



HAL
open science

Présentation de la BiblioBase

Elsa Courant, Jordi Brahamcha-Marin, Gaëlle Lafage, Francesca Guglielmi

► **To cite this version:**

Elsa Courant, Jordi Brahamcha-Marin, Gaëlle Lafage, Francesca Guglielmi. Présentation de la BiblioBase. 2023. hal-04332215

HAL Id: hal-04332215

<https://hal.science/hal-04332215v1>

Submitted on 12 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Présentation de la *BiblioBase*.

Principes d'édition d'une base de données bibliographiques littéraires
fondée sur la *Bibliographie de la France* (1811-1900)

Version abrégée

par

Elsa Courant (CNRS / CELLF)

Jordi Brahamcha-Marin (CNRS / CELLF)

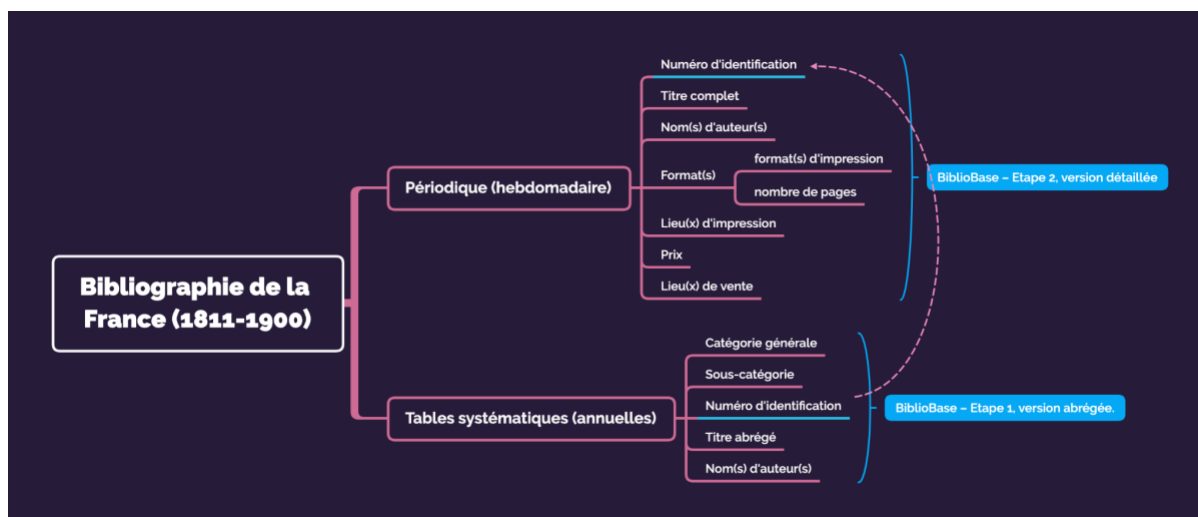
Gaëlle Lafage (CNRS / CELLF)

Francesca Guglielmi (CELLF)

I. Définition et cadre général

La *BiblioBase* est une base de données bibliographiques sur la littérature du XIX^e siècle. Elle s'appuie sur la *Bibliographie de la France*, périodique fondé sous le Premier Empire, en 1811ⁱ. Ce journal a pour vocation de recenser l'ensemble des ouvrages dont la diffusion a été autorisée après leur dépôt légal. Parce que les informations qu'elle contient dépendent des déclarations initiales des imprimeurs et des procédures de contrôle de l'imprimerie, la *Bibliographie de la France* n'est pas une source infaillible ni parfaitement exhaustive sur la production littéraire qu'elle recenseⁱⁱ. Elle est cependant l'une des sources les plus complètes pour constituer des corpus ciblés, notamment par domaines et par genres littéraires.

En effet, la *Bibliographie de la France* recense les publications déclarées semaine après semaine. Ces publications, sous un format abrégé, font l'objet d'un classement méthodique à la fin de l'année dans un sommaire intitulé « Tables systématiques ». À chaque ouvrage est associé un numéro qui permet la circulation du lecteur entre les « Tables systématiques » comprenant une version abrégée des éléments d'identification des œuvres, et le périodique hebdomadaire comprenant une version détaillée qui rassemble l'ensemble des informations paratextuelles. Le schéma ci-dessous résume ce principe de circulation, ainsi que les étapes de développement de la *BiblioBase*.



