

## Cortana, Siri et Google Now vont-ils changer l'univers de la recherche d'information ?

[Retour au sommaire de la lettre](#)

<b>Domaine :</b>	Recherche	<b>Référencement</b>
<b>Niveau :</b>	Pour tous	<b>Avancé</b>

*Les "assistants numériques personnels" comme Siri, Google Now ou Cortana (Microsoft) sont aujourd'hui présents sur tous nos smartphones et envahissent petit à petit nos ordinateurs de bureau. La recherche vocale est en train de s'insérer dans nos habitudes quotidiennes. Le SEO va donc devoir s'adapter et prendre en compte ces nouvelles façons de rechercher l'information sur le Web pour présenter des réponses correspondant aux demandes et attentes des internautes, qui vocalisent de plus en plus leurs requêtes...*

Depuis l'apparition de Siri sur les iPhones d' Apple en 2011, plusieurs systèmes concurrents ont été lancés : Google Now, Cortana de Microsoft, et même un système équivalent (baptisé S Voice) chez le constructeur Samsung. Ces nouveaux outils logiciels démocratisent depuis 3 ans la reconnaissance vocale grâce à leur implantation sur les smartphones les plus populaires.

Ils sont également, et petit à petit, portés sur tous les autres dispositifs connectés à internet, depuis les montres connectées jusqu'à l'ordinateur de bureau. Et parler à son objet connecté devient rapidement un mode d'interaction "normal" pour les utilisateurs.

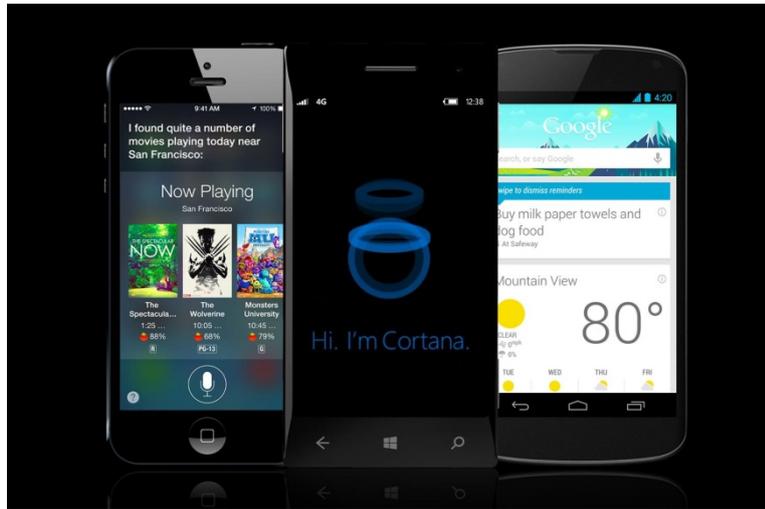
L'adoption rapide des interfaces vocales n'est pas sans conséquences sur les usages des internautes, en particulier sur la façon dont ils utilisent les moteurs de recherche. Ce changement ne restera probablement pas sans influencer le fonctionnement futur des outils de recherche, et la façon d'optimiser les sites pour les moteurs dans un monde où les nouvelles interfaces sans clavier deviendraient la norme.

Après avoir décrit chacun des membres de la famille en rapide évolution des "assistants personnels numériques", nous allons voir que ce qui semble une "révolution" pour les observateurs les moins attentifs n'est en réalité que le résultat d'une lente et constante évolution des technologies de reconnaissance automatique de la parole, et qu'il était logique que cette interface soit un jour proposée sur les smartphones.

Puis nous nous intéresserons à la façon dont Google en particulier a adapté le fonctionnement de son moteur pour être plus efficace dans la recherche d'informations avec une interface vocale. Nous verrons que le moteur Google est déjà prêt pour cette évolution. Mais l'êtes-vous aussi ?

### ***Panorama des assistants numériques***

Quatre ans après le lancement de Siri d'Apple, tous les grands acteurs du monde des smartphones ont lancé leur propre outil concurrent : qu'il s'agisse des éditeurs de systèmes d'exploitation pour mobiles, comme Google ou Microsoft, ou les constructeurs comme Samsung. La famille s'agrandit sans cesse, et il a fallu lui trouver un nom. Le plus répondu est "*personal digital assistants*", le terme ne désignant plus un appareil de type PDA, mais un logiciel, une application.



Source de l'illustration : Digital Trends

## Apple Siri

Siri d'Apple a été la première application lancée. La première version datant de 2009 était vendue sur l'apple Store sous la forme d'une application indépendante. Siri a été inventé par le Stanford Research Institute (SRI), une organisation fondée par l'Université de Stanford. Le projet Siri est un *spin off* d'un projet du SRI financé par l'agence de Défense américaine DARPA, baptisé CALO, dans le cadre de leur programme PAL (*Personal Assistant that Learns*). L'application a été ensuite développée dans le cadre d'une filiale commerciale du SRI, baptisée SIRI Inc. Apple a fini par acheter la société SIRI Inc. en avril 2010, et a intégré Siri dans son système d'exploitation iOS, à partir de la version 5. Siri a été présenté comme l'innovation majeure accompagnant l'iPhone 4S.



