

SYSTRAN

Technologies du langage

Le moteur de traduction Hybride SYSTRAN

Règles linguistiques ou traitements statistiques : pourquoi choisir ?

En matière de traduction automatique (TA), l'approche linguistique ou statistique présente chacune des avantages.

✓ Les règles linguistiques et les dictionnaires assurent le respect de votre terminologie métier et une bonne qualité de traduction, même sans personnalisation du logiciel.

✓ Les composants statistiques permettent l'apprentissage automatique et rapide à partir de corpus monolingues et de textes déjà traduits et validés (corpus multilingues).

Alors pourquoi devriez vous choisir entre les deux ? Le moteur de traduction hybride mis au point par SYSTRAN concentre les avantages des deux technologies dans un moteur unique et très performant, capable d'effectuer des traductions de haute qualité dans tous les domaines.

Les autres technologies du langage

Parce que le développement de moteurs de traduction automatique s'appuie sur des technologies de pointe, SYSTRAN demeure à l'avant-garde des technologies de traitement du langage naturel depuis plusieurs décennies. La maîtrise de ces technologies nous permet de proposer :

L'analyse morphologique complète

La segmentation de texte et analyse syntaxique profonde

Le développement de composants d'apprentissage automatique spécialisés par langue

La gestion de corpus bilingues et monolingues

Les règles linguistiques restent au cœur du moteur hybride qui bénéficie du même niveau de performance, de vitesse et de robustesse qui a fait la réputation de SYSTRAN depuis de nombreuses années. Les modules statistiques agissent ensuite à chaque phase du processus (analyse, transfert, post-édition) pour améliorer la qualité de traduction.

Adapté à votre langage métier

L'apprentissage constant et la personnalisation permettent aux logiciels SYSTRAN de s'adapter à votre activité. Vous pouvez créer des dictionnaires et construire des modèles de traduction à partir de données existantes. A partir de traductions déjà réalisées par des humains, le moteur est capable d'apprendre à traduire correctement des mots et expressions spécifiques à votre métier. Parce qu'il combine règles linguistiques et traitements statistiques, le moteur d'entraînement du moteur hybride de SYSTRAN nécessite beaucoup moins de données qu'un moteur purement statistique.

La construction de dictionnaires

L'extraction et filtrage de documents

La reconnaissance optique de caractères (OCR)

La reconnaissance automatique de la parole (ASR)



Reconnaissance Automatique de la Parole (ASR)

Parallèlement au développement des technologies de traduction, SYSTRAN a fortement investi dans le développement de technologies de type « ASR » qui permettent d'analyser la parole humaine et de la transcrire en format texte.

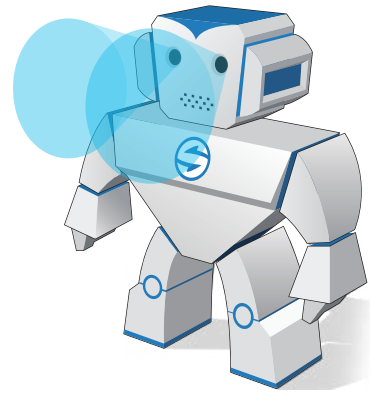
En combinant nos technologies ASR avec notre moteur de traduction hybride, nous ouvrons la voie à de multiples applications dont par exemple :

Intégration de la reconnaissance vocale dans les appareils électroniques domestiques

Aide aux touristes asiatiques pour communiquer (Coréen/Chinois/Japonais/Anglais)

Reconnaissance de plus de deux millions de noms de sites historiques et touristiques (coréen uniquement)

Extraire d'une conversation des phrases spécifiques à partir de mots clés (coréen uniquement)



Reconnaissance Optique de Caractères (OCR)

En combinant nos technologies de pointe avec les dernières innovations dans ce domaine, nous avons pu réaliser d'importants progrès en matière de technologies OCR.

La technologie OCR de SYSTRAN s'appuie sur le moteur open-source Tesseract auquel nous avons intégré nos propres technologies de traitement d'image afin de permettre par exemple la suppression des images d'arrière-plan, le redressement de document, et la traduction de chaînes de caractères.

Grâce à notre moteur de traduction hybride et nos ressources linguistiques, nous prévoyons également de proposer des services spécifiques combinant technologies OCR et traduction automatique comme par exemple :

Extraction de texte à partir d'images pour offrir une traduction de pages Web complète

Localisation d'interface utilisateur (traduction de boutons sous forme d'image)

Traduction instantanée de documents papier scannés ou photographiés