

## Onomastique et Big Data



Elian Carsenat / Fondateur de NamSor Applied Onomastics / September 17th, 2013

progress and growth

technology and business

*Aujourd'hui, le principal domaine d'application professionnelle de l'onomastique est le choix du meilleur nom pour qu'une entreprise ou un produit se démarque sur le marché international. Onoma – racine grecque du mot – est aussi une marque déposée de Nomen, l'agence de création de marques fondée par Marcel Botton en 1981, celle-là même qui a exploité le « Radoteur », générateur de noms inventé par Roland Moreno, et créé de nombreuses marques mondialement connues telles que Vinci, Clio ou Amundi. Pour autant, même lorsque leur entreprise a été baptisée, les dirigeants n'en ont pas terminé avec l'onomastique. La mondialisation, le numérique et le Big Data sont autant de facteurs susceptibles de perturber les secteurs de la vente et du marketing, de la communication, des ressources humaines, de la gestion des risques. Ils ouvrent en même temps de nouveaux champs d'expérimentation pour des technologies innovantes. Et, bien que l'analyse par les noms comporte un risque d'abus non négligeable, elle peut aussi trouver des usages positifs et inattendus, comme d'aider à développer certaines régions dans le monde.*

Chaque jour, notre cerveau interprète des noms, dans une langue que nous comprenons, une culture que nous connaissons, une région que nous visitons : le menu d'un restaurant, le nom d'une entreprise... même le nom d'un animal de compagnie peut révéler quelque chose sur son propriétaire. Les noms (prénom, nom, pseudo) sont porteurs de sens qui varient en fonction de la langue et de la culture d'un individu mais constituent souvent une part essentielle de son identité.

### L'art d'extraire le sens des noms

S'il était possible de programmer un ordinateur pour extraire le sens des noms, nous fournirait-il de précieux renseignements dans le domaine des affaires ? Aux Etats-Unis, un certain nombre de personnes en sont convaincues. La CIA (Agence centrale de renseignement) a une longue expérience en la matière. Dans les années 1980, elle a utilisé le logiciel de reconnaissance de nom LAS pour aider à identifier les espions russes, démasquer de fausses identités, traquer l'influence soviétique ; le logiciel était alimenté par une base de données d'un milliard de noms recueillis par l'agence de renseignement, soit l'équivalent de la totalité de la population développée à l'époque.

Après avoir prospéré avec le renforcement de la sécurité et des budgets alloués au renseignement après le 11-Septembre, LAS s'est diversifié vers d'autres marchés : marketing, conformité des services financiers (notamment KYC, c'est-à-dire Know Your Customer), etc. IBM a fait l'acquisition de LAS en 2006. En 2011, les agences de sécurité américaines ont fait appel aux services de l'organisation [MITRE Corporation](#) pour renforcer l'innovation dans les technologies ayant un intérêt pour le gouvernement fédéral. « Le principal défi a consisté à travailler sur le muticultural name matching, une technologie à la base de l'appariement d'identité, qui consiste à comparer des bases de données se rapportant à des individus. Elle permet notamment de vérifier l'adhésion à la sécurité sociale, l'identification et le regroupement des familles dans les opérations de secours en cas de catastrophe, la

fusion ou l'élimination des doublons dans les bases de données. L'appariement des noms peut également servir à améliorer la précision et la rapidité des recherches de papiers d'identité, l'analyse des réseaux sociaux et d'autres tâches encore dans lesquelles une personne peut être désignée par plusieurs versions et orthographe d'un nom. » Un nom traduit plus – ou autre chose – qu'une simple nationalité. Ainsi, le nom de famille des frères Tsarnaev, les auteurs de l'attentat de Boston (avril 2013), a-t-il une terminaison en « v », caractéristique des noms slaves mais peut être reconnu comme plus précisément originaire du Caucase. Après la tragédie, les médias ont rapporté que le FBI ignorait tout du voyage effectué en 2012 par Tamerlan Tsarnaev au Daghestan, région sensible de la Fédération de Russie, car « son nom avait mal été orthographié sur ses documents de voyage ». Cette information n'a toutefois pas été confirmée et semble peu probable en regard des immenses investissements effectués par les Etats-Unis dans ce domaine.

