

# SYSTRAN

## Language Technologies

### SYSTRAN's Hybrid Translation Engine

#### Rule-based or statistical? Get the best of both worlds!

When it comes to machine translation (MT), rule-based and statistical methods both have their advantages.

✓ Rule-based components guarantee predictable and consistent translations, compliance with corporate terminology and out-of-domain usability.

✓ Statistical components learn from existing monolingual and multilingual corpora, reducing customization costs and improving domain-specific quality.

But why choose between the two? With SYSTRAN's new hybrid engine, you can get the best of both worlds all in a single, powerful, high-precision engine, achieving quality translations in any domain.

SYSTRAN's renowned rule-based translation software is the backbone of its hybrid MT and provides a solid translation framework. Through new statistical techniques, the engine learns from existing monolingual and bilingual data to improve every phase of translation and enhance the customization process.

#### Adapts to your business language

SYSTRAN software is constantly learning and customizing itself to your business. It creates dictionaries and builds translation models from existing data. And it learns how words and phrases should be translated for specific domains based on existing human translations. Because SYSTRAN's hybrid MT solutions combine rule-based and statistical techniques, the amount of data required to train the software is much less than statistically driven MT solutions.

### Other language technologies

Because MT development requires high-end technologies, SYSTRAN has spent decades at the forefront of linguistic and natural-languages processing. By mastering these technologies, we're able to offer you:

Comprehensive morphological analysis

Part-of-speech tagging and deep syntactic analysis

Development of specialized language-specific machine-learning components

Management of bilingual and monolingual corpora

Dictionary building

Document filtering

Optical Character Reader (OCR)

Automatic Speech Recognition (ASR)



## Automatic Speech Recognition (ASR)

SYSTRAN not only works on interpretation and translation, but also invests heavily in ASR technology to analyze human speech and convert it into text. By combining our ASR capabilities with our hybrid translation engine, we're able to deliver optimal results across all types of applications.

Here are a few examples of how SYSTRAN's ASR capabilities improve services:

Enables consumer electronics and smart devices to be operated via voice recognition

---

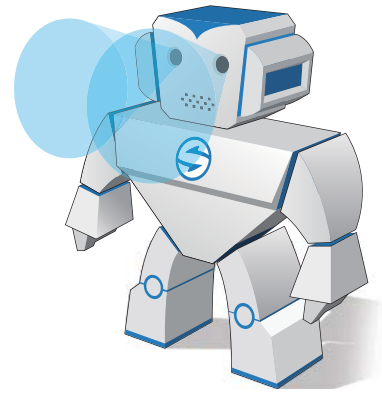
Helps Asian travelers overcome communication barriers (Korean/Chinese/Japanese/English)

---

Recognizes more than two million landmarks and destinations as language commands (Korean)

---

Extracts specific sentences in the dialogue or sentences that include key words (Korean)



## Optical Character Reader (OCR)

By working to bring together industry innovation with our own state-of-the-art technologies, we've been able to make exciting progress in OCR technology.

SYSTRAN OCR technology is based on Tesseract's reliable open-source OCR Engine, and we've made improvement to the engine by applying various self-developed image-processing technologies, such as removing background image, document de-skewing, and section-character string translation.

We are also planning to provide various specialized OCR technologies and translation services through our hybrid translation engine and language resources.

Application field of OCR and Natural Language Processing Technology:

Extract text from images for easy Web translation

---

UI localization (localization of the buttons that are composed of images)

---

Instant translation of hardcopy documents that are scanned or photographed