse, non un résultat acquis.

P2 — La montée de l'influence académique d'une famille est un marqueur datable d'émergence disciplinaire, distinct de sa pertinence rétrospective. La couche *influence académique*, soumise à l'anti-anachronisme, sépare deux questions souvent confondues : une théorie éclaire-t-elle la période, et existait-elle comme école constituée *dans* la période ? Sur le pilote, le constructivisme passe d'une influence académique *rejetée* avant son émergence à 89.25 en 1991–2001 : ce saut date son institutionnalisation disciplinaire, indépendamment de sa valeur comme grille de lecture rétrospective. *Condition de réfutation* : le signal bibliométrique externe et indépendant (comptes de publications par famille, via OpenAlex) doit corrélérer avec la couche académique ; une montée de l'influence académique sans montée correspondante des publications invaliderait le marqueur. La proposition exige cependant une précaution : l'anti-anachronisme place *par construction* l'influence académique à « rejeté » avant l'émergence d'une école, de sorte que le saut observé est en partie inscrit dans la règle de codage, et non purement découvert. P2 n'est non triviale que si la *datation* de l'émergence et la pente de la montée sont elles-mêmes sourcées plutôt que décrétées – ce que seul le calibrage bibliométrique externe peut garantir. Sans cela, la proposition risque de ne faire que refléter sa propre règle.

P3 — La sensibilité d'un classement aux pondérations est une propriété de la période, non un artefact de mesure. Une famille dont le rang synthétique reste stable sous toutes les pondérations raisonnables signale une période interprétativement *réglée* ; une famille dont le rang bascule signale une contestation interprétative réelle. Sur le pilote 1945–1962, le réalisme est robuste (*delta_max* 3.45) tandis que la géopolitique est modérément sensible (7.40) : cette différence informe sur la structure du débat propre à la période, elle n'est pas du bruit. *Condition de réfutation* : si un re-codage sourcé annulait la sensibilité (toutes les familles devenant robustes), l'hypothèse selon laquelle la sensibilité porte une information sur la période serait infirmée.

Ces trois propositions sont, sur le présent pilote, des *preuves de concept* : les données sont un seed non vérifié (*draft_unverified_seed*), et leur test au sens fort exige le calibrage sourcé d'au moins une période et un codage inter-codeurs indépendant. Ce qui est acquis dès maintenant n'est pas leur vérité empirique, mais leur *énonçabilité* : le dispositif transforme des intuitions

diffuses en énoncés que l'on peut désormais chercher à réfuter.

3 Statut métrique des scores

Le point le plus vulnérable du modèle est aussi son point le plus visible : l'usage de scores numériques. Il faut donc poser explicitement leur statut métrique. Les scores de *Strates RI* ne sont pas des mesures cardinales fortes. Un score de 80 ne signifie pas qu'une théorie est deux fois plus présente qu'une théorie scorée à 40. L'écart entre 40 et 50 n'a pas nécessairement la même signification que l'écart entre 80 et 90. Les scores sont des jugements ordonnés et ancrés, traités comme quasi-intervalles pour permettre comparaison, audit et visualisation.

Cette hypothèse de quasi-intervalle est une convention méthodologique, non une propriété du monde. Elle est acceptable seulement si trois conditions sont respectées. Premièrement, les ancres de codage doivent être explicites. Deuxièmement, les résultats doivent être soumis à une analyse de sensibilité. Troisièmement, les conclusions doivent éviter les interprétations arithmétiques fortes. Il est légitime de dire qu'une théorie obtient un score nettement plus élevé qu'une autre selon le codebook ; il est illégitime de dire qu'elle est exactement deux fois plus explicative.

On objectera que tout l'appareil – moyenne pondérée des critères, normalisation, écarts de synthèse, `delta_max` – effectue précisément les opérations arithmétiques que ce statut ordinal devrait interdire. L'objection est fondée et doit être assumée plutôt qu'esquivée. Le parti retenu est délibéré et borné : les opérations cardinales servent à *ordonner et trier* des jugements, non à mesurer des quantités du monde. Un écart de synthèse de 18 points entre deux entités, ou de 32 points entre deux familles sur une couche, ne se lit pas comme « telle théorie est tant de fois plus présente » ; il se lit comme un rang assorti d'une marge – une hiérarchie nette et stable, non une distance métrique. La pondération par défaut (0.25, 0.40, 0.35) n'estime aucun poids réel : c'est une règle de combinaison transparente, dont l'analyse de sensibilité teste justement la fragilité. L'arithmétique est donc ici un instrument de classement contrôlé, pas une prétention métrologique ; sa légitimité ne tient pas à une propriété d'intervalle du réel, mais à ce qu'elle reste falsifiable par re-pondération et révisable par re-codage. Les écarts chiffrés du pilote doivent être lus dans ce régime ordinal-avec-marge, et non comme des mesures.

La grille d'ancrage générale est la suivante :

Score	Interprétation générale
0–20	Pertinence absente, marginale ou purement rétrospective sans capacité discriminante réelle.
21–40	Pertinence faible ou locale ; la théorie éclaire certains aspects mais ne structure pas l'interprétation de la période.
41–60	Pertinence moyenne ; la théorie explique des dimensions significatives mais non dominantes ou fortement disputées.
61–80	Pertinence forte ; la théorie explique des structures ou pratiques importantes et répétées, avec éléments probants solides.
81–100	Pertinence centrale ; la théorie capte une logique structurante majeure de la période, avec forte capacité discriminante et éléments probants robustes.

Le score brut est prioritaire. Il exprime l'intensité analytique attribuée à une théorie pour une couche donnée. Le score normalisé est dérivé :

$$S_{norm}(i) = \frac{S_{raw}(i)}{\sum_{j=1}^n S_{raw}(j)} \times 100$$

Cette normalisation facilite les graphiques, mais elle est conceptuellement dangereuse. Les théories ne sont pas mutuellement exclusives. Plusieurs familles peuvent devenir plus pertinentes en même temps. Par conséquent, aucune conclusion analytique ne doit être tirée uniquement d'un score normalisé. Toute interprétation substantielle doit revenir au score brut correspondant.

4 Taxonomie et matrice de non-confusion

La V1 de *Strates RI* utilise sept familles agrégées, dans le prolongement des cartographies disciplinaires reçues (Battistella, 2012; Acharya and Buzan, 2019) : réalisme ; libéralisme ; marxisme, dépendance et structuralisme ; constructivisme ; École anglaise ; critiques, postcoloniales et féministes ; géopolitique, impérialisme et raison d’État. Cette taxonomie est un compromis. Elle est assez compacte pour permettre des visualisations historiques de longue durée, mais assez large pour intégrer les principales traditions de la discipline.

Deux précautions sont indispensables. Premièrement, la famille « critiques, postcoloniales et féministes » n’est pas une équivalence théorique. Elle regroupe en V1 des traditions différentes pour des raisons de visualisation. Le modèle doit conserver des sous-familles distinctes : théorie critique, poststructuralisme, postcolonial/décolonial, féminisme RI, race et empire dans les RI. Deuxièmement, la famille « géopolitique, impérialisme et raison d’État » chevauche partiellement le réalisme et l’économie politique internationale. Elle n’est justifiée que si elle est traitée comme une lecture spatiale et matérielle du pouvoir : routes, mers, frontières, profondeur stratégique, ressources, glacis, empires territoriaux et contraintes géographiques.

En V1, chacune des sept familles est déjà décomposée en sous-familles dans la taxonomie canonique (`theory_taxonomy.json`), activables lorsqu’une période le justifie ; elles ne sont pas encore scorées séparément, mais sont déclarées pour interdire la lecture d’une famille comme un bloc homogène. La famille « critiques, postcoloniales et féministes » l’illustre : elle agrège six traditions distinctes qu’il ne faut pas confondre.

Sous-famille (identifiant)	Tradition
<code>critical_theory</code>	théorie critique (École de Francfort, Cox, Linklater)
<code>poststructuralism</code>	poststructuralisme
<code>postcolonialism</code>	approches postcoloniales
<code>feminist_ir</code>	féminisme des relations internationales
<code>decoloniality</code>	décolonialité
<code>critical_security_studies</code>	études critiques de sécurité

La part normalisée éventuelle de la famille agrégée ne doit donc jamais être lue comme une grandeur théorique unique : c’est une commodité de visualisation regroupant des programmes hétérogènes. La liste complète des sous-familles des sept familles est versionnée en annexe (`annexe_pilot_subfamilles.csv`).

