

## La production et le partage du savoir scientifique sur internet : analyse du discours médiatique sur "l'ordinateur quantique"

### Introduction :

Depuis quelques années, nous entendons de plus en plus parler du fameux « ordinateur quantique » aux capacités surprenantes. Il semble souvent se rapprocher de l'un de ces mots « à la mode » dont beaucoup de gens parlent sans vraiment savoir de quoi il s'agit. Et pour cause, son fonctionnement n'est pas des plus aisé à saisir, particulièrement pour des personnes n'ayant pas reçu de formation à certains principes scientifiques très techniques. L'informatique quantique est une science à la croisée de l'informatique et de la mécanique quantique, c'est à dire la physique des particules. Il s'agit d'un ordinateur qui se base sur les propriétés quantiques de la matière. Pour rappel, l'ordinateur classique est construit en fonction de « bits », les unités de mesure les plus basiques qui soient, pouvant prendre 2 valeurs : le 1 ou le 0. L'informatique que nous connaissons est correspond donc à une énumération de 1 et de 0 qui dicte son comportement. L'ordinateur quantique, bien que fonctionnant toujours sous le système binaire de 1 et de 0 ne fonctionne pas en fonction de bit mais de « qubits ». Le qubit est donc la plus petite unité de mesure de l'ordinateur quantique et évolue à une échelle microscopique. Du fait de l'état « quantique » de la matière à cette échelle, le qubit répond au principe de superposition qui fait qu'une particule peut se trouver à deux endroits au même moment. On ne parle alors plus de « suite de un et de zéro » mais d'une superposition de ces uns et zéros. En d'autres termes, ils sont autant un que zéro<sup>1</sup>. Du fait de cet état superposé, les capacités des ordinateurs quantiques seraient alors décuplées.

Bien évidemment, la description ci-dessus correspond à une grossière vulgarisation et ne reflète peut être pas la réelle nature de l'ordinateur quantique. Cependant, le but de notre travail n'est pas de présenter un exposé du fonctionnement de cette technologie, ni même faire un état des lieux de l'avancement de la recherche dans ce domaine. En réalité, au même titre que les algorithmes, l'intelligence artificielle où encore les crypto-monnaies, la production d'articles touchant à l'ordinateur quantique connaît une profusion de plus en plus forte et de nombreux champs de la recherche s'y intéressent. De ce fait, il se place comme un phénomène parfait pour effectuer un bilan des manières de communiquer sur un sujet entre les différents acteurs-producteurs qui peuplent le web.

Malgré la difficulté qu'entraîne sa compréhension, il semble que les articles sur le sujet soient très populaires. On peut évidemment imputer cela au fonctionnement même d'internet, qui permet un accès à n'importe quel type d'information, même les plus obscures. L'ordinateur quantique n'est évidemment pas la seule notion technique dont la documentation est rendue plus facilement accessible grâce à internet : nous aurions très bien pu nous intéresser à n'importe quel terme à la mode comme ceux que nous avons cités précédemment. Cependant, ce sujet spécifique est particulièrement intéressant à traiter dans la mesure où il « englobe » les enjeux de nombreux autres termes tout aussi techniques. Il se place dans les discours comme un atout majeur, plein de promesses au regard de différents domaines que constituent l'IA, l'algorithmique, les crypto-monnaies ainsi que pour d'autres domaines moins spécifiques tels que la sécurité informatique, l'automobile, la santé etc. Il s'agit finalement d'un sujet qui, s'il se concrétise, pourrait remettre en question la manière dont les individus perçoivent et manipulent l'informatique et les technologies qui y sont liées. Si l'on s'intéresse à l'avenir des sciences et techniques, à en croire les discours

---

1 Nous nous sommes basés sur la page Wikipédia du calculateur quantique afin d'élaborer cette définition. Cela pour deux raisons : d'une part il semblait nécessaire de ne pas utiliser de site faisant parti du corpus que nous allons traiter et d'autre part parce que Wikipédia transcrit une parole élaborée en commun et donc assez objective sur la question.

actuels sur le net, l'ordinateur quantique est une pierre angulaire d'un réseau qui relierait de très nombreuses thématiques et domaines différents.

Comment se construit l'explication d'un sujet de niche pour le grand public sur le net ? Concrètement, nous nous demanderons si il existe des types de sites qui s'expriment plus ou moins sur les ordinateurs quantiques et nous observerons la manière dont ils le font.

Le corpus que nous allons étudier correspond à un ensemble d'articles – scientifiques, presse ou blogs – récupérés sur internet et portant sur l'ordinateur quantique. Nous ajoutons à ce corpus des données relatives aux sites qui hébergeaient les articles en question et nous les catégorisons en fonction de plusieurs caractéristiques. Les articles portent sur des champs du monde social et professionnel bien différents et sont tous réunis par cet intérêt pour l'ordinateur quantique.

Nous supposons comme point de départ que les discours sur les ordinateurs quantiques sont plus ou moins construits par chaque type d'acteurs dans la mesure des intérêts personnels qu'ils peuvent trouver dans leur domaine propre. Ainsi, chaque site aura tendance à s'approprier un vocabulaire concernant l'ordinateur quantique qui sera en cohérence avec les sujets qui l'intéressent particulièrement. Cette hypothèse suppose que l'individu cherchant à se renseigner sur l'ordinateur quantique se retrouvera face à des types de discours différents en fonction de l'orientation principale du site sur lequel il se trouve.

Afin de vérifier cette affirmation, nous développerons une analyse en 3 étapes. Dans un premier temps, nous décrirons la méthodologie que nous utiliserons au cours de notre étude et surtout le regard que nous allons porter sur les articles et les sites internet. Nous tenterons d'éviter du mieux que l'on peut le biais du « chercheur déconnecté de sa population » en procédant par mimétisme de recherche.

Ensuite, nous procéderons à une analyse cartographique des sites internet pour comprendre la manière dont ils s'articulent sur le web. Nous décrirons avec plus de précision les caractéristiques du corpus et effectuerons un premier profilage des sites qui fournissent ce type d'information.

Enfin, la troisième partie portera sur une analyse des discours entourant les ordinateurs quantiques. Nous nous demanderons ici comment le sujet est traité en fonction des profils que nous aurons tiré précédemment et si l'on peut identifier des « tendances de communication » ou des tendances sémantiques pertinentes.

## **Méthodologie :**

Le principe de cette étude est de comprendre en partie comment se traduisent les processus de recherche de l'information technique à l'ère de l'internet. Dans un monde où l'accessibilité à l'information scientifique est favorisée par la prolifération de sites internet spécialisés, on peut se demander si la recherche personnelle sera structurée par les sites eux-mêmes ? Comment les processus de recherche se traduisent-ils dans cet écosystème de prolifération d'information ?

### 1/ Le regard de « l'utilisateur lambda »

Nous n'aborderons pas la recherche d'information d'un point de vue de chercheur mais plutôt du point de vue d'un utilisateur « moyen » que nous allons essayer de décrire. Cet utilisateur réunit certaines caractéristiques : il n'est pas particulièrement scientifique mais n'est pas non plus hermétique à la rédaction scientifique. Il est tout autant capable de lire un article scientifique qu'un article de presse et va même tenter de diversifier ses connaissances et sa vision en cherchant ces deux types d'éléments. Cet individu ne cherche pas non plus à privilégier un type de canal d'information particulier et va plutôt passer de sites en sites afin de réunir un panel d'information. Lors de sa recherche, il explore donc des blogs, des sites de presse, des sites scientifiques, des sites publiant des podcasts ou des vidéos, ou encore, des sites très généraux de définitions.

Imaginons à présent la manière dont cet individu effectue ses recherches : nous partirons du principe que l'individu moyen utilise le moteur de recherche Google pour sa recherche. En effet, selon des statistiques établies par le site « Le blog du modérateur », les parts de marché du moteur de recherche Google s'élèvent à 94.1 % pour la France<sup>2</sup> (3,4% pour Bing et 1,8% pour Yahoo!). Cet utilisateur appartient à ce que l'on appelle le « grand public », il n'est pas du tout spécialiste de la thématique de l'ordinateur quantique. Nous nous mettrons donc à la place de cet individu qui, n'y connaissant rien, souhaite réunir des informations sur le sujet.

Nous récupérerons donc les articles en nous basant sur le moteur de recherche de Google. Nous partons dans un premier temps du postulat qu'une recherche générale commence avec des mots clés très généraux. Le premier mot clé qui servira à la récupération des sites est donc simplement « ordinateur quantique » sans aucun ajout d'autres termes. Afin de diversifier les articles et les sites que nous réunissons, il est nécessaire de ne pas se laisser enfermer par les premiers résultats renvoyés par l'algorithme de Google et aller voir aussi les sites référencés dans les pages suivantes. En ne nous limitant pas à la page des « plus populaires », nous tentons donc de limiter le biais du *page rank* en récupérant les articles apparaissant sur les 5 premières pages de résultats.

Évidemment, ces résultats ne suffisent pas à caractériser l'ensemble des types de recherches qui peuvent être effectués par des néophytes du sujet. Bien que nous considérons qu'il s'agit d'individus n'ayant aucune notion de l'ordinateur quantique, il est logique de penser que beaucoup s'y intéressent parce qu'ils en ont entendu parler dans un contexte particulier. De ce fait, la recherche sera d'emblée « orientée » par un élément extérieur ayant entraîné l'intérêt de l'individu. Afin de reproduire le processus d'orientation de la recherche et que les résultats correspondent à une certaine réalité, nous décidons de nous laisser guider par les mots clés les plus souvent associés au terme « ordinateur quantique » sur Google. Nous avons donc procédé de manière basique et avons écrit dans le moteur de recherche le terme « ordinateur quantique » accompagné d'un espace afin de voir les termes qui apparaissent en premier. Nous procédons ensuite de la même manière pour nos recherches avec ces nouveaux mots clés (récupération des 5 premières pages en prenant soin de ne pas prendre en compte les doublons). Les mots clés que nous avons croisés avec « ordinateur quantique » sont donc : « prix », « Google », « IBM », « explication », « pour les nuls ». Nous décidons de nous arrêter au 5<sup>e</sup> mot clé apparaissant car, à partir de ce cinquième mot,

---

2 <https://www.blogdumoderateur.com/chiffres-google/>

les sites et articles que nous trouvons par cette recherche ont quasiment tous déjà été référencé à partir des mots clés précédents. À partir de ce 5<sup>e</sup> mot, nous avons en quelque sorte fait le tour des possibilités d'articles sur lesquels notre individu moyen peut tomber lorsqu'il effectue une recherche basique.