En Europe, le cadre juridique permettant de tirer parti de ces outils varie d'un pays à l'autre, mais il est généralement très strict. La [directive](#) 95/46/CE du Parlement européen et du Conseil de l'Europe du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, l'article 8 stipule que « les États membres interdisent le traitement des données personnelles révélant l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ». En principe, cette directive s'applique aussi aux agences de sécurité, mais il existe des exemptions que les États membres peuvent interpréter différemment.

En utilisant les termes de « profilage ethnique discriminatoire » au lieu de « profilage ethnique » pour décrire la pratique de fonder les décisions de faire appliquer la loi exclusivement ou principalement sur la race, l'ethnie ou la religion d'un individu, l'Union européenne [reconnait la nécessité](#) pour les forces de sécurité de comprendre les relations complexes qui existent entre la nationalité, la géographie et des concepts plus subjectifs tels que l'origine ethnique, le milieu culturel, la civilisation, la religion. La manière dont ces informations peuvent être utilisées, ces données collectées, touche de façon drastique à la sécurité nationale. Le Royaume-Uni et la France, par exemple, n'ont pas le même point de vue sur la question. Quoi qu'il en soit, ce qui est fait en pratique par les organismes de lutte contre le terrorisme n'est pas une information d'ordre public.

La sécurité, le contrôle des frontières, etc. constituent un secteur d'activité à part entière. Qu'en est-il du reste?

### **Le renseignement appliqué à la clientèle : potentiel commercial et questions éthiques**

Dans le domaine de la vente et du marketing, l'onomastique peut servir à enrichir une base de données clients grâce à des informations extraites des noms qui ne seraient pas disponibles autrement. Les détaillants et les marques de luxe – en particulier dans les domaines de l'alimentation, des vêtements et des cosmétiques où l'ethnicité joue un rôle important – peuvent améliorer leurs connaissances client et utiliser ces données pour opérer plus efficacement par le biais des réseaux en ligne. Déjà, au début du XXe siècle, l'homme d'affaires américain John Wanamaker (1838-1922) déclarait : « La moitié de l'argent que je dépense en publicité est gaspillée ; le problème, c'est que je ne sais pas laquelle. » Des entreprises comme L'Oréal qui dépensent plusieurs milliards de dollars chaque année en communication et en publicité cherchent à améliorer l'efficacité de leur ciblage.

Quant aux projets de partenariat public-privé (PPP) dans le secteur minier, l'énergie ou les infrastructures, ils sont susceptibles d'avoir des impacts sociaux importants sur un espace donné et d'y soulever diverses questions politiques, économiques ou sociales. Bien cerner et définir correctement les typologies de géographie humaine et connaître les intérêts des communautés qui cohabitent sur ce territoire peut s'avérer

déterminant pour obtenir l'adhésion de toutes les parties prenantes. L'onomastique, combinée avec une segmentation géo-démographique, permet d'accéder aux complexités spatiales des territoires urbains, des métropoles et des périphéries, à tous les espaces à intérêt majeur géopolitique. Elle est utile pour créer des cartes géographiques destinées aux processus décisionnaires ou à des fins de communication. La technologie d'association automatique des noms (automatic name clustering) permet de décrypter les identités complexes présentes dans les territoires quelque soit leur échelle (un continent, une rue) afin d'aider à répondre à des questions difficiles (quel devrait être le tracé d'une ligne de tramway dans une région multi-ethnique ? Comment redistribuer les revenus pétroliers off-shore ?) Outil de décision et d'analyse, elle permet de gérer les tensions et frustrations inévitables par le biais d'une communication adaptée.

