

L'intelligence artificielle : quand elle doit comprendre la langue d'un nouveau monde

[En 1957 Chomsky expose la séquence des mots](#) « les idées vertes dorment furieusement » comme étant un exemple d'aberration sémantique. De nos jours, cette séquence des mots nous semble tout à fait compréhensible, voire banale, si nous sommes dans une thématique autour de l'écologie. Tout change très vite, même le langage et ses pratiques dans le monde de l'IA. Et qu'en est-il du traitement automatique du langage en ces temps de bouleversement lié à la pandémie ?

Ces dernières années, les technologies d'Intelligence Artificielle se sont [fortement démocratisées](#), notamment dans le domaine du « Traitement Automatique du Langage Naturel » (TALN, *Natural Language Processing*, NLP en Anglais). Aujourd'hui, entraîner un assistant virtuel, un analyseur d'emails, un détecteur de sentiments, est à la portée de tous, même sans formation dans le domaine. Avec cette facilité, les *bots* fleurissent sur toutes les applications, mais souvent avec une qualité et un périmètre limités. Ils sont loin d'être de véritables assistants virtuels capables d'adresser plus de 80% des demandes utilisateurs, comme ceux à l'état de l'art.

Même si l'Intelligence Artificielle a fortement progressé, elle n'a pas encore passé le stade de la maîtrise absolue de la langue, et s'appuie, que ce soit à travers des règles ou bien de l'apprentissage automatique, sur des algorithmes très différents suivant ce qu'on lui demande de faire.

Ces algorithmes, finalement, définissent la façon de penser de cette Intelligence Artificielle, et ainsi la méthode d'enseignement, que ce soit par l'exemple (connexionnisme) ou par la théorie (symbolisme).

Alors, comment devenons-nous experts dans l'art de mettre en place ces algorithmes de Langage Naturel ?

Considérées comme le pétrole du XXI^e siècle, les données massives présentes sur le Web ne seraient pas d'un très grand intérêt sans un outil pour leur donner « du sens ». Les méthodes statistiques et l'apprentissage automatique se sont avérés, notamment ces dernières années, très efficaces pour aborder des problématiques de traitement du texte, de la voix ou de la vision.

L'apprentissage profond gagne de plus en plus de terrain dans les laboratoires de recherche, les startups comme les grands groupes du domaine. Mais des résultats souvent pauvres en « intelligence » (comme la confusion d'un chien avec un muffin, ou d'un perroquet avec du guacamole, dans le domaine du traitement de l'image ; ou bien, les assistants virtuels qui ne « comprennent » pas les demandes utilisateurs les plus simples comme « bonjour »), donnent non seulement des sujets de détournements moquant l'IA sur Internet, mais également des vrais défis à

relever pour les concepteurs de ces modèles cognitifs.

Les résultats obtenus à l'aide des méthodes d'apprentissage profond sont pourtant impressionnants lorsqu'on reste sur des « situations » proches de celles apprises par le modèle cognitif lors de son entraînement. Mais dès qu'on s'en éloigne, les résultats peuvent paraître ridicules. Les modèles qui sont définis dans la phase d'entraînement sont des modèles d'approximation : lorsqu'on les met en place, on laisse les statistiques décider de prendre en compte certains aspects de la réalité ou de la langue, et d'ignorer d'autres, en fonction de la fréquence d'apparition ou des liens qui peuvent être établis entre les données.

Même si les machines sont capables, d'une part d'enregistrer une quantité massive d'informations sans les altérer et d'autre part d'exécuter des calculs d'une manière très rapide, on ne peut pas affirmer qu'elles « comprennent » les données qu'elles sont en train de traiter ou les tâches qu'elles sont en train d'exécuter. Par conséquent, l'une des principales difficultés, quand on veut créer un modèle cognitif est d'indiquer de manière claire et non ambiguë ce que l'on veut.

Pour entraîner efficacement un modèle cognitif, il faut lui fournir les « bonnes » données. Mais on ne dispose pas toujours d'un ensemble de données adéquat à la problématique que l'on cherche à modéliser. Toute la « réalité » qui nous entoure et qu'on utilise pour interpréter ce que l'on voit, entend ou lit dans la presse, ne peut pas être capturée dans un grand corpus de données, si énorme soit-il. En ce qui concerne les corpus pour le traitement de la voix, obtenir une quantité considérable des conversations spontanées est d'une difficulté encore plus grande.