## 2/ Cartographier le web : une richesse d'information

Nous décidons d'appréhender les résultats en fonction de deux approches différentes : d'une part nous observerons l'aspect « réseau » relationnel entre les différents éléments du corpus et d'autre part, un aspect plus marqué par « l'individuel » à travers une analyse des manière d'évoquer une thématique spécifique. De nombreuses problématiques ont été rencontrées au cours de la récupération et de la mise en forme des données. Ces problématiques étaient principalement d'ordre techniques, même si certains biais d'analyse peuvent être notés pour la partie portant sur l'analyse sémantique.

Dans un premier temps, nous avons utilisé le logiciel Hyphe Browser, développé par le Medialab de Science-po afin de cartographier la manière dont les sites se placent les uns par rapport aux autres sur internet. N'ayant pas accès au serveur du Medialab, nous avons utilisé la version « démo » de Hyphe, qui ne permet qu'une profondeur de 1 dans la récupération du réseau de liens des sites. Cela signifie que chacun des liens sont des liens directs : le logiciel ne crée un lien entre deux sites uniquement si le site d'arrivée est directement et explicitement cité par le site de départ. Pour simplifier, il ne va pas créer « d'individu supplémentaire » si il trouve un lien hypertexte menant à un site qui n'était pas référencé dans le corpus de base. Dans le cadre de notre étude, nous n'avons pas vraiment besoin de plus dans la mesure où nous souhaitons uniquement observer les relations entre les sites internes au corpus et non pas les relations qu'ils peuvent avoir avec des sites extérieurs.

Ensuite, la méthode de récupération avait initialement été prévue à l'aide d'un script basé sur le langage Python et qui aurait permis une récupération automatisée des différents articles internet. Cependant, par la structure très différente des pages internet, l'écriture d'un tel script ainsi que les compétences actuelles dont nous disposions aurait demandé un travail immense et aurait coûté bien plus de temps qu'il n'en aurait fait gagner. En effet, si la récupération des données d'une page HTML est facilement réalisable, la récupération d'informations

```
<!DOCTYPE html>
<html class="no-js" prefix="og: http://ogp.me/ns# fb: http://ogp.me/ns/fb#" lang="fr"
<#head>
</head>
<body class="post-template-default single single-post postid-10846 single-format-st
<#header id="header" class="gridlove-site-header hidden-md-down gridlove-header-sh
  <div class="gridlove-header-4">
    <div class="gridlove-header-wrapper">
      <div class="gridlove-header-bottom">
        <nav class="gridlove-slot-c">
          <div class="gridlove-main-navigation">
            <ul class="gridlove-actions gridlove-menu">
              </ul>
            </div>
          </div>
        </div>
      </div>
    </div>
  <div id="gridlove-header-responsive" class="gridlove-header-responsive hidden-lg-
  <div id="cover" class="gridlove-cover-area gridlove-cover-single">
  <div id="content" class="gridlove-site-content container gridlove-sidebar-right">
    <div class="row">
      <div class="gridlove-content gridlove-single-layout-7">
        <div class="gridlove-sidebar">
          <div id="mks_author_widget-2" class="widget gridlove-box mks_author_widget">
            <div id="fbw_id-2" class="widget gridlove-box widget_fbw_id">
            <div id="gridlove_posts_widget-2" class="widget gridlove-box gridlove_posts_
            <div id="gridlove_category_widget-2" class="widget gridlove-box gridlove_cat
          </div>
        </div>
      </div>
    </div>
  </div>
</body>
</html>
```

```
<!DOCTYPE html>
<!--
<html lang="fr" itemscope itemtype="http://schema.org/NewsArticle" itemid="https://a
quantique--100-millions-de-fois-plus-rapide-qu-un-ordinateur-classique.htm" xmlns:co
-->
<html class="pw-locale-fr-ra1-pw-desktop" xmlns:fb="http://www.facebook.com/2008/fb
<!--s#dvD#head-->
<#img#>
<div id="img7" style="position: absolute; height: 0px; width: 0px;" src="https://
32C322curr_pnge_cnt%22%3A1%7D%7D">
<script type="text/javascript" src="//cdn.tradelab.fr/seq.js?add=653753"></script>
<script type="text/javascript" src="//cdn.tradelab.fr/seq.js?add=946931"></script>
<script type="text/javascript">
  if ( typeof extra_info === 'undefined') var extra_info = {'order_id':''}; if (t
  injectScript(elem, function(){}); }
</script>
<script type="text/javascript">
  if ( typeof tldc === 'object' ) { if ( typeof tldc.vis === 'object' ) { if ( t
  function(){}); } }
</script>
<script type="text/javascript">
</script>
<script type="text/javascript">
  if ( typeof tldc === 'object' ) { if ( typeof tldc.vis === 'object' ) { if ( t
  function(){}); } }
</script>
<script type="text/javascript" src="//cdn.tradelab.fr/conv/723451.js"></script>
<script type="text/javascript" src="//cdn.tradelab.fr/conv/629913.js"></script>
<script type="text/javascript" src="//cdn.tradelab.fr/conv/629914.js"></script>
<script type="text/javascript">
  <script type="text/javascript">
  <script type="text/javascript" src="//ib.adnxs.com/getuid?/its.tradelab.fr/7tvp
  <script type="text/javascript" async="" src="//asqlesechos.nuqqad.net/rc?nuqqe8
  </script>
<script async="" src="https://connect.facebook.net/en_US/fbevents.js"></script>
html.pw-locale-fr-ra1-pw-desktop head script
```

contenues dans pleins de balises différentes se révèle un processus difficile à automatiser. Le principal problème vient du fait que l'organisation d'internet n'est absolument pas homogène. Chaque site internet possède sa propre structure, sa propre manière d'agencer les informations, ses propres balises. Il était alors difficile d'imaginer un processus d'automatisation capable de reconnaître dans chaque cas la balise HTML contenant le texte de l'article. C'est cette organisation déstructurée qui fait d'internet l'un des espaces les plus riches à l'étude mais aussi l'un des plus complexes à analyser. Afin d'éviter une perte de temps inutile et compte tenu de la taille relativement faible du corpus que nous comptons construire (environ une centaine

d'articles), nous avons décidé de récupérer le contenu des articles à la main à l'aide de la fonction copier/coller.

Une fois les données récupérées sur des pages textes, nous avons construit un nouveau document dans lequel nous n'avons conservé que le contenu de l'article, en prenant soin de supprimer des informations telles que les noms des auteurs de l'article, les titres, les images ou tout ce qui pouvait brouiller l'analyse sémantique des discours.

### 3/ L'information sur le web : pluralité des formes

Il faut à présent évoquer un autre type de problématique qui ne se rapporte pas vraiment à un aspect technique à savoir le fait que tous les sites qui ont été récupérés n'ont pas pu être gardés dans le cadre de l'analyse textuelle effectuée en 3<sup>e</sup> partie. En raison de la diversité des résultats récupérés au cours de la recherche<sup>3</sup>, certains sites ne se prêtaient pas au jeu de l'analyse textuelle du fait d'être des podcasts, des liens vers des vidéos ou tout simplement dans certains cas des articles ou dépêches beaucoup trop courts pour être analysés au même titre que les autres.

Il y a donc une différence entre les sites ayant servi à l'analyse textuelle et les sites cartographiant l'écosystème internet des ordinateurs quantiques. Les raisons pour lesquelles nous avons décidé de retirer un site de l'analyse textuelle ou de la cartographie sont les suivantes :

Tout d'abord, nous avons retiré des sites beaucoup trop généraux qui venaient en quelque sorte « biaiser » l'analyse en concentrant la quasi totalité des liens. Selon cette logique, nous avons donc supprimé de la cartographie les sites Wikipédia et Youtube. Wikipédia a été exclu pour deux autres raisons : d'une part nous nous en sommes servis afin de définir les termes du sujet, et d'autre part parce qu'il ne s'agit pas d'un article de presse mais d'un article collaboratif avec un statut trop différents du reste du corpus. De même, nous avons éliminé le site de la Fnac, ainsi que les autres sites qui correspondaient à des sites de vente en ligne d'un produit<sup>4</sup>. Si ces sites ont été retirés, c'est parce qu'ils n'avaient pas une vocation informative et de ce fait, l'individu n'y reste généralement pas longtemps, ou alors pas dans une optique d'apprentissage d'un sujet.

Mis à part ces éléments, la quasi-totalité des sites a été gardé pour la cartographie car quel que soit le contenu de la page, il s'agit d'une information sur lequel notre individu moyen peut tomber au cours de sa recherche et qu'il peut utiliser pour se renseigner. C'est pour cette raison que nous avons conservé les sites de forums comme jeuxvidéos.com. Les forums constituent encore aujourd'hui une source de partage et d'information importante pour les individus. Bien que rattrapés pendant un temps par l'essor des réseaux sociaux, ils constituent un dispositif ancien et fortement ancré dans les pratiques. En parallèle au ralentissement de réseaux comme Facebook depuis quelques années, l'utilisation des forums semble connaître un renouveau, notamment chez les adolescents<sup>5</sup>.

En ce qui concerne l'analyse textuelle, on peut dire pour résumer que tous les sites de podcasts (vidéo ou audio) ont été supprimé du corpus. De la même manière, les sites ne fournissant qu'un article incomplet ou alors une liste de plusieurs articles ont été ignorés. Nous nous sommes ensuite débarrassé des liens qui correspondaient à des fichiers PDF (pour un sujet très scientifique comme l'ordinateur quantique, nous en avons croisé à plusieurs reprises) ainsi que les forums du fait de la difficultés d'analyser une prise de parole libre et une multitude de commentaires au même titre que des articles rédigés. Enfin, nous avons éliminé les sites qui étaient des doublons d'autres sites et les sites qui étaient des cours entiers et complets. Pour ce dernier cas, nous estimons qu'un individu qui veut se renseigner ne va pas se plonger dans un cours mais plutôt piocher les informations à droite et à gauche pour avoir une information complète. Ces phénomènes de

---

3 Rappelons que cette recherche a été effectuée en se mettant à la place d'individus lambda, qui ne cherchent pas forcément à cibler les informations. Nous avons *scrapé* ces pages web au « lance pierre » sans nous soucier dans un premier temps de ce qu'elles contenaient.

