

First strategies for integrating hybrid approaches into established systems

Panel on Hybrid Technologies

Jean Senellart

John S. White

Systran Software



Limits to RBMT and SMT

- Known practical and theoretical limits to RBMT and SMT
- RBMT
 - Overtly expresses definition of infinite set of phenomena
 - As rule base grows, undesired (and unanticipated) rule interactions occur
 - Reality is infinite, but non-deterministic
- SMT
 - Empirically discovers the non-absolute nature of the phenomena
 - Requires huge parallel resources
 - Treats the set of phenomena as if it were finite

Statistical Methods at SYSTRAN

- Long established MT developer
 - Since 1968, 40+ LPs
 - Purely mainframe service-oriented for many years
 - Migrated through the platforms (MF, C/S, desktop, SOA, handheld)
 - If you've ever hit the "translate" link, you've used SYSTRAN
- Rule Based Systems
 - Highly Knowledge Intensive
 - Vast Lexical and Grammatical stores
 - Can leverage off cognate and/or similar languages, BUT
 - Need Agility for new language pairs
 - Ditto new lexical domains
 - Especially where there is little available expertise

[Use of Statistics at SYSTRAN]

Three approaches:

- Massive terminology extraction for aligned corpus
- Use of language model for alternative meaning selection
- Mix rules and statistics