L'équipe des « inventeurs » de Siri : de g. à dr. Cheyer, Kittlaus et Gruber

Siri réutilise la technologie de reconnaissance automatique de la parole de la société Nuance. Ses principales caractéristiques sont :

- une interface "conversationnelle" : le système imite le bavardage entre humains, et embarque quelques blagues et surprises pour amuser l'utilisateur ;
- l'exploitation du "contexte personnel" : le système récupère les informations de contexte (localisation, informations personnelles, historique des demandes) pour améliorer la pertinence de son comportement ;
- la "délégation de service": le système Siri est une interface qui traduit les commandes vocales en commandes ou en requêtes qui sont ensuite adressées sous une forme classique à d'autres services, comme un moteur de recherche ou un site internet. La réponse est ensuite restituée *via* l'interface vocale et/ou affichée sur le smartphone.



*L'agent conversationnel de Siri est programmé pour répondre de manière amusante à certaines questions...*

## Google Now

Google a alors très vite réagi face à l'engouement engendré par Siri en déployant deux applications innovantes : Google Voice Search et surtout Google Now.

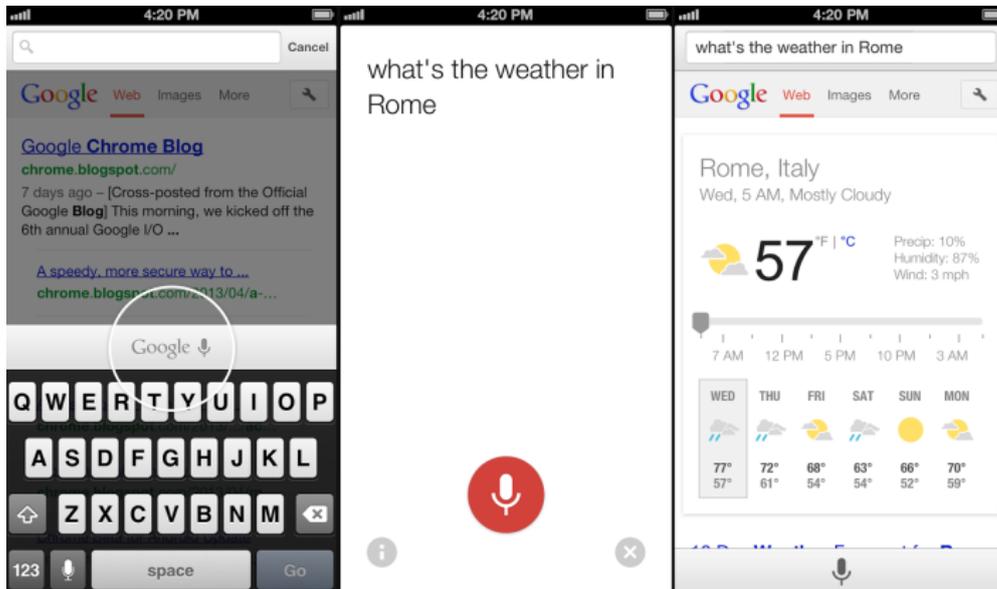
**Google Voice Search** est l'interface de reconnaissance automatique de la parole développée par Google. Apparue d'abord sur les smartphones Android, elle a ensuite été étendue sur d'autres dispositifs, notamment les ordinateurs de bureau et le navigateur Chrome. Google Voice Search reconnaît à la fois des commandes prédéfinies, et est capable de reconnaître une requête qu'il peut ensuite passer au moteur de recherche... Sur Android existe aussi l'interface "Voice Action" qui permet de commander son smartphone à la voix.

Google a développé sa propre technologie de reconnaissance automatique de la parole, dès 2004. Ses premières tentatives d'implémentation sur un smartphone sur iOS et Android datent de 2008.

Lors de la sortie de l'iPhone 4S, Google travaillait déjà sur sa propre solution sous le nom de code Majel (le futur **Google Now**). L'approche de Google s'est révélée différente de celle d'Apple sur plusieurs points :

- une technologie de reconnaissance automatique de la parole plus efficace (confirmée par de nombreux tests) ;
- Google Now n'est pas qu'une interface vocale, en fait l'application est d'abord un assistant personnel capable de vous "pousser" l'information dont vous avez besoin quand vous en avez besoin, et *via* l'interface la plus appropriée. L'interface vocale est un plus, elle n'est pas au centre de l'application ;
- une approche conversationnelle de la recherche d'informations, mais pas d'agent conversationnel programmé pour agir comme votre compagnon (Google Now est sérieux, ne fait pas de blagues) ;
- une exploitation très avancée du contexte personnel : Google Now présente, sous la forme de "cartes", les informations dont vous avez probablement besoin en ce moment en fonction de votre géolocalisation, de votre historique, de vos paramètres etc. ;
- la délégation de services est limitée : les informations poussées viennent des services de Google pour l'essentiel.

