

L'Effet des

MÉTADONNÉES

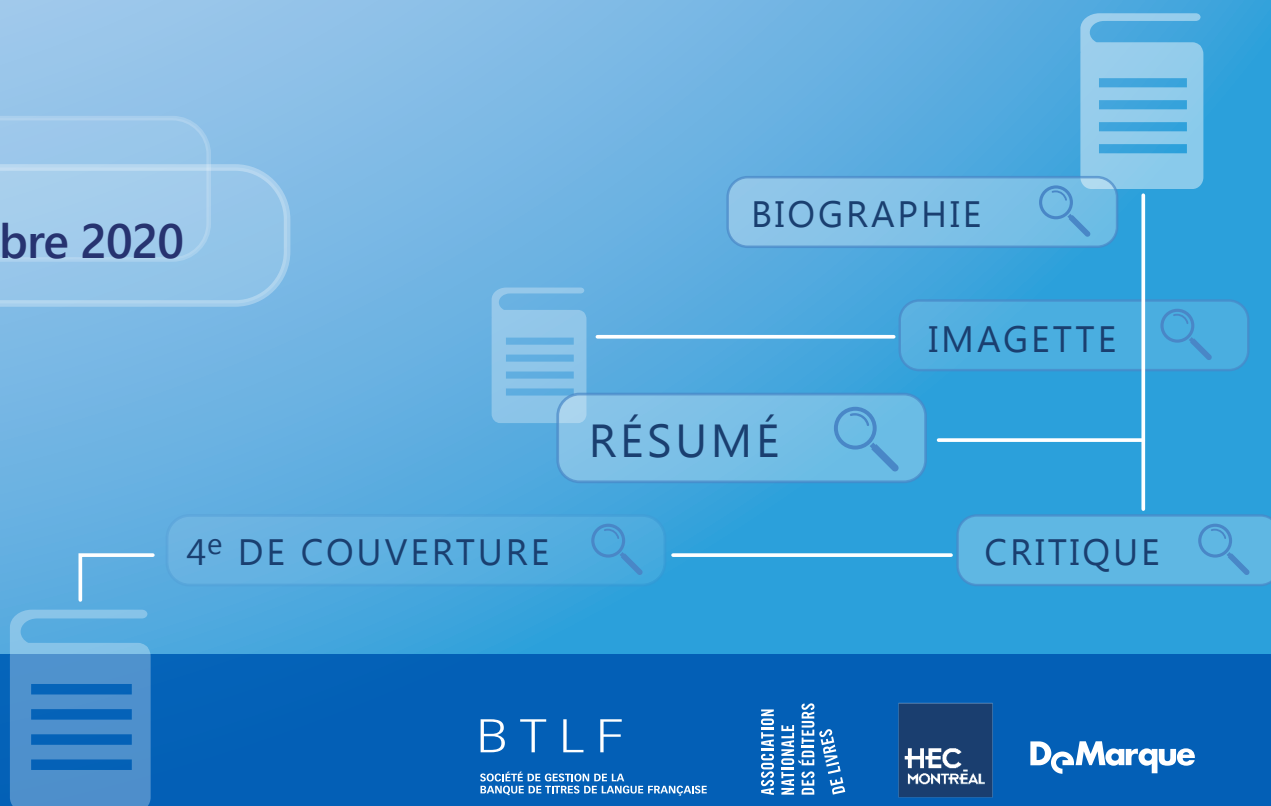


sur les ventes de livres

Analyse de la relation entre les métadonnées d'enrichissement et les ventes de livres

Une étude réalisée par la **Société de gestion de la BTLF** en collaboration avec l'**Association nationale des éditeurs de livres** et le Professorship en données massives pour les arts et la culture de **HEC Montréal**

Septembre 2020



Crédits

Analyste

Yves Leblond, enseignant et analyste senior au Département de sciences de la décision de HEC Montréal

Rédaction

Yves Leblond, enseignant et analyste senior au Département de sciences de la décision de HEC Montréal

Renaud Legoux, professeur titulaire du Professorship en données massives pour les arts et la culture de HEC Montréal

Assemblage et préparation des données

Michel Cervellin, directeur TI, BTLF

Marco Omiccioli, responsable des opérations de Gaspard, BTLF

Mélissa Haquenne, Experte produit Cantook Hub, De Marque

Coordination

Isabelle Gaudet-Labine, chargée de projets, BTLF

Supervision

Patrick Joly, directeur général, BTLF

Administration

Eveline Favretti, chargée de projet, ANEL

Serge Grenier, comptable, BTLF

Révision

Isabelle Gaudet-Labine, chargée de projets, BTLF

Coralie Piotr, attachée de direction, BTLF

Graphisme et mise en page

Nathalie Duperré, La Boîte de Pandore

Remerciements

La Société de gestion de la BTLF remercie l'Association nationale des éditeurs de livres pour sa collaboration à l'idéation et à l'administration du projet.

La Société de gestion de la BTLF et ses collaborateurs remercient le Conseil des Arts du Canada de son soutien.



Conseil des arts
du Canada

Canada Council
for the Arts

Table des matières

Contexte	4
Résumé	5
Introduction	6
1. Méthodologie	7
1.1 Livres imprimés	7
1.2 Livres numériques	9
2- Résultats pour les livres imprimés	10
2.1 Présence des métadonnées	10
2.2 L'effet de la présence d'au moins une métadonnée	11
2.3 L'effet de l'imagette	12
2.4 L'effet de la présence des métadonnées avant la parution	13
2.5 L'effet du résumé avec et sans imagette	14
2.6 L'effet du nombre de métadonnées d'enrichissement	15
2.7 Analyse par catégorie de livres.....	16
3- Analyse pour les livres numériques	17
3.1 L'effet de la présence des métadonnées avant la parution	18
Conclusion	19

Contexte

Les métadonnées bibliographiques, d'enrichissement et commerciales liées à un livre sont essentielles à sa visibilité et à sa découvrabilité. Plusieurs intervenants du milieu du livre doivent coopérer afin de produire et diffuser ces métadonnées. Afin de minimiser les écarts entre les meilleures pratiques et les métadonnées telles qu'elles sont émises et propagées, il importe à la Société de gestion de la Banque de titres de langue française (BTLF) et à l'Association nationale des éditeurs de livres (ANEL) de collaborer afin de développer des outils permettant aux intervenants de mieux coopérer.

Ainsi, grâce à une subvention du Fonds Stratégie numérique du Conseil des arts du Canada, la BTLF, en partenariat avec l'ANEL et le Professorship en données massives pour les arts et la culture de HEC Montréal, a entrepris de concevoir une trousse d'outils sur les métadonnées qui transigent par la BTLF à l'intention des éditeurs, des distributeurs, des diffuseurs, des libraires, des bibliothécaires et des auteurs. L'Étude *L'Effet des métadonnées d'enrichissement sur les ventes de livres – analyse de la relation entre les métadonnées d'enrichissement et les ventes de livres* est le premier outil de cette trousse.