La première étape de développement de la *BiblioBase* consiste en l'élaboration d'une base de données qui rassemble toutes les informations bibliographiques abrégées contenues dans les « Tables systématiques » littéraires sur la période 1811-1900, à l'exception de la période 1857-1865, au cours de laquelle la *Bibliographe de la France* cesse de publier ses « Tables systématiques » annuelles. Sont concernées les catégories « Belles-Lettres », « Littérature française », « Littérature ancienne » et « Littérature étrangère », selon la terminologie évolutive des tables. Aucune modification n'a été apportée sur les choix de classement effectués par les bibliographes. La *BiblioBase* reflète donc une conception de la littérature implicitement adoptée, et souvent redéfinie, par les rédacteurs de la *Bibliographie de la France*. Les ouvrages communément considérés comme littéraires de nos jours, mais recensés par les bibliographes dans des catégories concurrentes (« Religion », « Livres d'éducation », « Philosophie », etc.) n'ont pas été ajoutés à la *BiblioBase*. Inversement, des ouvrages pouvant relever d'autres domaines que la littérature y figurentⁱⁱⁱ.

Le corpus général de la *BiblioBase* contient environ 156 000 références.

La présente notice s'applique à la version abrégée de la *BiblioBase*^{iv}.

La *BiblioBase* est le résultat du projet ANR Jeunes Chercheurs ses « e-BdF » (2023-2026), coordonné par Elsa Courant^v.

II. Méthode de collecte des données

Une première saisie diplomatique des références bibliographiques contenues dans les « Tables systématiques » littéraires a été confiée à un prestataire^{vi}. Les listes d'ouvrages ont fait l'objet d'une océrisation complétée par une double saisie comparée afin de minimiser les risques d'erreurs. Les données obtenues ont fait l'objet d'une première parcellisation livrée en mars 2023. Un repérage de la langue des titres a également été effectué par le prestataire.

Dans un second temps, un groupe de travail constitué au sein de l'équipe ANR e-BdF s'est attaché au nettoyage et à l'enrichissement manuel des données bibliographiques. Ce groupe constitué par Elsa Courant, coordinatrice de l'ANR e-BdF, a mobilisé deux chercheurs, Jordi Brahamcha-Marin et Elsa Courant, et deux ingénieures, Francesca Guglielmi et Gaëlle Lafage. La finalisation des données s'est effectuée en 3 étapes :

1. **Nettoyage des données** : correction et vérification de l'ensemble des noms d'auteurs avec suppression des prépositions (« ~~par~~ V. Hugo ») ; correction et vérification des titres abrégés.

2. Ajout de 6 colonnes :

- une colonne « **Livraison** », indiquant dans quelle livraison hebdomadaire du périodique l'ouvrage a été annoncé ;
- une colonne « **Date** », correspondant au jour de parution de la livraison ;
- une colonne « **Posture** », proposant un étiquetage des signatures des ouvrages suivant les principes exposés plus bas ;
- deux colonnes « **Traducteur** » et « **Para-auteur** », permettant de dissocier l'auteur du texte de l'ensemble des pratiques autoriales marginales ou paratextuelles, dans les cas d'œuvres traduites, mais aussi revues, annotées, commentées ou réécrites.
- une colonne « **Commentaire** », permettant au groupe de travail d'indiquer les incohérences ou erreurs soulevées par la vérification des données bibliographiques issues des tables de la *Bibliographie de la France*, en particulier les erreurs de numérotation, ou certaines erreurs de graphie des titres ou des auteurs effectuées par les bibliographes.

L'objectif de ces catégories descriptives, et en particulier des propositions qualitatives contenues dans la colonne « Posture », est de faciliter la sélection de corpus bibliographiques, et de définir des tendances statistiques — en tenant compte des limitations inhérentes au corpus ainsi qu'aux choix de méthode exposés dans la présente notice.

Dans un troisième temps, les données vérifiées ont été converties au format CSV, puis publiées sur la plateforme de dépôt Nakala^{vii}. Conformément aux principes FAIR et afin de faciliter l'interopérabilité des données, la licence CC-BY-NC a été choisie.

III. Catégories d'analyse et parcellisation

Les données bibliographiques de la *BiblioBase* dans sa version abrégée sont décrites selon les critères ci-dessous. Dans la suite de cette notice, le titre du périodique *Bibliographie de la France* est abrégé sous la forme : *BdF*.

1. Datation

Pour une datation plus fine des annonces de parution des ouvrages, nous avons effectué un travail de repérage à partir des numéros de notice figurant dans les « tables systématiques » de la *BdF* et renvoyant à la notice détaillée de la référence bibliographique. Une datation de l'annonce de parution officielle des ouvrages recensés est donc désormais possible avec une précision hebdomadaire, et non plus seulement annuelle.

