

INTELLIGENCE ARTIFICIELLE

SYSTRAN : 50 ANS D'INNOVATION ET DE SAVOIR-FAIRE AU SERVICE DE LA TRADUCTION AUTOMATIQUE

Le développement de la technologie autour de l'intelligence artificielle ouvre de nouvelles perspectives dans le secteur de la traduction automatique. Le point avec Jean Senellart (92) Global CTO du groupe **SYSTRAN** et Directeur Général de SYSTRAN S.A.S.



Jean Senellart (92)

BIO EXPRESS

Jean Senellart (92) est titulaire d'un Doctorat en Informatique Linguistique de l'Université de Paris VII. Il a débuté sa carrière comme chercheur et a enseigné à l'École Polytechnique et à l'Université de Marne la Vallée. En 1999, il rejoint SYSTRAN France en tant que chef de projet avant de prendre la direction des équipes R&D avec lesquelles il lance 4 générations de produits. Il a mis au point plusieurs technologies innovantes comme l'« Intuitive Coding™ » en 2005 permettant le codage massif de données, la post-édition statistique (SPE) en 2008 permettant de combiner des approches statistiques et à base de règles et plus récemment la technologie « PNMT® » (Pure Neural Machine Translation). En 2014, il est nommé Global CTO du groupe et depuis juillet 2015, il occupe également le poste de Directeur Général de SYSTRAN S.A.S. Il a publié plus de 50 papiers scientifiques à portée internationale et intervient régulièrement dans de nombreuses conférences.

SYSTRAN, dont le cœur de métier est la traduction automatique, va fêter son cinquantième anniversaire en 2018. Quelles sont les principales évolutions connues par la société ?

SYSTRAN a initialement vu le jour dans un contexte de Guerre Froide pour fournir de la traduction du russe vers l'anglais. Notre métier a ensuite évolué et a été porté par le développement d'Internet et des outils de communication. De nouveaux besoins de traduction professionnelle sont apparus. Nous sommes restés très présents dans le monde de la défense et de la sécurité où la traduction joue un rôle primordial dans le traitement massif et l'analyse de données provenant de multiples sources.

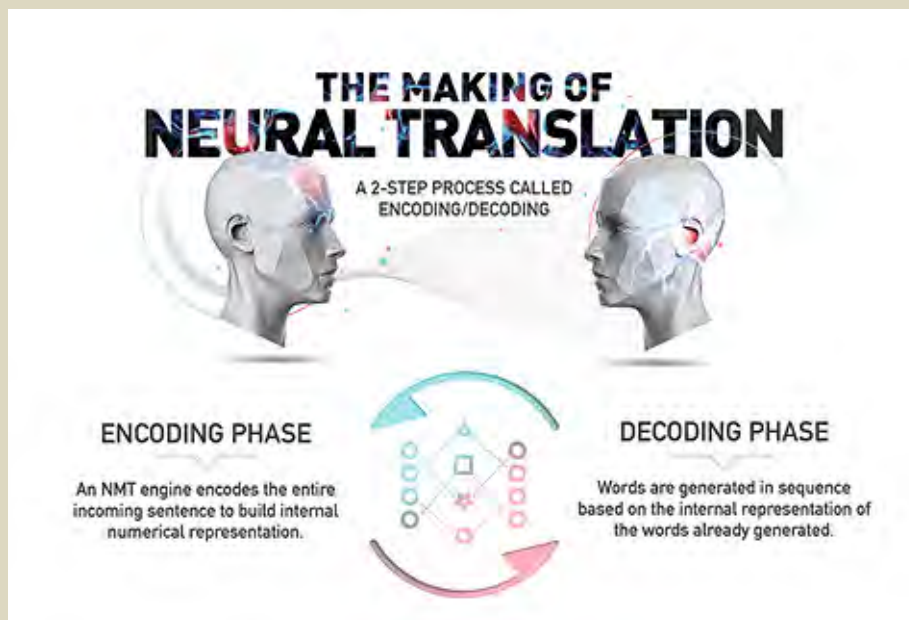
Nous opérons aussi largement dans le secteur privé. Nous avons notamment lancé le premier moteur de traduction en ligne Babel Fish et nous

avons équipé, entre autres, les portails de traduction automatique de Google et Yahoo. Jusqu'en 2007, la traduction en ligne, c'était essentiellement SYSTRAN !

Dès 2005, nous avons commencé à être utilisés comme un outil de productivité par les services de traduction. Grâce à notre technologie qui avait atteint une certaine maturité, les traducteurs pouvaient accélérer la vitesse de traduction, par la post-édition de traduction automatique et l'utilisation de mémoires de traduction.

En parallèle, au sein des entreprises, la communication est devenue de plus en plus multilingue avec un besoin de solutions de traduction en temps réel et intégrées aux systèmes bureautiques et de messagerie avec une prise en compte de la terminologie propre à la société.

La particularité de SYSTRAN est en effet de proposer une traduction spécialisée. Nous proposons




AN ARTIFICIAL NEURAL NETWORK (ANN) IS COMPOSED OF LAYERS OF ARTIFICIAL NEURONS.

THE LAYERS ARE CONNECTED TOGETHER WITH WEIGHTS KNOWN AS PARAMETERS.


THE NEURAL NETWORK IS UNIQUE IN THAT IT CAN AUTOMATICALLY CORRECT ITS PARAMETERS DURING THE TRAINING PHASE (LASTING A FEW WEEKS) BY SENDING CORRECTIVE FEEDBACK TO THE ENGINE BASED ON THE GENERATED OUTPUT AND THE EXPECTED OUTPUT (REFERENCE).