4 la plupart du temps un livre portant sur l'ordinateur, l'informatique ou la physique quantique

5 CORROY, Laurence. La culture médiatique des jeunes : des objets populaires méprisés ?, *Jeunes et médias - Les Cahiers francophones de l'éducation aux médias* - 2015, n°1

recherche « au hasard » sur internet ont d'ailleurs souvent été étudiés par les sciences sociales. Nous pourrions assimiler ce processus à une forme de sérendipité : l'utilisateur « erre » de site en site afin de réunir ses informations et fini par trouver plus que ce qu'il ne cherchait à la base. Les moteurs de recherche ainsi que leurs options d'auto-complétion permettent alors à l'internaute de préciser une recherche qu'il a parfois lui même du mal à formuler<sup>6</sup>. En d'autres termes, ces méthodes permettent aux utilisateurs de « trouver en cherchant à côté ».

---

6 Ghislaine Gallenga et Gilles Raveneau, « Dynamiques temporelles et sérendipité dans les recherches contemporaines », *Temporalités* [En ligne], 24 | 2016, mis en ligne le 29 mars 2017

## Cartographie des acteurs-producteurs

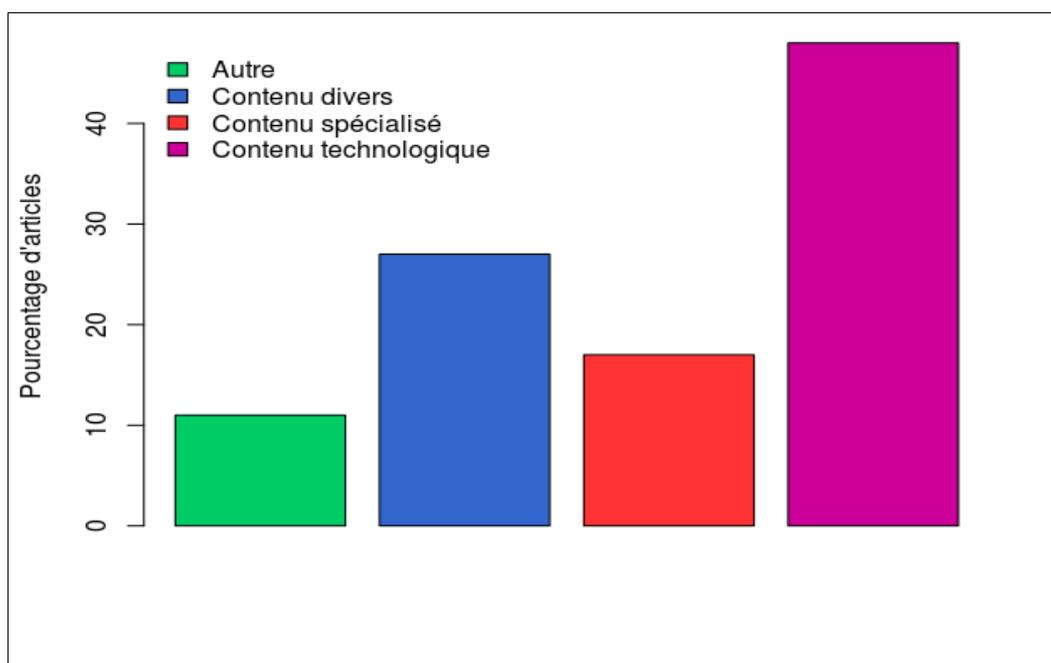
Rentrons à présent dans le vif du sujet à travers l'établissement d'un premier état des lieux sur notre corpus. Il faut se demander comment nous pouvons cartographier ces espaces d'information. De quelle manière circule l'information en fonction des caractéristiques de chaque site. S'agit-il d'un cercle fermé, dans le sens où il y aurait un fort taux de citation des sites entre eux ? Ou alors au contraire les sites évoquant la thématique de l'ordinateur quantique possèdent-ils une forte indépendance et ne semblent pas vouloir « s'affilier » à d'autres acteurs ?

### 1/ Des profils identifiés : une forte activité des acteurs liés au numérique

Le corpus que nous conservons pour la partie cartographique est donc composé de 103 sites internet, ayant chacun leur domaine de prédilection. Il ne s'agit pas seulement de sites d'actualités technologique même si ce domaine arrive évidemment en tête. Afin de simplifier la compréhension des différents types de sites, nous les avons réparti en fonction de 5 catégories ou « types » de contenu :

- Contenu technologique
- Contenu divers (pour parler des sites sans thématique principale, il s'agit ici principalement de sites d'information et d'article de presse)
- Contenu spécialisé (cette catégorie réunit tous les types de contenus qui sont tournés vers un domaine spécifique autre que le domaine technologique) : nous rangeons par exemple dans cette catégories les articles issus de sites principalement tournés vers l'économie, la finance, les automobiles ou encore tout type de sites scientifiques dont l'intérêt principal ne serait pas les nouvelles technologies).
- Contenu de type « autre » : nous réunissons ici le reste des sites, qui se caractérise par une originalité dans la manière de structurer le discours. Il s'agit pour la plupart de sites de blogs ou de forums, ce que nous ne pouvons donc pas placer dans les mêmes catégories que les articles de presse ou scientifique.

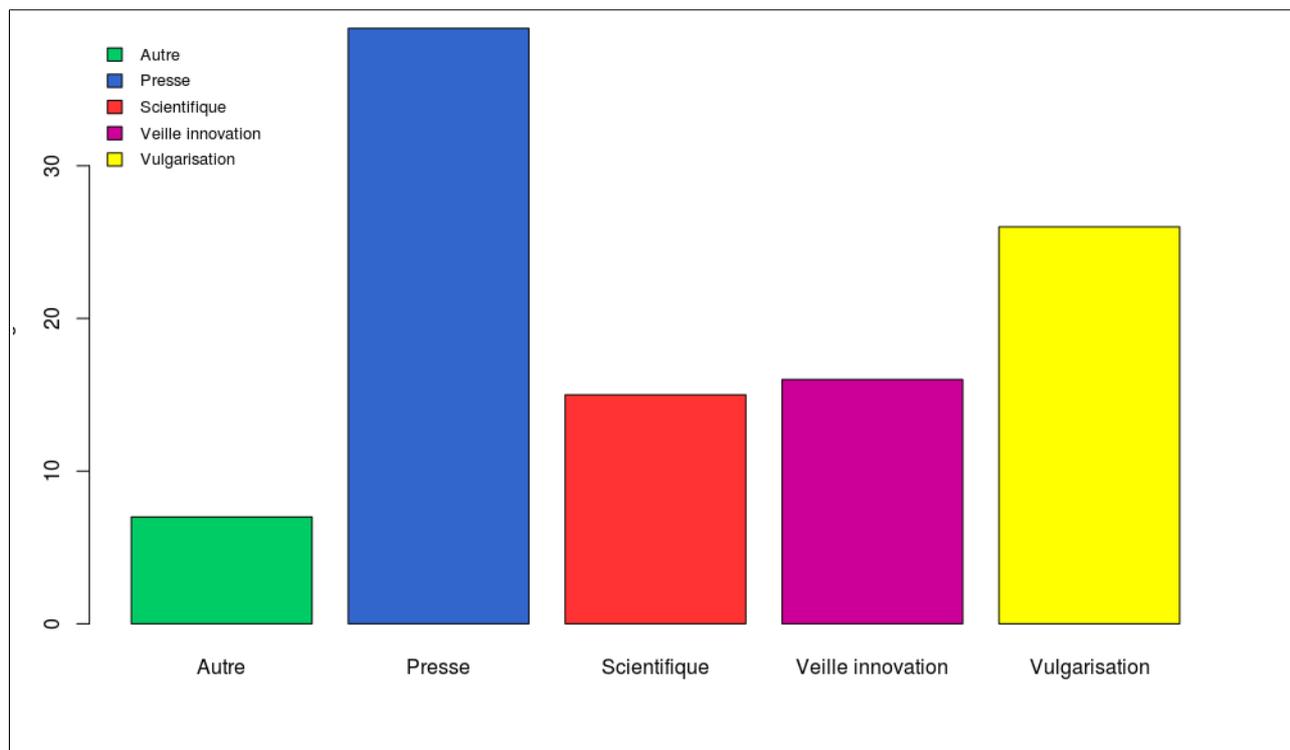
Figure 1 : Répartition des sites du corpus en fonction de leur contenu



Ce sont surtout des sites tournés vers l'innovation technologique qui s'intéressent à la thématique des ordinateurs quantique avec 46 % du corpus composé de ce type de site internet. Néanmoins, environ un quart du corpus a en réalité une activité très diverse, ce qui n'est pas négligeable si l'on considère le caractère très pointu que revêt l'ordinateur quantique. De plus, une part tout de même importante de sites spécialisés s'y intéressent. Cela s'explique par les grandes espérances que portent les ordinateurs quantiques. En effet, à travers la révolution informatique qu'il laisse entrevoir, la plupart des domaines scientifiques et professionnels risquent de se trouver fondamentalement transformés par l'arrivée d'un tel outil. Les discours scientifiques – hors de ce corpus évidemment – appuient sur les capacités imaginées des ordinateurs quantiques. Leur rapidité de calcul dépasserai largement les systèmes informatisés d'aujourd'hui. Qu'il s'agisse du domaine de la santé dans la mise en place des calculs de détection des maladies, de la finance pour la manipulation rapide d'actions ou encore de la sécurité informatique avec les possibilité de décryptage quasi-instantané des clés les plus perfectionnées, l'ordinateur quantique a toute les raisons d'intéresser à tous les niveaux.

Nous avons caractérisés les sites en fonction de plusieurs modalités permettant de déterminer son « style » de communication. Nous nous demandons par exemple si l'article contient des photos pour accompagner son contenu. De même, si il possède une vidéos, nous évaluons la taille de l'article en termes de mots, nous déterminons si l'article est accompagné d'un schéma pour appuyer son explication<sup>7</sup>. Enfin, nous regardons si l'article possède un auteur défini où si il est anonymisé. Nous considérons par exemple que si un article ne possède pas d'auteur, il aura moins de renommé ou de crédibilité.