L'« **Année** » correspond à l'année de l'annonce de publication.

La « **Livraison** » désigne le numéro du journal dans lequel paraît la référence complète au cours de l'année. À quelques rares exceptions près (comme en 1826 et 1827, où paraissent deux numéros par semaine), cette publication était hebdomadaire et comportait 52 ou 53 livraisons par an.

La « **Date** » correspond à la date précise à laquelle est parue la livraison du journal dans laquelle figure la référence complète. Elle apparaît seulement dans l'en-tête de la première page de la livraison.

Le « **Numéro** » correspond à celui donné à la référence bibliographique dans la *BdF*. Un numéro ne correspond pas nécessairement à un ouvrage complet. Les ouvrages publiés en livraisons ou en plusieurs tomes sont associés à plusieurs numéros, car annoncés dans la *BdF* au fur et à mesure des parutions. Lorsque le numéro indiqué dans la table systématique était erroné, il a été corrigé. Les bibliographes ont parfois utilisé une numérotation à part pour les ouvrages étrangers (notamment pour les années 1855, 1856, 1874, 1875). Ainsi, des publications différentes portent le même numéro de référence dans la *BdF*.

2. Situation des ouvrages dans les « Tables systématiques »

La colonne « **Page** » indique le ou les numéros de pages où figure la référence bibliographique dans les « tables systématiques ». En effet, lorsque la référence apparaît dans plusieurs catégories ou sous-catégories des tables, les différents numéros de pages sont reportés dans la colonne.

Les colonnes « **Catégorie** » et « **Sous-catégorie** » correspondent aux intitulés de classement choisis par les bibliographes pour rassembler et décrire les ouvrages identifiés comme littéraires dans les « Tables systématiques ».

Ce classement est un système bibliographique régulièrement modifié et parfois justifié par des notes, avant-propos ou textes critiques au sein de la *BdF*. Les dénominations des catégories et les nombres de subdivisions peuvent drastiquement varier d'une année à l'autre, en particulier au milieu du XIX^e siècle. La BiblioBase restitue ces évolutions sans intervenir sur les formulations d'origine.

3. Informations relatives aux titres des ouvrages

La colonne « **Langue** » correspond à la langue ou aux langues dans lesquelles est écrit l'ouvrage. Par défaut, on indique la langue du titre. Lorsque le titre de l'œuvre suggère un texte bilingue (par exemple, s'il s'agit d'une traduction « avec texte en regard »), on indique les deux langues.

Le cas échéant, on utilise la catégorie « Langues régionales ». Confrontés à la difficulté d'identifier précisément l'ensemble des langues relevant de cette catégorie, et considérant certains cas d'incertitude terminologique, nous avons renoncé à une désignation plus précise.

La colonne « **Titre** » indique le titre de l'œuvre tel qu'il est donné dans les « Tables systématiques ». Sont également indiquées les fonctions précises des para-auteurs (ex. : « avec des notes de... ») et des renseignements sur l'opération de traduction (ex. : « trad. de l'anglais par... »).

Dans le cas des titres latins, ceux-ci sont reproduits tels quels dans la colonne « Titre », avec le nom de l'auteur sous sa forme latine (ex. : « M. T. Ciceronis oratio in Verrem, de suppliciis »), et le nom de l'auteur est repris dans la colonne « auteur » sous sa forme française habituelle (ex. : « Cicéron »).

4. Données auctoriales

a. « Posture »

La colonne « **Posture** », perfectible et expérimentale, propose une qualification des effets de réception des signatures des ouvrages, ou des choix de contournement de l'identification auctoriale dans le cas des pseudonymes ou publications anonymes. Cette « posture », lisible dans les signatures, peut correspondre à une stratégie de l'auteur^{viii}, à un choix de l'éditeur ou à un choix des bibliographes, qui modifient parfois les désignations actoriales figurant sur les pages de titre^{ix}.

La « Posture » peut concerner le ou les auteur(s), le ou les traducteur(s) et/ou le ou les para-auteur(s).

Le décompte des postures ne s'est pas fondé sur le nombre d'entités nommées.