Cette analyse a été rendue possible grâce aux deux systèmes d'information gérés par la BTLF : Memento et Gaspard. La base de données Memento inclut un ensemble de métadonnées sur le livre francophone publié au Canada. Gaspard permet de mesurer la performance économique d'un livre. L'implication de De Marque a aussi permis d'appliquer l'analyse au livre numérique.

Jusqu'ici, le milieu ne disposait pas de référence pouvant démontrer de manière conséquente l'impact du travail des intervenants de la chaîne du livre en matière de métadonnées. Cette étude permet de mesurer concrètement l'effet de la production de métadonnées enrichies, transmises avant la parution, sur les ventes de livres. Les éditeurs, mais aussi les diffuseurs, constateront ainsi l'importance d'intégrer la production des métadonnées dans leur stratégie commerciale. Nous espérons que ces nouvelles connaissances favoriseront l'éclosion de meilleures pratiques, au bénéfice des éditeurs, des créateurs et de l'ensemble des intervenants du livre.

Résumé

Ce document présente une analyse statistique de la quantité de livres vendus dans les différents points de vente enregistrés à la BTLF en fonction de la présence ou non de certaines métadonnées d'enrichissement. L'objectif est d'offrir une mesure de l'effet des métadonnées sur les ventes de livres. Une première analyse porte sur les titres imprimés publiés par des éditeurs québécois¹ au cours de la période du 1^{er} janvier 2016 au 19 mai 2018. La méthode retenue est l'utilisation de tests statistiques de différence sur les ventes avec et sans métadonnées en contrôlant² l'impact du nombre de titres publiés annuellement par l'éditeur et le nombre d'exemplaires vendus dans le passé par auteur.

La conclusion générale de l'analyse est que la quantité vendue de titres avec métadonnées d'enrichissement est significativement plus élevée que la quantité vendue de titres similaires sans métadonnées d'enrichissement. Plus spécifiquement, les résultats révèlent que l'effet de la métadonnée est particulièrement présent pour les éditeurs à moyen et à grand volumes de publication annuels. L'effet est également plus présent pour les auteurs qui sont considérés comme des moyens vendeurs ou des grands vendeurs (c'est-à-dire pour les auteurs dont les ventes enregistrées sont moyennes et grandes, selon la catégorisation proposée en p. 8 et p. 9 de l'étude).

Le type de métadonnée qui semble avoir préséance en termes d'effet est l'imagerie, particulièrement pour les éditeurs à moyen et à grand volumes de publication et pour les auteurs qui sont des moyens vendeurs et des grands vendeurs. Cette imagerie a plus d'effet lorsqu'elle est fournie et diffusée avant la parution du livre qu'après. Toutefois, cet effet pré-parution est présent surtout pour les livres d'auteurs dont sont vendus habituellement moins d'exemplaires.

En ce qui concerne les autres métadonnées d'enrichissement, l'ajout seul du résumé lorsqu'il n'y a pas d'imagerie n'a pas d'effet significatif sur les ventes. Par contre, lorsque l'imagerie est déjà présente, l'ajout du résumé augmente les ventes de façon significative pour les auteurs qui sont de grands vendeurs et pour les éditeurs à moyen et à grand volumes de publication.

Pour les éditeurs à moyen et à grand volumes de publication, la quantité moyenne d'exemplaires vendus est significativement plus élevée lorsque quatre des cinq métadonnées sont disponibles par rapport à la moyenne observée lorsque seulement l'imagerie est présente.

Finalement, une analyse similaire, mais de moindre envergure, a été effectuée pour les livres numériques. Comme pour les livres imprimés, on observe que la quantité moyenne d'exemplaires vendus par titre est supérieure lorsque les métadonnées d'enrichissement sont fournies avant la date de parution plutôt qu'après, et ce pour les éditeurs à moyen et à grand volumes de publication.

¹ Dans cette étude, « éditeur québécois » fait référence à une maison d'édition située au Québec qui publie et vend sur le marché québécois des titres en langue française sous une ou plusieurs marques. Pour l'étude, on y inclut les maisons d'édition hors Québec (Canada) qui publient des titres en langue française. Pour Gaspard, chaque marque compte pour un éditeur.

² « Le contrôle statistique consiste à analyser les variations d'un processus et à distinguer les causes spéciales des variations des causes principales. » Qualité Management, « Contrôle statistique », dans <http://www.definition-qualite.com>, [En ligne], <http://www.definition-qualite.com/contrrole-statistique.htm> (Page consultée le 28 avril 2020).

Introduction

Ce rapport montre l'effet de la présence des métadonnées d'enrichissement sur les ventes de livres en français dans les librairies et autres points de vente enregistrés à la BTLF³. À cette fin, la BTLF a fourni une banque de données relatives aux titres d'éditeurs québécois ayant enregistré au moins une vente dans la période du 1^{er} septembre 2015 au 15 mai 2019, pour un nombre total de 19 558 titres.

Nous avons retenu la définition du concept de métadonnées du Grand dictionnaire terminologique de l'Office québécois de la langue française: «Ensemble structuré de données accompagnant un ouvrage et servant notamment à en décrire le contenu et le format, à assurer son indexation dans les moteurs de recherche et les bases de données, et à faciliter la gestion des droits d'auteur qui y sont liés.» Cette définition a entre autres été mobilisée dans le rapport sur l'*État des lieux sur les métadonnées relatives aux contenus culturels* de l'OCCQ en 2017.

Les métadonnées d'enrichissement ici considérées correspondent à la présence ou non d'un résumé, d'une 4^e de couverture, d'une imagerie (image de couverture), d'une biographie et d'une critique. La date d'apparition des métadonnées pour la 4^e de couverture, l'imagerie et la critique est également fournie et permet ainsi de mesurer l'effet sur les ventes de la disponibilité de ces informations avant la date de parution des livres. Une analyse similaire, mais de moindre portée, a été effectuée pour les livres numériques.

³ Grâce à son outil d'information sur les ventes Gaspard, la BTLF recueille les données de ventes d'environ 60% du marché du livre franco-canadien. Les détaillants enregistrés sont des librairies indépendantes, des coopératives en milieu scolaire, des chaînes de librairies et des grandes surfaces.

1. Méthodologie

1.1 Livres imprimés

Les données relatives aux métadonnées bibliographiques d'enrichissement proviennent du catalogue en ligne des produits de l'édition francophone Memento de la BTLF, alors que les données de vente proviennent de Gaspard. Les titres retenus pour fin d'analyse sont tous ceux publiés par un éditeur québécois⁴ entre le 1^{er} janvier 2016 et le 19 mai 2018 et ayant enregistré au moins une vente. Un nombre total de 12 852 titres ont ainsi été retenus pour l'analyse. Pour chaque année étudiée, c'est, pour les éditeurs, le nombre de titres publiés l'année précédente qui a été pris en compte et, pour les auteurs, le nombre d'exemplaires vendus depuis 2010 jusqu'à l'année examinée.