Figure 2 : Répartition des articles en fonction de leur thématique principale



La figure 2 montre que l'essentiel de notre corpus est composé d'articles de presse. Dans une moindre mesure, on retrouve une certaine quantité d'articles de vulgarisation, qui ont tout de même vocation à expliquer un aspect plus poussé que de simples articles de presse. La « veille innovation » est composée d'articles courts, ayant surtout pour objectif de faire un résumé spécialisé

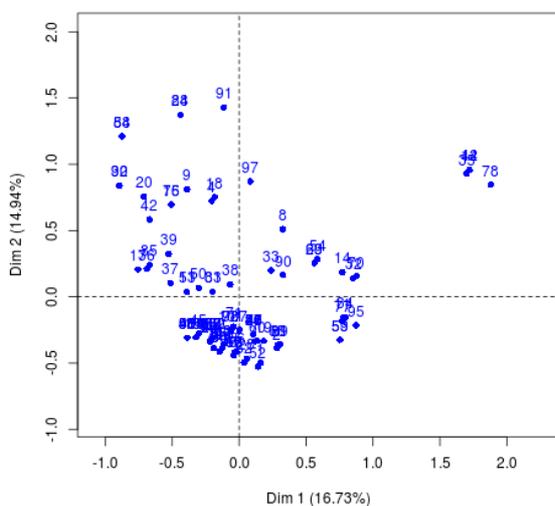
<sup>7</sup> Par « schéma » nous entendons une image différente d'une simple photo illustrative : une image ayant pour objectif d'illustrer une notion.

sur un certain domaine. Les articles scientifiques quant à eux sont généralement assez longs, et ne cherchent pas particulièrement à simplifier l'explication pour le grand public mais plutôt à être le plus complet possible. Enfin, la catégorie « autre » correspond une fois encore aux sites comme les forums ou les blogs, où il s'agit surtout de discussion et où la forme n'est pas structurée comme un article.

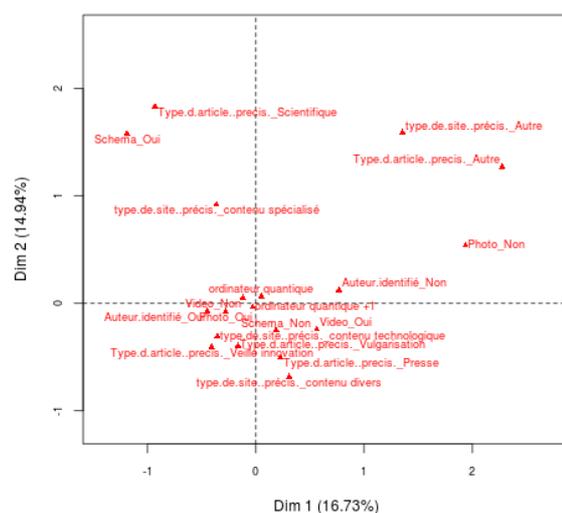
## 2/ similarité et opposition dans les styles d'écriture

Jusqu'à présent, nous avons considéré des catégorisations très subjectives, par observation brutes, dans le sens où l'objectif était de se mettre à la place d'un individu pas particulièrement renseigné sur le sujet. Nous allons à présent procéder à une catégorisation plus objective à l'aide d'une analyse factorielle. En effet, nous allons représenter graphiquement les différents sites internet en fonction de toutes les modalités qu'elles possèdent et déterminer la meilleure classification possible en fonction de ces modalités.

Graphique 1 : Répartition des sites en fonction de leur proximité caractéristique



Graphique 2 : Répartition des caractéristiques en fonction de leur influence



Le graphique 1 représente la répartition des individus (en l'occurrence, les sites internet) entre eux en fonction de leur proximité caractéristique. Plus des points sont proches les uns des autres, plus ils auront des caractéristiques communes. Rappelons que les différentes caractéristiques sont : l'orientation thématique du site internet, le « style » de l'article publié, le fait d'avoir un auteur identifié, le fait de contenir une photo, une vidéo ou encore le fait de contenir un schéma explicatif. La construction de ce « nuage » d'individus se lit en fonction des deux principales dimensions, deux principaux types de profils qui vont orienter la répartition des individus dans l'espace.

La répartition que l'on observe peut donc s'interpréter ainsi : la première dimension influence la répartition des individus à hauteur de 16,73 % et la deuxième dimension à hauteur de 14,94 %. Avec une différence de cet ordre, nous considérons que les deux dimensions ont une influence à peu près similaire sur la répartition des individus. L'enjeu à présent est de caractériser les deux dimensions afin de comprendre quels sont les deux principaux traits spécifiques qui vont caractériser les articles du corpus. Pour cela, il faut regarder la répartition des modalités sur le même graphique :

La figure 4 représente la manière dont les différentes modalités de la base répartissent les individus en fonction de leurs caractéristiques. Commençons par l'analyse de l'axe 1 : sur la droite de l'axe, nous retrouvons des modalités comme le fait de ne pas avoir d'auteur identifié, ne pas avoir de photo pour accompagner l'article, être un article de catégorie « Autre » (donc issu d'un blog ou d'un forum) ou, avec moins d'importance, le fait de comporter une vidéo.

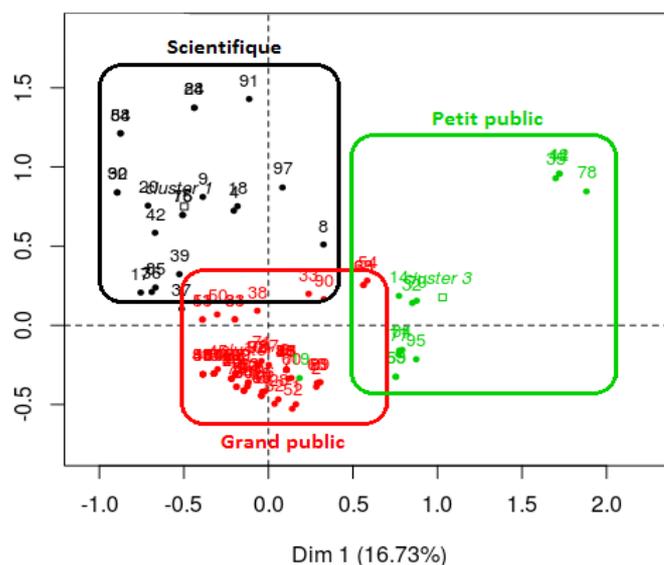
En ce qui concerne l'axe deux, nous voyons que, tout en haut du graphe, les modalités que l'on retrouve sont les types d'articles plutôt scientifiques, le fait de contenir des schéma explicatifs, le fait d'être un site spécialisé dans un domaine. À l'opposé de cet axe, nous retrouvons des types d'articles de vulgarisation, de presse, ainsi que des contenus divers. Nous pouvons difficilement considérer les autres modalités comme jouant significativement sur la répartition des individus dans la mesure où elles sont trop regroupées pour être réellement significatives. Nous pouvons ajouter que, en se concentrant de nouveau sur l'axe 1, le fait de comporter un schéma se place sur la gauche.

Au regard de ces dispositions, les deux axes semblent répartir les articles en fonction de si ils sont des articles privilégiant le fond, la forme ou la technique de rédaction. Expliquons-nous : selon cette interprétation, l'axe 1 correspondrait à un style d'écriture ou de rédaction (type « article » pour la partie gauche, type « déstructuré » pour la partie droite), cela est en cohérence avec le fait de retrouver de ce côté des blogs indépendants ou de la discussion de forum. Précisons tout de même ici que le style d'écriture ne concerne pas du tout le contenu de l'article mais la manière de présenter les arguments. L'analyse du contenu correspond plutôt à l'axe 2 : en effet, toute la partie supérieur de l'axe 2 rassemble des articles de fond, qui se basent sur des informations précises et un désir d'explication très précise. Cela expliquerait pourquoi nous retrouvons des articles de presse et de vulgarisation sur la partie inférieure du graphe.

Afin de résumer ces répartitions, nous pouvons dire que les individus se trouvant sur la partie supérieur gauche du graphique 1 correspondent à des articles plutôt scientifiques avec une vocation à donner des informations très spécifiques et poussées ainsi qu'une logique de transmission d'un réel enseignement. L'amas d'individu que l'on retrouve sur le bas du graphique correspond à des articles plutôt grand public, principalement des articles de presse ou de vulgarisation dont l'objectif n'est pas forcément d'enseigner mais plutôt d'informer. Enfin, les quelques individus que l'on retrouve sur la droite correspondent à des articles rédigés par des personnes anonymes, n'ayant pas de renommé ni de légitimité particulière pour informer (en effet si l'on regarde l'individu extrême numéro 78 il s'agit d'un blog de dénonciation, voir conspirationniste).

Afin de résumer la caractérisation des sites, nous utilisons un algorithme de classification à l'aide du logiciel R. Comme nous l'avions prévu, l'algorithme classe les individus globalement de la même manière que nous l'avions fait. Il détecte donc trois catégories de sites que nous appellerons articles destinés aux scientifiques (en noir sur le graphe suivant), articles destinés aux grand public (en rouge sur le graphe suivant) et articles destinés au « petit public » c'est à dire fait par des citoyens pour des citoyens (en vert sur le graphe suivant).