Pour ce qui est des ressources humaines, j'ai eu récemment l'occasion de m'entretenir avec un dirigeant d'une grande banque européenne. Ce dernier déplorait les choix faits dans la sélection des employés expatriés pour travailler sur un projet important d'acquisition dans un des pays BRIC, une mauvaise stratégie qui avait coûté plusieurs centaines de millions d'euros en dépréciations à l'entreprise. Parmi les milliers d'employés travaillant au siège européen, la banque aurait pu cibler les noms de ceux susceptibles d'accepter d'aller travailler dans leur pays d'origine. Avoir plus de salariés connaissant les deux langues et les deux cultures d'entreprise aurait aidé à combler le fossé interculturel entre la direction locale et les expatriés représentant le siège, et permis d'économiser plusieurs millions d'euros.

Dans le monde numérique, l'onomastique apporte un nouvel angle de vue à l'analyse des graphes sociaux : elle peut mettre en évidence des communautés en ligne, établir le profil de leaders d'opinion en fonction de leur audience internationale. Sur Twitter, par exemple, il est possible de créer plus vite un canal de communication ciblé sur une communauté particulière (expatriés d'affaires, touristes, migrants, investisseurs, experts internationaux de l'énergie etc.).

Considérons maintenant un usage provocateur et controversé de l'onomastique qui touche à l'éthique. Les différences de culture, de nationalité, de milieu social impliquent différents comportements en matière d'argent et de prise de risques : gagner de l'argent, épargner, dépenser, jouer, investir, faire un don, risquer la mort et tout perdre... Or, c'est un fait : les individus dotés d'un patronyme aristocratique gagnent plus d'argent et obtiennent un crédit plus avantageux que ceux dont le nom de famille traduit des origines modestes ou immigrées de fraîche date. En allant plus loin, on pourrait imaginer qu'une banque ajuste le taux d'un crédit selon le nom de l'emprunteur, qu'une compagnie d'assurance automobile en fasse autant avec l'évaluation du risque (y compris les risques de fraude, de conduite dangereuse, etc.) en fonction du nom du sociétaire. Elles mesureraient mieux leurs risques et pourraient donc proposer des tarifs plus compétitifs à certains de leurs clients. Des clients qu'elles pourraient en outre mieux cibler commercialement...

Une telle utilisation de l'onomastique est très controversée car elle soulève la question de l'égalité et la discrimination. Pourtant la discrimination est un fait dans la société comme dans les entreprises, et l'onomastique peut permettre de mieux comprendre son fonctionnement. Pourquoi les individus seraient-ils jugés sur un nom à particule avant de l'être sur leurs compétences ? Il est révélateur que Casanova ait choisi de se faire appeler chevalier de Seingalt et s'interroge au XVIIIe siècle déjà si d'Alembert aurait pu atteindre le même niveau de notoriété en conservant son simple nom de Jean Le Rond.

Favorable à l'égalité et à la justice, j'ai cependant élaboré un algorithme basé sur les noms. NamSor est un logiciel de DataMining qui applique l'onomastique pour l'analyse des flux mondiaux de capitaux, d'idées, de personnes. Comme toute nouvelle technologie puissante, elle comporte des risques potentiels d'abus, mais je

suis persuadé qu'elle peut aussi avoir un usage positif.

L'onomastique joue traditionnellement un rôle important en géo-démographie, l'analyse de la sociologie d'un territoire (y compris les origines culturelles et ethniques de ses habitants) à partir de sources publiques et des données de recensement. La géo-démographie peut être un outil utile pour, par exemple, garantir à tous un accès équitable aux services publics. La société [Experian](#) est l'un des leaders dans ce domaine bien développé au Royaume-Uni.

L'utilisation efficace des Big Data et Open Data est considérée comme un catalyseur essentiel dans l'essor des villes intelligentes de demain : permettre une bonne distribution des ressources, une utilisation plus efficace de l'énergie, une réaction rapide en cas de crise, etc. La combinaison des réseaux sociaux et des applications mobiles avec des dispositifs géo-localisés ouvre de nouvelles possibilités. La connaissance de la diversité des populations qui cohabitent dans un espace et un temps donnés peut aider à concevoir des villes et des réseaux de transports plus accueillants. Les marqueurs pour percevoir les populations peuvent aider à dessiner une image claire des mesures à prendre pour prévenir la discrimination et désamorcer des bombes à retardement sociales ici et là.