Pour chacun de ces titres, les variables considérées sont la présence ou non de chacune des cinq métadonnées retenues (la présence ou non d'un résumé, d'une 4^e de couverture, d'une imagerie, d'une biographie et d'une critique), la date d'apparition de la métadonnée imagerie et le nombre d'exemplaires vendus (au détail et aux collectivités) lors de la première année suivant la date de parution.

Dans ce rapport d'étude, la mention « titre n'ayant aucune métadonnée » signifie la non-présence des cinq métadonnées retenues. De plus, afin de pouvoir isoler, dans la mesure du possible, l'effet de la métadonnée sur les ventes, le nombre d'exemplaires vendus dans le passé par auteur et le nombre de titres publiés annuellement par éditeur ont été considérés. Puisque ces deux variables ont un impact certain sur la quantité d'exemplaires vendus et que les pratiques en termes de métadonnées sont influencées par le type d'auteur et d'éditeur, il était impératif de contrôler leur effet dans l'analyse.

À cette fin, une catégorisation des auteurs et des éditeurs a été créée. Pour les éditeurs, la catégorisation est basée sur la distribution du nombre de titres publiés lors de l'année précédant l'année à l'étude. Ainsi, pour les titres publiés en 2017, les éditeurs ayant publié trois titres ou moins en 2016 appartiennent au premier quartile de cette distribution et sont qualifiés d' « éditeurs à petit volume de publication ». Ceux ayant publié de quatre à vingt titres se situent entre le premier et le troisième quartile⁵ et sont identifiés comme des « éditeurs à moyen volume de publication ». Les autres, ayant publié plus de vingt titres en 2016, sont considérés comme des « éditeurs à grand volume de publication ».

Une classification similaire des auteurs a été effectuée, basée sur le nombre d'exemplaires vendus depuis 2010 jusqu'à l'année à l'étude. Ainsi, pour l'année 2017, les auteurs considérés comme de « petits vendeurs » sont ceux dont 0 à 302 exemplaires ont été vendus entre 2010 et 2016; les « moyens vendeurs » sont ceux dont 302 à 6526 exemplaires ont été vendus et les « grands vendeurs », ceux dont plus de 6526 exemplaires ont été vendus. Précisons que les désignations « petit vendeur » et « éditeur à petit volume » ont aussi été assignées aux titres dont l'auteur ou l'éditeur n'a eu aucune publication dans le passé.

Il importe de retenir que, tant pour les auteurs que pour les éditeurs, les termes « petit, moyen, grand » utilisés dans ce rapport le sont à des fins de regroupements statistiques uniquement et ne comportent aucun jugement de valeur.

⁴ Rappelons que les éditeurs franco-canadiens hors Québec ayant vendu au moins un titre pendant la période visée dans l'un des points de vente enregistrés à la BTLF sont inclus dans nos données.

⁵ Un quartile est chacune des trois valeurs qui divisent les données triées en quatre parts égales.

Pour les besoins de l'analyse, chaque titre s'est ainsi vu associer un type d'éditeur et un type d'auteur.

Par exemple, pour les titres publiés en 2017, la catégorisation obtenue des éditeurs et des auteurs est présentée dans le tableau 1 alors que les tableaux 2 et 3 montrent différentes statistiques concernant chacun des types d'éditeurs et d'auteurs. Remarquons, dans les deux cas, que le pourcentage de titres avec au moins une métadonnée augmente selon le type d'éditeur ou d'auteur.

Tableau 1

Définition des types d'éditeurs et d'auteurs, pour les titres publiés en 2017

	Petit	Moyen	Grand
Auteur (nb exemplaires vendus de 2010-2016)	0 à 302	303 à 6 526	plus de 6 526
Éditeur (nb titres publiés en 2016)	3 ou moins	4 à 20	plus de 20

Tableau 2

Statistiques sur les types d'éditeurs, pour les titres parus entre le 1^{er} janvier 2016 et le 19 mai 2018

Type d'éditeur	Nb d'éditeurs	Quantité moyenne vendue par titre	Nombre de titres parus	% titres avec au moins une des 5 métadonnées
Petit volume	389	274	1 423	76%
Moyen volume	178	461	3 349	82%
Grand volume	80	720	8 084	90%

Tableau 3

Statistiques sur les types d'auteurs, pour les titres parus entre le 1^{er} janvier 2016 et le 19 mai 2018

Type d'auteur	Nb d'auteurs	Quantité moyenne vendue par titre	Nombre de titres parus	% titres avec au moins une des 5 métadonnées
Petit vendeur	3 710	409	6 545	84%
Moyen vendeur	2 285	415	3 357	85%
Grand vendeur	809	1 246	2 954	94%

Plutôt que de commenter des moyennes de manière strictement descriptive, des tests statistiques ont été effectués afin de déterminer si la différence entre le nombre moyen d'exemplaires vendus par titre avec et sans métadonnées (ou selon le nombre de métadonnées présentes) est significative. Ces différences ont été jugées comme statistiquement significatives à l'aide d'un test de différence, le test de t de Satterthwaite⁶. Le seuil de significativité a été fixé à 5%, ce qui est la norme pour ce type d'étude. Toutefois, le fait qu'une différence ne soit pas jugée statistiquement significative ne signifie pas pour autant qu'il n'y a aucune différence, mais seulement qu'elle ne peut être démontrée avec les données présentes. Cette absence de significativité peut être due à une trop grande variabilité des ventes et/ou à un échantillon de trop petite taille pour détecter une différence.

1.2 Livres numériques

Pour les livres numériques, les données utilisées proviennent de De Marque, agrégateur, distributeur et diffuseur de contenus de livres numériques. Les titres retenus sont tous ceux de format numérique publiés par un éditeur franco-canadien entre le 1^{er} janvier 2016 et le 19 mai 2018. Les métadonnées considérées sont les mêmes que pour les livres imprimés. Les ventes de ces titres ont été colligées par format (formats PDF et ePub principalement) pour les douze mois suivant leur date de parution. Toutefois, aucune analyse de l'effet de la présence ou non des cinq métadonnées d'enrichissement retenues n'a été effectuée compte tenu de l'existence de ces métadonnées pour la quasi-totalité des titres numériques. Par contre, l'information relative à la date de disponibilité des métadonnées a permis d'analyser l'effet de la présence des métadonnées avant la date de parution (plutôt qu'après) sur les ventes.

Cette analyse a été réalisée par type d'éditeur. Ces types ont été définis selon la même méthodologie que pour les livres imprimés.

⁶ Test statistique d'égalité des moyennes de deux populations à partir d'échantillons de tailles différentes comportant de possibles inégalités des variances.

2. Résultats pour les livres imprimés

2.1 Présence des métadonnées

Le tableau 4 expose les métadonnées d'enrichissement pour les 12 852 titres considérés. Il se dégage de ces résultats que 13,6% de ces titres n'ont aucune des cinq métadonnées retenues. Ce pourcentage décroît en fonction du volume de publication de l'éditeur, c'est-à-dire que plus l'éditeur publie de titres annuellement, moins il a de titres n'ayant aucune des cinq métadonnées retenues. En effet, alors que près d'un titre sur quatre ne possède aucune des cinq métadonnées considérées chez les éditeurs à petit volume, ce ratio est de un sur dix pour les titres publiés par les éditeurs à grand volume.