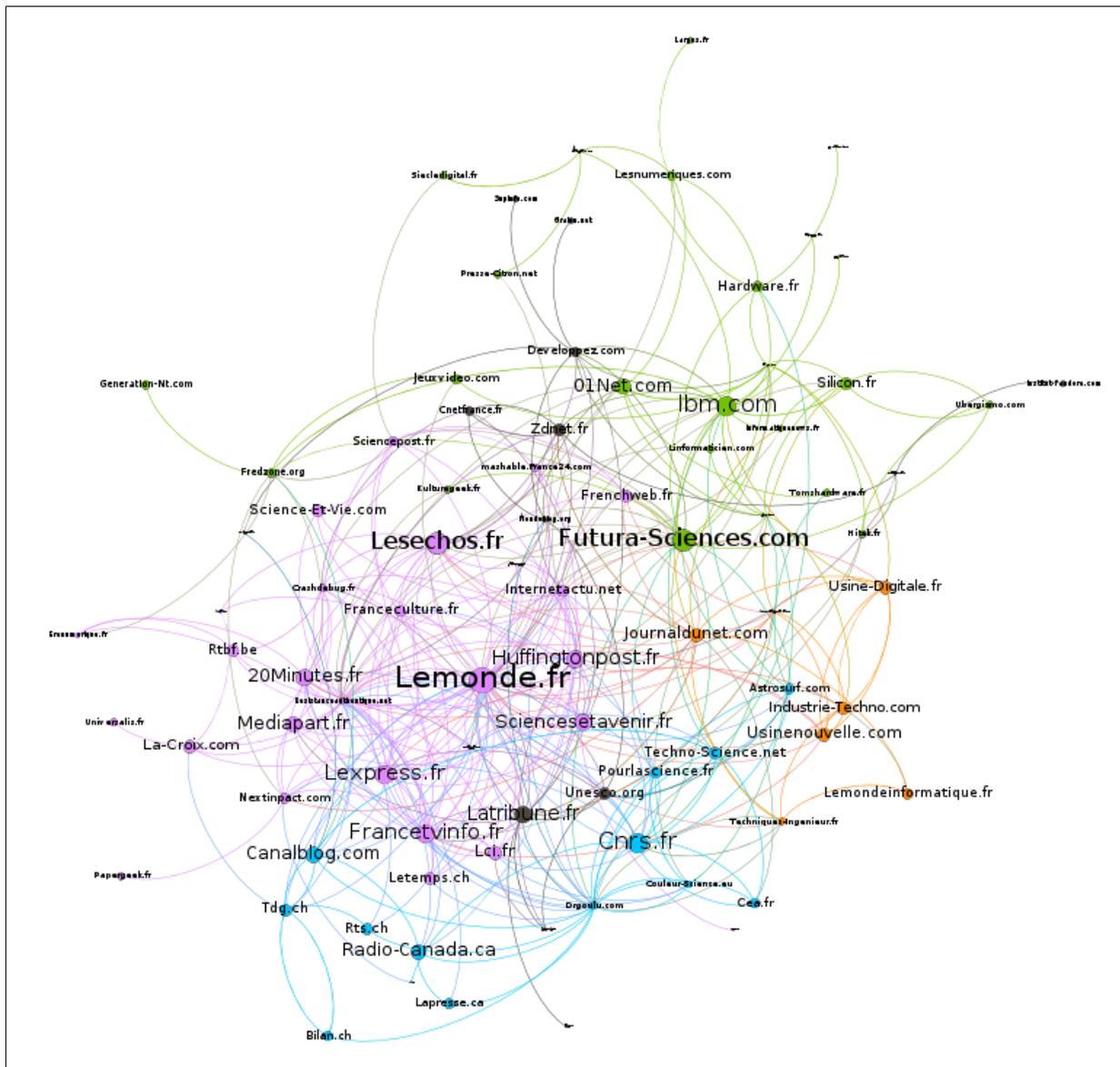
**Figure 5 : Classification des individus en fonction de leurs caractéristiques**



### 3/ Le partage de la communication et de l'information à travers la citation des pairs

Nous allons à présent représenter les relations que les différents sites du corpus ont entre eux. Est ce que la légitimité de la parole intervient lorsque l'on s'exprime sur les ordinateurs quantiques ? Est ce que tout le monde (journalistes, scientifiques, individus anonymes) a les compétences pour parler de ce sujet ? Comment se passe l'échange de l'information entre les sites ?

**Figure 6 : réseau des citations entre les sites internet du corpus**



Les sites du corpus sont très fortement liés entre eux. Les liens qui sont représentés doivent être interprétés comme le fait de trouver sur le site un lien hypertexte menant à un autre site. Un trait entre deux points lorsqu'il évolue dans le sens des aiguilles d'une montre représente la présence d'un lien hypertexte et donc d'une citation. L'analyse d'un tel réseau internet se traduit par le fait que la fréquence des citations sur un tel sujet est très fréquente, ainsi, nous pouvons dire que l'ordinateur quantique est souvent traité de manière assez scientifique, en citant de nombreuses sources. Le sens de citation doit se lire dans le sens des aiguilles d'une montre, ainsi, si la courbe de lien d'un site internet est courbée vers le bas, il faut lire la direction de droite à gauche (le site de

droite cite le site de gauche) et inversement si la courbe de lien est courbée vers le haut. Si nous prenons par exemple le lien entre Canalblog.fr et RadioCanada (en bas à gauche du graphique), nous voyons c'est c'est le site Canalblog qui a cité RadioCanada. En revanche, si l'on regarde le lien entre le site Nextimpact et Papergeek.fr, c'est Nextimpact qui a cité Papergeek : du fait de l'orientation inférieure de la courbe, le sens des aiguille d'une montre correspond à un mouvement de la droite vers la gauche.

Cette représentation nous permet de visualiser quels sont les sites qui ont le plus de poids dans l'écosystème web de l'ordinateur quantique. Les nœuds principaux sont surtout des sites de presse : LeMonde.fr, Lexpress.fr, lesEchos.fr, Huffinghtonpost. Cependant, il y a aussi une part assez important de sites spécialisés qui se retrouvent avec une importance forte. Ainsi, des sites comme FuturaScience.fr ou IBM.com sont parmi les plus souvent cités. Cette représentation nous permet de mettre en lumière un éléments important de la circulation de l'information sur internet : les sites classifiés dans la catégorie « scientifique » se trouvent plutôt vers les extrémités du réseau et donc on moins tendance à citer des sources venant d'internet que les sites de presse. Au contraire, ils seront fortement cités par ces derniers. En effet, nous supposons que ces sites publient en général des articles de fond se basant sur de nombreuses recherches effectuées dans des contextes de recherche académique. Lorsque l'on parle d'article de presse, on imagine tout aussi facilement que la plupart des journalistes n'ont pas tellement le temps d'effectuer de véritables recherches académiques et s'en remettent donc à ces instituts qui font office de référence sur des sujets techniques pointus comme l'ordinateur quantique. Marc Bassoni, Maître de conférences en sciences de l'information et de la communication, décrit d'ailleurs fort bien ce phénomène a l'air du web 2.0 : alors que les plate-formes de discussion et d'information se développent, on voit apparaître de nouvelles formes de co-construction des articles journalistico-scientifique se basant sur le principe de collaboration et de partage de connaissance<sup>8</sup>. Par ailleurs, en ce qui concerne les sites de presse, nous remarquons qu'ils sont souvent aussi cités qu'ils ne citent eux même. À travers cela, nous comprenons qu'un sujet comme l'ordinateur quantique n'appelle pas à s'en remettre à quelques acteurs ayant une autorité sur le sujet. Tout le monde s'y intéressent en se citant les uns les autres, ce qui illustre un désir de mise en commun de la connaissance.

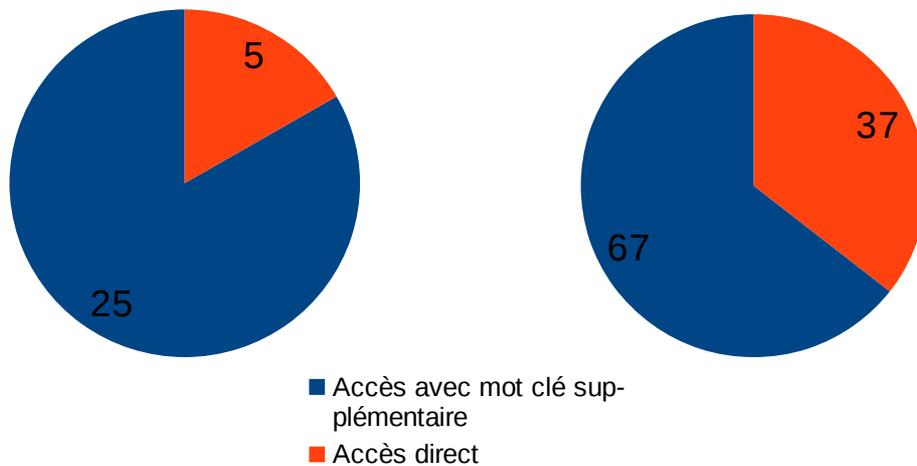
Notons tout de même que quelques sites s'excluent de ce réseau du fait de leur non citation<sup>9</sup>. En effet, des sites comme ZoneNumérique, Lebigdata.fr ou encore Leprogrammeurmarocain.fr forment une petite ceinture isolées autour du réseau sans être reliés à aucun autre site. Cela signifie qu'il ne sont ni cités par des sites, et qu'ils ne citent aucun autre site du corpus non plus. Cela semble plutôt rentrer dans la catégorie des sites indépendants, de type « blogs », qui ne se soucient pas spécialement de la légitimité de leur parole. Néanmoins, il ne faut pas en conclure que le discours développé sur ces sites n'a aucune valeur. En effet, si l'on se réfère au graphique de répartition des sites que nous avons étudié dans la partie précédente, rappelons que la modalité principale qui caractérisait ces sites se trouvait plus dans le style d'écriture et de mise en forme que dans l'information concrète transmise. Ainsi ce n'est pas forcément parce qu'un site est un blog qu'il ne sera pas pertinent en terme d'information, il aura juste une légitimité moins marquée. Cela se vérifie d'ailleurs si l'on étudie statistiquement ces sites :

---

8 Marc Bassoni. L'information spécialisée au défi de la blogosphère : vers de nouvelles formes de médiation journalistique ?. 2015

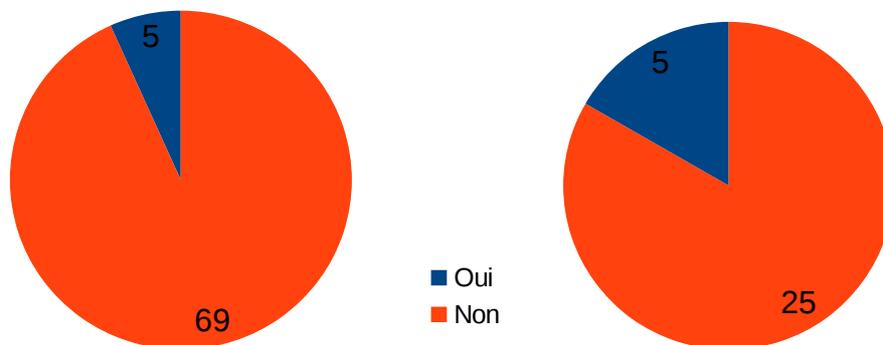
9 Nous les avons d'ailleurs isolés de la représentation afin que le réseau soit plus facilement lisible.

