

Séries chronologiques sur un corpus de presse écrite et littérature scientifique : statistiques inter-textuelles, infra-textuelles et supra-textuelles.

Le cas d'une innovation technologique controversée : la RFID

Alexandre Delanoë¹, Laura Draetta²

¹ Telecom ParisTech – alexandre.delanoë@telecom-paristech.fr

² Telecom ParisTech – laura.draetta@telecom-paristech.fr

Abstract

This paper presents the textual and statistical data study of a public debate that deals with the introduction in daily use of the innovation called RFID (Radio Frequency Identification). The analysis focuses first on the mainstream press and shows how specific textual forms mark the series. This *inter-textual* methodology is mobilized to highlight a shift in trend from a textual key date : that shift is called « controversy ». The changeover date is observed by following methods to analyze the corpus :

- *Infra-textual* method is used to study the textual dynamics in a section of the corpus itself.
- *Supra-textual* method is used to study the textual and normalized dynamics outside the initial corpus.

Inter-textual shows the first correlation that can not be interpreted as an explanation yet. *Infra-textual* analysis allows to highlight the dynamics of public debate while the *supra-textual* analysis reveals the diachronic appropriation of the RFID term by different areas: economic and managerial fields, the fields of public debate, and the fields of engineering.

Résumé

Cette communication présente l'étude statistique et textuelle d'un débat public, celui de l'introduction de l'innovation RFID (Radio-fréquences Identification) dans les usages quotidiens. L'analyse se concentre d'abord sur la presse grand public par une approche *inter-textuelle* qui vise à étudier les dynamiques textuelles normalisées à l'intérieur du corpus. Cette première exploration met en valeur un basculement de tendance textuelle à partir d'un événement que nous appelons « controverse ». La date de basculement est ensuite évaluée en suivant 2 autres référentiels :

- La méthode *infra-textuelle* vise à étudier les dynamiques textuelles normalisées dans une section raisonnée du corpus initial.
- La méthode *supra-textuelle* vise à étudier les dynamiques textuelles normalisées à l'extérieur du corpus initial.

L'analyse inter-textuelle met en valeur une première corrélation qui ne peut avoir valeur d'explication à ce stade de l'analyse. La contextualisation *infra-textuelle* permet de préciser la dynamique du débat public tandis que la l'analyse *supra-textuelle* révèle l'appropriation diachronique du terme RFID par les différents champs d'action des acteurs : les champs économiques et gestionnaires, le champs du débat public, et le champs des sciences de l'ingénieur.

Mots-clés : débat public, séries chronologiques textuelles, lexicométrie, sociologie, RFID radiofréquence identification

1. Introduction à l'identification par radio-fréquences

L'identification par radio-fréquences, présentée sous l'acronyme RFID, est une technologie de lecture et d'écriture à distance, par radiofréquences. Grâce à une propriété innovante, celle de réaliser une communication sans contact, elle apparaît au début du XXI^{ème} siècle comme une technologie émergente dans la gestion des flux d'une économie globale en particulier dans le domaine de la traçabilité des objets, des animaux, mais aussi celui de l'authentification des personnes.

Cependant, ses usages apparaissent discutables et sont vivement discutés. Une multiplicité d'arguments critiques dénonce (Boltanski, 1984) publiquement cette technologie notamment pour l'atteinte à la vie privée et d'autres risques (Draetta, Delanoë, 2012). Des « lanceurs d'alerte (Chateauraynaud et Torny, 1999) », voire des « entrepreneurs de morale (Becker, 1963) » interviennent publiquement tantôt pour dévoiler les risques de dérives et de mésusages, tantôt pour s'opposer radicalement à la technologie en elle-même.

L'histoire de la RFID est controversée dans un débat public où les acteurs font multitude. En conséquence, l'étendue de ce terrain d'enquête nécessite une analyse statistique textuelle des corpus de presse écrite et littérature scientifique pour observer son développement historique.