### **Cibler les diasporas : une nouvelle donne pour le développement ?**

Mais l'utilisation la plus prometteuse de logiciels tels que NamSor pourrait être ailleurs. Les régions du monde moins développées font de leur faiblesse (la pauvreté) une force (une main-d'œuvre moins chère) pour attirer les investissements. L'idée est de déclencher un cercle vertueux – créations d'emplois, développement des infrastructures, meilleure éducation, inversion des flux migratoires, etc. – qui finit par attirer encore plus les investissements directs étrangers (IDE). La région devient plus attrayante et se hisse progressivement vers le haut dans la chaîne de valeur mondiale. Perdant sa compétitivité en terme de main-d'œuvre bon marché du fait de la nouvelle richesse de sa population, elle développe une autre économie, basée sur l'innovation, les services, le tourisme, la consommation.

La plupart des pays mettent en œuvre une politique de développement économique dirigée vers les régions les plus pauvres, comme un moyen de préserver leur cohésion et leur intégrité territoriale. Ces politiques sont plus efficaces quand elles se combinent avec des initiatives privées couronnées de succès. L'objectif de nombreuses agences de promotion des investissements n'est donc pas tant d'attirer en masse des capitaux que d'accueillir une grande entreprise, source d'emploi et de croissance locale. Mais il existe une rude concurrence mondiale entre les différents territoires.

Les régions les plus pauvres ont une autre faiblesse susceptible de devenir une force : leur émigration. Considérée dans un premier temps comme une perte en terme d'opportunités, le phénomène finit par engendrer une diaspora, laquelle peut servir de levier pour attirer localement des investissements.

Prenons le cas de l'Irlande. Au début des années 1980, le pays a pris des mesures décisives pour renouer de manière active avec ses émigrés ou des hommes d'affaires prospères d'origine irlandaise. [Comme le note Rebekah Berry](#), « en 1986, l'Irlande était l'un des pays les plus pauvres de l'Union européenne, mais [en 2002], c'est devenu l'un des plus riches. Le moteur de cette nouvelle prospérité a été l'investissement direct étranger. [Entre 1986 et 2002], les Irlandais ont attiré d'immenses capitaux d'Amérique – en grande partie grâce à un siècle d'attaches personnelles et familiales – et ils ont utilisé cet argent pour construire des usines. »

En Chine, les régions du Ningxia, du Gansu et du Qinghai comptent parmi le moins grand nombre de millionnaires du pays. Mais s'ils pouvaient renouer avec ceux d'entre eux qui ont réussi à Pékin, Shanghai ou même à l'étranger, cela ouvrirait

peut-être de nouvelles perspectives et opportunités de développement.

Dans cette veine, l'onomastique appliquée a démontré son utilité au service de la stratégie de développement d'un pays : la Lituanie.