AN ARTIFICIAL NEURAL NETWORK WITH MORE THAN 2 LAYERS IS CALLED A DEEP NEURAL NETWORK (DNN).

NEURAL NETWORKS USED FOR NATURAL LANGUAGE PROCESSING HAVE 3 TO 20 LAYERS.




3 MAIN INGREDIENTS




WORD EMBEDDINGS

POWERFUL LANGUAGE MAP



RECURRENT NEURAL NETWORKS (RNN)

CONTEXTUAL KNOWLEDGE



ATTENTION MODEL

ATTENTION CAPACITY

à nos clients du sur-mesure, une solution adaptée à leurs métiers en reprenant leur terminologie interne, leur style et en nous alignant sur leurs critères de qualité. Pour cette raison, les dernières technologies permettant de spécialiser automatiquement la traduction sur la base de textes déjà traduits ont permis de faire un bond énorme en termes de qualité et de réduire les temps nécessaires à la production de moteurs spécialisés.

Comment s'inscrivent les évolutions technologiques dans l'histoire de SYSTRAN ?

Aujourd'hui, le Big Data, la performance des algorithmes et la puissance des machines, offrent de nouveaux leviers de développement et de performance pour la traduction automatique. Cependant, c'est sur la technologie elle-même que les changements majeurs sont apparus. Au fil des années, nous nous sommes développés en conservant toujours un temps d'avance technologique.

Les années 70 ont été marquées par la multiplication des sociétés de traduction automatique. Parmi toutes ces sociétés, nous sommes les seuls à avoir traversé les décennies.

Dans cette première phase, la technologie utilisée pour la traduction automatique était celle des systèmes experts, qui s'appuyaient sur des règles décrivant les langues. Ces règles étaient ensuite implémentées dans un programme. Il s'agissait d'une technologie coûteuse qui néces-

sitait des financements sur plusieurs années et beaucoup de ressources humaines. Pendant plus de 20 ans, c'est ce système que nous avons développé et qui a notamment été utilisé par la Commission Européenne.

Trente ans plus tard, en 1998, IBM lance une nouvelle technologie de traduction purement statistique se basant sur des textes déjà traduits pour reconstruire des traductions par fragments. En 2007, Google poursuit sur cette même voie, en réalisant un exploit technologique à l'échelle du Web avec une seconde génération de Google Translate qui fait entrer les acteurs du marché dans la conquête des données (the bigger, the better).

De son côté, SYSTRAN se positionne en concevant une traduction hybride qui combine les deux technologies misant sur la qualité adaptée à chaque domaine.

Il y a environ deux ans, votre positionnement a complètement évolué avec la découverte de la traduction neuronale...

En effet, les résultats de recherche menée par l'Université de Montréal en traduction neuronale ont ouvert de nouvelles perspectives que nous n'avons pas hésité à saisir : la traduction neuronale qui s'appuie sur l'apprentissage profond, une des techniques de l'intelligence artificielle. En septembre 2016, nous étions, avec Google les premiers à lancer des systèmes opérationnels basés sur ces moteurs neuronaux. Alors que la

CHIFFRES CLÉS

- 50 ans d'existence
- 3 bureaux à Paris, San Diego et Séoul
- 200 collaborateurs
- 50+ langues couvertes

recherche sur ce domaine continuait de progresser à une vitesse considérable, nous avons décidé de contribuer à cet essor et de lancer en parallèle un framework open source « OpenNMT ».

Après avoir accumulé plus de 48 ans de technologie, nous avons donc rendu notre technologie publique... OpenNMT a été développé en collaboration avec Harvard NLP et met à disposition des développeurs et chercheurs des algorithmes neuronaux à l'état de l'art du traitement du langage naturel. Ce framework a déjà reçu des prix académiques et fédère plusieurs centaines de membres actifs : passionnés, chercheurs, développeurs et industriels.

Même si nos clients peuvent avoir accès librement à cette technologie, ils font appel à nous car ils recherchent une solution clé en main. Nous sommes avant tout un éditeur de solutions logicielles et détenteurs d'un véritable savoir-faire.

Quels sont vos principaux défis ?

Aujourd'hui, nous sommes sur un positionnement inédit : en misant sur l'Open Source, nous avons en quelque sorte ouvert le marché à de nouveaux acteurs. Nous allons accompagner les prochaines évolutions. En ambitionnant d'être un leader sur ce marché, nous ne pouvons pas nous refermer sur nous-mêmes. Concrètement, pour maintenir notre positionnement sur ce domaine à la pointe de la recherche, nous nous appuyons sur une équipe dédiée d'une dizaine d'ingénieurs docteurs. Aujourd'hui, nous entendons beaucoup parler de l'intelligence artificielle pour le traitement de la voix, des images et des textes. Très vite, les nouvelles découvertes technologiques vont sûrement nous permettre de pouvoir traiter ces 3 dimensions ensemble. Dans cette optique, il est aussi important de comprendre que l'intelligence artificielle n'est pas seulement une question de bon paramétrage des réseaux de neurones. Ce sont des modèles qui doivent être entraînés sur le long terme. Le prochain défi est donc de développer et d'accompagner sur plusieurs années cet apprentissage infini de nos modèles neuronaux et « d'élever » des IA multilingues. ■