2. Historique du traitement de la RFID dans la presse grand public

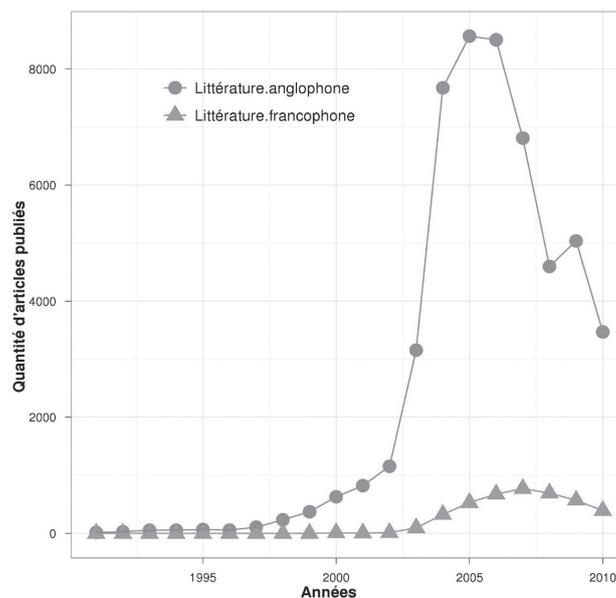


Figure 1 : Évolution historique et comparée des corpus anglophone et francophone. Légende : Le graphique montre la distribution par année des articles anglophones et francophones, traitant de la RFID, publiés entre 1991 et 2010. (Source : notre élaboration à partir de l'exploitation de la base de données Factiva.)

À partir de l'équation de recherche centrée sur les mots-clefs « RFID and radio-frequency » (et sa traduction française), deux corpus d'articles de presse ont été extraits depuis la base de données FACTIVA, l'un anglophone et l'autre francophone. La distribution chronologique des articles contenus dans ces corpus permet d'évaluer l'entrée puis l'émergence de la RFID dans le champ de la presse écrite (cf. Figure 1) de 1990 à 2010. Ce corpus est significatif, il comporte 55 510 articles dont 51 433 sont anglophones et 4 077 francophones. Mais ce corpus reste soumis à fluctuation car son exhaustivité dépend de la couverture de la base de données FACTIVA, elle-même dépendante de ses relations commerciales avec ses fournisseurs de données.

3 phases peuvent alors être repérées:

- 1) Les premiers pas de l'usage « RFID » dans l'espace public : ce sont les premières applications de la technologie ;
- 2) Une augmentation exponentielle d'articles à partir de 2003 : un événement déclenche de nombreux commentaires en faveur mais aussi à l'encontre de la technologie.
- 3) Une baisse radicale de la quantité d'articles à partir de 2005/2006.

Cette communication s'attache essentiellement à étudier la baisse de publication d'articles à partir de 2005/2006. Nous renvoyons à (Draetta et Delanoë, 2012) pour les deux premières phases. Il s'agit en effet, ici, de comprendre le basculement de la tendance observable à partir de l'année 2005.

3. L'impact de la critique révélé par les spécificités chronologiques : statistiques inter-textuelles

Le corpus de presse écrite comporte des textes relatant l'espérance en la technologie mais aussi des textes critiques (ou faisant référence à des propos de la critique) exprimant une certaine méfiance à l'égard de la technologie et de ses usages. Or il existe des formes élémentaires de « méta-information (Lebart et Salem, 1994) » à partir desquelles les acteurs se retrouvent pour exprimer leurs perspectives économiques ou socio-politiques. Suivant cet postulat à notre recherche, les articles comporteraient des mots-clés propres à chaque collectif d'acteurs. Ce serait le cas du terme emblématique *privacy*. En effet, à la suite d'une enquête qualitative dans le milieu industriel de la RFID, nous faisons l'hypothèse que ce terme est le marqueur d'un renversement de tendance dans l'appréciation de la technologie RFID.