Tableau 4

Pourcentages de titres avec métadonnées, par type d'éditeur

Type d'éditeur	Aucune métadonnée	Imagette et résumé seulement	Imagette sans résumé seulement	Imagette, résumé et autres	Autres métadonnées mais pas d'imagette	Total
Petit volume	24,0%	12,4%	9,6%	45,8%	8,2%	100%
Moyen volume	18,0%	9,7%	6,3%	61,9%	4,1%	100%
Grand volume	9,9%	18,8%	6,4%	58,2%	6,6%	100%
Total	13,6%	15,8%	6,7%	57,8%	6,1%	100%

Un autre constat est la forte représentation de l'imagette dans les métadonnées. En effet, l'imagette est présente pour un peu plus de 80% des titres de l'ensemble des éditeurs. Enfin, près de 58% des titres ont comme métadonnées l'imagette, le résumé ET au moins une autre des trois métadonnées considérées (critique, biographie ou 4^e de couverture).

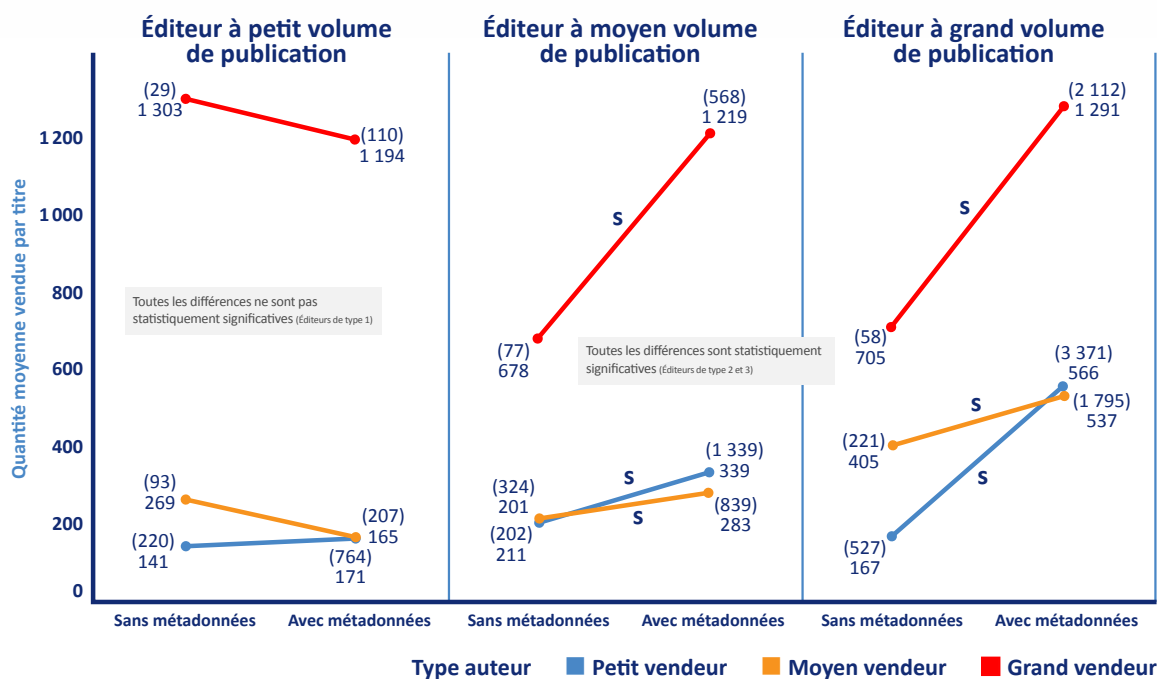
2.2 L'effet de la présence d'au moins une métadonnée

Une première analyse consiste à comparer, selon le type d'éditeur et selon le type d'auteur, le nombre d'exemplaires vendus des titres ayant au moins une métadonnée avec le nombre d'exemplaires vendus des titres sans aucune métadonnée. Les résultats présentés dans le graphique 1 illustrent que pour les éditeurs à moyen et à grand volumes de publication, on remarque une différence positive significative (test t sous le seuil de 5%) du nombre de titres vendus avec au moins une métadonnée, et ce peu importe le type d'auteur. Par exemple, pour les titres publiés par des éditeurs à grand volume de publication et dont les auteurs sont peu vendeurs, les 3 371 titres avec au moins une métadonnée ont été vendus en moyenne à 566 exemplaires lors de leur première année de parution comparativement à 167 exemplaires pour les 527 titres sans aucune métadonnée. Pour les éditeurs à petit volume de publication, les écarts observés ne sont pas statistiquement significatifs.

Dans les graphiques, les droites annotées avec la lettre « S » signifient une différence statistiquement significative. Pour les droites sans annotation, les différences ne sont pas jugées statistiquement significatives⁷.

Graphique 1

Quantité moyenne d'exemplaires vendus par titre avec au moins une métadonnée et sans aucune métadonnée, par type d'auteur et type d'éditeur*



* Les nombres entre parenthèses représentent la quantité de titres traités pour l'ensemble du groupe alors que les nombres qui ne sont pas entre parenthèses représentent la quantité moyenne d'exemplaires vendus par titre.

⁷ Rappel : le fait qu'une différence ne soit pas jugée statistiquement significative ne signifie pas pour autant qu'il n'y a aucune différence, mais seulement qu'elle ne peut être démontrée avec les données présentes. Cette absence de significativité peut être due à une trop grande variabilité des ventes et/ou à un échantillon de trop petite taille pour détecter une différence.