Avec l'arrivée d'Android 4.1 (Jelly Bean) sur le Google Nexus en juin 2012, Google Voice Search a été "fusionné" avec Google Now. C'est à partir de cette date que l'on considère que Google Now réunit les attributs lui permettant de rivaliser avec Google Siri (en dépit des différences de concept entre les deux approches).



*La version iOS de Google Voice Search en action*

### **Cortana de Microsoft**

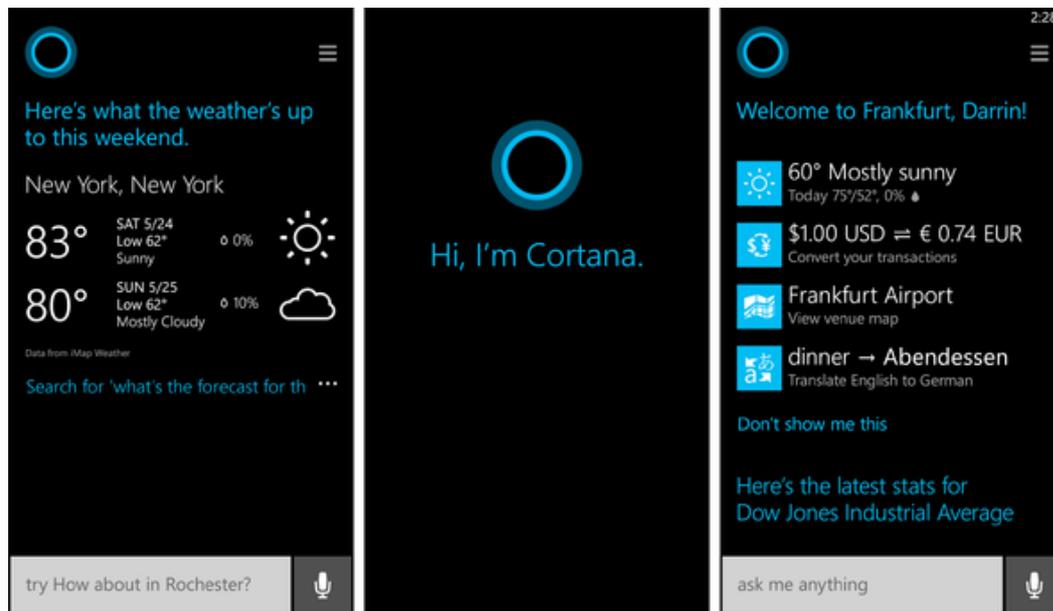
Microsoft avait déjà un long historique dans le développement d'applications de reconnaissance automatique de la parole, ainsi que dans le lancement d'assistants personnels (on se souvient de Clippy l'assistant d'office en 2003). La sortie de Cortana en 2014 n'a donc été une surprise pour personne.

Cortana est donc la réponse de Microsoft à Siri. On retrouve dans l'outil de Microsoft le concept d'agent conversationnel, le "compagnon" capable de bavarder avec l'utilisateur (et même chanter dans le cas de Cortana).

Le nom de Cortana était au départ un nom de code, tiré du personnage du jeu Halo.

Pour le moment, Cortana est disponible en anglais, et en chinois, sur les smartphones équipés de la dernière version de Windows Phone, et sur Windows 8.1. La disponibilité de Cortana dans d'autres pays et d'autres langues devrait rapidement croître, la version française étant attendue pour fin 2014/ début 2015.

Cortana tire parti des services apportés par le moteur de recherche Bing Search, et de son intégration poussée dans le système d'exploitation. Comme pour Google Now, ceci dote Cortana d'une meilleure capacité de réponse que Siri dans certains contextes.



*Quelques écrans typiques de Cortana, le nouvel assistant de Microsoft*

### **Les autres assistants numériques**

Doter un smartphone d'une interface vocale était une évidence pour les constructeurs depuis la disparition des claviers physiques au profit des écrans tactiles. En ce sens, le succès des iPhones a lancé le déclenchement d'une véritable compétition entre constructeurs, éditeurs de systèmes d'exploitation pour smartphones, et spécialistes de la reconnaissance vocale. Parmi les plus notables, on peut citer :

- les solutions de Nuance (Dragon notamment) ;
- S Voice de Samsung ;
- Voice Mate de LG (le successeur de Quick Voice, basé sur l'assistant personnel Maluuba) ;
- Ask Ziggy (l'ancêtre de Cortana chez Microsoft) ;
- Assistant de Speaktoit.

### ***La longue histoire de la reconnaissance automatique de la parole***

La démocratisation rapide des assistants numériques basés sur la reconnaissance automatique de la parole pourrait laisser penser qu'il s'agit du résultat d'une brutale avancée technologique. Comme souvent, il n'en est rien : ces technologies sont anciennes, et progressent lentement depuis de nombreuses années.