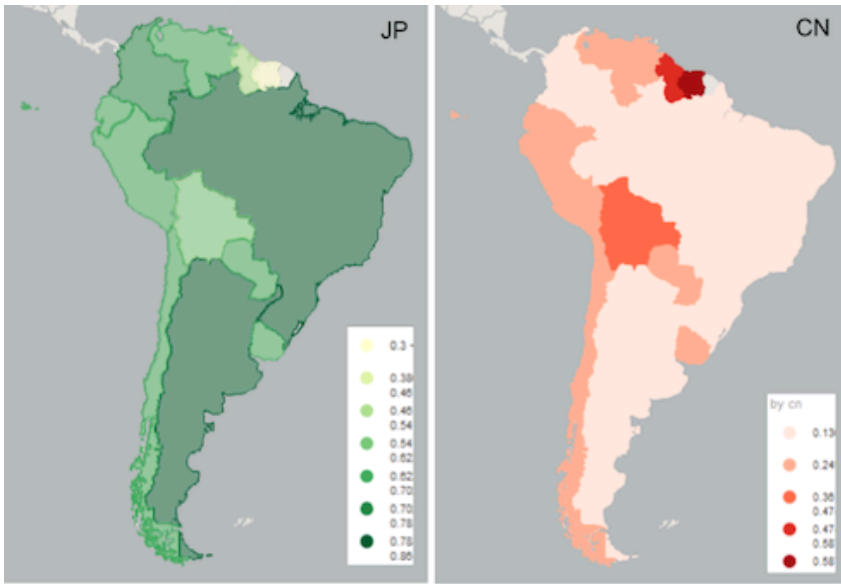
**InvestLithuania** est la première agence de promotion des investissements (IPA) à utiliser la reconnaissance de nom pour attirer des investissements directs étrangers. Le pays compte trois millions d'habitants et près d'un million d'émigrés d'origine lituanienne hors de ses frontières, soit un grand potentiel de liens personnels et familiaux à mettre à profit pour attirer de nouveaux projets d'investissement. Le **logiciel de reconnaissance de nom NamSor** a aidé à faire ressortir ces liens. Autre méthode pour identifier de nouvelles pistes d'investissements : mieux connaître et mieux exploiter le réseau existant des hommes d'affaires étrangers dans le pays même. Domas Girtavicius, consultant senior chez InvestLithuania, a déclaré avoir été « impressionné par la précision du logiciel de reconnaissance de nom : il a déterminé de manière fiable le pays d'origine avec un taux d'erreurs tout à fait gérable ».

Le succès de ce projet avec InvestLithuania a permis d'intervenir lors d'une table ronde au Forum mondial économique lituanien, organisé à Vilnius le 3 juin 2013 par les dirigeants lituaniens mondiaux (GLL). GLL est une association à but non lucratif indépendante dont la mission est de renouer avec les Lituaniens et les amis de la Lituanie à l'étranger. Cette excellente initiative apporte une richesse d'expertise internationale dans tous les domaines (politique, éducation, culture, affaires, etc.) et aide à combler certaines lacunes culturelles. Plus précisément, GLL contribue à introduire en Lituanie des éléments de culture anglo-saxonne, comme l'esprit d'entreprise et le réseautage professionnel.

Alors que certaines diasporas, en particulier celles provenant de l'aire méditerranéenne, ont une culture millénaire des affaires et des réseaux personnels, d'autres pays ont du mal à s'adapter à leur nouvelle situation ou n'y pensent pas. Pour la diaspora libanaise, la valeur d'un réseau social comme LinkedIn est faible. Quel meilleur outil de communication à Marseille que le bouche à oreille pour lancer Massilia Mundi, qui vise à devenir le réseau social de cette ville de forte culture d'immigration et d'émigration ? Mais pour beaucoup d'agences de promotion des investissements, LinkedIn est un outil essentiel. Par exemple, dans la culture traditionnelle lituanienne, les liens familiaux et amicaux sont très forts, au détriment souvent de ces relations professionnelles ou de ces contacts occasionnels qui constituent un réseau d'affaires. De nombreux pays sont dans la même situation, où une organisation spécifique pourrait aider les individus d'exception partis à l'international à renouer des liens avec leur pays. Dans de tels cas, des outils comme les réseaux sociaux, les bases de données professionnelles et l'onomastique peuvent faire la différence.

