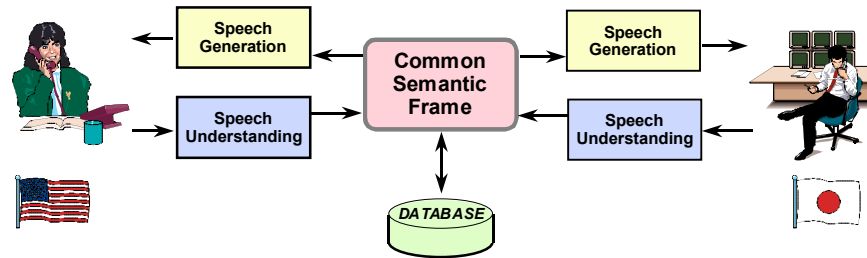


MIT Multilingual Conversational System Research

James Glass and Stephanie Seneff
Spoken Language Systems Group, LCS



- Explore language-independent approaches to speech understanding and generation
- Develop necessary human-language technologies to enable porting of conversational interfaces from English to Japanese
- Use existing Jupiter weather-information domain as test case

LCS
Spoken Language Systems Group

MIT Progress to June 1999

- **Language generation:**
 - Created initial translation capability for weather domain
 - Estimate 70% correct translation of weather reports, with remainder being mostly comprehensible, but slightly ill-formed
 - Required minor alterations to original meaning representation structure (e.g., preposition 'in' in locative vs temporal reference)
- **Data collection:**
 - Created initial 600 sentence set in Mokusei domain by translating English queries, and soliciting examples from NTT colleagues
- **Language understanding:**
 - Developed initial context-free grammar to parse 600 sentence set
 - Estimate 50% of sentences are parsable, with remainder consisting of both out-of-domain queries and more complex queries which we expect to parse in the future

LCS
Spoken Language Systems Group



Six Month Plan

- **Continued refinement of Mokusei components**
 - 1 **Improve translation of system responses, for more natural phrasing**
 - 2 **Expand language understanding coverage, and add robust parsing**
 - 3 **Develop initial speech recognition capability using available telephone-quality corpora and seeding from Jupiter models**
 - 4 **Perform read-speech data collection from native speakers in U.S. and Japan, using 600 sentence set to create initial templates**
 - 5 **Begin wizard-based data collection efforts to obtain spontaneous speech for improved language modeling and understanding**