#### **Une vieille idée**

Les premiers travaux en matière de reconnaissance de la parole datent du début du XXe siècle, et la première application concrète considérée comme réussie date de... 1952 (projet des laboratoires Bell). L'approche numérique a remplacé l'approche analogique à partir de 1960. La première application commercialisée date de 1972. Dès la fin des années 70 apparaissent des cartes électroniques capables de comprendre les commandes vocales. En 1983, Dassault commercialise le premier système de commandes vocales dans un avion de chasse !

Comment fonctionne une application de reconnaissance automatique de la parole ?

Il faut bien séparer le problème de la synthèse vocale ("vocaliser" un texte à l'aide d'une voix synthétique), qui est un problème relativement simple et aujourd'hui parfaitement au point (ou presque), de la reconnaissance automatique de la parole (ou RAP : la "compréhension" des sons produits par une voix humaine pour les traduire en texte ou en commandes) qui s'est avéré être un problème redoutable et très complexe à résoudre.

Il faut dans un premier temps "isoler" les sons émis par la voix humaine des autres "bruits" qui peuvent être captés par un micro. Pour réussir cela, on s'appuie sur le fait que la voix humaine couvre des fréquences caractéristiques (différentes pour les femmes et les hommes) ce qui permet d'éliminer tout bruit émis à d'autres fréquences. Néanmoins, on s'aperçoit facilement à l'usage que les systèmes de RAP fonctionnent très mal dans un environnement bruyant, ou pire, quand plusieurs personnes parlent en même temps.

L'autre difficulté consiste à reconnaître les "sons", les "phonèmes" émis par la voix humaine. On est très vite confronté à la très grande variabilité des prononciations d'une langue à l'autre, d'un accent à l'autre et d'une personne à l'autre.

L'une des solutions pour améliorer le système est bien sûr de l'adapter à une langue, et même un accent en particulier. On améliore très efficacement les systèmes de reconnaissance de commandes vocales en passant par une phase d'apprentissage : on fait lire à l'utilisateur un texte afin que le système puisse disposer d'une référence de phonèmes prononcés par un utilisateur particulier. Hélas, cela signifie que l'on crée un système qui n'est pas universel, ce qui pose des problèmes sérieux pour des applications grand public.

Le problème le plus complexe vient ensuite : le flux des sons émis par une voix humaine peut être continu, il n'est pas découpé en "mots". Dans ces conditions, reconnaître qu'une suite de deux phonèmes correspond à un mot, et n'est pas une partie d'un autre mot, n'a rien de trivial. Il faut pour y arriver, imiter le fonctionnement du cerveau humain en s'appuyant :

- sur une mémoire de sons très importante ;
- sur la grammaire : une analyse syntaxique va aider à désambiguïser une suite de sons dans le contexte et aider à identifier à quelle forme de quel mot on a affaire ("est" ou "ait" ?) ;
- sur la connaissance du contexte.

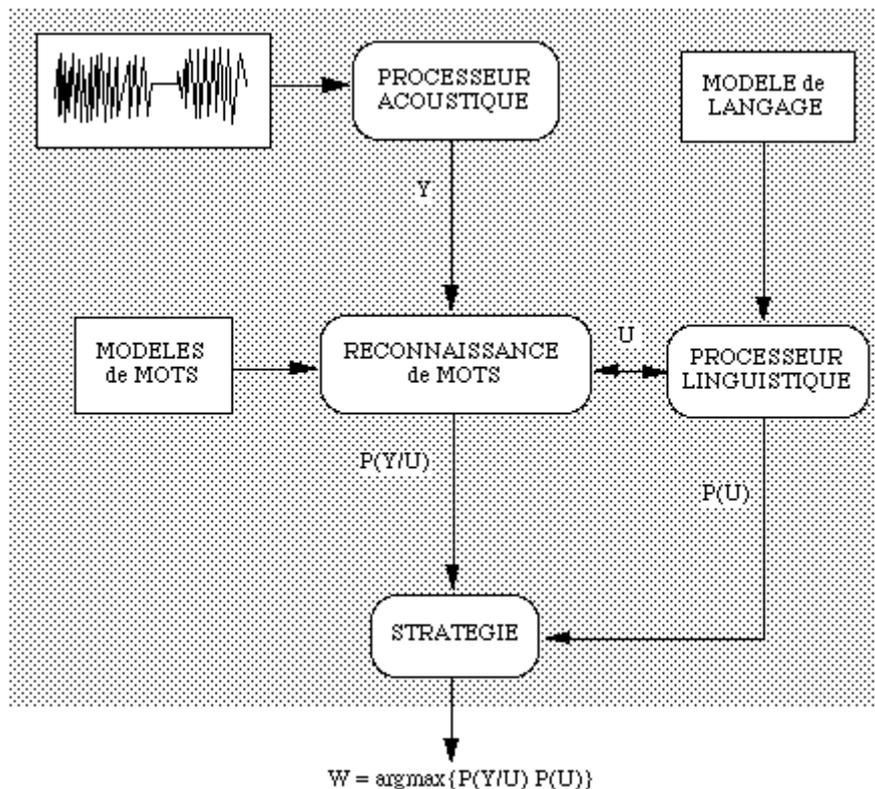


Schéma simplifié d'un système de reconnaissance automatique de la parole

Dans la pratique, les applications de reconnaissance automatique de la parole s'appuient sur les probabilités pour identifier le "mot" le plus probable correspondant aux sons prononcés. On peut ainsi déterminer :

