

Adscriptor

Ads & Marketing Translator - www.Adscriptor.com - le Web à la loupe

Confiez-moi vos idées, je les traduirai en mots. C'est mon métier !

samedi, juillet 29, 2006

Google RS2, traducteur automatique de deuxième troisième génération

Préambule

Cadre de la situation

Qu'est-ce que Google RS2 ? (selon moi :-)

P.S. Qu'est-ce que Google h9 ? (toujours selon moi :-)

* * *

Préambule

Dans mon précédent [billet](#), je vous invitais à lire une [info de Zorgloob](#) reprenant un [scoop de Tony Ruscoe](#) sur une panoplie de nouveaux services testés par Google.



Pas question de revenir sur le sujet, mais plutôt de tenter d'approfondir l'un de ces services, nom de code **Google RS2**, déjà déniché par [Tony Ruscoe](#) en juin dernier sans avoir la moindre idée de ce dont il s'agissait, et qui nous dit aujourd'hui que ce pourrait être une fonctionnalité en rapport avec le système de traduction automatique sur lequel travaille la firme, puisque le lien pointe vers [Google Translate](#) :

The link in the 'My Account' page points to Google Translate, so could it be anything to do with the statistical machine translation system that we know Google's working on?

[\[Début\]](#)

* * *

Cadre de la situation

Étant passionné de traduction autant par goût que par métier, j'avais déjà eu l'occasion d'aborder [Google et la traduction automatique](#), un billet dans lequel je traçais grossièrement les contours du système de Google, dont l'approche :

consiste à gaver les machines de milliards de mots de texte (ce n'est pas la matière première qui lui manque !), ce qu'on appelle la linguistique de corpus (parallèles, alignés ou non), en associant des corpus (ou corpora pour les puristes) monolingues à des bi-textes (en prenant par exemple un site bilingue, ou tri- ou n-lingue, dont les textes sont segmentés puis alignés afin de fournir une

mémoire de traduction) pour y appliquer ensuite des techniques d'apprentissage statistiques permettant de construire des modèles de traduction.

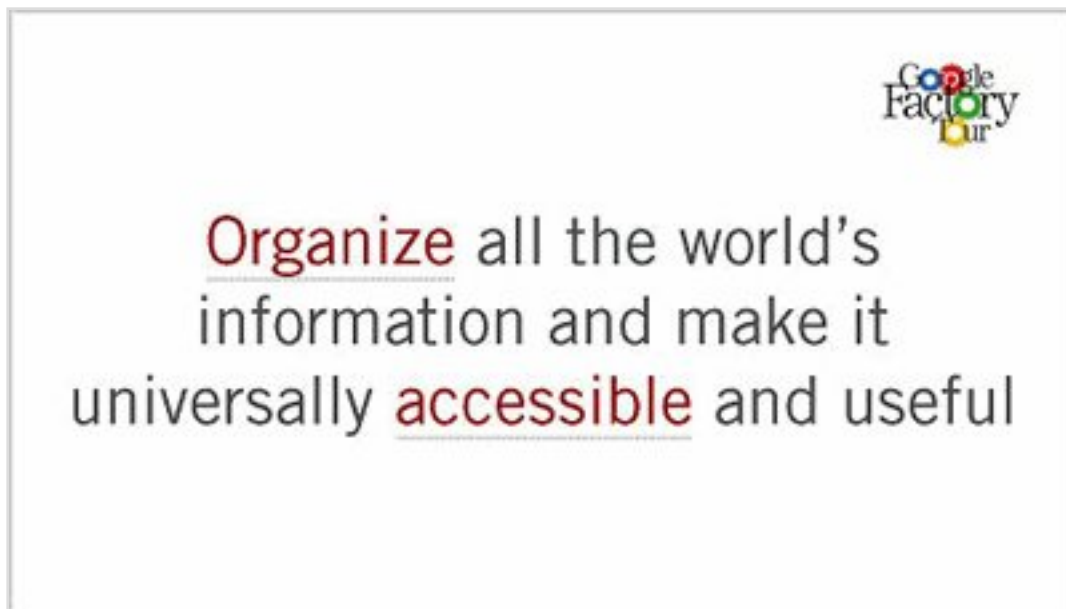
Et de constater que Google était vraiment très fort, puisque, dernier arrivé dans le domaine de la TA (où les places sont chères, par ailleurs), la société se classait au premier rang (note : 0.5131) en devançant IBM (0.4646), etc., et en se détachant très nettement de Systran (0.1079), aussi bien pour la paire linguistique chinois-anglais qu'arabe-anglais, selon les résultats officiels de l'évaluation sur la traduction automatique faite par le National Institute of Standards and Technology : [NIST 2005 - Machine Translation Evaluation Official Results](#).