Figure 7 : Comparaison de la répartition entre les sites accessible avec la simple requête « ordinateur quantique » et les sites nécessitant des mots clés. À gauche, le corpus des points isolés, à droite le corpus entier



La figure 7 illustre que la part des sites accessibles en accès direct diminue énormément dans le sous corpus des sites « isolés » du réseau. Cela est facilement compréhensible dans la mesure où à partir du moment où ils ne citent aucun autre sites et ne sont pas cités, ils perdent de la valeur aux yeux du moteur de recherche de Google, qui ne va pas favoriser leur visibilité auprès du public. En effet, là où les sites en accès direct composent Un bon tiers du corpus original, ce taux tombe à 1/5 pour le sous corpus des points isolés.

Figure 8 : Taux d'utilisation de schémas dans le corpus total et le corpus des points isolés



Si le taux d'utilisation de schémas était d'environ 1/14 pour le corpus total, il passe à 1/5 sur le corpus des points isolé. Il y a donc bien une tendance plus forte à utiliser ce type d'élément dans le développement des articles isolés du réseau. Ces sites que l'on pourrait qualifier de « dissidents » n'ont finalement pas l'air d'avoir le même objectif de communication. Cela se remarque grâce à un élément simple mais significatif : la présence d'un ou plusieurs schémas explicatifs dans l'article. En effet, la présence d'un tel élément est un marqueur d'un désir de compréhension concret. On ne souhaite pas transmettre la forme de l'information mais faire comprendre le fond à la personne qui la reçoit. Cette démarche s'ancre dans une logique plus d'enseignement que d'information. Ainsi, les sites isolés du corpus ne doivent pas être compris comme des sites à information faible mais plutôt comme des sites trop dissidents pour être cités ou se soucier de citer les autres : ils utilisent un type de communication moins *mainstream*.



Cette première représentation ancre le discours global concernant l'ordinateur quantique dans un discours qui semble très scientifique. En effet, en fonction des mots les plus utilisés dans l'ensemble du corpus, c'est avant tout un discours très pratique qui s'illustre. Le mot qui est le plus souvent prononcé est « qubits ». Si l'on regarde aux alentours de ce terme, nous retrouvons un vocabulaire fortement lié aux ordinateurs de manière générale tel que « algorithme », « informatique », « calcul », « machine », etc. Le centre du nuage correspond tout de même à des termes très généraux, qui ne rentrent pas dans les détails du fonctionnement de l'ordinateur quantique (à l'exception de « qubit »).

Le terme « qubit » se place donc comme une sorte de « mot clé » qui sera cité de toutes parts lorsque l'on expliquera le fonctionnement de l'ordinateur, pourtant aucun article ne rentre concrètement dans des détails permettant de saisir la globalité de ce que représente cette notion. Nous avons à faire ici à une notion poussée qui, à force d'être citée et répétée dans l'espace médiatique, se vulgarise en un terme banal, ne voulant plus dire grand-chose mais avec une forte représentation ancrée dans l'imaginaire. Nous qualifierons ce processus comme une forme de « banalisation du technique ». Cette logique est assez courante dans le monde médiatique : lors de polémiques par exemple, les médias peuvent facilement se saisir d'un ou deux termes dont le contenu exact est assez obscur, tout en lui imposant une définition très superficielle qui deviendra alors la norme que le grand public comprendra sans avoir besoin de voir plus loin pour se construire un avis sur la question. Par exemple, nous pourrions faire un parallèle avec la polémique des OGM dont la définition exacte lors de son expansion n'était pas forcément connue par tous : bien que personne ne sache réellement la signification du terme, cela a rapidement été catégorisé comme « mauvais ».

Plus l'on s'écarte du centre du nuage, plus les termes utilisés semblent vagues. Nous repérons par ailleurs 3 stades de discours. Le premier stade est donc le cadre technique avec les termes directement liés à un fonctionnement de l'appareil. Le second stade peut être apparenté à un discours plutôt pratique et des exemples des possibilités entraînées par l'ordinateur quantique avec des termes comme « permettre », « recherche », « physique », « chercheur », « exemple », « résoudre ». Enfin un troisième stade de discours se caractérise plutôt par les opportunités offertes par l'ordinateur quantique avec un vocabulaire comme « intelligence », « cryptographie », « sécurité », « microsoft », « application » etc.

Le processus informatif global peut se résumer comme mettant l'accent sur 3 caractéristiques des ordinateurs quantiques : une description technique de l'appareil, une description pratique sur la recherche effectuée autour de cet appareil et enfin les opportunités qu'il pourrait amener. Cette analyse présente le sujet de l'ordinateur quantique de manière très scientifique et au final ne met pas en valeur l'idée des représentations fortes qui se sont construites autour de cet outil.

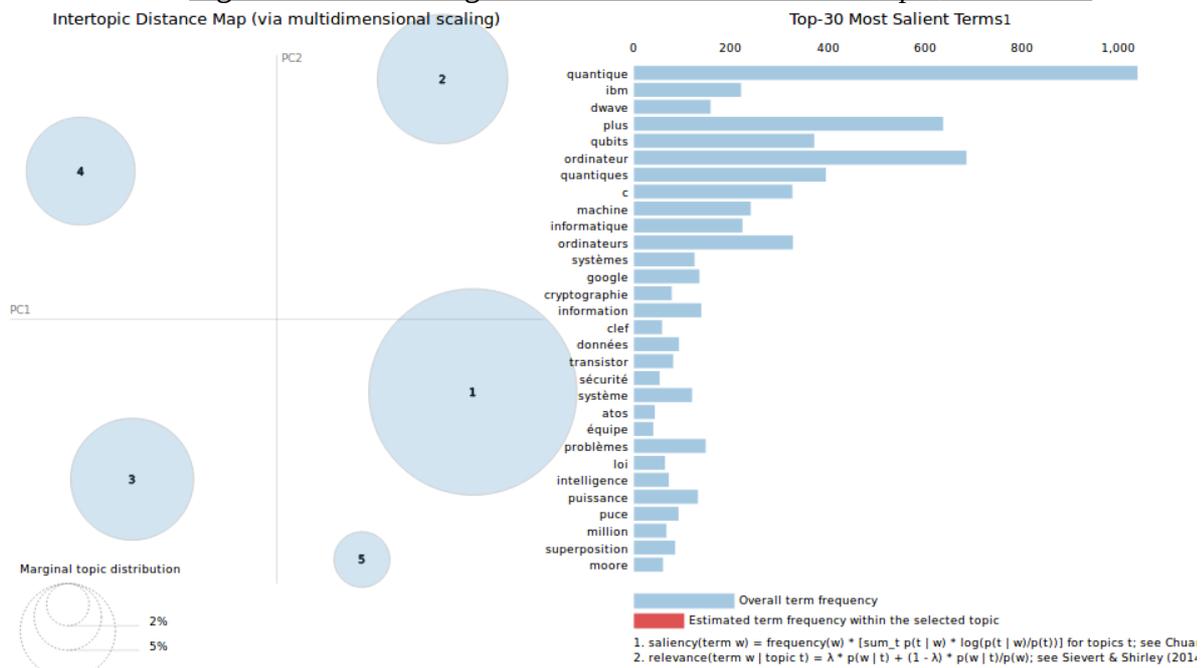
## 2/ L'émergence de sujets dans le discours : entre temporalité et valeur

La manière de s'exprimer sur une thématique particulière peut être influencée par de nombreux éléments. Dans le cas des articles sur l'ordinateur quantique par exemple, l'orientation thématique du site et sa vision du sujet va participer à l'élaboration d'une certaine structure du discours en fonction de plusieurs éléments thématiques. Pour déterminer ces structures, nous utiliserons un modèle de statistique probabiliste permettant d'identifier des sujets ou des thèmes abstraits dans le corpus. Ces thèmes abstraits peuvent être résumés par la notion de *topic*.

Le modèle répartit donc les différents mots du corpus en fonction de 5 « thématiques » principales : Nous demandons au calcul de ne pas dépasser 5 *topics* car au-delà de ce nombre, la cohérence du *topic* devient trop imprécise pour être réellement considérée comme valable. À partir de 6 *topic*, le 6<sup>e</sup> apparaît comme ayant une distribution marginale de moins de 2 %, c'est à dire que ce *topic* se retrouve dans moins de 2 % des articles qui font partie du corpus. L'analyse se limitera

donc aux *topic* apparaissant dans plus de 2 % des articles de notre corpus du fait du nombre réduit de notre base<sup>10</sup>.

Figure 10 : Résultats globaux de la construction de 5 *topic* sur Cortex



Sur la figure 10, les *topic* sont répartis sur l'ensemble du plan factoriel et donc éloignés les uns des autres. Cela signifie qu'ils sont assez imperméables : la construction des thématiques est très nette et les articles traitant de l'ordinateur quantique abordent quelques sujets de manière très autonome, sans forcément l'intégrer à un discours qui serait plus général.

Les barres bleues du graphe représentent la fréquence simple d'un mot dans un corpus particulier. Ainsi, nous retrouvons les observations initiales du nuage de mot qui font ressortir les termes : « quantique », « ordinateurs », « ibm », « qubits », « machine » etc. Vérifions à présent l'exclusivité des termes du corpus par rapport au *topic* particulier : ce que nous observons correspond donc à la pertinence des termes dans la construction des *topic*.

Les mots initialement très récurrents dans tous les corpus prennent une toute autre dimension lorsque l'on analyse la répartition des termes les uns en fonction des autres. Si l'on augmente donc le taux de cohérence, les *topic* prennent une toute autre allure. Les termes « ordinateur » et « quantique » n'apparaissent alors plus dans aucun des *topic* lorsque l'on analyse la cooccurrence et non plus la simple occurrence.

10 À la vue du nombre d'article, 2 % signifie moins de 2 articles. Il n'est donc pas significatif de s'intéresser à des topics qui apparaissent dans moins de 2 articles.