- quelle est la probabilité de voir apparaître une certaine suite de mots dans une langue (modèle de langage) ;
- la probabilité qu'une prononciation particulière pour une syllabe donnée soit utilisée dans un contexte donné (modèle de prononciation) ;

- la probabilité qu'une séquence de vecteurs acoustiques corresponde à une suite de mots (modèle acoustico-phonétique) ;

La plupart des logiciels modernes de RAP utilisent analyse combinée de ces trois probabilités pour fonctionner.

### Pourquoi maintenant ?

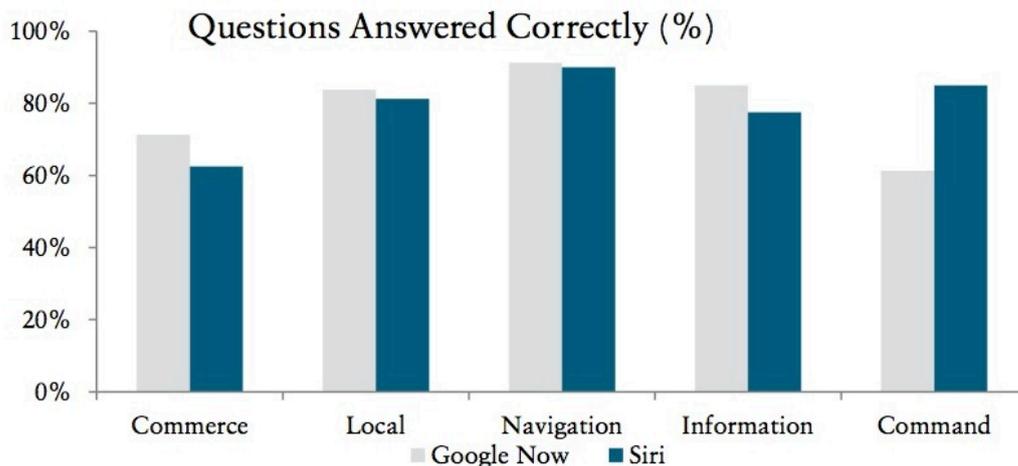
Les progrès de la RAP ont été lents et chaotiques. En effet, la reconnaissance automatique de la parole peut fonctionner avec un taux de reconnaissance proche de 100% dès lors que l'on limite le champ sémantique à reconnaître, et les variables d'environnement... Mais à chaque fois que l'on veut rendre l'application plus "universelle", le taux recule, y compris pour les "mots" que le système reconnaissait parfaitement avant.

Pour améliorer le système, l'exploitation de bases de données gigantesques est indispensable. L'analyse de la prononciation, dans toutes les langues à couvrir, avec tous les accents, dans tous les contextes possibles, de tous les termes utilisés est nécessaire.

Dans ces conditions les progrès de la RAP suivent les progrès technologiques en informatique : plus les processeurs sont puissants, plus les capacités de stockage d'information sont importants, plus il est possible de proposer des systèmes de RAP performants. Il faut également que les progrès suivent dans le domaine des algorithmes, car l'amélioration de la couverture d'un système de RAP crée de nouveaux challenges à chaque fois.

Mais après quarante ans, les programmes automatiques de reconnaissance automatique de la parole peuvent enfin être installés sur des appareils aussi compacts que dans un smartphone, en apportant un véritable service à l'utilisateur.

Ceci dit, le taux de reconnaissance des assistants personnels numériques à la Siri est certes acceptable, mais on est encore loin des 100%. Gageons que la compétition technologique actuelle aboutisse à une diminution notable du taux d'erreurs, mais l'histoire a démontré que les progrès dans ce domaine ne pouvaient pas être rapides, et qu'il faudra plusieurs années avant que l'on puisse considérer ces systèmes comme fiables et réellement capables de nous faire gagner du temps.