La règle de non-confusion est la capacité discriminante. Une théorie ne reçoit pas un score élevé parce qu’elle peut absorber vaguement un phénomène. Elle reçoit un score élevé si elle explique mieux que les autres ce qui est en jeu. La matrice suivante sert de guide :

Phénomène	Réalisme	Géopolitique	Marxisme/IPE	Libéralisme/ES
Sphère d'influence militaire	élevé	élevé	faible/moyen	faible
Glacis territorial	élevé	élevé	faible	faible
Contrôle de routes maritimes	moyen	élevé	moyen	moyen si institution- nalisé
Extraction coloniale	faible/moyen	moyen	élevé	faible
Capitalisme impérial	moyen	moyen	élevé	faible
Institutionnalisation de règles communes	faible/moyen	faible	moyen	élevé
Normes de souveraineté et reconnaissance	moyen	faible	moyen	élevé pour École an- glaise/constructivisme
Production savante de hiérarchies impériales	faible	moyen	moyen	élevé pour cri- tique/postcolonial

Cette matrice n'est pas un automate. Elle force seulement l'analyste à ne pas confondre proximité sémantique et pouvoir discriminant. Son échelle est volontairement qualitative (« élevé / moyen / faible ») là où le reste du dispositif est numérique : à l'échelle d'un croisement phénomène-famille, chiffrer la capacité discriminante donnerait une précision factice. La granularité ordinale grossière est ici un choix de prudence, non une inconséquence : elle réserve le codage chiffré au niveau, mieux documenté, des couches.

Le chevauchement réalisme/géopolitique est le plus exposé à la confusion ; le pilote permet de le tester sur pièces. Les deux familles restent distinctes *en chiffres*, et l'écart se creuse là où l'enjeu n'est pas seulement la puissance brute mais la production savante de lectures spatiales. Sur les trois périodes pilotes, l'écart de score brut entre réalisme et géopolitique ([annexe_pilot_realism_vs_geopolitics.csv](#)) va de 2 points en 1815–1848 (réalité stratégique : 81.25 contre 79.25, où les deux lectures convergent) à un maximum de 32 points en 1991–2001 sur l'influence académique (réalisme 81.25 contre géopolitique 49.25) : le retour théorique du réalisme structurel après la guerre froide n'est pas un regain de la géopolitique. En 1945–1962, même là où les deux scores sont élevés sur la réalité stratégique (93.25 contre 79.25), l'écart de 14 points puis de 25 points sur l'influence académique montre que la matrice discrimine au lieu d'absorber.

5 Périodisation et degré d'intégration du système

La périodisation n'est jamais neutre. Elle impose les unités temporelles à l'intérieur desquelles les scores sont calculés, et arbitre implicitement entre le temps court de l'événement et la longue durée des structures ([Braudel, 1958](#)). Découper en périodes, c'est déjà décider de ce qui fait rupture et de ce qui fait continuité. La V1 retient quatorze périodes : 1815–1848 ; 1848–1871 ; 1871–1914 ; 1914–1919 ; 1919–1939 ; 1939–1945 ; 1945–1962 ; 1962–1979 ; 1979–1991 ; 1991–2001 ; 2001–2008 ; 2008–2014 ; 2014–2022 ; 2022–2026.

Ces ruptures sont contestables. C'est précisément pourquoi elles doivent être justifiées. 1815 marque le Congrès de Vienne et la recomposition de l'ordre européen ([Schroeder, 1994](#) ; [Kissinger, 1957](#)). 1848 signale les révolutions européennes et les fissures de l'ordre conservateur. 1871 combine l'unification allemande, la reconfiguration de l'équilibre européen et le départ d'une phase impériale et militarisée qui mène à 1914 ([Kennedy, 1987](#) ; [Clark, 2012](#)). 1919 est le moment de Versailles et de la Société des Nations, ainsi que la crise de l'internationalisme libéral de l'entre-deux-guerres ([Carr, 1939](#) ; [MacMillan, 2001](#) ; [Tooze, 2014](#)). 1945 ouvre l'ordre ONU/Bretton Woods et la bipolarisation. 1962, après Cuba, signale une stabilisation relative de la dissuasion et une phase de détente conflictuelle. 1979 concentre plusieurs bascules : révolution

iranienne, invasion soviétique de l’Afghanistan, crise de la détente et inflexion néolibérale. 1991 marque la disparition de l’URSS. 2001 reconfigure les doctrines de sécurité autour de la guerre contre le terrorisme. 2008 fragilise la confiance dans la mondialisation libérale. 2014 cristallise la rupture ukrainienne, la compétition sino-américaine et le retour plus visible de la coercition territoriale. 2022 ouvre une phase de guerre de haute intensité en Europe, de sanctions massives, de militarisation et de fragmentation économique partielle.

Le modèle doit aussi reconnaître que le « système international » n’a pas le même degré de globalité en 1815 et en 2026. Un indicateur de degré d’intégration systémique – faible, moyen, élevé ou très élevé – est donc attaché à chaque période. En 1815, l’ordre est fortement européen et impérial, avec des interactions globales médiées par les empires. En 2026, les interdépendances financières, technologiques, énergétiques et informationnelles sont beaucoup plus denses. Comparer les deux périodes exige de ne pas faire comme si l’objet était parfaitement stable.

6 Unités d’analyse et couches de scoring

Le modèle distingue trois couches : réalité stratégique, doctrine politique et influence académique. La première porte sur ce qui structure les rapports de force – polarité, capacités militaires, alliances, conflits, flux économiques, infrastructures, institutions effectives, hiérarchies impériales et dépendances – au sens où Aron faisait de la configuration des puissances et de la conduite diplomatico-stratégique l’objet premier de l’analyse (Aron, 1962). La seconde porte sur ce que les acteurs disent, codifient ou revendiquent : doctrines d’État, discours diplomatiques, textes stratégiques, traités, votes, pratiques institutionnelles et langage de légitimation. La troisième porte sur le champ intellectuel – livres, articles, manuels, débats universitaires, écoles, citations, programmes et reconnaissance disciplinaire – c’est-à-dire sur la position d’une théorie dans le *champ scientifique*, sa consécration et son autorité, et non sur sa vérité (Bourdieu, 1976). Mesurer l’influence académique, c’est cartographier un rapport de forces savant, pas trancher un débat épistémique.

Ce dernier point n’est pas un ornement théorique : il commande l’opérationnalisation de la couche. Si l’influence académique mesure une position dans le champ scientifique, alors ses indicateurs ne sont pas la validité d’une théorie mais les marqueurs de sa consécration – manuels et programmes qui la canonisent, citations et écoles qui l’instituent, revues et postes qui la reproduisent (Bourdieu, 1976). Trois conséquences en découlent, directement inscrites dans le codage. D’abord, une théorie peut être féconde et faiblement consacrée, ou consacrée et contestée : la couche ne préjuge pas de sa vérité. Ensuite, la consécration est datable et cumulative, ce qui fonde l’anti-anachronisme *sur cette couche précise* – une école ne pèse pas avant d’exister comme position dans le champ – et étaye la proposition P2. Enfin, le champ savant ayant sa temporalité propre, l’influence académique peut durablement diverger de la réalité stratégique : c’est cette divergence que le dispositif rend visible au lieu de la lisser. Les autres références continentales (Aron, Braudel, la sociologie des RI de Merle, Badie et Devin) jouent ici un rôle plus circonscrit, de positionnement et d’ancrage disciplinaire ; seule la théorie du champ est mobilisée comme principe opératoire d’une couche.

Cette distinction évite une confusion majeure. Une théorie peut être analytiquement utile pour expliquer une période sans exister comme doctrine des acteurs ou comme école académique de l’époque. Le constructivisme peut éclairer rétrospectivement la souveraineté ou la légitimité au XIXe siècle ; il ne peut pas être compté comme influence académique à cette date. Le néoréalisme peut éclairer la bipolarité de la Guerre froide ; il ne peut pas être doctrine de Metternich. L’anti-anachronisme n’interdit pas l’analyse rétrospective, mais il oblige à étiqueter la couche concernée.

Les unités d’analyse incluent le système global, les grandes puissances, les empires, les alliances ou blocs, les régions stratégiques, les espaces coloniaux ou postcoloniaux, les institutions internationales et certains acteurs transnationaux. V1 traite principalement États, empires, blocs

et institutions. Les firmes multinationales, ONG, réseaux financiers, diasporas, groupes armés ou organisations terroristes sont prévus dans l’architecture, mais ne doivent pas être sur-promis tant que leur codebook spécifique n’est pas stabilisé.