- Quand plusieurs postures sont pertinentes pour un même ouvrage, elles sont cumulées. Il suffit que l'une de ces trois fonctions implique une certaine posture pour que celle-ci soit indiquée dans la colonne « Posture ». Ainsi, la posture « Nom féminin » peut indiquer une signature féminine de l'auteur, du traducteur ou du para-auteur (ou des trois, ou de deux d'entre eux).
- Quand plusieurs noms relatifs aux données actoriales se rapportent à une même posture, celle-ci n'est mentionnée qu'une seule fois.

Cinq « postures » fondamentales ont été utilisées pour décrire le corpus :

- la posture « **Nom féminin** » s'applique lorsque la *BdF* indique une signature interprétable comme féminine, pour l'auteur, le para-auteur ou le traducteur. Cette posture est utilisée pour des signatures féminines explicites : celles qui contiennent par exemple « Mme », « Mlle », « Mrs. », « comtesse », etc., ou un prénom féminin non épïcène et non abrégé.

À titre d'exception, dans le cas d'une discordance entre la posture genrée du pseudonyme et l'identité générique réelle de l'auteur.rice, on indique en cumul de posture l'identité générique réelle si et seulement si on peut établir qu'elle est connue du public de l'époque au moment de la parution de l'ouvrage. Pour ce faire, on se fonde sur une publication contemporaine dans la presse et/ou un dictionnaire des pseudonymes. À titre d'exemple, « George Sand » appelle la posture « Pseudonyme masculin;Nom féminin ». Ce choix de compromis concerne 0,17 % du corpus global, soit 271 références identifiées dans le corpus général d'environ 156 000 références^x.

Nous avons initialement décidé d'indiquer aussi la posture « Nom masculin ». Mais nous avons abandonné cette idée, considérant que : le genre des auteurs, para-auteurs et traducteurs sera finalement accessible, dans un état ultérieur de la base, grâce au lien vers des notices d'autorité ; la distinction entre « M. » signifiant « Monsieur » et « M. » comme initiale d'un prénom n'est pas toujours possible^{xi} ; le projet *e-BdF* cherche notamment à évaluer l'importance quantitative de textes minoritaires à différents égards, parmi lesquels les textes signés par des femmes ou sous un nom féminin. À ce titre, les signatures sous un nom exclusivement masculin ou supposé tel n'ont pas fait l'objet d'une catégorisation.

- La posture « **Anonyme** » s'applique aux ouvrages dont le nom de l'auteur n'est pas mentionné dans les « Tables systématiques » de la *BdF*. On ne distingue pas les cas où cet

anonymat est réel (l'auteur est inconnu des bibliographes, n'apparaît pas sur la page de titre de l'ouvrage) et ceux où il est un artefact bibliographique, lié à la grande notoriété de l'ouvrage qui dispense les bibliographes d'en indiquer l'auteur supposé connu (ex. : *La Henriade*, sans indication d'auteur [Voltaire]).

Deux cas de figure justifient le recours à la posture « Anonyme » : lorsqu'il n'y a aucune indication auctoriale (la case « Auteur » est alors laissée vide) ; lorsqu'il y a une indication auctoriale réduite à une ou plusieurs initiales (dans ce cas, la posture « Anonyme » peut porter sur l'auteur, le traducteur ou le para-auteur).

- Faute d'un terme plus général, la posture « **Pseudonyme** » renseigne les formes de désignation auctoriale obliques ou indirectes. Il peut s'agir :
 - *d'une désignation périphrastique (« les élèves du collège de Meaux », « un républicain »...);
 - *d'un pseudonyme transparent (« Pangloss »);
 - *d'un pseudonyme connu, identifié comme tel par les auteurs de la *BiblioBase* (Voltaire, Molière, George Sand, P.-J. Stahl, comtesse Dash, etc.).

On utilise la catégorie « Pseudonyme » y compris lorsque l'altération du nom d'état-civil porte seulement sur le prénom (ex. : « J.-K. Huysmans » pour « Charles-Marie-Georges Huysmans »). La représentativité de cette catégorie pour des études statistiques est cependant fragile, en raison des difficultés inhérentes à l'identification des pseudonymes.

- La posture « **Collectif** » concerne les références qui comportent une forme de collaboration auctoriale (auteur, traducteur ou para-auteur) ; elle n'inclut pas les anthologies. Elle concerne en revanche les ouvrages rédigés à plusieurs mains, sous la direction d'un auteur unique. Elle décrit aussi les correspondances croisées (lettres d'Héloïse et Abélard, etc.).