*Une étude comparée des résultats de Google Now et de Siri. Siri obtient de meilleurs résultats que Google Now en reconnaissance de commandes vocales, mais Google Now est meilleur pour toutes les autres requêtes.*

	Google Now	Siri	Cortana
"Where is the NASDAQ today?"	4.79 seconds	5.9 seconds	6.14 seconds
"What's the traffic like around Washington Avenue?"	5.97 seconds	8.82 seconds (launches Apple Maps)	6.05 seconds
"What's the weather like?"	4.26 seconds	4.33 seconds	4.22 seconds
"What's the population of Germany?"	4.82 seconds	4.51 seconds	5.86 seconds
"What's the time in London?"	4.63 seconds	4.1 seconds	6.29 seconds
Overall average	Averaging 4.894 seconds	Averaging 5.53 seconds	Averaging 5.712 seconds

Comprendre une requête en langage naturel et renvoyer la bonne réponse nécessite des calculs importants et cela prend du temps, sans compter le temps nécessaire pour extraire les résultats. Encore aujourd'hui, des délais de plusieurs secondes sont la norme.

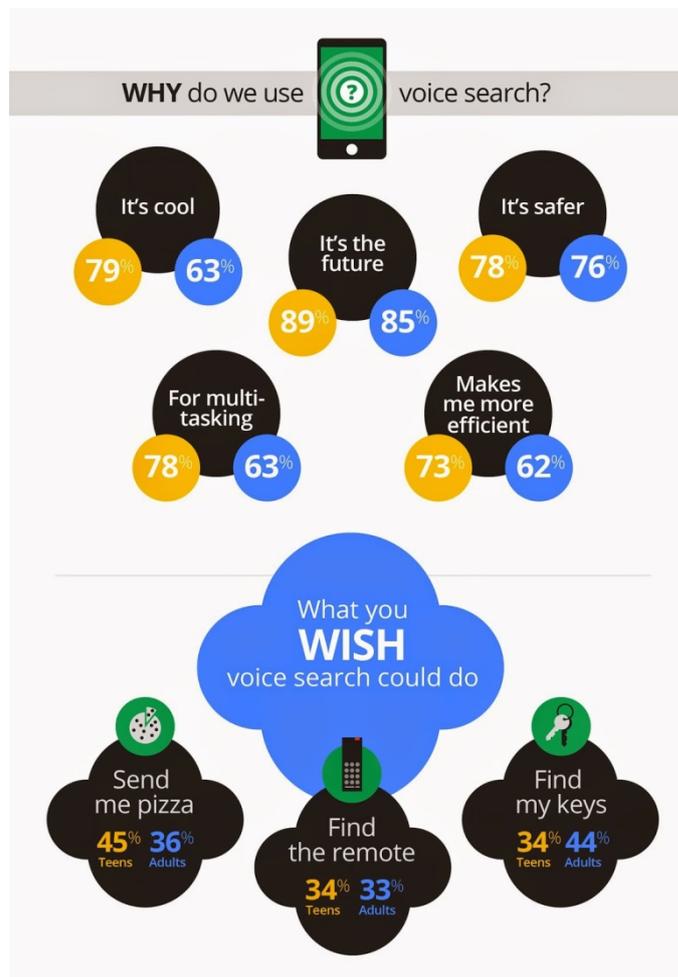
### ***Les assistants personnels numériques changent les usages des moteurs de recherche***

En attendant, l'adoption par les utilisateurs de ces nouvelles interfaces vocales commence d'ores et déjà à changer sérieusement la façon dont ils cherchent l'information à l'aide d'un moteur de recherche comme Google.

#### **L'adoption de cette interface par les utilisateurs est en plein boom**

Récemment, Google a communiqué les premiers chiffres sur l'usage de Voice Search : selon cette étude publiée le 14 octobre 2014, 55% des adolescents et 41% des adultes américains utilisent la recherche vocale quotidiennement.

Le contexte d'utilisation le plus fréquent semble être la recherche d'informations pendant que les utilisateurs regardent la télévision. Plutôt que regarder un "deuxième écran" (ce qui est évidemment difficile), commander à la voix un deuxième dispositif semble devenir une habitude fréquente.



Les usages de la recherche vocale d'après l'étude publiée par Google le 14 octobre 2014

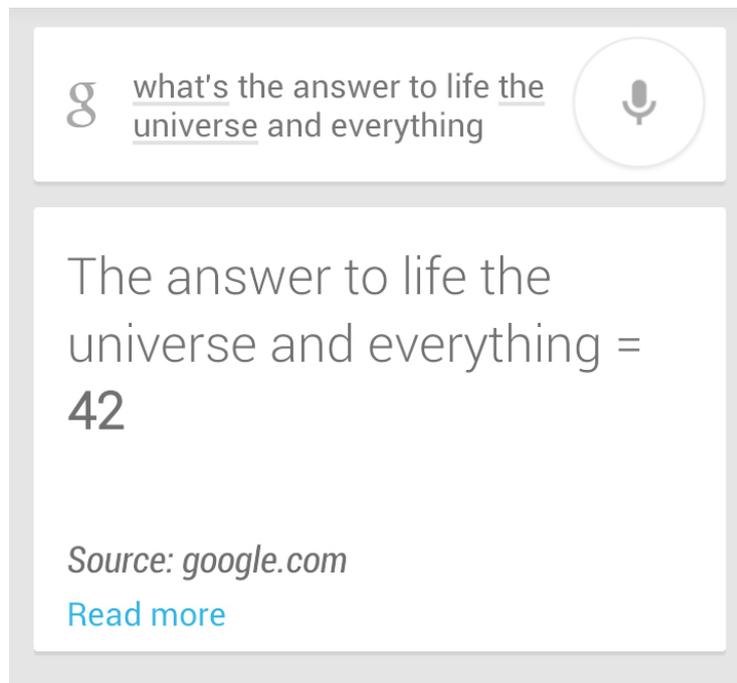
### Les requêtes en langage naturel explosent