Le corpus de presse anglophone ne comporte que 10% (environ) de ses articles contenant le terme *privacy*. Ceci implique que 90% des articles du corpus ne citent pas ce terme. Par contre, en comparant les variations annuelles du nombre d'articles publiés sur « RFID et *privacy* » avec celles des articles portant sur la RFID en général (et ne pas faisant référence à *privacy*), il est possible d'observer l'influence de la critique sur le traitement public de la technologie RFID. En effet, l'augmentation de la quantité d'articles n'évoquant pas la *privacy* est négativement corrélée à l'augmentation des articles évoquant l'atteinte à la vie privée (Figure 2).

La Figure 2 présente l'évolution comparée du corpus citant textuellement (ou non) le terme « *privacy* ». Mais cette corrélation ne peut avoir valeur d'explication. Comme aime à se le rappeler R. Boudon, en citant Lazarsfeld (Assogba, 2005), ce n'est pas parce qu'il y a comparativement plus d'enfants et de cigognes dans les cantons alsaciens que les enfants sont

apportés par les cigognes. La corrélation entre vols de cigogne et natalité en Alsace ne prouve donc pas la causalité. Sensibles à cet aspect méthodologique, il convient alors d'étudier de manière comparative l'énergie performative de tous ces articles.

Une analyse des spécificités textuelles permet de porter un jugement sur la fréquence de chacune des unités textuelles dans chacune des parties du corpus (Salem, 1988) (Salem, 1994). Ce travail statistique consiste à révéler les termes spécifiques à chacune des parties: celle comportant le terme *privacy* et l'autre partie qui ne contient pas le mot *privacy*.

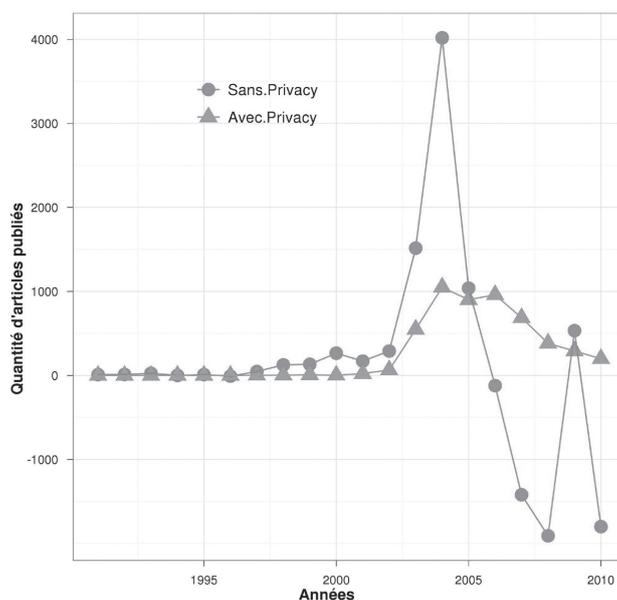


Figure 2 : Variation annuelle du nombre d'articles publiés ne citant pas *privacy* comparé à la quantité d'articles publiés par année citant *privacy*. (Source : Notre élaboration à partir de l'exploitation de la base de données Factiva.)

La figure 3 a pu être réalisée par la méthode de l'échantillonnage qui consiste en la sélection d'une partie dans un tout. L'échantillonnage est une notion importante en métrologie : lorsqu'on ne peut pas saisir un événement dans son ensemble, il faut effectuer des mesures en nombre fini, afin de représenter l'événement. Suivant cette méthode, précisée dans la légende de la figure, nous obtenons alors deux ensembles de terminologies bien distinctes qui révèlent les acteurs principaux de chaque partie, le contexte d'élaboration des textes et l'enjeu. Les consommateurs apparaissent sous le terme *consumer* et les clients finaux des industriels apparaissent sous le terme *customer*. Cette différence de perspective a un impact sur le déploiement de l'industrie RFID.