7 Protocole de scoring

Le score brut combine trois blocs de critères, chacun étant la moyenne des critères qui le composent : critères communs (C), critères spécifiques à la théorie (T) et adéquation à la couche (L). Le profil de pondération par défaut, tel qu’implémenté dans l’application (`default_criteria_weights`), est :

$$S_{raw} = 0.25 C + 0.40 T + 0.35 L$$

Le bloc commun C regroupe trois critères réellement codés : pouvoir explicatif, centralité historique et robustesse des éléments probants. Le bloc spécifique T comprend deux critères synthétiques : l’adéquation au noyau théorique (`fit_to_theory_core`) et la capacité discriminante (`discriminating_value`). Le bloc spécifique reçoit le poids le plus élevé (0.40). Ces deux critères condensent les dimensions propres à chaque famille : compétition de puissance, dilemme de sécurité, polarité, militarisation, alliances et intérêt national pour le réalisme ; institutions, interdépendance, démocratie, coopération, droit et régimes pour le libéralisme ; normes, identités, légitimité, reconnaissance et socialisation pour le constructivisme ; société internationale, souveraineté, ordre, droit et diplomatie pour l’École anglaise ; capitalisme mondial, impérialisme économique, centre-périphérie, production, finance et hégémonie matérielle pour l’IPE/marxisme/dépendance ; domination, colonialité, genre, race, savoir-pouvoir et violences structurelles pour les approches critiques, postcoloniales et féministes ; territoire, routes, mers, ressources, frontières et profondeur stratégique pour la géopolitique. Le bloc L mesure l’adéquation à la couche considérée (`layer_fit`). Le score brut n’est donc pas saisi librement : il est la moyenne pondérée de ces trois blocs, et une garde au build (tolérance 0.5 point) rejette tout score qui divergerait de ses propres critères.

Le protocole suit sept étapes :

1. Définir la période et la couche.
2. Lister les éléments historiques majeurs avant de scorer.
3. Transformer ces éléments en éléments probants structurés.
4. Associer chaque élément probant à plusieurs théories possibles.
5. Scorer les critères communs et spécifiques.
6. Documenter au moins deux éléments contradictoires pour tout score supérieur ou égal à 85.
7. Calculer le score, la confiance et la sensibilité aux pondérations.

Cette procédure sert à limiter la circularité. L’analyste ne doit pas partir de la théorie dominante présumée pour sélectionner seulement les faits qui la confirment. Il doit d’abord stabiliser les éléments probants, puis scorer les théories.

8 Cas pilote : le système international de 1945 à 1962

Le cas pilote applique le protocole au système international global de 1945–1962. Cette période est choisie parce qu’elle est historiographiquement centrale, fortement documentée et théoriquement discriminante. Elle combine bipolarisation, institutionnalisation, décolonisation initiale, rivalité idéologique, dissuasion nucléaire et structuration économique occidentale. Elle permet donc de tester si le modèle évite de réduire la période à une seule théorie.

8.1 Éléments probants principaux

Les éléments probants retenus sont les suivants :

1. bipolarisation États-Unis/URSS ;
2. monopole puis dyarchie nucléaire progressive ;
3. doctrine Truman et containment ;
4. plan Marshall et reconstruction occidentale ;
5. création de l'ONU ;
6. institutions de Bretton Woods ;
7. création de l'OTAN ;
8. guerre de Corée ;
9. crises de Berlin ;
10. crise de Cuba ;
11. décolonisation initiale ;
12. Bandung et montée du non-alignement ;
13. révolution chinoise et guerre civile chinoise ;
14. consolidation d'un ordre économique occidental.

Ces éléments ne soutiennent pas tous la même théorie. La bipolarité, la dissuasion, les alliances et la guerre de Corée soutiennent fortement le réalisme. L'ONU, Bretton Woods, l'OTAN et le plan Marshall soutiennent aussi le libéralisme institutionnel, quoique souvent sous condition hégémonique. La décolonisation, Bandung et les asymétries économiques soutiennent les lectures postcoloniales et dépendantistes. Les normes de souveraineté, d'autodétermination et de reconnaissance soutiennent le constructivisme et l'École anglaise. Cette pluralité est précisément ce que le protocole doit conserver.

8.2 Scores pilotes par couche

Le tableau suivant présente les scores bruts pilotes, tels qu'appliqués dans l'atlas *ir-strata* (l'application versionnée du protocole). Ils sont codés par l'auteur selon le codebook de cet article ; ils ne constituent pas une validation inter-codeurs, mais servent à tester la cohérence interne du protocole. Conformément à la règle d'anti-anachronisme appliquée strictement (voir plus bas), une théorie non encore émergée comme doctrine ou comme école à l'époque n'est pas scorée faiblement : elle est *rejetée* (exclue) sur la couche concernée. Le constructivisme (formalisé vers 1989), l'École anglaise (vers 1959) et les approches critiques/postcoloniales/féministes (champ des RI vers 1978) sont donc rejetés en doctrine politique et en influence académique pour 1945–1962, et ne subsistent qu'en réalité stratégique, au titre d'une lecture rétrospective.

Famille théorique	Réalité stratégique	Doctrine politique	Influence académique
Réalisme	93.25	87.25	84.25
Libéralisme	69.25	74.25	69.25
Marxisme/dépendance/IPE	67.25	71.25	74.25
Constructivisme	44.25	rejeté	rejeté
École anglaise	64.25	rejeté	rejeté
Critiques/postcoloniales/féministes	44.25	rejeté	rejeté
Géopolitique/impérialisme/raison d'État	79.25	71.25	59.25

La synthèse par défaut (poids 0.60, 0.25 et 0.15 pour réalité stratégique, doctrine politique et influence académique) n'est calculée que pour les familles présentes sur les *trois* couches. Les familles rejetées en doctrine et en académique n'ont donc pas de synthèse ; elles ne figurent qu'en réalité stratégique. La synthèse, normalisée sur les quatre familles synthétisables, donne :

Famille théorique	Score brut synthétique	Score normalisé
Réalisme	90.40	29.7
Géopolitique/impérialisme/raison d'État	74.25	24.4
Libéralisme	70.50	23.2
Marxisme/dépendance/IPE	69.30	22.8

L'interprétation n'est pas : « 1945–1962 est réaliste à 29.7 % ». Cette phrase serait méthodologiquement fautive. L'interprétation correcte est : le réalisme obtient le plus haut score brut synthétique, avec forte capacité discriminante pour la réalité stratégique ; les scores géopolitique et libéral restent élevés, ce qui interdit une lecture exclusivement réaliste ; le marxisme/dépendance reste significatif à l'échelle globale, tout en pouvant être beaucoup plus fort sur des entités ou espaces postcoloniaux spécifiques. La normalisation ne porte que sur les quatre familles synthétisables : elle exprime une part relative *dans cet ensemble restreint*, non une part du réel. Le constructivisme, l'École anglaise et les approches critiques conservent par ailleurs une présence notable en réalité stratégique (44.25, 64.25 et 44.25), comme grilles de lecture rétrospectives de structures réelles (souveraineté, société internationale, décolonisation).

8.3 Exemple détaillé : réalisme, réalité stratégique

Pour le réalisme dans la couche réalité stratégique, les trois critères communs sont codés ainsi : pouvoir explicatif 99, centralité historique 94, robustesse des éléments probants 92. Moyenne : 95. Les deux critères spécifiques sont : adéquation au noyau théorique 94, capacité discriminante 89. Moyenne : 91.5. L'adéquation à la couche (`layer_fit`) est 94. Le score brut, dérivé par la moyenne pondérée des trois blocs, est :

$$S_{raw} = 0.25(95) + 0.40(91.5) + 0.35(94) = 93.25$$

Ce score (93.25) n'est pas saisi indépendamment : il est exactement celui que l'application dérive de ses propres critères, et une garde au build rejette tout écart supérieur à 0.5 point entre le score et ses critères (voir la section *Reproductibilité*).

Les éléments probants principaux sont la bipolarité, le containment, l'OTAN, la guerre de Corée, les crises de Berlin, la dissuasion nucléaire et Cuba. Les éléments contradictoires sont l'institutionnalisation libérale occidentale, l'ONU, Bretton Woods, la décolonisation et les normes d'autodétermination. Ces contradictions n'annulent pas le score réaliste ; elles empêchent seulement d'en faire une théorie totale.