Lorsque plusieurs para-auteurs sont indiqués, l'ouvrage est classé comme « collectif », sauf si la description dans la table systématique indique que les deux para-auteurs ont des fonctions différentes et n'ont pas réellement collaboré au même travail. Par exemple, la posture « Collectif » n'est pas utilisée si l'un des para-auteurs est préfacier, et l'autre annotateur.

- La posture « **Compilation** » concerne les ouvrages recueillant les textes de plusieurs auteurs, sans dialogue entre eux, dans la mesure où c'est par un geste éditorial extérieur que ces textes se retrouvent rassemblés. C'est le cas, par exemple, des anthologies, des manuels scolaires, des recueils, des éditions d'un discours de réception académique avec sa réponse, etc. Dans ces cas-là, lorsque la case « Auteur » est vide, la posture « Compilation » se substitue à la posture « Anonyme ». Mais on indique bien « Compilation ; Anonyme » lorsque la compilation est réalisée par un para-auteur indiqué par de simples initiales.
- Certaines de ces postures sont susceptibles de se combiner. Si la posture « Nom masculin » a finalement été abandonnée, on a créé les postures « **Anonyme masculin** » et

« **Anonyme féminin** » distinctes d'« Anonyme » (neutre), et « **Pseudonyme masculin** », « **Pseudonyme féminin** » et « **Pseudonyme collectif** » distinctes de « Pseudonyme » (neutre), pour pouvoir continuer à identifier finement les postures genrées dans ces deux cas spécifiques. Ainsi :

* « le comte de R*** » appelle la posture « Anonyme masculin » ;

* « Mme F. de J. » appelle la posture « Anonyme féminin » ;

* « le bibliophile Jacob » appelle la posture « Pseudonyme masculin » (mais « G. Sand » appelle « Pseudonyme ») ;

* « comtesse Dash » appelle la posture « Pseudonyme féminin » ;

* « les élèves du collège de Meaux » appelle la posture « Pseudonyme collectif ».

Dans le cas d'« Erckmann-Chatrian », il s'agit d'un « Pseudonyme », et non d'un « Pseudonyme collectif » ; cependant on indique « Pseudonyme ;Collectif » (sur le modèle de « Pseudonyme ;Nom féminin ») dans la mesure où il était notoire qu'il s'agissait d'une signature collective.

b. « Auteur », « Traducteur » et « Para-auteur »

Pour les colonnes « **Auteur** », « **Traducteur** » et « **Para-auteur** », les graphies fautives ou excentriques employées dans les tables systématiques ont été corrigées dans le cas où elles l'étaient également par la notice détaillée. En ce cas, la colonne « **Commentaire** » le précise. Dans les autres cas, nous avons respecté les graphies indiquées dans les tables systématiques, y compris quand elles s'écartent de l'usage actuel (ex. La Martine pour Lamartine, Lafontaine pour La Fontaine, Shakspere ou Shakspeare pour Shakespeare).

La colonne « **Auteur** » renseigne le ou les noms d'auteurs tels qu'ils sont indiqués dans la table systématique. Lorsqu'aucun nom d'auteur n'est renseigné pour un ouvrage célèbre comme *L'Henriade* ou *La Jérusalem délivrée*, nous avons laissé la case « Auteur » vide.

La colonne « **Traducteur** » indique le nom du traducteur de l'ouvrage lorsqu'il est mentionné. En revanche, dans le cas où le titre indique un ouvrage en français « traduit en vers » dans la même langue, nous avons considéré qu'il s'agissait d'un texte nouveau, et le signataire de la traduction en vers figure dans la colonne « Auteur ».

La colonne « **Para-auteur** » réunit les noms de ceux qui ont participé à la publication mais n'en sont pas l'auteur principal. Cela comprend les éditeurs scientifiques, directeurs de publication, annotateurs ou responsables de la révision d'un texte.

5. Annotations

Dans la colonne « **Renvoi** », sont reportées les indications des bibliographes données dans la table systématique renvoyant à d'autres sections de la publication (« Table alphabétique » ou autres sections des « Tables systématiques »).

La colonne « **Commentaire** » contient les remarques de l'équipe d'édition de la *BiblioBase* sur les informations de la notice saisie (ex. : coquilles dans le titre ou le nom d'auteur).

IV. Perspectives de développement

Plusieurs perspectives de développement de la *BiblioBase* sont à l'étude. Un hébergement des données bibliographiques sur la plateforme d'édition de manuscrits numériques EMAN doit permettre de faciliter l'exploration, l'exploitation et la visualisation des données via une interface dédiée. Par ailleurs, une nouvelle version de la *BiblioBase*, en cours de réalisation, sera fondée sur les notices détaillées, et restituera l'ensemble des informations bibliographiques qu'elles contiennent.