Figure 3 : Représentation graphique des spécificités textuelles du corpus anglophone issu de Factiva (51 433 articles). Légende : La grosseur de la typographie des termes est corrélée avec leur niveau de spécificité dans chaque sous-corpus. Schéma réalisé avec Lexico, Awk/bash/shuf, Wordle, Gimp et Inkscape. La sphère « RFID » rassemble tous les articles citant RFID dans leur contenu. Parmi ces articles, près de 10% évoquent la *privacy*. En décomposant le corpus en deux parties, discriminées par le terme *privacy*, une analyse des spécificités fut réalisée sur 6 échantillons prélevés de manière aléatoire. Pour rester cohérents avec notre sujet d'étude, le script de génération de l'aléatoire, *shuf* pour *suffle*, s'est inspiré d'un algorithme de chiffrement destiné à préserver la confidentialité des données numériques *privacy*). La moyenne des scores de spécificités a été calculée puis traduite en occurrences dans un fichier interprétable par le site Internet www.wordle.net. (Source : Notre élaboration à partir de l'exploitation de la base de données Factiva.)

Deux mondes de justification se distinguent clairement (Boltanski et Thevenot, 1991) (Figure 3). Jusqu'en 2010, 90% de l'ensemble des textes publiés sur la RFID concernent le monde industriel et sa logique marchande, tandis que les 10% restants concernent une cité que nous appelons « critique ». Celle-ci dénonce la RFID principalement pour son atteinte (potentielle ou effective) à la vie privée. Les termes du corpus relatif à la RFID et ne citant pas *privacy* font plutôt référence au développement économique et industriel centré sur le client, le *customer*. Les produits, la compétition, le profit, la hiérarchisation (*rank*) et les applications apparaissent dans ces articles qui manifestent une logique d'optimisation du service et de la logistique. Toutefois, une certaine incertitude apparaît du fait du *risque* des investissements réalisés: le marché tarde à se révéler...

Plus qu'une simple puce sans contact, la puce RFID devient une boîte noire invisible et omniprésente si bien que son usage généralisé risque de peser sur la vie quotidienne, sur la vie privée, sur la vie professionnelle mais aussi sur la santé. C'est pourquoi les acteurs se

mobilisent publiquement soit pour la promouvoir, soit pour la contrer. Elle devient un problème public, au sens de (Gusfield, 1981), à savoir un enjeu de réflexion et de protestation publique (Cefai et Pasquier, 2003) contre de nouveaux risques. La technologie RFID franchit la frontière qui sépare le public du privé (Dewey, 1927) et la portée de ses conséquences appelle le contrôle, soit par prohibition, soit par promotion soit par l'encadrement des pratiques. Le public se constitue d'acteurs attentifs à l'enjeu de la vie privée et se mobilise donc pour l'orientation de la législation en matière de nouveaux usages émergents ou en perspective : l'opinion a ses publics (Brugidou, 2008). Or la multiplicité des interventions brouille le message et complexifie le travail analytique de la recherche. D'où la nécessité d'orienter la recherche dans deux directions différentes.

4. La contextualisation infra et supra textuelle

Pour retracer le débat public, 2 approches apparaissent complémentaires :

- 1) Tracer les termes du débat et les acteurs à l'intérieur du corpus lui-même, il s'agit de spécifier chronologiquement les formes textuelles les unes par rapport aux autres dans une découpe non pas aléatoire (comme dans le cas de l'échantillonnage) mais raisonnée du corpus. Nous appelons cette approche la *contextualisation infra-textuelle*.
- 2) Tracer la contextualisation des termes eux-mêmes à l'extérieur des textes, dans un champs de production textuel plus vaste, ce qui suppose un marquage de forme supplémentaire. Nous appelons cette approche la *contextualisation supra-textuelle* :

Ces 2 approches nécessitent un nouveau travail sur les corpus d'analyse :

- 1) Pour l'approche infra-textuelle, le corpus de presse écrite sera réduit au débat sur les risques socialement construits autour de la technologie RFID en France. Cette section vise à vérifier si le point de basculement de la tendance observé en 2005 se confirme également sur une section plus limitée du corpus.
- 2) Pour l'approche supra-textuelle, le corpus de presse écrite sera complété par un corpus de littérature scientifique de sciences de gestion, et des sciences de l'ingénieur. Nous étudierons alors l'évolution chronologique de chacun des corpus.