8.4 Exemple détaillé : constructivisme, influence académique

Pour le constructivisme dans la couche influence académique, l'application ne retient pas un score faible : elle *rejette* le score (statut `deprecated`, motif anachronisme). Ce n'est pas parce que les normes ou identités seraient absentes de la période ; c'est parce que le constructivisme comme école académique des RI n'est pas encore formalisé. La couche académique exige des textes, débats et écoles reconnus dans le champ, or les travaux constructivistes majeurs sont postérieurs (Onuf, 1989; Kratochwil, 1989; Wendt, 1999). Deux traitements de l'anachronisme sont possibles : attribuer un score faible (bande 0–20) en reconnaissant d'éventuels précurseurs, ou exclure entièrement la couche tant que l'école n'a pas émergé. L'atlas *ir-strata* retient la seconde option, plus stricte : un score doctrine/académique antérieur à l'émergence de la théorie

est rejeté et exclu des graphiques et de la synthèse, plutôt que dilué dans un score bas. Le constructivisme demeure en revanche admissible en réalité stratégique (44.25), comme lecture rétrospective des normes de souveraineté et de légitimité de la période.

9 Analyse de sensibilité

L'analyse de sensibilité teste si la synthèse dépend trop des pondérations de couches. Elle porte sur la *synthèse brute* – conformément au primat du score brut défendu plus haut – recalculée sous les cinq profils de pondération de l'application : défaut (0.60/0.25/0.15), stratégique dur (0.75/0.15/0.10), doctrinal (0.45/0.40/0.15), académique (0.35/0.20/0.45) et équilibré (0.50/0.30/0.20). Le `delta_max` est l'écart maximal de la synthèse brute entre profils ; le statut suit les seuils robuste (au plus 5 points), modérément sensible (6–10), sensible (11–15) et très sensible (plus de 15).

Famille	Défaut	Strat. dur	Doctrinal	Acad.	Équil.	delta_max
Réalisme	90.40	91.45	89.50	88.00	89.65	3.45
Géopolitique	74.25	76.05	73.05	68.65	72.85	7.40
Libéralisme	70.50	70.00	71.25	70.25	70.75	1.25
Marxisme/dépendance	69.50	68.55	69.90	71.20	69.85	2.65

L'analyse ne porte que sur les quatre familles synthétisables : le constructivisme, l'École anglaise et les approches critiques sont rejetés en doctrine et en académique pour cette période et n'entrent donc pas dans la synthèse (ils ne varient qu'en réalité stratégique). Le résultat principal est robuste : le réalisme reste premier dans tous les profils, avec un `delta_max` de 3.45 points (statut robuste). La géopolitique reste deuxième mais plus sensible (`delta_max` de 7.40 points, modérément sensible), reflétant un poids stratégique fort et une influence académique plus faible. Libéralisme et marxisme/dépendance sont très proches et très stables (`delta_max` inférieurs à 3 points). Cette analyse montre l'utilité du dispositif : il ne suffit pas d'afficher un score, il faut montrer comment il réagit aux hypothèses de synthèse. Elle est reconduite plus loin sur 1991–2001.

10 Fiabilité inter-codeurs : protocole de validation

Le cas pilote ci-dessus est un codage auteur unique. Il ne prouve donc pas la fiabilité inter-codeurs. Le papier ne doit pas prétendre le contraire. La validation académique complète exige un test indépendant.

La cohérence de la démarche impose ici un retour réflexif sur le codeur lui-même – faute de quoi l'invocation de l'objectivation participante resterait creuse. Le pilote a été codé par un unique analyste, formé dans une tradition disciplinaire donnée, qui n'est neutre ni envers le réalisme ni envers les approches critiques. Ses jugements de « capacité discriminante » sont donc situés : ils portent la trace de sa propre position dans le champ qu'il prétend cartographier. Le dispositif ne dissout pas ce biais ; il l'expose et le rend attaquant, en consignait critères, éléments probants, contradictions et niveaux de confiance, de sorte qu'un tiers puisse identifier où le codeur a vu de la centralité là où un autre verrait de la marge. C'est en ce sens précis, et seulement en ce sens, que l'exercice relève d'une objectivation participante : il objective les schèmes de classement d'un codeur situé, sans prétendre s'en affranchir. La levée effective du biais ne viendra que du codage indépendant décrit ci-dessous.

Le protocole proposé est le suivant : trois codeurs reçoivent le même codebook, la même liste d'éléments probants de base et les mêmes références minimales. Ils codent indépendamment trois périodes : 1815–1848, 1945–1962 et 1991–2001. Pour chaque période, ils codent sept familles théoriques sur trois couches. L'échantillon minimal comprend donc 63 scores par codeur. La

mesure principale est l'écart absolu moyen. Un score est considéré comme suffisamment fiable si l'écart moyen inter-codeurs est inférieur ou égal à 10 points. Les écarts de 11 à 15 points déclenchent discussion ; les écarts supérieurs à 15 points créent un statut *contested*. Une mesure complémentaire peut discrétiser les scores en cinq classes correspondant aux ancrs 0–20, 21–40, 41–60, 61–80 et 81–100, puis calculer un accord inter-codeurs.

La résolution des divergences ne doit pas effacer le désaccord. Elle doit produire un journal de révision indiquant le score initial, le score révisé, les raisons du changement, les éléments probants ajoutés ou retirés et l'impact sur la confiance. Cette procédure est plus importante qu'une apparence d'accord artificiel.

Ce protocole est désormais outillé dans l'application. Un harnais calcule, dès qu'au moins deux codages indépendants existent, l'écart absolu moyen, l'ICC(3,1) et l'alpha de Krippendorff par bloc (période et couche), avec un seuil d'alerte à 10 points. Une console de codage permet à un spécialiste de re-coder le sous-ensemble protocole à *l'aveugle* : le score de l'auteur n'est jamais affiché, seuls le contexte et les éléments probants le sont, et le score brut est dérivé en direct des critères saisis. Tant qu'aucun codage indépendant n'est déposé, l'outil reste en mode informatif et les données conservent leur statut de *seed* non vérifié : l'infrastructure de validation est prête, la preuve empirique reste à produire par des codeurs humains.

11 Extension du pilote : trois périodes

Le cas 1945–1962 démontre l'applicabilité du protocole, mais une période unique ne teste ni sa robustesse temporelle ni son comportement face à l'anachronisme. Cette section étend donc le pilote à deux périodes supplémentaires choisies pour leur difficulté méthodologique : 1815–1848, qui teste l'anachronisme et l'ordre européen pré-disciplinaire, et 1991–2001, qui teste un cas controversé d'ordre libéral hégémonique. Les scores ci-dessous ne sont pas recodés *ad hoc* : ils sont exactement ceux servis par l'application *ir-strata* (couches brutes, scope global), de sorte que l'extension reste auditable et reproductible.

11.1 1815–1848 : un cas où la synthèse n'existe pas

La période du Concert européen et de l'ordre conservateur (Schroeder, 1994; Kissinger, 1957) produit un résultat instructif précisément parce que la méthode *refuse* d'y calculer une synthèse. Aucune des sept familles n'existe comme école académique des RI à cette date : les émergences disciplinaires (géopolitique vers 1890, marxisme/IPE vers 1902, libéralisme vers 1919, réalisme vers 1939) sont toutes postérieures à 1848. La couche `academic_influence` est donc *rejetée pour toutes les familles*, et comme la synthèse n'est définie que pour les familles présentes sur les trois couches, **aucune synthèse globale n'est calculée pour 1815–1848**.

Famille théorique	Réalité stratégique	Doctrines politique	Influence académique
Réalisme	81.25	77.25	rejeté
Géopolitique/impérialisme/raison d'État	79.25	75.25	rejeté
École anglaise	57.25	rejeté	rejeté
Constructivisme	37.25	rejeté	rejeté
Libéralisme	29.25	27.25	rejeté
Critiques/postcoloniales/féministes	21.85	rejeté	rejeté
Marxisme/dépendance/IPE	13.25	rejeté	rejeté

En réalité stratégique, la lecture rétrospective place le réalisme (81.25) et la géopolitique/raison d'État (79.25) en tête, ce qui est cohérent avec un ordre fondé sur l'équilibre des puissances, les sphères dynastiques et la gestion concertée des crises. L'École anglaise obtient un score notable

(57.25) comme grille rétrospective de la société internationale et de ses institutions primaires (souveraineté, diplomatie, équilibre) (Bull, 1977; Watson, 1992). En doctrine politique, seules les familles dont les idées sont disponibles aux acteurs de l'époque sont admises : raison d'État et géopolitique (héritage ancien), libéralisme (post-1789). Le constructivisme (37.25 en réalité stratégique) éclaire rétrospectivement les normes de légitimité dynastique sans jamais pouvoir être doctrine de Metternich ni école académique. Ce cas montre que l'anti-anachronisme strict n'est pas un détail technique : il change la nature même du résultat, en assumant explicitement qu'il n'existe pas de pondération synthétique là où une couche n'a pas encore d'existence historique.