Or l'édition 2006 s'est terminée hier, comme indiqué sur le [Formulaire d'enregistrement](#) au *2006 NIST Machine Translation Evaluation* :

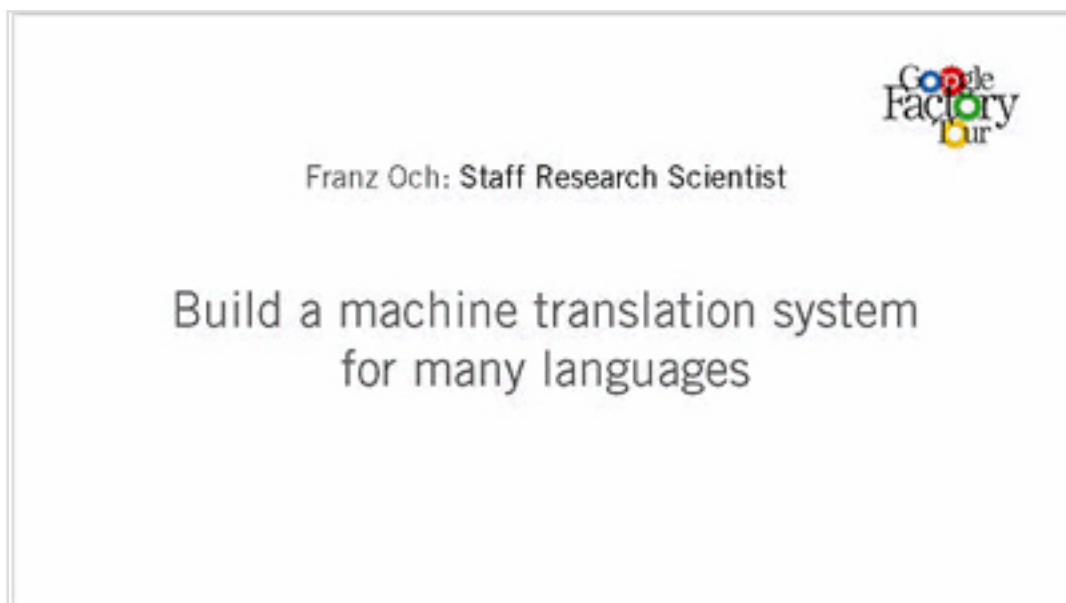
MT06 Evaluation	
July 2006	Evaluation of MT: <ul style="list-style-type: none"> • Chinese to English • Arabic to English
01/31/06	Cutoff date for training data. Systems to be evaluated must not use training data created after 1/31/2006. This cutoff date guarantees that the evaluation data will be new data, never before seen by the system.
05/31/06	Registration to participate deadline. All participants are required to complete the NIST 2006 MT Evaluation Registration form.
07/24/06	Evaluation test data E-mailed to participants.
07/29/06	System translations due at NIST by 12 noon EDT.
09/02/06	Preliminary release of results to the participants.
tentative: 09/6-7/06	The evaluation workshop, open to participants only, will be held in the Baltimore/Washington area.
10/01/06	Official public release of results.

Donc, les participants ont reçu le test confidentiel de traduction automatique par courriel le 24 et l'ont rendu hier à 16h (12h [EDT](#)). Selon le [protocole](#), le test devrait globalement se composer de textes (environ 40 mille mots), mais aussi de [discours](#), il s'agit donc d'une **évaluation données-voix**. Documentation supplémentaire [ici](#).