Un tel niveau d'adoption ne peut que changer la nature des requêtes envoyées au moteur de recherche. Pour le moment, aucune information n'a filtré sur le nombre de requêtes en langage naturel que les interfaces vocales envoient à un moteur de recherche comme Google. Mais on peut parier que ce nombre est en train d'exploser.

Grâce aux interfaces de type Google Voice Search, il est devenu beaucoup plus simple pour un utilisateur de formuler une question en langage naturel. Or de telles requêtes sont beaucoup plus riches en informations que l'entrée de deux ou trois mots clés isolés. L'analyse de la syntaxe de la question permet en effet de comprendre beaucoup mieux quel est l'objet de la requête, et donc de fournir une réponse plus pertinente. Si, en plus, le moteur dispose d'informations de contexte personnel, comme celles exploitées par Google Now, notamment l'historique des recherches et du comportement de l'utilisateur, les chances de fournir une information utile en réponse à une requête augmentent sérieusement.

### Une base de faits très large est nécessaire

Mais en même temps, fournir ces réponses pertinentes demande de disposer d'une base de faits très large. Par exemple, si vous demandez via Google Voice Search "Quel âge a Nicolas Sarkozy" une seule réponse sera vous sera renvoyée par synthèse vocale: "59 ans". Cette information est tirée du Knowledge Graph de Google qui représente l'un des atouts majeurs de Google Now pour lutter contre un concurrent comme Siri.



*La base de faits de Google permet à Google now de trouver la bonne réponse à cette question tirée du roman « le guide du Routard Galactique ».*

### **L'apport de l'indexation des concepts (Hummingbird)**

Mais certaines questions sont plus larges, moins précises, et n'appellent pas une réponse formulée sous une forme aussi immédiate et précise. Par contre, une question formulée en langage naturel permet d'identifier le "concept" auquel la question fait allusion.

Par exemple la question "où manger la meilleure Pizza de Paris" appelle des réponses qui évoquent des lieux où on peut manger des Pizzas, donc des Pizzerias, pas une page donnant la recette de la meilleure Pizza.

Pour exploiter ce supplément d'informations apporté par une question formulée en langage naturel, il faut être capable d'identifier les concepts demandés, et de filtrer les réponses données en fonction de ce concept.

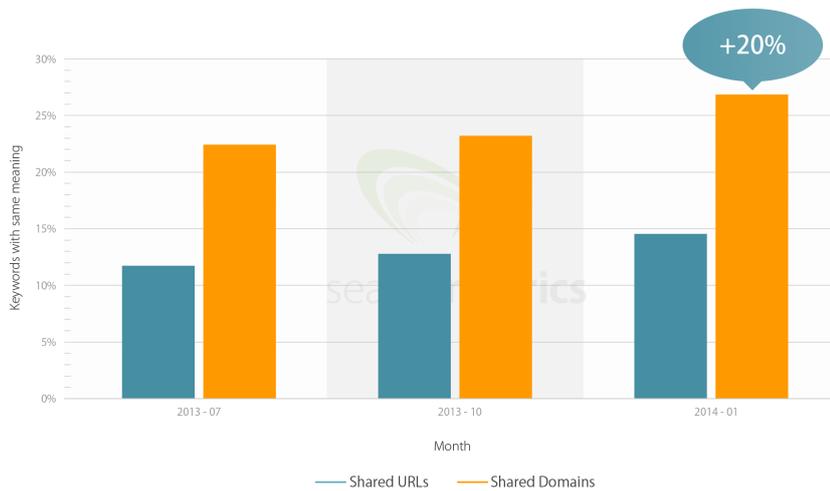
L'une des supériorités actuelles de Google Now par rapport à Siri réside moins dans une application de reconnaissance de la parole plus efficace (même si certains affirment, tests à l'appui, que c'est le cas) que dans la bonne qualité de la reconnaissance des concepts demandés.

Pour améliorer la qualité des résultats renvoyés, il restait aussi à stocker l'association entre les pages et les concepts. C'est la révolution apportée par Hummingbird, la nouvelle version de l'architecture d'index de Google lancée en septembre 2013, et qui autorise l'indexation des concepts, et permet de retourner les pages associées à un concept lors d'une requête.

