

Extrait du Spyworld Actu

<http://www.spyworld-actu.com/spip.php?article2948>

# **Le Pentagone veut se doter d'une super-machine à traduire**

- Renseignement - International -



Date de mise en ligne : mercredi 8 novembre 2006

---

**Spyworld Actu**

---

La DARPA, agence du département américain de la Défense à l'origine de l'ancêtre d'Internet à la fin des années 60, s'attelle à un nouveau projet : mettre au point d'ici 2010 un logiciel capable de traduire quasi-simultanément l'arabe ou le mandarin avec une précision de 90% à 95%, pour le plus grand profit des services de renseignement.

Pour eux, un des plus grands défis est de reconnaître rapidement des informations importantes dans des langues étrangères. C'est pourquoi l'armée américaine aimerait se doter d'une machine capable d'écouter émissions de télévision et conversations téléphoniques ou de lire des sites Web en arabe et en chinois, de traduire ces contenus et d'en résumer les éléments-clés.

L'an dernier, l'Agence des projets de recherche avancée de la Défense (DARPA) a lancé le programme Gale, acronyme anglais d'Exploitation globale autonome du langage, qui vise à concevoir un tel logiciel de traduction en temps réel.

Son objectif d'arriver à une précision de 90% à 95% pourrait s'avérer impossible. Même l'homme pourrait être incapable d'atteindre un tel degré d'exactitude, étant donné les risques de malentendus entre interlocuteurs, d'incompréhension d'expressions idiomatiques ou d'interprétation erronée du message.

Heureusement pour les chercheurs de Gale, ils n'ont pas eu à s'élever à ce niveau d'excellence tout de suite. La première année, ils devaient traduire à l'oral l'arabe et le mandarin avec une précision de 65%, le but étant fixé à 75% pour le texte.

La DARPA a engagé trois sociétés en concurrence les unes avec les autres pour le projet : IBM, l'institut de recherche à but non lucratif SRI International, et la firme BBN Technologies. Chaque année, leurs travaux seront évalués et l'équipe la moins performante pourrait être éliminée.

Pour BBN, qui a permis à l'agence Associated Press d'avoir un rare aperçu des activités habituellement confidentielles menées pour la DARPA, l'enjeu est de taille. Cette société réalise 80% de son chiffre d'affaires avec l'armée, et un gros contrat comme Gale lui a déjà rapporté 16 millions de dollars (12,6 millions d'euros) la première année.

Anciennement appelée Bolt, Beranek and Newman, du nom de ses fondateurs, la firme est connue pour ses travaux sur les réseaux informatiques dans les années 60, qui débouchèrent plus tard sur Internet. A la fin de cette décennie, la DARPA a mis au point le réseau Arpanet, ancêtre du Web.

BBN est également spécialisée dans les technologies de reconnaissance vocale et de traduction, mais a dû recruter des chercheurs extérieurs pour la mission Gale. "Nous n'avons jamais eu un projet de cette complexité", soulignait le chercheur de BBN Owen Kimball en avril. "Il y en a qui vont s'arracher les cheveux."

Avant le lancement du projet, BBN affirmait pouvoir traduire automatiquement les journaux télévisés étrangers avec une précision supérieure à 80%. Mais la DARPA s'intéresse non seulement à ces programmes au langage policé et de bonne tenue, mais aussi aux interviews du citoyen lambda dans la rue et aux "chats" (forums) utilisant un ton familier sur Internet.

C'est là où les choses se gâtent : bruits de fond, dialectes, accents, argot, mots escamotés... un véritable cauchemar pour les ingénieurs. Sans parler du fait que l'arabe et le mandarin ont une structure très différente de l'anglais, ce qui rend la tâche encore plus ardue.

## Le Pentagone veut se doter d'une super-machine à traduire

---

Les trois équipes de chercheur en concurrence ont stocké une pléthore d'enregistrements et de textes en arabe et chinois dans leurs ordinateurs. Le but est de régler avec précision le processus informatique, baptisé algorithme, qui réalise l'analyse du langage.

La DARPA a réalisé le premier test d'évaluation en juin : des milliers d'heures d'enregistrements audio et des millions de pages à traduire. Elle a jugé les traductions au nombre de corrections humaines nécessaires pour donner aux phrases leur sens réel.

Résultat : BBN a obtenu à l'écrit une précision de 75% en arabe et mandarin, et à l'oral 69% pour l'arabe et 67% pour le mandarin. IBM a obtenu un meilleur score à l'écrit pour arabe et le SRI a fait mieux pour le mandarin.

Les objectifs de première année fixés par la DARPA ont été largement remplis. Aucune équipe n'a été éliminée, au moins pour le moment. Mais elles vont continuer à être testées et à devoir faire leurs preuves pendant encore quatre ans maximum.

*Post-scriptum :*

<http://fr.news.yahoo.com/08112006/5...>