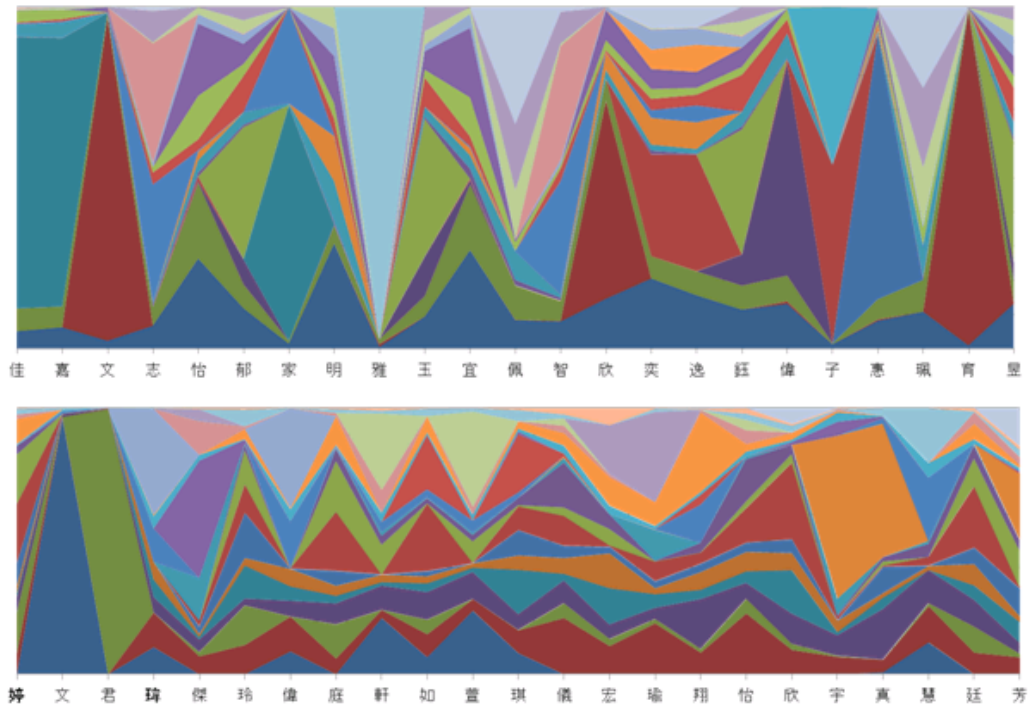
Cela fonctionnerait-il aussi en Chine ? En 2005-2009, alors que je travaillais pour un cabinet de conseil, j'ai eu l'occasion de gérer un projet dans le secteur bancaire, avec des équipes françaises et chinoises : l'une, à Paris, comprenait plusieurs jeunes diplômés de ParisTech d'origine chinoise, l'autre se trouvait à Shanghai. Je me souviens de l'excitation et du plaisir de tous à l'idée de travailler avec la Chine, de se rendre à Shanghai, de déguster des spécialités des différentes provinces chinoises, de découvrir la culture chinoise. Plusieurs personnes de cette équipe, françaises et chinoises, sont maintenant en Chine. Jing, devenue une amie, est retournée à Shanghai en 2009 et je me souviens à quel point elle restait attachée à Xiangtan, sa ville natale dans le Hunan, qu'elle était prête à aider à tout prix. Aussi je suis convaincu que s'il existait une organisation telle que Ningxia, Gansu and Qinghai Leaders, elle pourrait sans problème demander de l'aide, de l'argent, une expertise. Il est clair qu'une telle organisation s'avérerait utile pour combler le retard de ces trois provinces avec le reste du pays en matière de développement.