Topic 1 : Quantique/sci- entifique	Topic 2 : Expérimentati- on/tests	Topic 3 : Humanité/repr- ésentation	Topic 4 : Enjeux économiques	Topic 5 : Collaboration recherches
ion	volkswagen	clef	atos	européen
particule	développeur	cerveau	euro	atomique
spin	exécuter	publique	juillet	sécurité
probabilité	plateforme	quasiment	flux	développée
particules	outils	schéma	nasa	coût
superposition	future	connectés	cybersécurité	protection
intéressant	disponible	réseaux	conventionnel	europa
chat	avancée	moore	supercalculateur	former
onde	tutoriels	étude	canadienne	chercheur
shor	véritables	maximum	apprentissage	collaboration
atomes	signaux	stockage	rupture	miliard
schrödinger	instables	risque	simulateurs	industrielle
position	prochain	humain	dollar	distribution
pièce	médicaments	donnée	magnétiques	communication
explication	expérience	consommation	septembre	environnement

Parmi les *topic* créés, nous retrouvons un *topic* assez technique « science/matière quantique », un *topic* qui touche plutôt à la pratique « expérimentation/tests », un *topic* qui semble porter plus généralement sur les représentations humaines et les concepts créés autour de ce sujet, que nous appellerons « Humanité/représentations », un *topic* portant principalement sur les enjeux économiques qu'entraîne l'ordinateur quantique, que nous résumerons sans surprise par le terme « enjeux économiques » et enfin, un *topic* portant sur les idées de travaux généralisés des différents pays entre eux ainsi que sur l'idée de travaux en commun. Nous appellerons cette dernière catégorie « Collaboration dans la recherche ».

Ces différents *topic* semblent plus ou moins concorder avec les mots clés recherchés par les individus. En effet, parmi les mots clés associés à l'ordinateur quantique, nous retrouvons par exemple le mot « prix », que l'on apparente au *topic* des enjeux économiques. Le terme « explication » concorde avec les *topic* liés à l'aspect scientifique des ordinateurs quantiques. Les termes de type IBM ou Google s'ancrent dans le *topic* « expérimentation/test » dans la mesure où ce *topic* semble faire une sorte de veille afin de garder un œil sur les grands acteurs du numérique qui seraient susceptibles de développer des éléments liés aux ordinateurs quantiques.

Cette observation entraîne la question de la construction du savoir : est ce que le contenu des articles est influencé par un désir de connaissance spécifique des individus ? Ou alors est ce que les recherches des individus sont dirigées par les thématiques traitées dans le monde social, médiatique et scientifique ? Nous touchons ici à une question à laquelle il est difficile de répondre fondamentalement : les individus sont-ils construits par le social ou au contraire sont-ils les constructeurs de ce social ? Le point de vue le plus raisonné est de considérer qu'il s'agit d'une constante interaction entre les deux événements : les individus sont socialisés et structurés par un social qu'ils modèlent en permanence au cours de leur quotidien.

Nous décidons alors de décrire les axes du plan factoriel afin de délimiter clairement quelles sont les deux types de rhétoriques utilisées lorsque l'on parle de l'ordinateur quantique. Si l'on se base sur les titres que nous avons donné aux *topic*, les axes répartissent de manière très équilibrée ces derniers, mettant en lumière une opposition entre deux principaux arguments :

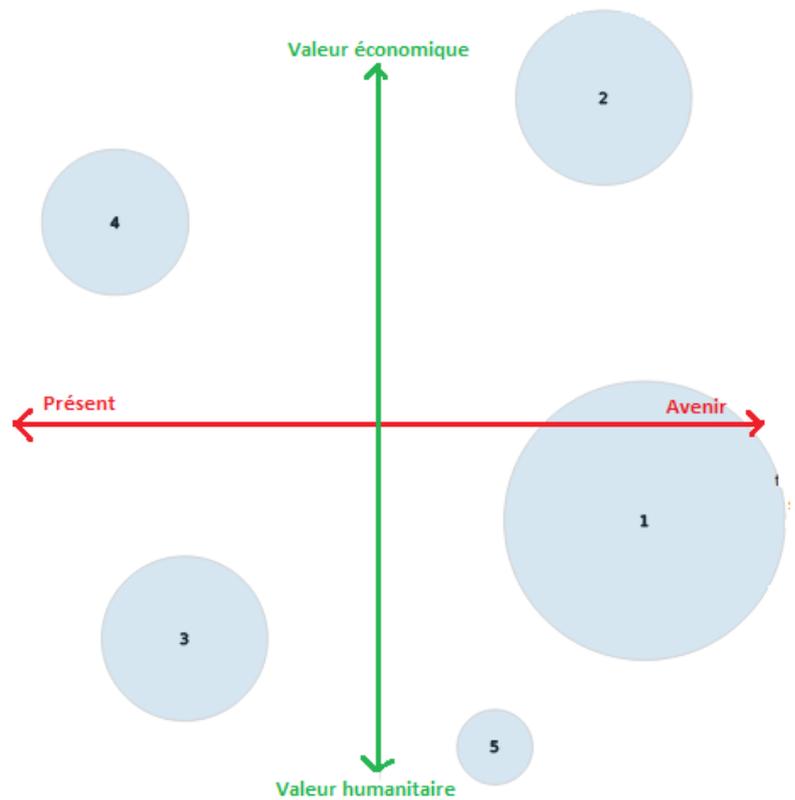
D'une part, on remarque que les *topic* « scientifiques » et « expérimentation » sont opposés aux *topic* « représentation » et « enjeux » sur l'axe horizontal. Cette tendance se traduit comme une opposition entre les « réalisations concrètes » d'une part (l'état des lieux actuel) et les « projets » d'autre part (les enjeux à développer à l'avenir).

D'autre part, les *topic* « expérimentation » et « enjeux économiques » sont opposés aux *topic* « scientifique » et « représentation » sur l'axe vertical. Cette tendance illustre une opposition entre une valeur concrète, économique et une valeur qui serait plus « réelle », plus objective en terme d'apport à l'humanité.

La dimension de la temporalité croise donc celle de la valeur dans les représentations du discours portant sur les ordinateurs quantique. On oppose un

état des lieux actuel à un projet à long terme en y ajoutant la notion de valeur économique pouvant se transformer en valeur « humaine ». Quel que soit l'axe choisi, l'opposition suppose une forme d'évolution. Évoquons pour conclure l'étrange placement du *topic* « collaboration dans la recherche » sur ce plan. Il semble illustrer l'idée d'une communication entre les pays afin d'accélérer un processus de recherche et de connaissance sur un sujet. Par ailleurs, il se trouve d'une part du côté de la temporalité à court terme, et d'autre part, du côté d'une valeur humaine, qui ne se limite pas à un aspect économique. Ces observations semblent donc logiques avec ce que l'on peut attendre d'un processus de communication internationale sur un domaine particulier : la nécessité de travailler immédiatement à un partage de connaissance afin d'enrichir les données que l'on possède sur un sujet. Cette idée recoupe d'ailleurs notre observation faite à l'étape cartographique, qui supposait un fort échange entre les sites internet sur les informations concernant les ordinateurs quantiques.

Nous en déduisons donc que l'imaginaire que l'on déduit des discours portant sur les ordinateurs quantiques peuvent être synthétisés ainsi : Si l'ordinateur quantique est actuellement très coûteux à mettre en place, autant par des moyens humains que des moyens économiques, il deviendra à l'avenir un élément crucial de l'évolution technologique et permettra de régler la plupart des problèmes auxquels nous faisons face aujourd'hui.



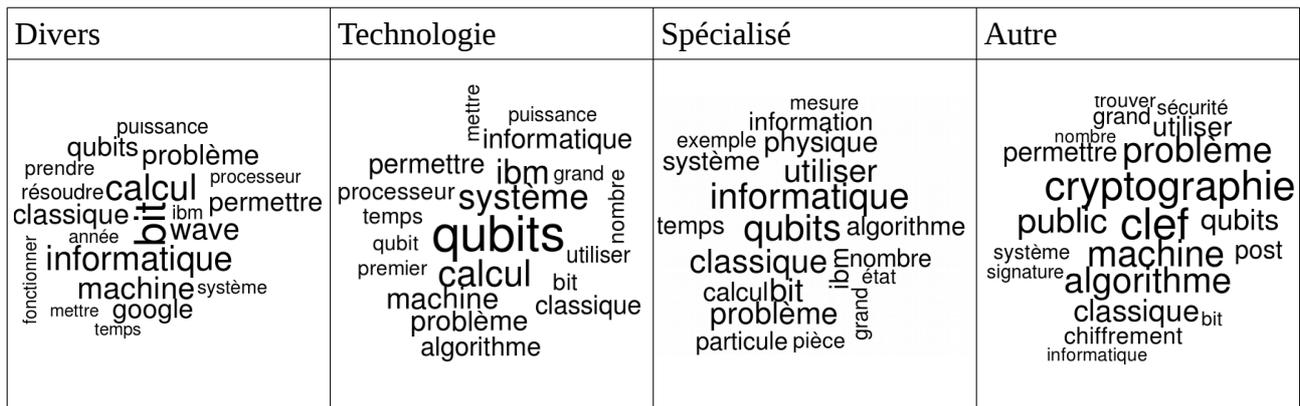
### 3/ Une spécificité de discours selon les profils ?

Nous allons à présent voir dans quelle mesure l'orientation principale d'un site internet peut influencer l'information qu'il transmet aux individus. Pour cela, nous procédons de nouveau à une construction de *topic*, cette fois en fonction des caractéristiques des sites et des articles de notre corpus.