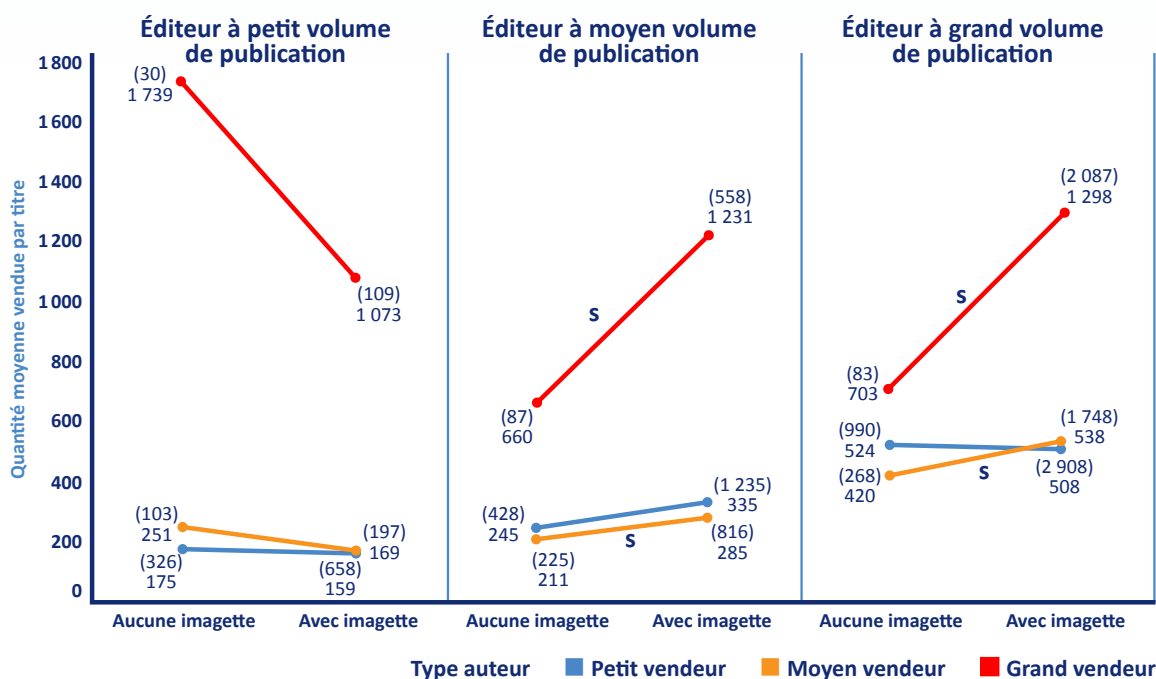
2.3 L'effet de l'imagette

Étant donné la présence importante de l'imagette en tant que métadonnée (pour plus de 80% des titres), la quantité moyenne d'exemplaires vendus avec imagette a été comparée à celle vendue sans la présence de l'imagette. Les résultats présentés dans le graphique 2 montrent que, pour les éditeurs avec des volumes de publication moyens et élevés et pour les auteurs qui sont de moyens et grands vendeurs, la moyenne d'exemplaires vendus lors de la première année de parution est significativement supérieure pour les titres avec imagette.

À noter que la différence en pourcentage entre les quantités moyennes d'exemplaires vendus avec et sans imagette est plus élevée pour les auteurs qui sont de grands vendeurs, avec une augmentation de près de 100% de ventes. L'effet est moindre, bien que significatif, pour les titres d'auteurs qui sont de moyens vendeurs (moyenne de 285 exemplaires vendus avec imagette vs 211 sans imagette pour les éditeurs à moyen volume, comparativement à une moyenne de 538 exemplaires vendus avec imagette vs 420 sans imagette pour les éditeurs à grand volume).

Graphique 2

Quantité moyenne d'exemplaires vendus par titre avec imagette et sans imagette, par type d'auteur et type d'éditeur



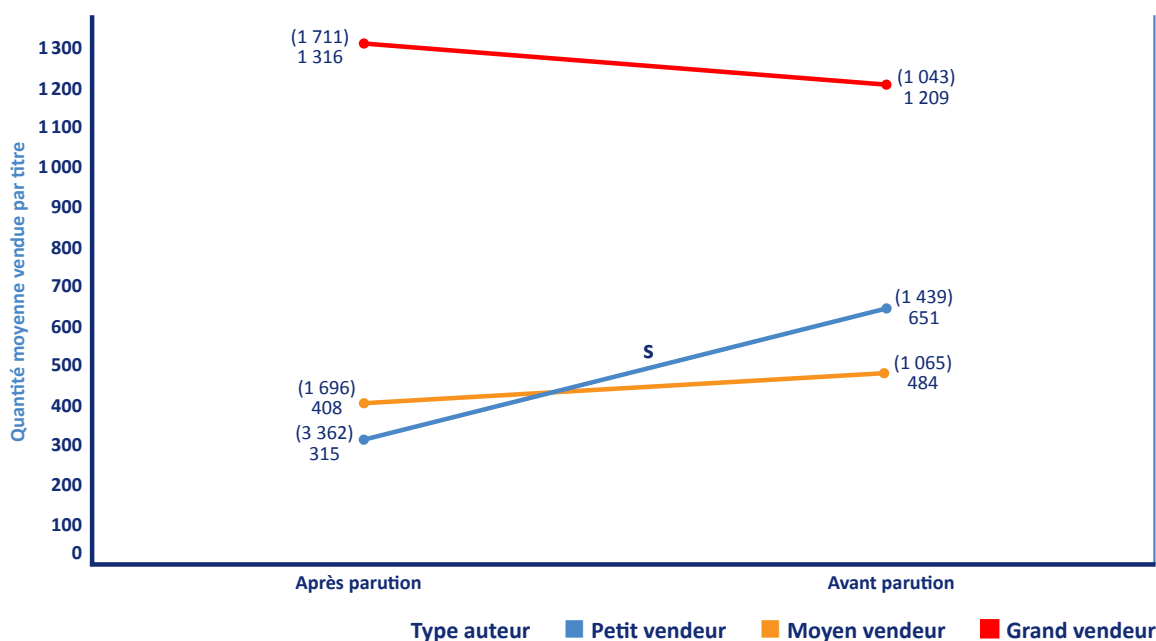
Les nombres entre parenthèses représentent la quantité de titres traités pour l'ensemble du groupe alors que les nombres qui ne sont pas entre parenthèses représentent la quantité moyenne d'exemplaires vendus par titre. La lettre « S » indique une différence statistiquement significative entre les deux points de la droite.

2.4 L'effet de la présence des métadonnées avant la parution

Une autre question d'intérêt est la mesure de l'effet sur les ventes du moment de l'apparition d'une métadonnée. Cette question a été analysée pour la métadonnée imagerie. L'ensemble des titres avec imagerie a été divisé en deux groupes: soit les titres dont l'imagerie en tant que métadonnée a été affichée **avant** la date de parution (34% des titres) et un autre groupe constitué des titres pour lesquels l'imagerie a été disponible seulement **après** la date de parution (66% des titres). Pour les titres avec auteur petit vendeur (catégorie avec le plus faible nombre de titres vendus), le nombre d'exemplaires vendus pour les titres avec apparition de l'imagerie **avant** la date de parution est significativement plus élevé que pour les titres dont l'imagerie a été disponible **après** la date de parution, et ce, peu importe le type d'éditeur (globalement: 651 exemplaires vendus en moyenne contre 315). Pour les titres d'auteurs identifiés comme moyens ou grands vendeurs, les différences ne sont pas globalement significatives, et ce, peu importe le type d'éditeur. Le graphique 3 présente les résultats par type d'auteur.