11.2 1991–2001 : un ordre libéral non réductible au libéralisme

La décennie post-soviétique est le cas controversé par excellence : moment unipolaire américain (Krauthammer, 1990), expansion institutionnelle et normative (Ikenberry, 2001, 2011), mais aussi hégémonie matérielle et usages de la force que les lectures critiques relisent comme reproduction d'asymétries (Cox, 1981; Said, 1978). Les sept familles sont ici synthétisables (le constructivisme, l'École anglaise et les approches critiques ont émergé comme écoles avant 1991).

Famille théorique	R. strat.	Doctrine	Acad.	Synthèse brute	Norm.
Libéralisme	89.25	93.25	87.25	89.95	19.3
Constructivisme	71.25	77.25	89.25	75.45	16.2
Réalisme	74.25	64.25	81.25	72.80	15.6
École anglaise	67.25	71.25	69.25	68.55	14.7
Critiques/postcoloniales/féministes	57.25	54.25	77.25	59.50	12.8
Géopolitique/impérialisme/raison d'État	54.25	47.25	49.25	51.75	11.1
Marxisme/dépendance/IPE	47.25	41.25	61.25	47.85	10.3

Le résultat n'est pas trivial. Le libéralisme domine la synthèse (89.95), ce qui correspond au diagnostic dominant d'un « moment libéral ». Mais la méthode interdit la lecture exclusivement libérale : le réalisme reste élevé en réalité stratégique (74.25), traduisant la condition hégémonique et les usages de la force (Golfe, Kosovo) que le score doctrinal libéral (93.25) tend à recouvrir ; le constructivisme est deuxième (75.45), tiré par une influence académique forte (89.25) qui reflète son essor des années 1990 (Finnemore and Sikkink, 1998; Wendt, 1999) ; les approches critiques et postcoloniales conservent une présence réelle (59.50), notamment en influence académique (77.25). L'écart entre une doctrine politique très libérale et une réalité stratégique plus disputée est précisément ce que les couches rendent visible : c'est l'apport discriminant du dispositif sur ce cas controversé.

11.3 Le global masque l'entité : 1945–1962

Le score global est une moyenne pondérée par le poids systémique des unités ; il lisse une hétérogénéité que le modèle conserve pourtant à l'échelle des entités. L'application *ir-strata* synthétise et normalise les scores *par entité* avec exactement le même protocole que le scope global. Le tableau ci-dessous décline la synthèse brute de 1945–1962 (part normalisée entre parenthèses, à ne jamais lire seule) pour six entités majeures, comparée au score global de référence. Les chiffres sont dérivés par script (`annexe_pilot_entities_1945_1962.csv`).

Entité	Réalisme	Géopolitique	Libéralisme	Marxisme/dép.
Système global (réf.)	90.40 (29.7)	74.25 (24.4)	70.50 (23.2)	69.30 (22.8)
États-Unis	96.32 (32.2)	64.57 (21.6)	78.82 (26.3)	59.62 (19.9)
URSS	96.32 (32.2)	64.57 (21.6)	60.82 (20.3)	77.62 (25.9)
Mondes en décolonisation	80.72 (28.5)	64.57 (22.8)	60.82 (21.4)	77.62 (27.4)
Europe occidentale	80.72 (28.5)	64.57 (22.8)	78.82 (27.8)	59.62 (21.0)
Chine	96.32 (32.2)	64.57 (21.6)	60.82 (20.3)	77.62 (25.9)
ONU	80.72 (28.5)	64.57 (22.8)	78.82 (27.8)	59.62 (21.0)

La divergence est nette et chiffrable. Le marxisme/dépendance, à 69.30 (22.8 %) au global, monte à 77.62 (27.4 %) sur les mondes en décolonisation et descend à 59.62 (19.9 %) sur les États-Unis : l'écart inter-entités atteint 18 points de score brut sur cette seule famille. Symétriquement, le libéralisme passe de 78.82 à l'Ouest (États-Unis, Europe occidentale, ONU) à 60.82 à l'Est (URSS, Chine, mondes en décolonisation). Le réalisme reste premier partout, mais plus haut sur les grandes puissances (96.32) que sur les institutions et les espaces périphériques (80.72). Le score global de 1945–1962 — réalisme dominant, marxisme/dépendance significatif mais quatrième — est donc une résultante, non un fait uniforme : lu seul, il masque que la dépendance domine presque les profils périphériques.

Deux gardes s'imposent. D'abord, ces scores d'entité sont eux aussi un *seed non vérifié* (**draft_unverified_seed**) : leur regroupement par blocs (Ouest/Est, grandes puissances/périphérie) reflète la génération de la graine, non une mesure historiographique ; ce qui est démontré ici est une *propriété structurelle* du dispositif — l'agrégat global lisse l'hétérogénéité — et non le détail des valeurs. Ensuite, les parts normalisées par entité restent des parts relatives *dans l'ensemble restreint des quatre familles synthétisables*, sans exclusivité mutuelle : aucune conclusion ne doit être tirée du seul pourcentage. La limite « le score global masque l'entité », autrefois seulement reconnue, est désormais quantifiée.

12 Sensibilité aux pondérations : 1991–2001

L'analyse de sensibilité est reconduite sur 1991–2001, selon les cinq mêmes profils et la même définition du **delta_max** (écart maximal de la synthèse brute entre profils ; seuils robuste au plus 5, modérément sensible 6–10, sensible 11–15, très sensible plus de 15). La période 1945–1962 a été traitée plus haut ; 1815–1848 n'a aucune synthèse.

1991–2001	Défaut	Strat. dur	Doctrinal	Acad.	Équil.	delta_max
Libéralisme	89.95	89.65	90.55	89.15	90.05	1.40
Constructivisme	75.45	73.95	76.35	80.55	76.65	6.60
Réalisme	72.80	73.45	71.30	75.40	72.65	4.10
École anglaise	68.55	68.05	69.15	68.95	68.85	1.10
Critiques/postcoloniales	59.50	58.80	59.05	65.65	60.35	6.85
Géopolitique	51.75	52.70	50.70	50.60	51.15	2.10
Marxisme/dépendance	47.85	47.75	46.95	52.35	48.25	5.40

Aucun score n'est sensible au sens fort (delta_max au-delà de 15). Les classements dominants survivent à tous les profils : le réalisme reste premier en 1945–1962, le libéralisme reste premier en 1991–2001. La sensibilité, lorsqu'elle existe, est interprétable plutôt qu'aléatoire : la géopolitique de 1945–1962 (delta_max de 7.4) chute sous profil académique, ce qui reflète son influence universitaire alors plus faible que sa réalité stratégique ; le constructivisme et les approches critiques de 1991–2001 (delta_max de 6.6 et 6.85) montent sous profil académique, traduisant leur poids disciplinaire des années 1990. La sensibilité mesure donc une propriété réelle de la période, pas un artefact.

13 Simulation de second codage

Le pilote reste un codage auteur unique. En l’absence de second codeur humain indépendant, on produit une `simulated_second_coding` explicitement marquée comme telle. **Cette simulation ne remplace pas un vrai test inter-codeurs** : elle sert seulement à détecter des zones de fragilité avant un codage humain indépendant. Le codeur simulé applique une règle déterministe et déclarée : escompter de 8 points la famille dominante de chaque période (méfiance envers la lecture dominante), relever de 6 points le marxisme/dépendance et les approches critiques (hypothèse d’une sous-pondération des hiérarchies mondiales à l’échelle globale), abaisser de 5 points le réalisme lorsqu’il n’est pas déjà dominant (méfiance envers le primat de la puissance). Seule la couche de synthèse est re-scorée ; 1815–1848 est exclue puisqu’elle n’a aucune synthèse.

1945–1962	Principal	Simulé	Écart abs.	Statut
Réalisme	90.40	82.40	8.00	aligné
Libéralisme	70.50	70.50	0.00	aligné
Marxisme/dépendance/IPE	69.30	75.30	6.00	aligné
Géopolitique	74.25	74.25	0.00	aligné

1991–2001	Principal	Simulé	Écart abs.	Statut
Libéralisme	89.95	81.95	8.00	aligné
Constructivisme	75.45	75.45	0.00	aligné
Réalisme	72.80	67.80	5.00	aligné
École anglaise	68.55	68.55	0.00	aligné
Critiques/postcoloniales/fém.	59.50	65.50	6.00	aligné
Géopolitique	51.75	51.75	0.00	aligné
Marxisme/dépendance/IPE	47.85	53.85	6.00	aligné

L’écart absolu moyen est de 3.55 points, l’écart maximal de 8 points, et aucun score ne franchit le seuil de désaccord (11–15) ni le seuil *contested* (au-delà de 15). Ce résultat doit être lu avec prudence : l’alignement est en partie une conséquence de la modestie, par construction, des ajustements simulés. Un codeur humain pourrait diverger davantage, en particulier sur les familles critiques et marxistes (dont l’ampleur dépend du choix d’unité d’analyse) et sur la décision structurante de *rejeter* plutôt que de scorer faiblement les théories non encore émergées. La simulation indique seulement que les classements dominants ne reposent pas sur un point d’équilibre fragile.