Pour preuve que Google a de la suite dans les idées, consultez les diapositives du [Google Inc. Factory Tour du 19 mai 2005](#), et notamment les numéros 135 :



qui réaffirme la [mission de Google](#), 136 :



puis de 137 à 140, qui donnent un aperçu de la précision du système (ici la n° 140) :



et enfin 145, dans le prolongement direct de ce qui précède et en rapport tout aussi direct avec ce qui suit...



* * *

Première constatation

Le système de TA de Google est la fine fleur des technologies de pointe, ce qui se fait de mieux en la matière (le moment venu, nous verrons d'ailleurs si les résultats 2006 confirment ceux de 2005, mais je ne doute pas un instant que, là aussi, Google se classera premier), et **n'a absolument aucun rapport avec ce qu'on trouve actuellement sur la page de Google translate**, qui n'est autre que Systran (bien que ce ne soit pas mentionné sur sa page de FAQ). Donc, pour l'instant, Christophe Asselin a encore raison :-)

J'en ai d'ailleurs la preuve ! J'avais précédemment tenté une expérience, relatée ici, en retraduisant du français vers l'anglais, avec l'outil de TA de Yahoo (Babelfish), un extrait (traduit de l'anglais vers le français) de mon billet sur la dérive publicitaire sur Internet et les incohérences de Google, afin de tester la pertinence de la traduction automatique sur Internet. J'ai donc soumis le même extrait à Google translate :



Le résultat est concluant :

<u>Mon billet traduit par</u> <u>Yahoo Babelfish</u>	<u>Mon billet traduit par</u> <u>Google Translate</u>

Did you already make a research on Internet lately ? Then you will undoubtedly have observed the lack of quality of the results, including at the competitors of Google such as Yahoo! or Ask.com ? And their new design changes nothing there. I am pained by it as much as you. Me also I remained wedged hours and hours in the reels of these farms of bonds, these splogs and other " scraper sites " : sites which seem to have contents, (...), whereas actually, generally it is recycled material coming without their knowledge of other content providers, assembled by automated processes.

Did you already make a research on Internet lately? Then you will undoubtedly have observed the lack of quality of the results, including at the competitors of Google such as Yahoo! or Ask.com? And their new design changes nothing there. I am pained by it as much as you. Me also I remained wedged hours and hours in the reels of these farms of bonds, these splogs and other "scraper sites": sites which seem to have contents, (...), whereas actually, generally it is recycled material coming without their knowledge of other content providers, assembled by automated processes.

The hurluberlus which create this kind of sites have one objective : you to bring to click on the sponsored bonds of Google AdSense.

The hurluberlus which create this kind of sites have one objective: you to bring to click on the sponsored bonds of Google AdSense.

Naturally, these sites do not respect of anything the commercial conditions provided by Google, which does not prevent that they hatch per million. Make just a research on "adsense ready web site " and you will have an idea of this miteuse industry, as discrete as an elephant in a porcelain store, which extracts a maximum of profits from the phenomenon.

Naturally, these sites do not respect of anything the commercial conditions envisaged by Google, which nevertheless they hatch per million. Make just a research on "adsense ready web site" and you will have an idea of this miteuse industry, as discrete as a bull in a china shop, which extracts a maximum of profits from the phenomenon.

How Google is it killing goose that lays the golden eggs.

How Google is it killing goose that lays the golden eggs.

The problems are numerous, and all the éradiquer could be expensive Google (when well even it would not be impossible). The first concerns the good old man plagiarism.

The problems are numerous, and all the éradiquer could be expensive Google (when well even it would not be impossible). The first concerns the good old man plagiarism.

The majority of these Spam boxes do not have that to attract the barge cliquor, but also to facilitate to the fraud with the clicks on a large scale (...)
Besides the public generally the purpose of no idea is from what occurs, even if the Net surfers more informed a little are well-informed. We are unaware of of them only the true proportions. Of aucuns advance a rather

The purpose of the majority of these boxes with spam are not only to attract the barge cliquor, but also to facilitate the fraud with the clicks on a large scale. (...)
Moreover the public generally does not have any idea what does without, even if the Net surfers informed a little more are well-informed. We are unaware of of them only the true proportions.

restricted percentage of clicks; others say half straightforwardly. In any event, Google, which has very to lose if the things are as badly as certain journalists citizens denounce it, minimizes the figures considerably.