Graphique 3

Quantité moyenne d'exemplaires vendus par titre en fonction de la disponibilité de l'imagerie avant ou après la date de parution, par type d'auteur



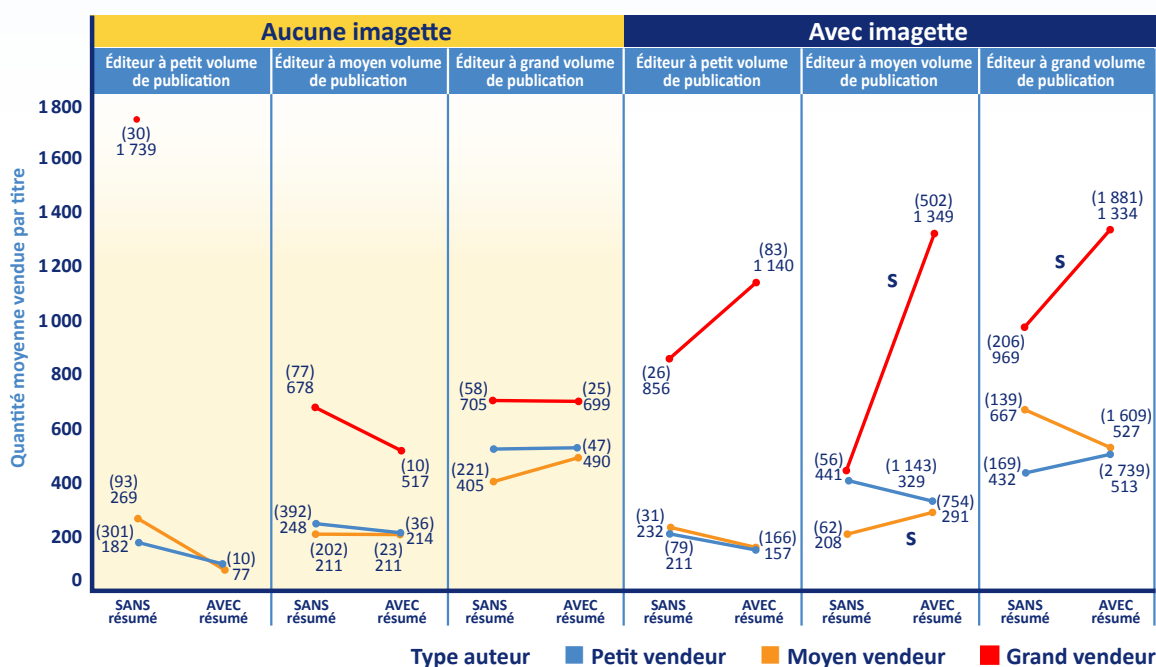
Les nombres entre parenthèses représentent la quantité de titres traités pour l'ensemble du groupe alors que les nombres qui ne sont pas entre parenthèses représentent la quantité moyenne d'exemplaires vendus par titre. La lettre « S » indique une différence statistiquement significative entre les deux points de la droite.

2.5 L'effet du résumé avec et sans imagerie

Une autre analyse effectuée a consisté à vérifier l'effet de la présence du résumé en tant que métadonnée selon la présence ou non de l'imagerie. Les résultats démontrent que l'ajout du résumé lorsque l'imagerie est absente n'augmente pas de façon significative le nombre d'exemplaires vendus, peu importe le type d'éditeur ou d'auteur. Par contre, lorsque l'on considère les titres avec imagerie, la quantité moyenne d'exemplaires vendus des titres auxquels le résumé a été ajouté est significativement supérieure à celle des titres sans résumé chez les auteurs grands vendeurs publiés par des éditeurs à moyen et à grand volumes de publication. L'effet est aussi présent pour les auteurs moyens vendeurs publiés par des éditeurs à moyen volume (voir graphique 4). La plus forte augmentation du nombre d'exemplaires vendus concerne les auteurs grands vendeurs publiés par des éditeurs à moyen volume de publication (près de 300%).

Graphique 4

Quantité moyenne d'exemplaires vendus par titre avec ou sans résumé selon la présence d'imagerie ou non, par type d'auteur et type d'éditeur



Les nombres entre parenthèses représentent la quantité de titres traités pour l'ensemble du groupe alors que les nombres qui ne sont pas entre parenthèses représentent la quantité moyenne d'exemplaires vendus par titre. La lettre « S » indique une différence statistiquement significative entre les deux points de la droite.

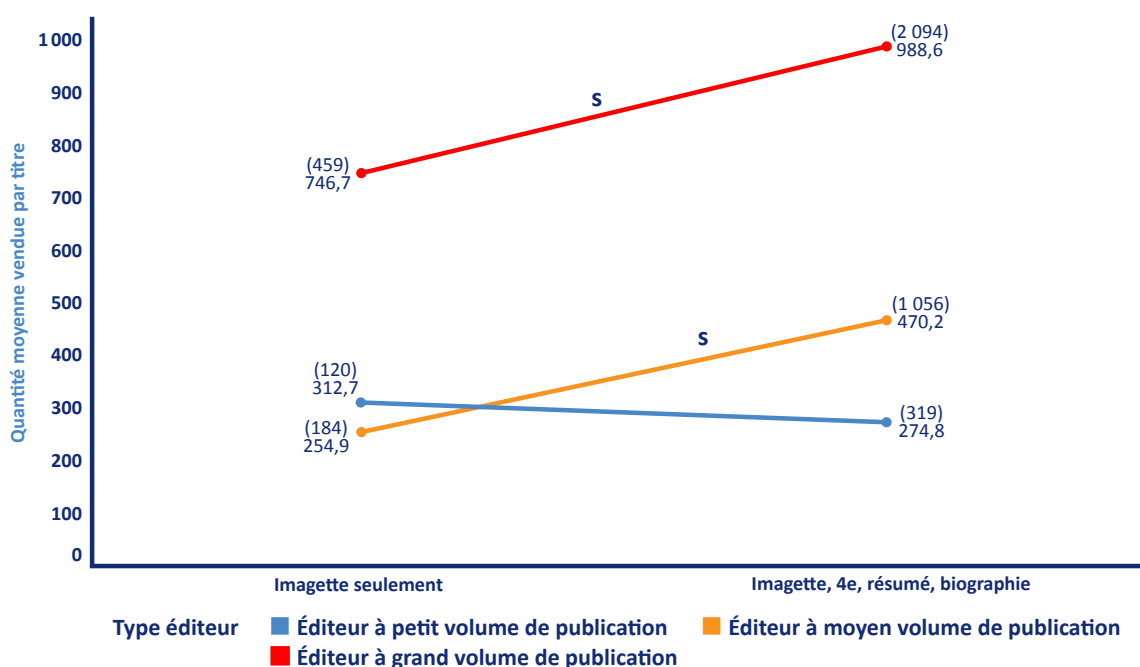
2.6 L'effet du nombre de métadonnées d'enrichissement

Une dernière analyse a consisté à vérifier l'effet du nombre de métadonnées présentes (une à quatre) sur la quantité d'exemplaires vendus par titre. Pour cela, on a comparé la vente moyenne des titres ayant seulement l'imagette comme métadonnée à celle des titres ayant l'imagette, la 4^e de couverture, le résumé et la biographie de l'auteur. Les résultats présentés dans le graphique 5 le sont par type d'éditeur (les conclusions quant à la significativité des différences par auteur ne se distinguaient pas de celles obtenues par type d'éditeur).

La conclusion principale est que l'on détecte une différence significative entre les quantités moyennes d'exemplaires vendus par titre selon le nombre de métadonnées d'enrichissement présentes chez les éditeurs à grand et à moyen volumes de publication (respectivement une hausse de 84% et 32% lorsque quatre métadonnées sont présentes).