14 Audit de cohérence méthodologique

L’extension à trois périodes permet de répondre explicitement aux questions d’audit que toute méthode de scoring doit affronter.

La méthode produit-elle des résultats triviaux ? Partiellement, et c’est attendu : « 1945–1962 = réalisme élevé » est trivial. Mais le dispositif ajoute le rôle non trivial du libéralisme institutionnel et de la géopolitique, la distinction entre réalité stratégique et doctrine, et surtout, pour 1991–2001, un résultat non trivial où le libéralisme domine la synthèse sans que la lecture libérale soit autorisée à devenir totale (réalisme stratégique à 74.25, présence critique à 59.50).

La méthode produit-elle des résultats absurdes ? Aucun score contre-intuitif manifeste n’apparaît. Le score le plus surprenant — marxisme/dépendance à 47.85 seulement en 1991–2001 — est défendable à l’échelle globale et explicitement signalé comme pouvant être bien plus élevé sur des entités ou espaces périphériques spécifiques.

Les résultats sont-ils trop sensibles aux pondérations ? Non : aucun `delta_max` ne dépasse 15 ; les sensibilités observées (6–7 points) sont interprétables historiquement.

Les catégories se chevauchent-elles trop ? Le risque est réel entre réalisme et géopolitique, et entre marxisme et approches critiques. La matrice de non-confusion et la capacité discriminante le limitent sans l'annuler : c'est une limite assumée de la taxonomie V1, non un défaut masqué.

L'anachronisme est-il contrôlé ? Oui, et le cas 1815–1848 en est la démonstration la plus forte : la méthode va jusqu'à ne produire aucune synthèse plutôt que d'attribuer une influence académique fictive à des écoles non encore nées.

Verdict : `partially_holds`.

Raisons : (1) la méthode produit des résultats cohérents, non triviaux et robustes aux pondérations sur les périodes synthétisables ; (2) le contrôle de l'anachronisme est effectif, jusqu'à refuser la synthèse quand une couche n'existe pas ; (3) mais la fiabilité reste seulement *simulée*, le score global masque la variation par entité, et le choix de rejeter plutôt que de scorer faiblement les théories non émergées est une hypothèse de modélisation forte qui conditionne une partie des résultats.

Conditions d'amélioration : (1) un vrai test inter-codeurs humain sur les 63 scores du protocole ; (2) une extension de la déclinaison par entité — amorcée ici sur 1945–1962, où elle quantifie déjà l'hétérogénéité masquée par le global — aux autres périodes synthétisables ; (3) une étude de robustesse du choix « rejet vs score faible » pour l'anti-anachronisme, idéalement avec les deux variantes publiées côte à côte.

15 Validité, falsifiabilité et modes d'échec

La validité conceptuelle dépend de la clarté du concept de « poids théorique ». Celui-ci ne signifie ni vérité d'une théorie, ni fréquence de citation, ni domination idéologique pure. Il signifie pertinence relative pour expliquer, doctrinaliser ou structurer académiquement une période donnée. La validité interne dépend de la cohérence entre critères et score. La validité externe dépend de la capacité du codebook à traiter des périodes très différentes sans changer de sens. La validité de conclusion dépend de la distinction entre score brut et normalisé.

Le modèle est falsifiable au sens faible : un score doit pouvoir baisser si des éléments probants contradictoires sont ajoutés, si une autre théorie a une meilleure capacité discriminante, si la période est redécoupée, si la sensibilité aux pondérations devient trop forte ou si le risque d'anachronisme a été sous-estimé. Un score irrévocable est une position doctrinale, pas un résultat. Cette exigence rejoint le principe plus général selon lequel l'inférence qualitative doit expliciter ses observations, ses hypothèses et ses opérations de sélection (King et al., 1994).

Les modes d'échec sont nombreux : fausse précision ; circularité ; biais occidental ; surpondération des grandes puissances ; taxonomie fourre-tout ; confusion doctrine/réalité ; normalisation abusive ; anachronisme masqué ; sous-traitement des acteurs non étatiques ; bibliographie décorative ; sur-ingénierie technique ; graphisme séduisant mais trompeur. Ces risques ne sont pas périphériques. Ils définissent le coût de la méthode. La quantification interprétative n'est défendable que si elle rend ses propres modes d'échec visibles.

16 Discussion : ce que montre le cas pilote

Le cas 1945–1962 montre trois choses. Premièrement, le réalisme domine bien la couche de réalité stratégique, mais l'ordre n'est pas réductible au réalisme : le libéralisme institutionnel et la géopolitique captent des dimensions importantes de l'institutionnalisation et des sphères post-1945. Deuxièmement, les approches critiques, postcoloniales et constructivistes ne doivent pas être réduites à zéro : leur influence académique est *rejetée* pour la période (l'école n'a pas encore émergé), mais elles restent des grilles de lecture rétrospectives valides en réalité stratégique pour éclairer la décolonisation, la souveraineté et les hiérarchies mondiales. La distinction entre couches change donc l'interprétation : une théorie peut être présente en réalité stratégique tout en étant absente comme doctrine et comme école de l'époque. Troisièmement, l'anti-anachronisme

strict (rejet plutôt que score faible) déplace la synthèse : seules les familles présentes sur les trois couches sont agrégées, ce qui assume explicitement qu’aucune synthèse globale n’est définie là où une couche n’existe pas encore.

L’extension à 1815–1848 et 1991–2001 fait apparaître ce que le seul cas de guerre froide ne pouvait montrer, et donne corps aux trois propositions de la section de contribution. Le contraste entre les deux périodes est instructif pour P1 : en 1815–1848, configuration multipolaire, la doctrine réaliste (77.25) reste *en deçà* de la réalité stratégique (81.25) ; en 1991–2001, configuration unipolaire, la doctrine libérale (93.25) surplombe une réalité stratégique plus disputée où le réalisme se maintient (74.25). L’écart doctrine/réalité n’est donc pas constant : il se creuse précisément là où un acteur prépondérant impose sa lecture, ce qui en fait un candidat-indicateur de condition hégémonique plutôt qu’une description de la nature de l’ordre. C’est l’énoncé le plus directement falsifiable du dispositif : il prédit l’absence d’inflation doctrinale dans les configurations sans hégémon, et un codage sourcé d’autres périodes multipolaires ou unipolaires suffirait à le mettre en défaut. De même, le saut d’influence académique du constructivisme entre les périodes (P2) et la sensibilité différenciée des familles aux pondérations en 1945–1962 (P3) ne sont lisibles que parce que les couches et les profils de pondération sont tenus séparés.

Le cas pilote montre aussi les limites du modèle. Les scores restent dépendants du codeur. La liste des éléments probants peut être contestée. Les bornes temporelles changent probablement certains résultats : inclure 1962 dans une période qui irait jusqu’à 1979 réduirait sans doute la centralité de la bipolarisation initiale et augmenterait le poids de la décolonisation, du non-alignement et de la dépendance. Enfin, le score global masque des variations d’entités : le Sud global, les États-Unis, l’URSS, l’Europe occidentale, la Chine et l’ONU ne reçoivent pas les mêmes profils théoriques, comme le quantifie la déclinaison par entité de 1945–1962 (jusqu’à 18 points d’écart sur le marxisme/dépendance entre mondes en décolonisation et États-Unis).

17 Reproductibilité et artefacts

Un protocole de ce type doit être accompagné d’artefacts versionnés : codebook, scores bruts, éléments probants, références, scripts de normalisation, exports CSV et journal de révision. L’intérêt de GitHub Pages ou d’un autre site statique n’est pas esthétique ; il est méthodologique. Il permet de publier simultanément les visualisations et les données qui les produisent. Mais l’outillage ne doit pas masquer la faiblesse éventuelle du codage. Un beau dashboard reposant sur des scores arbitraires reste arbitraire.