L'analyse infra-textuelle permet de mettre en valeur la dynamique du débat public tandis que la l'analyse supra-textuelle révèle l'appropriation diachronique du terme RFID par les différents champs d'action.

Ces deux approches permettent de qualifier le changement de cadre d'analyse opéré. En utilisant une analogie anglophone, on pourrait dire que la *contextualisation infra-textuelle* correspond à une fonction *zoom in* et la *contextualisation inter-textuelle* à celle du *zoom out*. Aussi, nous pourrions utiliser les termes *micro* et *macro*. Cependant ces mots renvoient à une échelle *spatiale* alors que les seuils de basculement textuels que nous étudions renvoient à une échelle *temporelle* ; c'est pourquoi la création des néologismes référentiels nous apparaissent justifiés pour leur valeur heuristique.

4.1. Le rôle de la contextualisation infra-textuelle

Trois catégories d'acteurs interviennent dans le débat public tel qu'il est relayé par la presse grand public : les promoteurs, les régulateurs et les acteurs de la critique.

En croisant les termes qui les désignent, souvent leur nom propre, avec l'année de parution de l'article qui les cite, on peut observer l'évolution de leur apparition dans notre corpus. Ainsi, les promoteurs apparaissent principalement par l'intermédiaire des entreprises qui font l'objet du plus grand nombre d'articles publiés: Wal-Mart, Gillette et Benetton. Du côté des régulateurs, le Parlement Européen et la CNIL sont surtout cités, le premier avec ses recommandations, la seconde avec ses interrogations. Quant aux acteurs de la critique, on retrouve principalement l'organisation américaine CASPIAN, le collectif grenoblois Pièces et Main d'Œuvre et le journaliste scientifique Michel Alberganti (cf. Figure 4), suivis par la catégorie socioprofessionnelle des caissières.

Ces acteurs n'interviennent pas au même moment dans la scène publique. Ils émergent au fur et à mesure, chacun avec des discours différents. Le débat public sur la RFID ne démarre qu'en 2003 (Figure 4). Avant cette date, lorsque les commentateurs et analystes commencent à évoquer la RFID, en 1998, les discours des promoteurs prévalent, en vantant les nouvelles applications possibles de la technologie. À cette époque, la critique n'est pas encore visible. Il faudra en effet attendre 2003 pour qu'elle se manifeste. Car 2003 est l'année où l'entreprise Wal-Mart, leader mondial de la grande distribution, introduit la RFID dans ses supermarchés à des fins logistiques. C'est le début des discours prospectifs des promoteurs qui voient dans la RFID une nouvelle technologie pouvant remplacer le code barre. Dans la même année, les marques Benetton et Gillette appliquent des puces RFID sur leurs produits à des fins opérationnelles. Des acteurs de la critique - CASPIAN en particulier - se manifestent dans la presse pour révéler un objectif d'antivol pour Gillette et un objectif de profilage pour Benetton : le débat sur le risque RFID s'amorce.

Si les entreprises Wal-Mart, Benetton et Gillette, font parler d'elles entre 2003 et 2005, ce sont ensuite les acteurs de la régulation et de la critique qui influent sur le débat (à partir de 2006). En effet, entre 2003 et 2008, l'attention de la presse semble se déplacer progressivement des interrogations sur la viabilité économique de la technologie à la régulation, en passant par la problématique de l'atteinte à la vie privée des consommateurs. Le seuil de basculement est bien l'année 2006 où les acteurs de la régulation apparaissent au même niveau que les promoteurs, sous le feu de la critique.

La tendance globale du corpus présentée au point 1, est donc confirmée par l'analyse des séries chronologiques textuelles centrée sur une partie du corpus : la contextualisation infra-textuelle.