Graphique 5

Quantité moyenne d'exemplaires vendus par titre avec imagette seulement ou avec quatre métadonnées (imagette, 4^e de couverture, résumé et biographie), par type d'éditeur



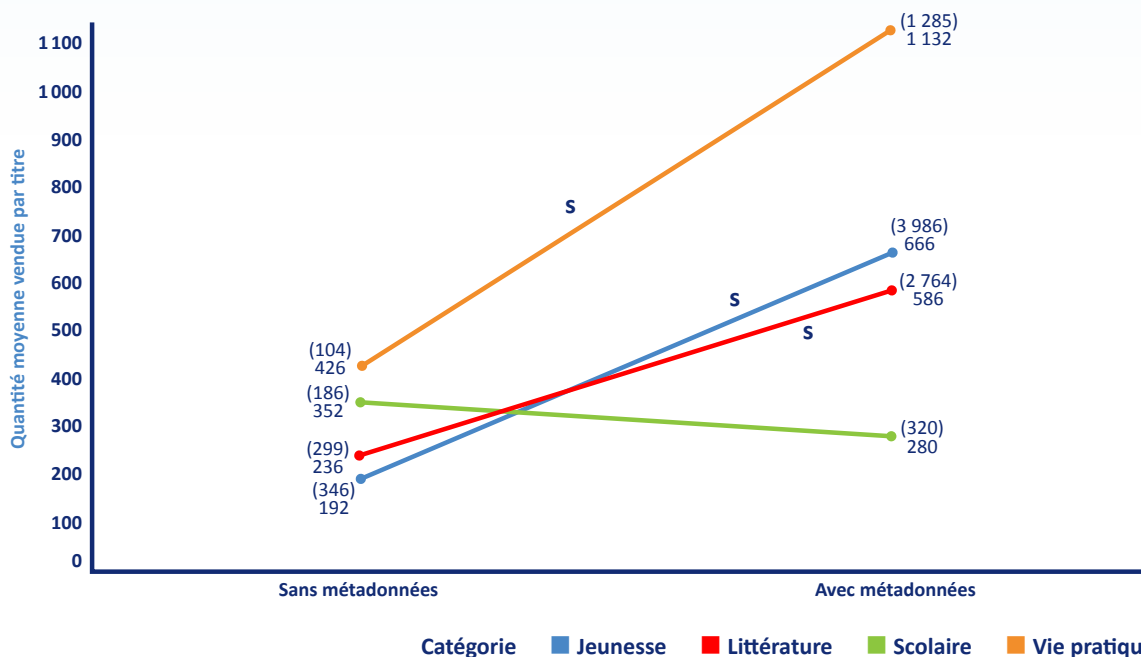
Les nombres entre parenthèses représentent la quantité de titres traités pour l'ensemble du groupe alors que les nombres qui ne sont pas entre parenthèses représentent la quantité moyenne d'exemplaires vendus par titre. La lettre « S » indique une différence statistiquement significative entre les deux points de la droite.

2.7 Analyse par catégorie de livres

Selon la nomenclature de la BTLF, chaque titre est associé à l'une des 22 catégories d'ouvrages définies dans Gaspard. Les quatre principales catégories en termes de nombre de titres publiés lors de la période de référence sont Jeunesse (34,5 % des titres), Littérature (24,4%), Vie Pratique (11,1%) et Scolaire (4,0%). Pour chacune de ces catégories, l'analyse de l'effet de la présence des métadonnées a été faite en considérant l'ensemble des éditeurs. Seule la catégorie Scolaire ne présente pas un nombre d'exemplaires vendus significativement supérieur avec la présence de métadonnées. Le ratio le plus élevé est pour la catégorie Jeunesse, où le nombre d'exemplaires vendus avec au moins une métadonnée est 3,47 fois plus élevé que pour les titres sans aucune métadonnée (676 vs 192).

Graphique 6

Quantité moyenne d'exemplaires vendus par titre sans aucune métadonnée et avec au moins une métadonnée, par catégorie



Les nombres entre parenthèses représentent la quantité de titres traités pour l'ensemble du groupe alors que les nombres qui ne sont pas entre parenthèses représentent la quantité moyenne d'exemplaires vendus par titre. La lettre « S » indique une différence statistiquement significative entre les deux points de la droite.

3. Analyse pour les livres numériques

Pour les livres numériques, une base de données fournie par De Marque regroupant tous les titres publiés entre le 1^{er} Janvier 2016 et le 19 mai 2018 a été utilisée afin d'analyser l'effet des métadonnées sur les ventes.

L'objectif premier de cette analyse était initialement de répliquer l'analyse effectuée pour les livres imprimés, pour chacun des types de support numérique (audio, PDF, ePub et mobile). La comparaison des ventes selon la présence ou non de métadonnées n'a pu être effectuée étant donné que la quasi-totalité des titres possèdent des métadonnées, ce qui induit un nombre insuffisant de livres numériques sans métadonnées comme point de comparaison (voir Tableau 5. Notons que l'imagerie n'apparaît pas dans ce tableau, car tous les titres l'ont comme métadonnée).

Tableau 5

Répartition des titres numériques selon le type de métadonnée et de support numérique

Métadonnées	Audio	ePub	Mobile	PDF	Total	Total %
Aucune		20	1	48	69	0,40%
Biographie	1	113	1	186	301	1,60%
Résumé	1	78	15	95	189	1,00%
Biographie et résumé	311	7 503	51	9 008	16 873	90,40%
Critique et biographie				1	1	0,00%
Critique et résumé				1	1	0,00%
Critique, résumé et biographie	7	60		83	150	0,80%
4 ^e couverture et biographie		1		7	8	0,00%
4 ^e couverture et résumé				7	7	0,00%
4 ^e couverture, résumé et biographie		382		610	992	5,30%
4 ^e couverture, critique et biographie				1	1	0,00%
4 ^e couverture, critique, résumé et biographie	1	25		38	64	0,30%
Total	321	8 182	68	10 085	18 656	100%

