Information Transfer from Dialogue Response Generation to Speech Synthesis

Paul Bagshaw
France Telecom
(as summarized by Dan Burnett)
Use case

• Syntactic, semantic, and discourse info held by dialogue system should be available during speech synthesis to improve quality
  − Examples are speech acts, theme/rheme localisation, semantic roles ("record")

• SSML today does not have a mechanism for this info to be provided
Requirements

• Any dialogue system or TTS system already has its own internal categorizations, so
  – Any interface between the dialogue and speech synthesis systems MUST NOT result in the need to change any of the mechanisms or representations internal to either of the systems

• It must be possible to map between these two different internal categorizations
Request to W3C

• DO NOT attempt to normalise categorizations/tag sets.
• Instead, provide a generic mechanism to map between categorizations/tag sets. Example:
  - Dialogue system cares about function of “water”: noun, adjective, verb
  - TTS system cares about form of “water”: e.g., uncounted noun