Of aucuns advance a rather restricted percentage of clicks; others say half straightforwardly. In any event, Google, which has very to lose if the things are as badly as certain journalists citizens denounce it, minimizes the figures considerably.

Another factor facilitating this explosion of spammé contents and fraud to the clicks is the seizure of Google on the level of research on Internet, which one can roughly speaking estimate at 60%, in fact a quasi-monopoly. For as much, the consistent argument with saying that the market is enough large to be corrected all alone on the matter is somewhat naive. Without competition and information, the markets do not correct anything the whole, and I do not think that there are today serious competitors able to make the weight. For the moment.

Another factor facilitating this explosion of spammé contents and fraud to the clicks is the seizure of Google on the level of research on Internet, which one can roughly speaking estimate at 60%, in fact a quasi-monopoly. For as much, the consistent argument with saying that the market is enough large to be corrected all alone on the matter is somewhat naive. Without competition and information, the markets do not correct anything the whole, and I do not think that there are today serious competitors able to make the weight. For the moment.

Take the case of the advertisers of AdWords which believe that their results are diluted because of the fraud to the clicks. Be sure that they know that the biddings on their key words should cost them less, but can be made hear? And can they allow it when there are of them thousands and thousands of other loans to be paid more because: (A) they are unaware of all the problem; (b) they know it only too, but are able thanks to the fraud with the clicks to recover a part of what they spend on other side?

Take the case of the advertisers of AdWords which believe that their results are diluted because of the fraud to the clicks. Be sure that they know that the biddings on their key words should cost them less, but can be made hear? And can they allow it when there are of them thousands and thousands of other loans to be paid more because: (A) they are unaware of all the problem; (b) they know it only too, but are able thanks to the fraud with the clicks to recover part of what they spend on other side?

I do not know the answers, and nor even all the questions. But if the problem is as serious as some fear it, it is all the model of income set up by Google which could suffer from it, even the whole of the business of the pay-per-click. While waiting for the explosion of the farms of bonds and the splogs shows clearly that the fraud with the clicks is juicy and flourishing, at the expense of all the Net surfers. Except of Google.

I do not know the answers, and nor even all the questions. But if the problem is as serious as some fear it, it is all the model of income set up by Google which could suffer from it, even the whole of the business of the pay-per-click. While waiting for the explosion of the farms of bonds and the splogs shows clearly that the fraud with the clicks is juicy and flourishing, at the expense of all the Net surfers. Except Google.

For the moment, there still.

For the moment, there still.

Never more !

Never more !

One can believe or not that all these sites bourriels which make trade of AdSense are honest or defrauders in mass, but me that of which I is sure it is that it is necessary to thank Google for this situation. There is indeed no reason to put on line this kind of sites if one cannot draw an income thanks to the model from it from business set up by Google. And with Yahoo! and the other loans to enter the dance, I do not believe that one will see the tendency to be reversed as soon as.

One can believe or not that all these sites bourriels which make trade of AdSense are honest or defrauders in mass, but me that of which I is sure it is that it is necessary to thank Google for this situation. There is indeed no reason to put on line this kind of sites if one cannot draw an income thanks to the model from it from business set up by Google. And with Yahoo! and the other loans to enter the dance, I do not believe that one will see the tendency to be reversed as soon as.

I hope only that all that predicts of a return to good old days, now that the blogosphère am likely to lose always more his importance, considering whom it starts to be co-opted by all the traffickers of AdSense and to be diluted quickly. Result : the information providers known for their qualite/fiability - of which much is undoubtedly blogs - will gain there in importance. And as soon as they carry out the

I hope only that all that predicts of a return to good old days, now that the blogosphère am likely to lose always more his importance, considering whom it starts to be co-opted by all the traffickers of AdSense and to be diluted quickly. Result: the information providers known for their quality/reliability - of which much is undoubtedly blogs - will gain there in importance. And as soon as they carry out the

capacity of their information, they will want less and less that Google money plunders them and makes on the back of the reputation which they will have put such an amount of labour and energy to be built.

capacity of their information, they will want less and less that Google plunders them and makes money on the back of the reputation which they will have put such an amount of labour and energy to be built.