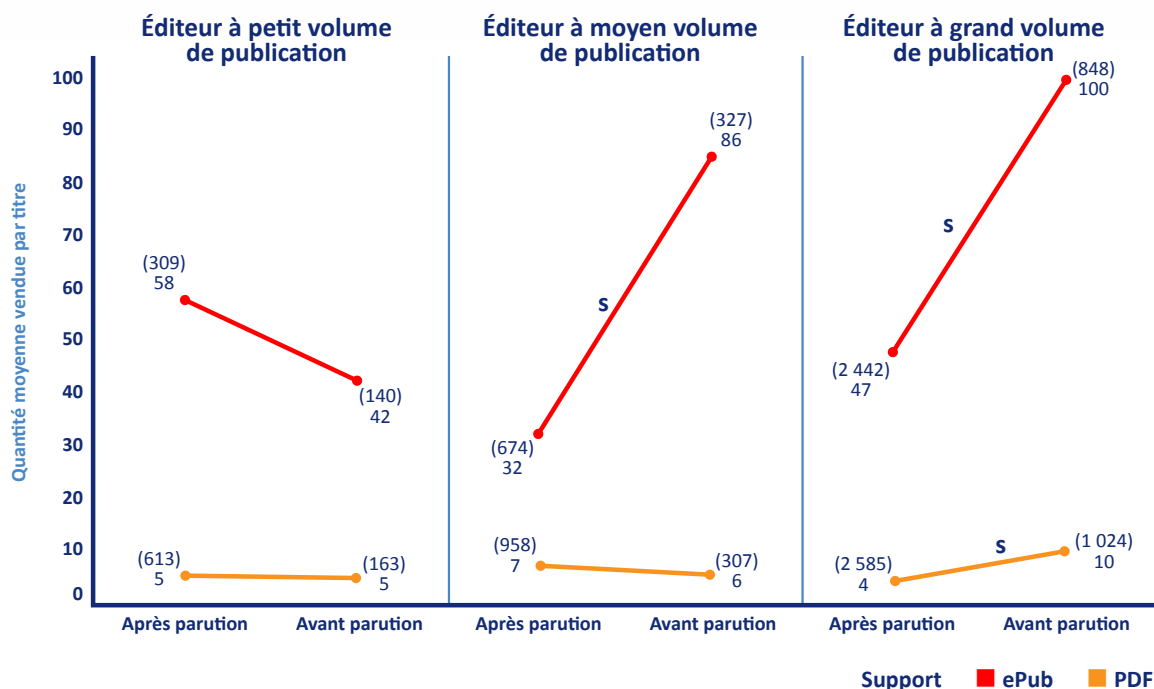
Nous remarquons ainsi que 90% des titres numériques ont l'imagerie, la biographie et le résumé comme métadonnées systématiquement disponibles.

3.1 L'effet de la présence des métadonnées avant la parution

La base de données de De Marque inclut comme information la présence ou non des métadonnées avant la date de parution du titre. Cette information a permis de comparer la quantité d'exemplaires vendus lorsque la disponibilité des métadonnées précède la date de parution à la quantité d'exemplaires vendus lorsque l'information est disponible après la date de parution. Cette analyse a été effectuée par type d'éditeur (défini de façon similaire aux titres imprimés) et type de format numérique (PDF et ePub⁸). Les résultats obtenus pour les éditeurs à moyen et à grand volumes de publication indiquent une différence significative du nombre moyen d'exemplaires vendus pour les titres de format ePub lorsque les métadonnées sont disponibles **avant** la date de parution: en effet, le nombre moyen d'exemplaires vendus par titre est plus que doublé lorsque comparé aux titres dont l'information relative aux métadonnées est disponible **après** la parution (voir graphique 6). Pour ce qui est du format PDF, la seule différence significative concerne les éditeurs à grand volume de publication.

Graphique 7

Quantité moyenne d'exemplaires vendus par titre numérique selon la date de disponibilité des métadonnées, par type d'éditeur et de format numérique



Les nombres entre parenthèses représentent la quantité de titres traités pour l'ensemble du groupe alors que les nombres qui ne sont pas entre parenthèses représentent la quantité moyenne d'exemplaires vendus par titre. La lettre « S » indique une différence statistiquement significative entre les deux points de la droite

8 Les formats PDF et ePub représentent 98% des titres.

Conclusion

Cette étude s'inscrit dans l'ensemble des efforts du secteur du livre entourant la découvrabilité des ouvrages québécois et franco-canadiens. Les résultats démontrent que des métadonnées riches et transmises à temps sont généralement liées à des ventes supérieures. L'étude fournit aussi des indications sur le contexte dans lequel les métadonnées sont les plus efficaces. Sans pouvoir identifier le mécanisme précis qui conduit à ces résultats positifs, nous pouvons avancer avec une relative certitude que les métadonnées sont un maillon important de la découvrabilité des livres et qu'elles ont un effet considérable sur les ventes. Une opportunité de recherche future serait d'identifier qui sont les utilisateurs de métadonnées tout au long de la chaîne de création du livre afin de mieux comprendre l'effet de la métadonnée que nous avons mesuré.

La présente étude est un complément à des recherches antérieures, notamment celles de Walter⁹ ainsi que de Breedt et Walter¹⁰ sur l'impact des métadonnées sur la vente de livres au Royaume Uni. En plus de valider les résultats sur le marché du livre québécois, notre étude, tout comme les précédentes, confirme un effet général des métadonnées sur les ventes ainsi qu'un impact spécifique de l'imagette et d'une information disponible avant la parution. Qui plus est, notre analyse décortique l'impact des métadonnées en fonction du volume de titres publiés annuellement par l'éditeur et en fonction du nombre d'exemplaires vendus dans le passé par auteur. Nous présentons aussi une analyse pour les livres numériques et pour les imprimés. Tout comme les rapports précédents, nous ne pouvons identifier qu'une relation de corrélation entre les pratiques de métadonnées et les ventes et non une causalité de l'effet.

Un élément clé de cette analyse est le fait que la dimension visuelle d'une métadonnée (l'imagette) semble un vecteur important de découvrabilité. Le fait qu'une métadonnée rendue disponible avant publication ait un impact aussi important témoigne aussi de la nécessité, pour les intervenants de la chaîne du livre, d'en gérer activement et stratégiquement le processus de production et de diffusion.

Notons que certains facteurs externes, qui n'ont pas pu être pris en compte dans cette étude, pourraient amplifier l'effet des métadonnées. Par exemple, la couverture médiatique entourant la sortie d'un livre, les dépenses de promotion ou encore le moment de l'année où un livre a été publié pourraient modérer l'effet des métadonnées.

Finalement, rappelons que la présente étude s'est intéressée à la performance commerciale des livres liée à de bonnes pratiques de production et de transmission des métadonnées d'enrichissement. Il serait intéressant de comparer l'effet des métadonnées sur les ventes aux particuliers et aux collectivités. De nouveaux efforts pourraient aussi viser à mieux comprendre l'effet de d'autres types de métadonnées sur les ventes, comme les sujets Thema, par exemple.

La BTLF, l'ANEL et De Marque souhaitent travailler de pair avec les éditeurs et avec les autres acteurs de la chaîne du livre afin d'améliorer la découvrabilité des ouvrages publiés en français et de soutenir leur commercialisation grâce aux métadonnées. Cette étude montre l'importance d'inclure la gestion des métadonnées dans toute stratégie commerciale liée au livre. La BTLF entend d'ailleurs raffiner les données de Gaspard et de Memento pour réaliser des analyses toujours plus fines, répondant aux besoins des intervenants de l'industrie.

⁹ David Walter, Nielsen BookScan, Nielsen Book UK Study: The Importance of Metadata for Discoverability and Sales, UK, 2016.

¹⁰ Andre Breedt and David Walter, Nielsen BookScan, The Link Between Metadata and Sales, UK, 2012.