Cette première version de la *BiblioBase* sert de preuve de conception. Elle montre l'intérêt de réaliser cette opération sur l'ensemble des domaines du savoir, tous représentés dans la *Bibliographie de la France* : Religion, Droit et Jurisprudence, Science politique, Histoire, Sciences naturelles, etc. Le principe guidant l'élaboration de cette base de données pourrait être étendu aux bibliographies nationales courantes d'autres pays, notamment pour mener des travaux de recherche comparatiste.

À ce titre, la *BiblioBase* est également un projet, perfectible et ouvert à de nouvelles collaborations. En cas d'anomalies constatées sur la base de données, l'équipe peut être contactée à l'adresse anr.ebdf@gmail.com.

ⁱ C. de Buzareingues, [in] *Bibliographie de la France. Journal de l'Imprimerie et de la Librairie*, numéro du cent-cinquantième, Paris, Au cercle de la Librairie, septembre 1961 ; I. de Conihout, « Police de la librairie et mesure du livre au XIX^e siècle : le dépôt légal, la *Bibliographie de la France* et la bibliométrie » [in] A. Vaillant (dir.), *Mesure(s) du livre*, Paris, Éditions de la Bibliothèque nationale, 1992 ; R.-J. Seckel, « La *Bibliographie de la France*. Survol historique, 1811-2011 », *Revue de la BNF*, 2011/3 (n. 39), p. 46 à 60, éd. Bibliothèque nationale de France ; A. Vaillant, *La Crise de la littérature. Romantisme et modernité*, Grenoble, UGA Éditions, 2019 ; etc.

ⁱⁱ Sur ce point, voir notamment M. Lyons, *Le Triomphe du livre : une histoire sociologique de la lecture dans la France du XIX^e siècle*, Paris, Promodis, 1987, p. 12 et 78.

ⁱⁱⁱ Ainsi, il apparaît que certaines éditions des œuvres de La Fontaine n'apparaissent pas dans la *BiblioBase* parce qu'elles sont classées ailleurs (par exemple, dans la catégorie « Livres d'éducation ») ; a contrario, certains manuels scolaires ou de rhétorique y trouvent souvent leur place.

^{iv} La seconde étape de publication de la *BiblioBase*, en cours d'élaboration, concerne la version détaillée des références bibliographiques.

^v Projet ANR-22-CE54-0011. La description du projet sur le site de l'ANR est accessible en ligne à : <https://anr.fr/Projet-ANR-22-CE54-0011>. Le carnet hypothèses du projet ANR est accessible en ligne à l'adresse : <https://ebdf.hypotheses.org/>. Enfin, le site vitrine de la base de données issu du projet ANR sera accessible en ligne, en 2024, à l'adresse : www.bibliobase.fr.

^{vi} Entreprise *WordPro*, dirigée par Yann Pitchal.

^{vii} ANR e-BdF (coord. E. Courant), J. Brahamcha-Marin, E. Courant, F. Guglielmi, G. Lafage, « *BiblioBase*. Tables littéraires de la *Bibliographie de la France* (1811-1900) – version abrégée », accessible sur Nakala à l'adresse : <https://nakala.fr/10.34847/nkl.586e98gl>. DOI : 10.34847/nkl.586e98gl.

viii Sur les questions de stratégies auctoriales, voir notamment R. Escarpit, *Sociologie de la littérature*, Paris, PUF, 1958.

ix Le choix du terme « posture » dans la *BiblioBase* s'inspire notamment des travaux de Jérôme Meizoz. Selon cet auteur, il désigne une « manière singulière d'occuper une "position" dans le champ littéraire ». J. Meizoz, *Postures littéraires. Mises en scène modernes de l'auteur*, Genève, Slatkine, 2007, p. 18. Sur ce point, voir également le numéro de la revue *ConTextes* consacré à « La posture. Genèse, usages et limites d'un concept » (2011, n°8), accessible en ligne sur : <https://journals.openedition.org/contextes/4692>.

x Les noms d'autrices principalement concernés par l'exception sont : George Sand, André Léo, Rachilde.

xi Cette limitation est soulevée en termes polémiques par Joseph-Marie Quérard, dans *Les Supercheries littéraires dévoilées*, « Avertissement pour la deuxième édition », Paris, Paul Daffis, 1869, t. I, p. 106.