* * *

Qu'est-ce que Google RS2 ? (My guess)

Selon moi, la première inférence, évidente, est que **Google RS2 n'a absolument rien à voir avec la page vers laquelle renvoie le lien, ni même avec la fonctionnalité de TA telle qu'on la connaît aujourd'hui (Systran)**. Voilà deux jours que je me perds en conjectures sur la signification possible de RS2, qui ne me dit absolument rien comme acronyme lié à la traduction. Et pourtant, j'en connais des sigles. Alors quoi, so what ?

Je vous livre la conclusion de mes élucubrations, ou, pour mieux dire, de mon intuition : RS2, nom de code de Google, = **RSS** (S2 = deux "S"), et le lien avec la traduction serait la mise à disposition par Google d'un **service de traduction automatique des flux de syndication, pour traduire à la volée des fils RSS ou de podcasts de/vers n'importe quelle langue** (à plein régime, puisque selon toute probabilité, cette fonction sera initialement "bridée" aux langues principales...).

Un service de cette nature existe déjà : [NATIVETEXT](#)

« Service gratuit qui traduit les flux RSS des blogs et des podcasts dans votre langue. Ce service s'opère grâce à la communauté car la traduction se fait par les membres et non par un système informatique. On parle de Syndication de Langue d'Origine (Native Language Syndication)

Utilisation très simple :

- créez votre compte
- ajoutez votre flux
- choisissez vos langues de traductions
- répondez à un petit questionnaire pour renseigner les traducteurs
- bienvenue au reste du monde !! »

Source : [Aurigance](#) (pour une fois, c'est pas moi qui traduit :-)

Conclusion

La différence tient en un mot : le service de Nativetext est **COMMUNAUTAIRE**, celui de Google **AUTOMATIQUE**.

Oui, un mot qui fait toute la différence. [Début]

[Jean-Marie Le Ray](#)

P.S.1 Deuxième constatation

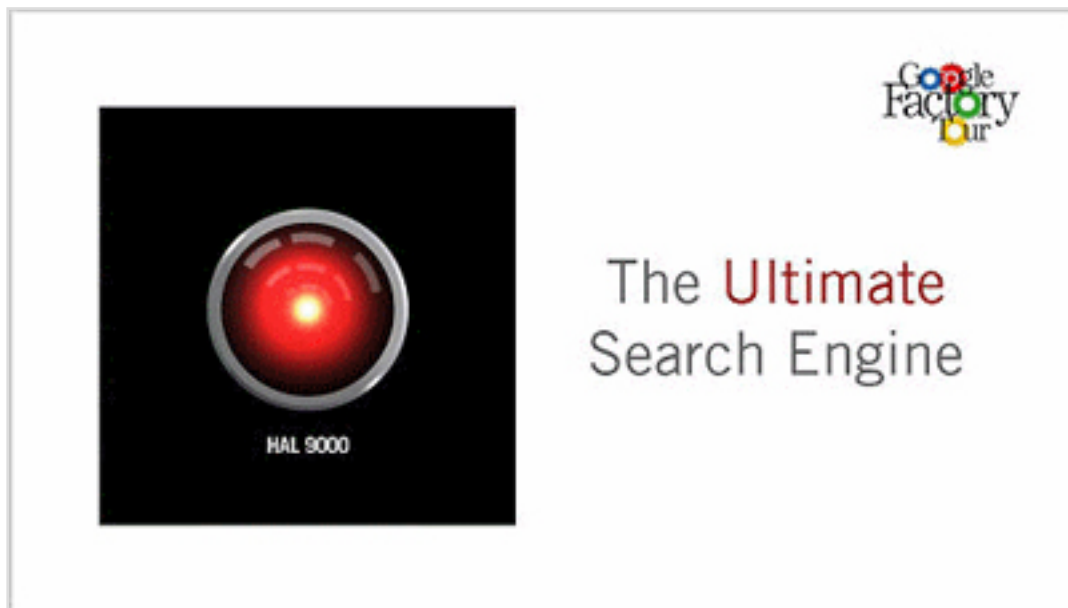
Cette hypothèse doit absolument être mise en rapport avec celle-ci : [Google et la recherche vocale](#),