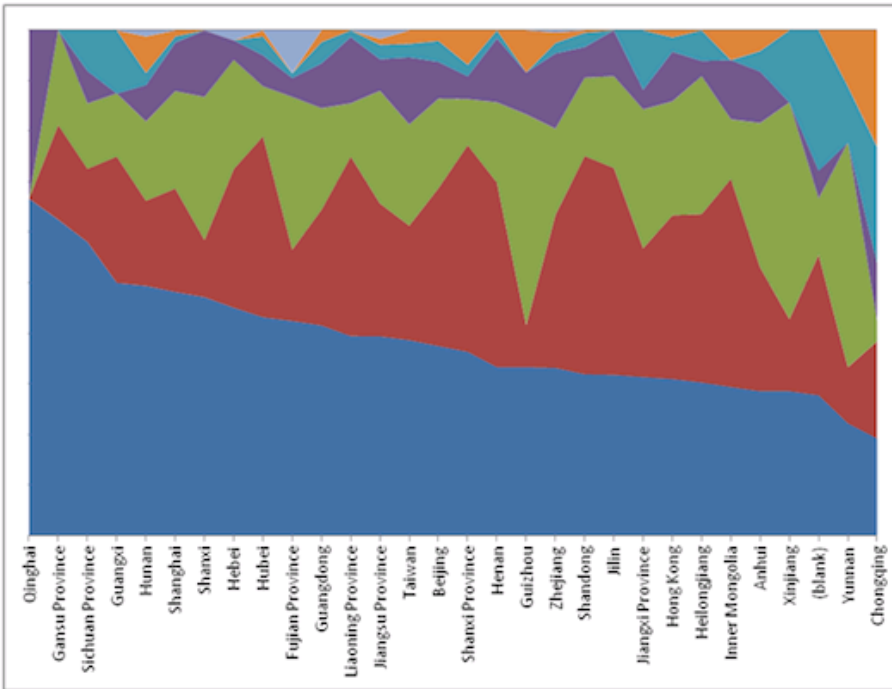
Sur le plan technique, les noms chinois sont clairement reconnaissables. Ainsi, grâce à une base de données professionnelle, il a été possible de réaliser une cartographie onomastique des chefs d'entreprise chinois. Les cartes ci-dessous figurent la densité des milieux d'affaires chinois et japonais en Amérique du Sud.



Source: Factiva DF Copyright 2013 NamSorts.com NomTriTM NamSorTM – All rights reserved

Combien de ces hommes d'affaires chinois (ou d'origine chinoise) ayant réussi sont originaires du Ningxia, du Gansu ou du Qinghai ? C'est là que l'onomastique appliquée peut se révéler intéressante. Toutes les questions ne sont pas résolues. À l'heure actuelle, dans l'analyse des caractères Kanji, le logiciel permet de détecter des phénomènes, sans les clarifier complètement. Voici, par exemple, deux visualisations de données produites dans le cadre de ces recherches.





Que voyons-nous ? Indiscutablement quelque chose d'intéressant. Un ensemble de signes qui doit être analysé et compris et qui peut se révéler de grande valeur pour qui cherche à localiser et identifier des investisseurs potentiels ou des décideurs. Les noms de famille chinois soulèvent des défis particuliers : ils sont en usage depuis de nombreux siècles et les moins communs ont eu tendance à disparaître au fil du temps de sorte qu'il ne reste aujourd'hui plus qu'une centaine de patronymes différents. Mais les prénoms traduisent encore des différences régionales ou d'autres éléments sémantiques. Les racines ne sont pas forcément très visibles mais l'onomastique peut toujours les rechercher. Et les résultats sont d'autant plus précieux que la quête a été difficile.

More on paristech review

On the topic

[Inde : la biométrie pour tous !](#) By Nandan Nilekani on January 31st, 2013

[Les Big Data de A à Z](#) By Henri Verdier on November 16th, 2012

[Big Data : le réel est mort, vive le réel !](#) By Jannis Kallinikos on November 16th, 2012

[Clics longs, cookies et publicité intrusive: quand Big Brother fait du marketing](#) By Knowledge@Wharton on April 27th, 2012

[Masses de données et marchés fragmentés, les nouveaux défis du marketing](#) By Knowledge@Wharton on November 25th, 2011

[La surveillance généralisée signe-t-elle la mort de l'intimité?](#) By ParisTech Review on September 15th, 2010

[Le téléphone portable redistribue-t-il les cartes?](#) By Vincent Blondel, Pierre Deville, Frédéric Morlot, Zbigniew Smoreda, Paul Van Dooren & Cezary Ziemlicki on November 15th, 2011

By the author

[Onomastique et Big Data](#) on September 17th, 2013

**ParisTech  
REVIEW:**