### **L'usage des assistants numériques diminue l'intérêt des requêtes par mots clés**

Pour le moment, la part des requêtes effectuées en langage naturel reste encore limitée, et la plupart des webmasters n'ont pas encore détecté les changements apportés à la structure des requêtes qui aboutissent sur leur site. Les changements apportés par Hummingbird aux pages de résultats de Google n'ont pas été vraiment remarqués. Pourtant, Hummingbird, d'après Google, a "redistribué" les classements sur 90% des requêtes, certes de manière subtile, mais le nombre de requêtes impactées est beaucoup plus important que pour une mise à jour de Panda et de Penguin.

Une étude de Searchmetrics a démontré qu'Hummingbird avait "rapproché" la liste des résultats sur certaines requêtes qui appelaient les mêmes concepts.



*Un graphe illustrant les résultats de l'étude de Searchmetrics : sur un échantillon de couple de requêtes différentes mais appelant le même concept, on trouve 20% de pages similaires en plus après Hummingbird dans les résultats de Google.*

## **Un changement d'usage qui influence le fonctionnement des moteurs et le référencement**

Cette double évolution (augmentation des requêtes en langage naturel, et indexation des faits et des concepts) produit progressivement un changement fondamental dans le fonctionnement des moteurs. Il s'agit moins pour Google ou Bing de chercher les pages qui contiennent des mots clés, que de trouver la réponse à des questions.

Et cela produit trois changements fondamentaux :

- les requêtes deviennent extrêmement variées dans leur formulation : la "longue traîne" des requêtes en langage naturel représente une part de plus en plus considérable des recherches, au détriment des requêtes sous formes de mots clés. Optimiser un site pour une expression clé particulière perd progressivement de sa pertinence (même si c'est encore pertinent aujourd'hui) ;
- les assistants à la Siri, Google Now et Cortana, veulent réaliser directement l'action demandée, ou renvoyer un fait en réponse à la question posée. Et de moins en moins renvoyer vers les pages d'un site. La part du trafic envoyée par les moteurs aux sites web qu'il référence, déjà en baisse tendancielle constante, va encore reculer à l'avenir avec ces nouveaux assistants numériques et leurs interfaces vocales ;
- les moteurs comprennent mieux ces questions en langage naturel, et exploitent de plus en plus et mieux les informations de contexte. Grâce à cela, ils identifient mieux quels sont les concepts auxquels les questions font allusion, et cherchent à renvoyer les meilleurs résultats associés à ces concepts, pas forcément les pages comportant les mêmes mots clés que la question. C'est donc (peut-être !) la fin de dictature du mot clé pour le référencement.

Est-ce que ces évolutions sonnent la fin du SEO ? Pas du tout. Certes, cela annonce la perte d'efficacité de certaines méthodes traditionnelles. Mais cela annonce surtout la nécessité de développer de nouvelles approches.... Et en particulier de se demander :

- comment faire en sorte que ses pages soient bien classées sur un concept donné dans un moteur post Hummingbird. Cela suppose d'identifier les concepts reconnus par Google, comment il associe pages et concepts et comment il classe les pages sur ces critères. C'est un nouveau champ de recherche et d'expérimentation passionnant pour les référenceurs ;
- comment proposer la réponse à une question posée par les assistants de type Siri, Google Now et Cortana. S'il s'agit de requêtes informationnelles, Bing et Google veulent donner la réponse directement, et chercher à être référencé sur ces questions appelant des faits restera peut être vain. Par contre s'il s'agit d'une requête navigationnelle ou transactionnelle, un nouveau champ d'optimisation s'ouvre pour l'avenir. Par exemple : 36% des utilisateurs américains de Google Now souhaitent que leur application permette de commander dans les années qui viennent.

En attendant, Google Now, Siri ou Cortana sont en rapide évolution. Et il est difficile de prédire aujourd'hui avec certitude quels autres usages et possibilités vont être intégrés dans ces assistants virtuels à l'avenir. Mais on peut parier sur le fait que nous ne sommes pas encore au bout de nos surprises... et d'autres changements impactant notre manière de chercher de l'information sur le Web sont à prévoir dans les 10 années qui viennent.

## Liens et bibliographie

Le site de Siri d'Apple :

<http://www.apple.com/fr/ios/siri/>

La rubrique du site de Microsoft sur Cortana :

<http://www.windowsphone.com/en-us/how-to/wp8/cortana/meet-cortana>

L'aide de Google sur Google Now :

<https://www.google.com/intl/fr/landing/now/>

et l'aide de Google sur Google Voice Search :

<http://www.google.com/insidesearch/features/voicesearch/index-chrome.html>

L'étude de Google sur l'usage de Voice Search :

<http://googleblog.blogspot.fr/2014/10/omg-mobile-voice-survey-reveals-teens.html>

**Philippe YONNET**, *Directeur Général de l'agence Search-Foresight, groupe My Media*  
(<http://www.search-foresight.com>).