Generic Architecture for Natural Language Generation in Human-Computer Dialogue

Vladimir POPESCU*, **, Jean CAELEN*, Corneliu BURILEANU**

*Grenoble Institute of Technology, France
**University “Politehnica” of Bucharest, Romania

May 10 2007
Context

- Task-oriented human-computer dialogue systems

- Pertinent user interaction $\Rightarrow$ natural language generation $\leftarrow$ important issue

- Language generation of most dialogue systems $\leftarrow$ template-based, inflexible, with little contextualization

- Language generation in monologue $\leftarrow$ thoroughly studied $\Rightarrow$ reference / standard architectures exist

- Language generation in dialogue $\leftarrow$ no principled, widely accepted, architecture exists; mostly ad-hoc approaches
Goals

- Design of a principled architecture suited for spoken language generation for dialogue

- Architecture compatible with de facto standards in language generation: Reiter & Dale’s architecture and RAGS (“Reference Architecture for Generation Systems”)

- Several processing levels, well separated through carefully designed interfaces

- Processing levels: independent of one another

- Interfaces independent of processing details
Motivations

- Design & implementation of a “sophisticated” language generation component in a generic spoken dialogue system

- Consideration of pragmatic aspects (viz. discourse structure, speech act types) in the language generation component

- Generic processing in the generation component (i.e., application-specific aspects well separated from application-independent ones)

- Consideration of ergonomic issues in user interaction (viz. illocutionary force, intonation)
Architecture Proposal

- Five processing levels:
  - **Logic level** ⇔ communicative intention computation
  - **Pragmatic level** ⇔ rhetorical structuring & illocutionary force tuning
  - **Linguistic level** ⇔ anaphora generation & surface realization
  - **Expressive level** ⇔ prosody computation & assignment
  - **Acoustic level** ⇔ speech synthesis

- Actually integrated in the generation component: only the *last four* from above
Key Points

1. Processing levels driven by pragmatic considerations:
   ▶ Rhetorical structuring ⇔ SDRT ("Segmented Discourse Representation Theory")
   ▶ Illocutionary force tuning ⇔ Speech Acts theory
   ▶ Anaphora generation ⇔ SDRT & semantics of the utterances

2. Firmly specified interfaces between processing levels
Interfaces Between Processing Levels

- Logic level → Pragmatic level: communicative intention, expressed in a first-order logic
- Logic level → Linguistic level: logic forms of the utterances in dialogue history & of the communicative intention (to be put in linguistic form)
- Pragmatic level → Linguistic level: discourse structure & illocutionary force
- Logic level → Expressive level: communicative intention, in logic form
- Pragmatic level → Expressive level: discourse structure & illocutionary forces
- Linguistic level → Expressive level: textual form of the communicative intention
- Linguistic level → Acoustic level: textual form of the communicative intention
- Expressive level → Acoustic level: pitch contour
Mappings with de facto Standards

- Reiter & Dale’s architecture vs. ours:
  - Content planning ↔ Logic generator (level) & Pragmatic generator (level)
  - Microplanning & Surface realization ↔ Linguistic generator (level)
  - ∅ ↔ Prosodic generator (Expressive level) & Acoustic generator (level)

- RAGS (Mellish et al, 1998-2006) vs. ours:
  - Conceptual level ↔ Logic level
  - Rhetorical & Document levels ↔ Pragmatic level
  - Semantic & Syntactic levels ↔ Linguistic level
  - Quotation level ↔ Pragmatic ∩ Linguistic levels
  - ∅ ↔ Expressive & Acoustic levels

- Conclusion:
  - differences with respect to Reiter & Dale and RAGS models
    ⇐ spoken dialogue aspects not considered in them
  - compatibilities exist at several levels ⇒ interoperability possible (e.g. 2/3 compatibility with RAGS, in representation formats)
Detailed View of the Architecture

LOGIC LEVEL

{Communicative intention}

PRAGMATIC LEVEL

{Discourse structure + illocutionary force}

LINGUISTIC LEVEL

{Text}

EXPRESSIVE LEVEL

{Pitch contour}

ACOUSTIC LEVEL

{Speech}
Conclusions & Prospects

▶ Generic architecture: well separated processing levels, clearly defined interfaces (refer to the paper for an extensive example)

▶ Interfaces independent of inner processing details

▶ Two approaches for the processing levels (currently in progress):
  1. Deterministic, logic-based algorithms (two papers submitted to TSD 2007 & SIGDIAL 2007)
  2. Stochastic, corpus-based (poster at SpeD 2007)