puisque'une fois que Google maîtrisera avec un degré de pertinence suffisant la traduction automatique de la voix, je vous laisse imaginer le reste...

RS2 P.S.2 Je me plante peut-être, et si vous avez autre chose à proposer, je suis preneur. Pour autant, je suis sûr de ne pas me tromper en affirmant que Google nous prépare quelque chose de révolutionnaire avec sa fonction de traduction automatique : texte-voix, Internet, vidéo, mobiles, [Adsenses personnalisés](#), etc., qui pourra concurrencer Google ? Pas moi, c'est sûr :-)

P.S.3 Tiens, juste parce que je me sens en forme, je vous livre une autre intuition à propos de Weaver, dont Zorgloob nous dit : « *aucune information sur ce service dont le lien pointe vers <http://www.google.com/h9> »*. D'après moi, ce n'est pas une allusion au tisserand (traduction de weaver en français), mais à Sigourney Weaver dans [Futurama](#) (in [Love and Rocket](#)), qui double un personnage [parodiant HAL 9000](#), le *Supercomputer*, « ordinateur exceptionnel doué d'intelligence et de parole », dans [2001 : l'odyssée de l'espace](#).

Or en langage Google, HAL 9000 (h9) est **The Ultimate Search Engine** :



Diapositive n° 131 du [Google Inc. Factory Tour du 19 mai 2005](#). On comprend pourquoi l'URL n'est plus accessible ! (moi, la photo, elle me fait penser à Big Brother...) [[Début](#)]

[Actualités](#), [Google](#), [Google RS2](#), [RS2](#), [traduction automatique](#), [traduction](#), [TA](#), [laboratoire de traduction](#), [Google h9](#), [h9](#), [Hal 9000](#), [Weaver](#), [Ultimate Search Engine](#), [OR](#), [outil de recherche](#), [moteur](#), [moteur de recherche](#), [laboratoire de réflexion](#), [Internet](#)

posted by Jean-Marie Le Ray at [29.7.06](#) _ [Del.icio.us](#)

0 Comments:

[Enregistrer un commentaire](#)

Links to this post:

[Créer un lien](#)

[<< Home](#)

|



[Suivi avec co.mments](#)

About Me



Jean-Marie Le Ray
Ardea, Roma, Italy

[View my complete profile](#)



www.Adscriptor.com
Le Web à la loupe

Rubrique Actualités

- [Actualités Adscriptor](#)
 - [Adscriptor, c'est quoi ?](#)
 - [Adscriptor, c'est qui ?](#)
-

Billets précédents

- [Google est la première marque mondiale en terme de progression](#)
- [Internet media à part entière : comparatif et mise en perspective](#)

- [Professionnel indépendant : grand questionnaire de l'été !](#)
- [Google communique les taux de clics invalides à ses annonceurs](#)
- [Google partenaire de Go Daddy dans le business des noms de domaine](#)
- [Le monde selon Google : coup de pub pour un best-seller !](#)
- [Internet 2010 : prospective et mise en perspective](#)
- [GDrive, nom de code : Platypus. Google Drive enfin sur la rampe de lancement](#)
- [Les noms de domaine et la longue traine](#)
- [Google, eBay, MySpace, McAfee : brèves et mises à jour](#)

Sur mon nuage

Categories



My **blog** is worth **\$9,032.64**.

[How much is your blog worth?](#)



[Abonnez-vous](#)

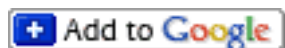


[Performancing](#)



Adscriptor via courriel :

Powered by Squeet.com



-->