L'analyse en fonction des *topic* semble moins pertinente lorsque l'on procède au même processus pour chacun des types de sites, en effet, ceux qui en découlent sont tous un petit peu mélangés et il est plus difficile d'en déduire des tendances de sujet. Ce manque de pertinence nous montre bien qu'il existe des manières de parler du sujet en fonction de certains domaines ou certains

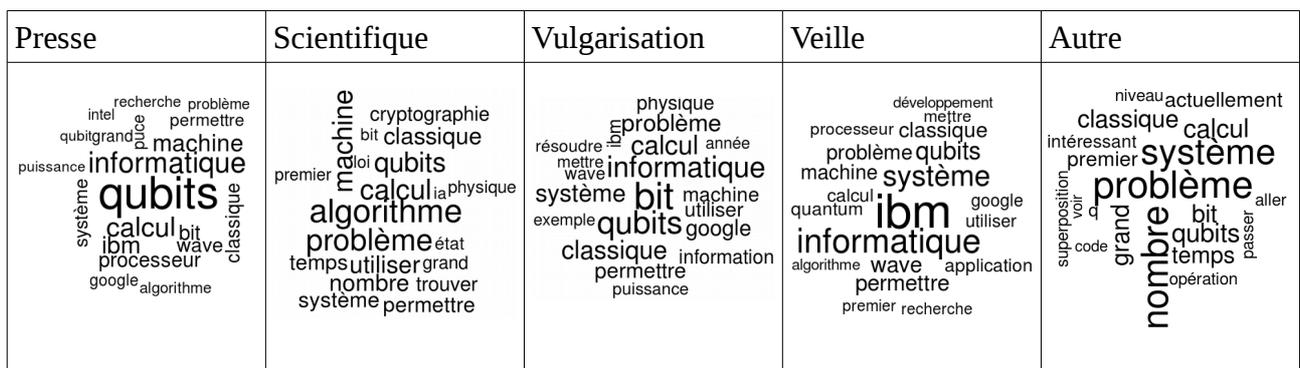
types de sites. Lorsque l'on procède à la même analyse seulement dans une catégorie particulière, il est normal de ne plus retrouver la construction de *topic* que nous avons au départ puisqu'à présent, le corpus est « homogénéisé » sur une thématique particulière. En effet nous trouvons avant des différences de type « traitement de sujets économiques, partage de la recherche, traitements de sujets techniques. Lorsque le corpus n'est composé que d'articles similaires, ces différences de discours n'apparaissent plus : les *topic* sont assez mélangés et homogènes entre eux car tous les éléments de ce nouveau corpus concordent. Il existe bien des manières de communiquer sur les ordinateurs quantiques en fonction des centres d'intérêts des différents sites internet qui traitent de cette problématique.

Figure 11 : Occurrence des mots en fonction du type de site :



Dans une telle situation, il faut oublier les méthodes de construction de *topic* et se limiter à une simple analyse des occurrences. Si l'on regarde les occurrences de mots en fonction des types de sites (Figure 11), la catégorie de site « autre » se démarque fortement du fait du poids légèrement moindre accordé au terme qubit. Ici, on ne se soucie pas forcément de résumer le principe de l'ordinateur quantique à travers un mot stéréotype lorsque l'on lit des articles sur l'ordinateur quantique. Ces sites de catégories « autre » semblent rentrer plus rapidement dans le vif du sujet en mettant l'accent sur des éléments plus concrets de l'ordinateur quantique (cryptographie, clef, algorithme, problème, etc).

Figure 12 : Occurrence des mots en fonction du type d'article :



Pour les articles scientifiques sur la figure 12, même si l'on retrouve le poids du terme « qubit », à la différence des autres, il y a beaucoup d'autres mots qui pèsent autant. D'une certaine manière, il semble que les articles scientifiques aient moins tendance à construire leur argumentation autour d'un mot clé particulier, qui se placerait alors comme le « mot à retenir », le « mot savant » autour duquel tout est développé : on s'assure que le plus de termes « savants » soit

évoqué de manière équivalente pour ne pas enfermer la compréhension dans un cliché de mots clés, ni le réduire à ces derniers.

Enfin, Les article de type « veille innovation » ont une forte récurrence du terme IBM, ce qui est significatif vis à vis de leur raison d'être : leur but est de tenir au courant sur les actualités de l'ordinateur quantique, notamment en identifiant les acteurs qui font des avancées sur le sujet.

Un effet de structure se fait sentir en fonction de l'orientation du site internet : Au regard des enjeux très variés que propose l'ordinateur quantique, il devient une sorte de thématique à la mode sur laquelle tout le monde communique un peu à sa manière en fonction de son intérêt propre.

## Conclusion :

Nous avons donc observé la production d'articles sur les ordinateurs quantiques en fonction de deux aspects : les « acteurs-producteurs » d'une part, et les « discours-information » d'autre part. Afin de retrouver une objectivité que nous n'avons pas forcément à la base, nous avons tenté de nous mettre à la place des internautes et d'adopter leurs comportements de recherche et de réception de l'information. De cette manière, nous nous assurons que le corpus que l'on récupère correspond bien à ce que n'importe qui sans aucune connaissance du sujet va se retrouver confronter. Notons tout de même un premier biais, à savoir que cette étude est ancrée dans le temps en raison du dynamisme d'internet. En effet, la popularité des articles (presse ou scientifique) est souvent temporaire et leur place sur les premières pages de Google peut donc s'amenuiser avec le temps. Les phases de récupération de données ont été effectuées au cours du mois de novembre 2017. Si nous procédions à la même récupération aujourd'hui, les articles seraient certainement différents.

Nous avons rapidement remarqué que les types d'articles étaient très variés et nous avons d'ailleurs dû affiner nos critères de sélection des sites afin d'avoir un corpus homogène pouvant être traité de manière cohérente.

Sans surprise, les premières informations révèlent que c'est avant tout les domaines s'intéressant aux technologies qui s'intéressent à cette thématique. Cependant, d'une part il ne s'agit pas forcément de domaines très scientifiques comme nous pouvions l'imaginer du fait de la technicité très poussée du sujet. D'autre part, il n'y a pas de chasse gardée dans la mesure où tous les champs du monde social semblent se sentir concernés d'une manière ou d'une autre par l'ordinateur quantique. La majorité des acteurs-producteurs est constitué de la presse grand public, ce qui montre bien l'intérêt que la société dans son ensemble y accorde et le désir de communiquer dessus.

Si nous devons résumer la production des articles portant sur l'ordinateur quantique, nous pourrions dire qu'elles répondent à deux critères de construction : un critère portant sur le style rédactionnel et un critère portant plutôt sur l'objectif de la communication vers le lecteur. Nous opposons donc d'une part les articles de type presse, courts et concis aux articles scientifiques ou blogs et d'autre part, les articles ayant une simple visée informative et les articles ayant une visée d'enseignement concret. Nous séparons donc les sites internet en 3 catégories que sont les articles scientifiques, les articles destinés au grand public et les articles destinés au « petit public ». L'analyse cartographique nous informe sur deux tendances dans la production d'articles : l'une est une forte tendance à la citation des pairs, c'est le cas surtout pour les articles de presse et dans une moindre mesure les articles scientifiques. L'autre tendance est justement une sorte d'hermétisme au partage de la connaissance sur le sujet avec des articles qui ne s'attardent pas forcément à citer d'autres sites internet. Cette catégorie surtout composée des sites tournés vers le « petit public » n'en est pas pour le moins illégitime dans son apport d'information.

Cette pluralité dans les propositions d'article semble refléter assez fidèlement le principe même d'internet : tous le monde a la possibilité de s'exprimer sur un sujet et tout le monde a la possibilité d'en entendre parler en fonction du canal et de la manière qui lui correspond le mieux. C'est lors de ce type de recherche désordonnée que la rationalité de l'utilisateur entre en jeu : il préfère souvent avoir une forte quantité de bruit afin de faire par la suite lui-même son tri en fonction des formats et des contenus d'articles qui lui ont plu. D'autant plus que cette méthode développe une possibilité de sérendipité très forte qui ne serait peut-être pas possible si l'individu se cantonnait à un domaine ou un champs particulier.

L'analyse sémantique à laquelle nous procédons révèle que, si l'ordinateur quantique semble un sujet très technique et scientifique, il est emprunt d'un fort imaginaire qui touchent les individus. En effet, le caractère visionnaire et innovant qui est décrit à travers les différents articles vont permettre de faire abstraction de cet aspect technique et les lecteurs vont alors dépasser la limite de leurs connaissances. Nous voulons dire par là que, de par les enjeux capitaux et quasiment métaphysiques qu'on lui octroie, la compréhension ou le manque de compétence pour suivre le

sujet ne devient plus un problème : les promesses sont plus fortes que l'incompréhension et la croyance l'est parfois plus que le raisonnement. Tout ce que les individus ont besoin de comprendre, c'est que c'est un sujet important.

Néanmoins, le système de « croyances » envers les promesses de l'ordinateur quantique n'est pas comparable au concept de foi par exemple dans la mesure où cette croyance est légitimée par un discours argumenté. La non connaissance de la partie très technique sur laquelle repose l'ordinateur quantique n'empêche tout de même pas de saisir les conséquences concrètes (et plus accessibles) que peut avoir l'ordinateur. Cette argumentation met l'accent sur deux dimensions que l'on peut résumer avec un seul et même concept. Les deux dimensions sont la temporalité et la valeur qui sont certainement les principaux axiomes de nos sociétés modernes. Les discours sur l'ordinateur quantique opposent donc un état actuel de la recherche à l'avenir que l'on réserve à l'outil d'une part et la valeur marchande qu'il possède à la valeur humanitaire qu'il développera à terme.

Comme nous l'avons précisé, ces deux ensembles se regroupent sous le concept de l'évolution, du fait de passer d'un état à l'autre. l'état temporel présent à un état incertain futur et la transformation de la valeur par l'investissement avec l'idée que le financement est nécessaire pour pouvoir profiter au maximum de ses capacités. L'enjeu, peut-être inconscient, de la production d'articles sur l'ordinateur quantique est finalement d'illustrer l'outil de manière subtile par ce qu'il est : une évolution technologique.

L'une des limites que nous pouvons poser concerne une question que nous nous posons au départ. Si nous avons pu montrer les processus de partage ou d'imperméabilité de l'information entre certains sites définis (les sites les « plus populaires » vis à vis de l'ordinateur quantique à un moment T) nous ne pouvons pas savoir dans quelle mesure il s'agit d'une bulle d'information fermée ou non. En effet, nous manquons de profondeur dans le crawl des pages web, ce qui a bloqué le crawl de sites que nous n'avons pas référencés à la base. Ainsi, nous pouvons imaginer que le réseau de citation est beaucoup plus large que celui que nous avons analysé. De plus, cela expliquerait pourquoi certains sites en apparence assez bien construits et légitimes n'étaient liés à aucun acteurs-producteur du réseau : peut être ces derniers vont chercher leurs sources à des endroits bien plus spécifiques, auquel l'utilisateur lambda n'a pas un accès spontané.