Le principe de reproductibilité est simple : le JSON canonique contient les *critères* élémentaires, les éléments probants, la confiance, les références et les statuts de maturité ; le score brut lui-même est dérivé des critères, puis les scores normalisés, la synthèse, les CSV et les graphiques en découlent. Toute modification d’un critère doit déclencher un journal de révision. Les données dérivées ne doivent jamais être modifiées à la main.

Ce protocole est mis en œuvre dans l’application *ir-strata*, atlas statique versionné (Docu-saurus, données JSON canoniques, scripts de validation et de génération, déploiement continu). L’application matérialise les garde-fous décrits ici : validation au build (bornes des scores, somme des poids, références présentes, anti-anachronisme, politique de maturité), audit de provenance des références (vérification d’existence via Crossref et OpenLibrary), audit anti-anachronisme (rejet des scores doctrine/académique antérieurs à l’émergence des théories) et rapport d’avancement du re-sourçage. Les chiffres des trois périodes du pilote (tableaux par couche, synthèses, sensibilité et simulation de second codage) sont exactement ceux servis par l’application : ils ne sont pas saisis à la main dans cet article mais *générés* par le script `scripts/generate_paper_artifacts.py`, qui dérive les CSV d’annexe (`annexe_pilot_*.csv`) des données canoniques et générées. La revendication de reproductibilité est ainsi outillée et vérifiable : chaque nombre du présent texte doit se retrouver dans ces annexes.

Trois garde-fous supplémentaires renforcent la rigueur interne. Premièrement, le score brut

n'est plus une valeur saisie indépendamment : il est *dérivé* des critères élémentaires par la formule de pondération, et une garde au build rejette tout score qui divergerait de ses propres critères. Les critères deviennent ainsi la seule source de vérité, et leur cohérence avec le score est vérifiable. Deuxièmement, l'indice de confiance interne est recalculé à partir de ses cinq dimensions selon une formule documentée, puis bloqué en cas d'incohérence :

$$C_{int} = 0.30 Q_{preuves} + 0.25 K_{consensus} + 0.20 S_{temporelle} + 0.15 (100 - W_{sensibilite}) + 0.10 (100 - A_{anachronisme})$$

de sorte que la confiance affichée découle bien de cette formule, et non d'un libellé arbitraire ; pour le réalisme en 1945–1962, on obtient $0.30(94) + 0.25(86) + 0.20(84) + 0.15(75) + 0.10(70) = 84.75$. Troisièmement, un signal bibliométrique externe et indépendant (comptes de publications par famille théorique) est confronté à la couche d'influence académique : une faible corrélation de rang signale une période à réexaminer. Ce dernier dispositif fournit un critère de falsifiabilité partiel, distinct du jugement de l'auteur.

Une analyse de sensibilité au choix « rejeter plutôt que scorer bas » l'anachronisme a par ailleurs été conduite : remplacer le rejet par un score faible ne modifie pas la famille dominante des périodes synthétisables ; le rejet ne fait que supprimer les synthèses des périodes pré-disciplinaires, ce qui est l'effet recherché. Le choix de modélisation apparaît donc robuste quant aux conclusions de premier ordre. Cette robustesse a toutefois une portée limitée : elle ne teste que la stabilité de la famille dominante, non celle des propositions P1 et P2, qui sont l'apport substantiel du papier. Or le rejet conditionne directement P2 (par construction) et déplace les parts normalisées qui nourrissent la lecture des écarts ; une étude de robustesse *ciblée sur P1 et P2* – non encore conduite – reste nécessaire avant d'en faire autre chose que des hypothèses.

18 Conclusion

La quantification interprétative des théories des relations internationales n'est défendable qu'à des conditions strictes. Elle doit renoncer à l'illusion d'une mesure objective forte. Elle doit reconnaître le statut quasi-ordinal de ses scores. Elle doit distinguer intensité brute et part normalisée. Elle doit justifier ses périodes. Elle doit documenter ses éléments probants, ses contradictions et ses controverses. Elle doit tester la sensibilité aux pondérations et préparer une validation inter-codeurs.

Le cas pilote 1945–1962 montre que la méthode peut produire des résultats plausibles sans réduire la période à une seule théorie. Mais l'apport ne se réduit pas à confirmer une famille dominante déjà connue : en tenant séparées les trois couches et en exposant la sensibilité aux pondérations, le dispositif rend énonçables des propositions que la lecture qualitative laisse implicites — l'écart doctrine/réalité comme signature d'une condition hégémonique, la montée de l'influence académique comme datation d'une émergence disciplinaire, la sensibilité d'un classement comme propriété interprétative de la période. Aucune n'est ici démontrée empiriquement : sur des données seed, ce sont des hypothèses que le protocole rend, pour la première fois, réfutables. Il montre aussi que la robustesse reste limitée tant qu'un codage indépendant et un calibrage sourcé n'ont pas été menés. La conclusion doit donc être dure : *Strates RI* ne devient un outil académique sérieux que si ses scores restent plus faciles à contester qu'à admirer. Si les chiffres ne sont pas reliés à des éléments probants, des controverses, une sensibilité et une procédure de révision, ils ne sont qu'une rhétorique numérique.

Le projet a donc une valeur à une condition : accepter que la rigueur réduise l'effet de séduction. Un atlas analytique sérieux doit parfois produire moins de graphiques, moins de certitude et plus de désaccord visible. Au fond, l'exercice relève moins de la mesure que de l'objectivation participante : il retourne sur la discipline elle-même les instruments qu'elle applique au monde, et expose le codeur à sa propre opération de classement (Bourdieu, 1976; Hoffmann, 1977). C'est le prix à payer pour quantifier sans objectiver – et la raison pour laquelle un tel dispositif a sa

place dans une sociologie réflexive des relations internationales plutôt que dans une scientisation de plus de la discipline.

Références

- Amitav Acharya. Global international relations (ir) and regional worlds. *International Studies Quarterly*, 58(4) :647–659, 2014.
- Amitav Acharya and Barry Buzan. *The Making of Global International Relations : Origins and Evolution of IR at Its Centenary*. Cambridge University Press, 2019.
- Robert Adcock and David Collier. Measurement validity : A shared standard for qualitative and quantitative research. *American Political Science Review*, 95(3) :529–546, 2001.
- Raymond Aron. *Paix et guerre entre les nations*. Calmann-Lévy, Paris, 1962.
- Richard K. Ashley. The poverty of neorealism. *International Organization*, 38(2) :225–286, 1984.
- Bertrand Badie. *L’État importé : essai sur l’occidentalisation de l’ordre politique*. Fayard, Paris, 1992.
- Bertrand Badie and Marie-Claude Smouts. *Le retournement du monde : sociologie de la scène internationale*. Presses de Sciences Po et Dalloz, Paris, 1992.
- Dario Battistella. *Théories des relations internationales*. Presses de Sciences Po, Paris, 4 edition, 2012.
- Andrew Bennett and Jeffrey T. Checkel. *Process Tracing : From Metaphor to Analytic Tool*. Cambridge University Press, 2015.
- Pierre Bourdieu. Le champ scientifique. *Actes de la recherche en sciences sociales*, 2(2) :88–104, 1976.
- Henry E. Brady and David Collier, editors. *Rethinking Social Inquiry : Diverse Tools, Shared Standards*. Rowman & Littlefield, 2010.
- Fernand Braudel. Histoire et sciences sociales : la longue durée. *Annales. Économies, Sociétés, Civilisations*, 13(4) :725–753, 1958.
- Hedley Bull. *The Anarchical Society : A Study of Order in World Politics*. Macmillan, 1977.
- Barry Buzan. *An Introduction to the English School of International Relations*. Polity, 2014.
- Fernando Henrique Cardoso and Enzo Faletto. *Dependency and Development in Latin America*. University of California Press, 1979.
- Edward Hallett Carr. *The Twenty Years’ Crisis, 1919–1939*. Macmillan, 1939.
- Christopher Clark. *The Sleepwalkers : How Europe Went to War in 1914*. Allen Lane, 2012.
- David Collier and Steven Levitsky. Democracy with adjectives : Conceptual innovation in comparative research. *World Politics*, 49(3) :430–451, 1997.
- Robert W. Cox. Social forces, states and world orders : Beyond international relations theory. *Millennium*, 10(2) :126–155, 1981.
- Guillaume Devin. *Sociologie des relations internationales*. La Découverte, Paris, 2002.