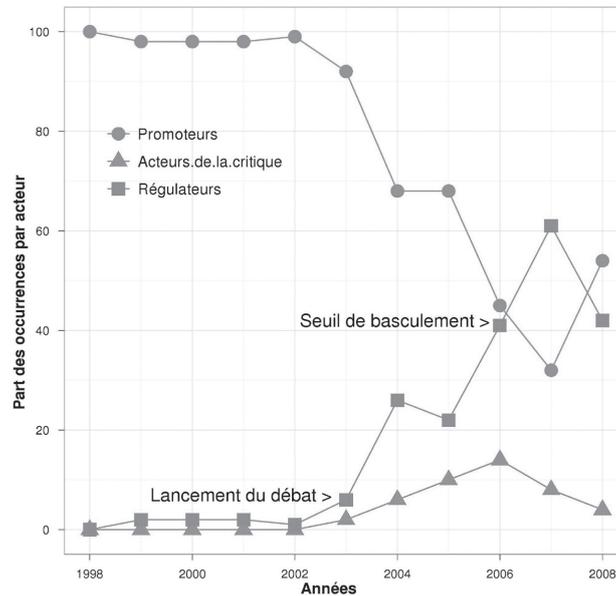


Figure 4 : Évolution annuelle des contributions des 3 catégories d'acteurs principales dans le débat : promoteurs, acteurs de la critique et régulateurs (Corpus francophone centré sur les « risques »).

Légende : La part de voix annuelle de chaque catégorie est obtenue en calculant le pourcentage de leurs occurrences sur l'ensemble des occurrences des acteurs de la même année. La catégorie des promoteurs est composée des principaux industriels cités : Wal-Mart, Gillette ou Benetton. La catégorie des régulateurs est composée par la CNIL et le Parlement Européen, principaux acteurs intervenant dans le débat pour la régulation. Enfin, les acteurs de la critique sont composés par les acteurs CASPIAN, K. Albrecht, Pièces et Main d'Œuvre, M. Alberganti et les caissières. (Source : Notre élaboration à partir de l'exploitation de la base de données Factiva.)

4.2. Le rôle de la contextualisation supra-textuelle

Concernant la contextualisation supra-textuelle, il s'agit de procéder à la contextualisation statistique et textuelle des écrits dans un champs plus vaste que celui délimité par le corpus initial. L'astuce consiste à conserver un marqueur du lieu de production du texte, de la communauté épistémique ou du champs de production du récit que ce soit la presse grand public, la littérature des sciences de gestion ou la littérature scientifique d'informatique et d'ingénierie. Cette démarche questionne implicitement l'historicité des écrits produits sur la RFID. Suivant cette méthodologie, nous avons extrait l'ensemble des articles publiés entre 1989 et 2010 citant « RFID » et « radio-frequency » en ciblant différents publics (grand public, public de sciences de gestion, public d'ingénierie) du fait de l'accès différencié aux bases de données.

Cette approche transversale est rendue difficile par au moins trois barrières. La première barrière est financière: l'accès à ces bases est coûteux et impose des collaborations institutionnelles pour y accéder. La seconde barrière est technique: pour accéder aux données, il convient de créer des robots pour télécharger automatiquement les notices sans pour autant surcharger les serveurs de requêtes ralentissant leurs services. Enfin, la troisième barrière, et non des moindres, est la barrière de la propriété intellectuelle qui limite une diffusion publique des informations aux droits d'auteurs restrictifs. Si ces barrières rendent difficile la recherche, en revanche, elles

justifient d'autant plus la spécialisation des champs de recherche. Si ce terrain n'était pas aussi difficile à constituer la valeur heuristique des résultats serait moindre.