De notre expérience dans le développement des projets cognitifs, on a pu constater que les techniques d'apprentissage automatique fonctionnaient mieux lorsque les algorithmes étaient dotés de pointeurs vers les informations pertinentes dans les données avec lesquelles ils étaient nourris. [En NLP](#) ces pointeurs sont souvent présentés sous la forme d'annotations (entités, relations entre entités, chaînes de coréférence). Quand ils sont enrichis de cette manière, les corpus sont plus adaptés à une exploration sémantique des données.

Telles qu'on les récupère, les données ne sont jamais « prêtes à l'emploi ». Un corpus comporte très souvent toutes sortes de bruits : phrases inachevées, fautes d'orthographe, répétitions, caractères parasites etc. Ces données ont besoin d'être pré-traitées, corrigées, sélectionnées, découpées, structurées et enrichies avec des métadonnées, de manière supervisée ou non.

Outre des données de « bonne qualité », un modèle cognitif efficace a besoin d'une bonne méthodologie de travail. Prenons l'exemple du découpage en mots d'un texte. La méthode la plus simple est de considérer un mot comme étant une suite de caractères délimitée par une espace ou un signe de ponctuation. De nombreux *pipelines* de traitement NLP proposent ce genre de segmentation en mots. Mais certains mots contiennent des signes à l'intérieur de leur composition : « chou-fleur », « aujourd'hui ».

Une telle technique de découpage ferait perdre de la nuance sémantique du texte analysé. Les mots composés, comme « pommes de terre », « arc en ciel », les expressions idiomatiques, les expressions verbales complexes comme : « en tenir bien compte », « faire le nécessaire » rendent cette tâche encore plus difficile. Comment analyser l'écriture inclusive ? Et qu'en est-il de cet utilisateur qui cherche des informations concernant le voyage qu'il a fait ou bien une actualisation

pour celui qu'il fera si l'assistant virtuel mis à sa disposition ne tient pas compte de la conjugaison du verbe ?

Il n'est pas suffisant de découper un texte en mots car un texte n'est pas la simple addition de mots. La plupart du temps nous avons besoin de pouvoir les identifier, les différencier, leur donner une valeur d'interprétation dans le contexte de leur utilisation.

La pandémie qui submerge le monde depuis une année n'est qu'un exemple de phénomène sociétal qui a rendu un certain nombre d'assistants virtuels obsolètes en l'espace une nuit. Cette nouvelle thématique de la COVID-19 devait être comprise par les différents systèmes cognitifs pas seulement en surface (à travers une corrélation des mots-clés) mais aussi dans les couches linguistiques des requêtes qu'ils étaient en train de consommer.

Lorsqu'un utilisateur fait une requête par rapport aux conditions de voyage, détecter une corrélation entre les mots *covid-Disney* et les mots *covid-Italie* relevant de la thématique *Voyage* par exemple, n'est pas suffisant. Des informations supplémentaires, parfois extralinguistiques, sont nécessaires afin de proposer une réponse pertinente. Savoir si la ville de départ ou la ville d'arrivée est une zone rouge, comprendre la raison de voyage ou bien la destination, être capable de faire la différence entre un voyage à l'intérieur du pays, un voyage à l'intérieur de l'espace européen, ne sont que quelques exemples dans ce sens.

A situation nouvelle des mots nouveaux. Une autre difficulté qui s'est rajoutée au traitement du langage ces derniers mois a été l'intégration de tous ces nouveaux sens des mots (*confinement, distance, geste, présentiel, masque, couvre-feu partiel, commerce essentiel...*) qui sont rentrés dans le lexique en l'espace de quelques semaines.

Un modèle cognitif avec un vocabulaire minimal se retrouvera face à énormément de situations d'incertitude, situation face auxquelles il n'aura pas de « bonne réponse ». Contrairement à un responsable de service client un assistant virtuel de base ne peut pas hausser les épaules, faire un regard résigné et dire « deux secondes, je consulte mon responsable et je reviens vers vous ».

Ce temps de l'Assistant Virtuel capable de s'adapter du jour au lendemain à des mutations du langage n'est pas encore arrivé, mais nous pouvons nous en approcher quand on rentre dans une démarche opérationnelle de supervision de ces systèmes. Ces IA, finalement entourées d'un savoir-faire très humain, dépendent de notre réactivité à leur fournir les nouvelles informations nécessaires pour comprendre ce nouveau langage, reflet d'un nouveau monde