- Michael W. Doyle. Kant, liberal legacies, and foreign affairs. *Philosophy & Public Affairs*, 12 (3) :205–235, 1983.
- Cynthia Enloe. *Bananas, Beaches and Bases : Making Feminist Sense of International Politics*. University of California Press, 1989.
- Frantz Fanon. *The Wretched of the Earth*. Grove Press, 1961.
- Martha Finnemore and Kathryn Sikkink. International norm dynamics and political change. *International Organization*, 52(4) :887–917, 1998.
- Andre Gunder Frank. The development of underdevelopment. *Monthly Review*, 18(4) :17–31, 1966.
- John Lewis Gaddis. *The Cold War : A New History*. Penguin Press, 2005.
- Alexander L. George and Andrew Bennett. *Case Studies and Theory Development in the Social Sciences*. MIT Press, 2005.
- John Gerring. *Case Study Research : Principles and Practices*. Cambridge University Press, 2007.
- Robert Gilpin. *War and Change in World Politics*. Cambridge University Press, 1981.
- Gary Goertz. *Social Science Concepts : A User's Guide*. Princeton University Press, 2006.
- Eric Hobsbawm. *Age of Extremes : The Short Twentieth Century, 1914–1991*. Michael Joseph, 1994.
- John A. Hobson. *Imperialism : A Study*. James Nisbet, 1902.
- Stanley Hoffmann. An american social science : International relations. *Daedalus*, 106(3) :41–60, 1977.
- G. John Ikenberry. *After Victory : Institutions, Strategic Restraint, and the Rebuilding of Order after Major Wars*. Princeton University Press, 2001.
- G. John Ikenberry. *Liberal Leviathan : The Origins, Crisis, and Transformation of the American World Order*. Princeton University Press, 2011.
- Paul Kennedy. *The Rise and Fall of the Great Powers*. Random House, 1987.
- Robert O. Keohane. *After Hegemony : Cooperation and Discord in the World Political Economy*. Princeton University Press, 1984.
- Robert O. Keohane and Joseph S. Nye. *Power and Interdependence : World Politics in Transition*. Little, Brown, 1977.
- Gary King, Robert O. Keohane, and Sidney Verba. *Designing Social Inquiry : Scientific Inference in Qualitative Research*. Princeton University Press, 1994.
- Henry A. Kissinger. *A World Restored : Metternich, Castlereagh and the Problems of Peace, 1812–1822*. Houghton Mifflin, 1957.
- Friedrich V. Kratochwil. *Rules, Norms, and Decisions*. Cambridge University Press, 1989.
- Charles Krauthammer. The unipolar moment. *Foreign Affairs*, 70(1) :23–33, 1990.
- Vladimir I. Lenin. *Imperialism, the Highest Stage of Capitalism*. Petrograd, 1917.

- Margaret MacMillan. *Paris 1919 : Six Months That Changed the World*. Random House, 2001.
- John J. Mearsheimer. *The Tragedy of Great Power Politics*. W. W. Norton, 2001.
- Marcel Merle. *Sociologie des relations internationales*. Dalloz, Paris, 4 edition, 1988.
- Andrew Moravcsik. Taking preferences seriously : A liberal theory of international politics. *International Organization*, 51(4) :513–553, 1997.
- Hans J. Morgenthau. *Politics Among Nations : The Struggle for Power and Peace*. Alfred A. Knopf, 1948.
- Nicholas Greenwood Onuf. *World of Our Making : Rules and Rule in Social Theory and International Relations*. University of South Carolina Press, 1989.
- Charles C. Ragin. *The Comparative Method : Moving Beyond Qualitative and Quantitative Strategies*. University of California Press, 1987.
- Edward W. Said. *Orientalism*. Pantheon Books, 1978.
- Paul W. Schroeder. *The Transformation of European Politics, 1763–1848*. Oxford University Press, 1994.
- Gayatri Chakravorty Spivak. Can the subaltern speak ? In Cary Nelson and Lawrence Grossberg, editors, *Marxism and the Interpretation of Culture*, pages 271–313. University of Illinois Press, 1988.
- Susan Strange. *States and Markets*. Pinter, 1988.
- Susan Strange. *The Retreat of the State : The Diffusion of Power in the World Economy*. Cambridge University Press, 1996.
- J. Ann Tickner. *Gender in International Relations : Feminist Perspectives on Achieving Global Security*. Columbia University Press, 1992.
- Charles Tilly. *Coercion, Capital, and European States, AD 990–1992*. Blackwell, 1992.
- Adam Tooze. *The Deluge : The Great War and the Remaking of Global Order, 1916–1931*. Allen Lane, 2014.
- Robert Vitalis. *White World Order, Black Power Politics : The Birth of American International Relations*. Cornell University Press, 2015.
- R. B. J. Walker. *Inside/Outside : International Relations as Political Theory*. Cambridge University Press, 1993.
- Immanuel Wallerstein. The rise and future demise of the world capitalist system. *Comparative Studies in Society and History*, 16(4) :387–415, 1974.
- Kenneth N. Waltz. *Theory of International Politics*. Addison-Wesley, 1979.
- Adam Watson. *The Evolution of International Society*. Routledge, 1992.
- Alexander Wendt. *Social Theory of International Politics*. Cambridge University Press, 1999.
- Odd Arne Westad. *The Global Cold War : Third World Interventions and the Making of Our Times*. Cambridge University Press, 2005.
- Martin Wight. *Systems of States*. Leicester University Press, 1977.

A Annexe A : codebook synthétique

Sources admissibles par couche :

Couche	Sources admissibles
Réalité stratégique	guerres, alliances, capacités, polarité, institutions effectives, flux économiques, empires, routes, infrastructures, dépendances, crises systémiques.
Doctrines politique	doctrines d'État, discours diplomatiques, traités, stratégies nationales, doctrines militaires, déclarations officielles, votes et pratiques institutionnelles.
Influence académique	ouvrages, articles, débats disciplinaires, manuels, écoles de pensée, programmes universitaires, réception et centralité historiographique.

B Annexe B : périodes et alternatives

La périodisation doit rester révisable. Les alternatives principales sont : 1815–1856 au lieu de 1815–1848 pour intégrer la guerre de Crimée ; 1871–1905 au lieu de 1871–1914 pour isoler la montée allemande avant les crises marocaines ; 1945–1955 au lieu de 1945–1962 pour isoler la formation institutionnelle initiale ; 1973–1991 au lieu de 1979–1991 pour donner plus de poids au choc pétrolier et à la crise économique mondiale ; 2008–2022 au lieu de 2008–2014/2014–2022 pour traiter la fragmentation comme une séquence longue. Chaque alternative doit être testée sur les scores bruts, pas seulement discutée narrativement.

C Annexe C : statut du pilote

Le pilote couvre désormais trois périodes (1815–1848, 1945–1962, 1991–2001), codées par l'auteur selon le codebook et alignées sur l'application *ir-strata*. Il démontre l'applicabilité interne du protocole, sa robustesse aux pondérations et son contrôle de l'anachronisme, mais *ne démontre pas* la fiabilité inter-codeurs : le second codage présenté est une *simulation déterministe explicitement marquée*, non un codage humain indépendant. Le verdict d'audit est **partially_holds**. Toute publication académique forte devra ajouter un second codage humain indépendant sur les 63 scores du protocole, un tableau d'écarts réel et un journal de résolution des divergences.

D Annexe D : artefacts du pilote

Les données du pilote sont versionnées sous forme de fichiers CSV : `annexe_pilot_scores_1815_1848.csv`, `annexe_pilot_scores_1945_1962.csv` et `annexe_pilot_scores_1991_2001.csv` (scores bruts par couche, synthèse, normalisation) ; `annexe_pilot_sensitivity.csv` (synthèse brute par profil de pondération, `delta_max` et statut) ; `annexe_pilot_intercoder_simulation.csv` (codage principal, codage simulé, écarts) ; `annexe_pilot_entities_1945_1962.csv` (synthèse brute et part normalisée par entité) ; `annexe_pilot_subfamilies.csv` (sous-familles déclarées des sept familles) ; `annexe_pilot_realism_vs_geopolitics.csv` (écart de score brut réalisme/géopolitique par couche et période, cas de divergence maximale marqué). Aucun chiffre de ces annexes n'est saisi à la main : tous sont produits par le script `scripts/generate_paper_artifacts.py`, qui les dérive des scores canoniques (`scores_raw.json`, `scores_entities_raw.json`), de la taxonomie (`theory_taxonomy.json`), des entités (`entities.json`), de la synthèse et de la sensibilité générées (`scores_synthesis_raw.json`, `scores_synthesis_normalized.json`, `scores_synthesis_sensitivity`) et des profils de `weight_profiles.json`. Le script est exécuté au build (`npm run generate:data`),

de sorte que toute évolution des données canoniques régénère automatiquement les annexes et le pilote reste reproductible.