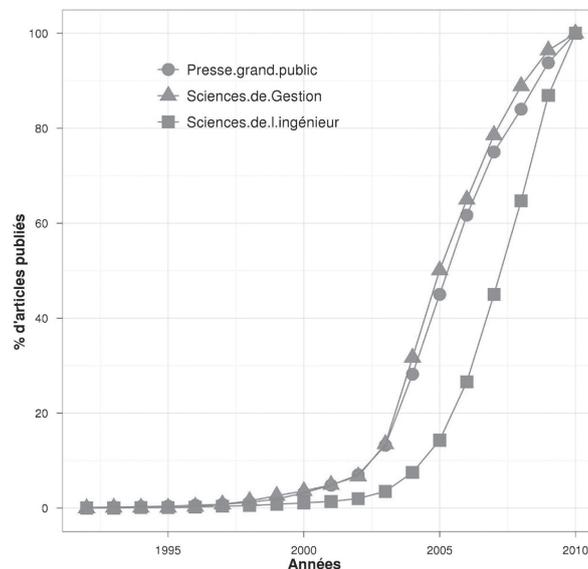


Figure 4 : Historique de l'origine des articles parus sur la RFID entre 1991 et 2010 (selon les bases de données de publication). Légende : Le graphique présente la part d'articles cumulés par année et par base de données. Les bases de données consultées sont: Factiva (Presse Grand Public) ; Ebsco (littérature de Sciences de Gestion); IEEE, Science Direct, Springer, Engineering Village, Siknowledge (littérature des Sciences de l'Ingénieur). Le graphique montre que la littérature des Sciences de Gestion publie (massivement) en premier sur le sujet RFID. Cette littérature est suivie de près par la Presse Grand Public. Enfin, la littérature des sciences de gestion publie massivement 2 ans après la Presse Grand Public. (Source: Notre élaboration à partir de l'exploitation des bases de données précédemment citées.)

Le comptage statistique normalisé permet de mettre en exergue les mouvements des flux d'informations pour répondre à deux questions de recherche fondamentale.

Dans la presse grand public, dans la littérature des sciences de gestion ou dans celle des sciences informatiques et de l'ingénierie quels auteurs s'emparent en premier, et massivement, de la chose RFID ? Quelle est la succession des problèmes publics évoqués par champ de production du savoir ?

Le sens commun pense que la science précède les applications qui précèdent le débat dans les « média de masse ». Or nous aboutissons à un tout autre résultat, contre intuitif.

En réalisant l'extraction dans les bases de données mentionnées, le corpus se compose de 106 607 documents publiés de 1989 à 2010. Ces documents sont des articles, des rapports ou des livres. Pour faciliter l'analyse, seules les notices ont été prélevées. Ces notices ne comportent pas le corps du texte mais le titre, le nom des auteurs, l'année de parution, l'éditeur, le résumé et éventuellement les mots-clés. Ce corpus se décompose ainsi en deux grandes parties : 51 097 articles de presse scientifique (dont la littérature de sciences de gestion et d'ingénierie) et 55 510 articles de presse grand public dont 51 433 articles anglophones et 4 077 francophones.

La temporalité du traitement du sujet RFID selon les bases de données apparaît de manière diachronique et révèle les publics identifiés dans leur historicité (Figure 5). La chose RFID est d'abord investie par les sciences de gestion qui y trouvent des applications pour gagner en efficacité dans leurs chaînes logistiques. Elle est relayée ensuite par la presse grand public pour enfin être étudiée par les publics scientifiques et spécialisés en informatique et en ingénierie¹.

Une exploration qualitative des données, accompagnée d'entretiens d'explicitation des résultats avec des promoteurs de la RFID, conduit à interpréter ces courbes en trois temps dialectiques comme nous le précisons dans la suite de cet ouvrage:

1. L'application logistique de la technologie contribue à la production d'une espérance gestionnaire en la technologie (RFID) et à la construction d'un imaginaire prophétique et économique;
2. Au moment du déploiement de la technologie dans la grande distribution apparaît une réaction dans la presse quotidienne, celle de la dénonciation des risques (que nous précisons plus loin). Le problème public se construit pour se déployer dans une controverse tandis que de nouveaux risques autour de cette technologie se révèlent;
3. Les scientifiques s'affairent à construire une technologie viable socialement par son encadrement réglementaire. Évidemment, ils connaissaient déjà la communication par radio-fréquences; ils ne l'ont pas découverte avec le débat public. Mais ce qu'ils ont découvert, c'est une nouvelle pratique de la science et de la technologie.

