

BONJOUR SIRI. PARLE-MOI DU TAL!

Publié le 2 mars 2015



Téléphone portable, ordinateur, tablette: tous nos outils informatiques utilisent le traitement automatique du langage (TAL) pour capter nos attentes, traiter des informations, restituer un résultat. Les progrès récents sont fulgurants.

« Avec Siri, nous parlons désormais à notre tablette ou à notre smartphone pour exprimer verbalement une requête », indique le Pr Cédric Fairon, directeur du CENTAL ([Centre de traitement automatique du langage](#)) à l'UCL.

A la frontière de l'informatique et de la linguistique

[Quand Siri nous « écoute »](#) c'est-à-dire quand cet outil informatique intégré dans les produits du fabricant américain Apple fonctionne, il analyse notre demande. Il tente de reconnaître les mots, les chaînes de caractères que nous exprimons pour les transformer en écrits et en identifier le sens. Le système consulte ensuite des bases de données internes ou en réseau et formule éventuellement la réponse que nous attendons.

« Il est donc question de reconnaissance de la parole, de synthèse vocale, de correction orthographique, de traduction automatique », précise le Pr Fairon. « Mais aussi de sémantique, de syntaxe, de lexique ». C'est exactement cela le TAL : une discipline à la frontière de l'informatique et de la linguistique. Avec une précision concernant Siri : « ce système ne fait pas d'analyse profonde de nos requêtes. Il travaille sur base de reconnaissance de mots-clés et de scripts potentiels », dit-il.

Test de Turing

Ce progrès technologique « dernier cri » est aussi vieux... que la guerre froide. « C'est à cette époque que les premiers travaux sur le traitement automatique du langage sont menés », précise le chercheur. « C'est [l'époque du test de Turing](#), qui vise à déterminer si l'interlocuteur auquel on a affaire est un être

humain ou une machine ». C'est aussi les débuts de la traduction automatique.

« *L'âge d'or du TAL remonte aux années 1948 à 1958* », précise le scientifique, « *avec l'apparition des premiers outils de TAL. On pensait pouvoir régler le problème de la traduction automatique en 2 mois... Dans les années 1960, la traduction automatique est remise en cause. On évolue plutôt vers l'élaboration d'outils d'aide à la traduction* ».

Coup d'accélérateur digital

Avec l'avènement du digital, le traitement automatique du langage va connaître de nouveaux développements rapides. Les ordinateurs, les réseaux comme internet, l'internationalisation des échanges, les succès industriels et les progrès de l'informatique couplés à la diminution des coûts vont être les clés de ce succès.

Les outils développés avant cette période ne sont pas pour autant délaissés. On les rhabille de vêtements neufs et on les utilise à d'autres fins: au filtrage de l'information par exemple, à l'extraction de données...

Outils disponibles en ligne

Car les enjeux actuels sont effectivement là. « *Apporter une plus-value aux bases de données disponibles et pas toujours très structurées est un enjeu majeur* », estime le linguiste.

Le CENTAL a saisi diverses opportunités à ce propos. Plusieurs outils de TAL développés en son sein sont disponibles en ligne. [Platon](#) est une aide à l'apprentissage de l'orthographe, [Amesure](#) analyse la complexité de textes administratifs, [recto-verso](#) transforme une phrase en fonction de la nouvelle orthographe...

Biographies automatiques et évolution des marchés

[La spin-off « Early Tracks »](#) développe également des outils ciblés. Il s'agit par exemple de l'extraction de données en vue de constituer des biographies automatiques. Le client était ici une agence de presse disposant d'importantes archives. L'outil mis au point a recensé et mis en relation des articles parfois vieux de plusieurs dizaines d'années reprenant un même nom propre en vue de constituer une biographie pertinente.

Dans le domaine médical, ce genre d'extraction de données pour chaque patient apporte également une plus-value. Idem en pharmacovigilance. Le secteur économique est également concerné. « *Le traitement automatique du langage pourrait dans ce cas aider à extraire de l'information grise disponible (celle qui est diffuse, peu structurée) des indices, des informations susceptibles de prédire l'évolution des marchés* », indique Cédric Fairon.

Passionnant? Poser la question, c'est y répondre. Paradoxalement, le CENTAL ne fait pas le plein de chercheurs... « *Comme nous sommes à la croisée de deux disciplines, la linguistique et l'informatique, les candidats chercheurs ne se pressent pas à notre porte* ». Etrange? Cela demande sans doute de passer au-dessus de certains automatismes liés à chaque discipline d'origine.

Bonjour Siri, parle-moi de l'avenir du TAL...

Ecoutez le Pr Cédric Fairoon parler de l'importance du TAL en matière de politique linguistique