Une nouvelle fois, et d'une manière différente, le seuil d'inflexion du débat apparaît en 2005/2006, cette fois-ci par une contextualisation infra-textuelle.

5. Conclusion

Entre les espoirs d'une promesse économique qu'elle procure et l'appel à la responsabilité qu'elle provoque, le débat public se poursuit et, à ce jour, la technologie RFID n'apparaît pas encore socialement acceptable. L'analyse du débat public par l'analyse statistique textuelle permet de mettre en valeur les spécificités chronologiques du seuil de basculement (2005/2006) à partir duquel la critique brise une dynamique propre à l'espérance du champ économique qui tendait à imaginer les potentialités de la technologie.

Cette communication met donc en valeur trois référentiels pour le déploiement de l'analyse et le cheminement de l'enquête statistique textuelle. L'analyse des spécificités textuelles sur les séries chronologiques permet de mettre en valeur une première représentation de la dynamique du débat public. Il s'agit du référentiel inter-textuel. La corrélation observée ne peut avoir valeur d'explication, c'est pourquoi l'enquête se poursuit. L'approche *infra-textuelle* oriente l'analyse vers la dynamique du débat public au sein du corpus lui-même tandis que la l'analyse *supra-textuelle* questionne l'appropriation diachronique du terme RFID au-delà du corpus initial c'est-à-dire par les différents champs d'action économiques, de presse grand public et des sciences de l'ingénieur. La correspondance des interprétations de ces 3 référentiels permet alors d'évaluer l'effet d'une critique sur l'évolution d'un débat public.

1 Toutes les courbes semblent avoir la même allure, ce qui laisse supposer une même équation d'évolution. La normalisation crée de la consistance et de la cohérence, mais n'induit en rien le comportement d'une courbe. Nous remercions Serge Galam pour son apport en rigueur et créativité mathématiques à ce stade de notre étude.

Références

- Assogba, Y. (2005). *La sociologie est-elle une science ? Entretien avec Raymond Boudon et systématisation de la démarche d'explication en sociologie*. Hors-Collection
- Becker, H. (1963). *Outsiders: Studies in the Sociology of Deviance*. New York: The Free Press.
- Boltanski, L. and Thévenot, L. (1991). *De la justification. Les économies de la grandeur*. Gallimard.
- Boltanski, L., Darré, Y., and Schiltz, M.-A. (1984). La dénonciation. *Actes de la recherche en sciences sociales*, 51(1):3–40.
- Brugidou, M. (2008). *L'opinion et ses publics. Une approche pragmatiste de l'opinion publique*. Presse de la fondation nationale des Sciences Politiques, Paris.
- Cefai, D. and Pasquier, D. (2003). *Les sens du public. Publics politiques, publics médiatiques*. PUF.
- Chateauraynaud, F. and Tornay, D. (1999). *Les sombres précurseurs. Une sociologie pragmatique de l'alerte et du risque*. École des Hautes Études en sciences sociales, Paris.
- Dewey, J. (1927). *Le public et ses problèmes (1927)*. Folio Essais. Editions Gallimard.
- Draetta, L. and Delanoë, A. (2012). *RFID : une technologie controversée. Ethnographie de la construction sociale du risque*. Hermes Science.
- Gusfield, J.-R. (1981). *Drinking-driving and the symbolic order. The culture of public problems*. The University of Chicago Press.
- Lebart, L. and Salem, A. (1994). *Statistique Textuelle*. Dunod.
- Salem, A. (1988). Approches du temps lexical, statistique textuelle et séries chronologiques. *Mots*, 17.
- Salem, A. (1994). La lexicométrie chronologique. L'exemple du père duchesne d'hébert. In *Langages de la révolution*, pages 1770–1815. 4ème Colloque de lexicologie politique.