Problematics

• What is the nature of the dialogue history inside a dialogue system?
  – What type of information has to be saved? Scope? Limits?
  – What are the roles of the dialogue history in the interpretation process?
  – How can a history take forgetting phenomena into account?

• What are the specificities of the dialogue history inside a multimodal system?
  – Are there separate histories for visual perception, speech, gesture?
  – What are the roles of the global history?

• How to represent information inside the dialogue history?
Topics

- Dialogue history and reference to objects
  - Introduction: the ‘reference domains’ model
  - Linguistic history, visual history, gestures history, task history
  - The global history as a model of short-term memory/forgetting
  - Illustration: the dialogue history inside two dialogue systems

- Multimodal meaning representation format
  - Introduction: linguistic and multimodal resources management
  - MMIL = MultiModal Interface Language
  - MMIL for representing information in dialogue history

Part I

Dialogue History
and Reference to Objects
Reference Domains: Definition

- The identification of the referents pass through the identification of the contextual subset to which they belong
- Several contextual sources intervene, each one implying a structuration of subsets
  - Reference domains linked to the linguistic context
  - Reference domains linked to the shared visual context
  - Reference domains linked to the task context
- Same structural properties ⇒ confrontation, integration
  [Landragin, Salmon-Alt & Romary 02]

Reference Domains: Example

\[ RD_1 \]
T: form
GF: coordination
DC: form
\[
\begin{array}{|c|c|}
\hline
\text{circle} & \text{triangle} \\
\hline
\end{array}
\]

\[ RD_2 \]
T: circle
GF: proximity
DC: x-axis
\[
\begin{array}{|c|c|}
\hline
\text{object}_2 & \text{object}_1 \\
\hline
\end{array}
\]
DC: colour
\[
\begin{array}{|c|c|}
\hline
\text{object}_2 & \text{object}_1 \\
\hline
\end{array}
\]

T = type
GF = grouping factor
DC = differentiation criterion

partitions
Reference Domains: Principles

Linguistic History

• Information useful to resolve references
  – Characteristics of the referring expression (determiner, quantifier, gesture...)
  – Predicate of the utterance
  – Information structure of the utterance (linguistic salience)
  – Prosodic aspects, communicative situation aspects...

• Which information must be kept in the linguistic history?
  – The referring expression, because of one-anaphora
  – The referent identity, because of evolutive referents and objects deletions
  – The reference domain in which the referring act occurs
Linguistic History (cont.)

- **Visual scene**
  
  ![Visual scene diagram]

- **Referring expressions**
  - The two big triangles
  - The left triangle
  - The right triangle
  - The small triangle

  ![Referring expressions diagram]

Visual History

- **Structuring the visual scene into perceptual groups**
  - Formalization using dendrograms

  ![Structuring visual scene diagram]

- **Other parameters**
  - Visual salience
  - Guiding lines

  ![Other parameters diagram]
Gesture History

- 2-D gestures: pointing, circling, targetting...

- Access to the referents
  - A pointing gesture can indicate a referent as well as a perceptual group
  - A circling gesture can delimitate a set of referents
  - A circling gesture can delimitate a ‘gestural reference domain’ in which the referent will be extracted

- Conclusions
  - The gesture history is a part of the visual history (visual perception and gesture work together to delimitate ‘visuo-gestural reference domains’)
  - Saving the trajectories is useful to better interpret a type of trajectory that the user produces several times

Task History

- Same example

- Task
  - Applies to all triangles
  - Treat/name/delete/tidy
  - Following an implicit order

RD1:
T: triangle
GF: task-linked
DC: treatment order
T1 | T2 | T3
DC: treatment order
T3 | T2 | T1

RD2:
T: form
DC: task-linked
treated | -treated
Nature of the Global History

- Components of the global history
  - Pointers to parts of local histories
  - Time stamps
  - Results of referring actions
  - A posteriori evaluations (e.g., relevance scores)
  - Reactions of the system

- Scope problem
  - Possible solution: to save only the information that have been the object of a computation during a previous phase of the dialogue

- The global history is a kind of coordinating agent

History = Model of Forgetting?

- Problematics
  - During a dialogue, the former information is forgotten step by step
  - But: what does ‘former information’ mean?
    - is it realistic to forget a piece of information that can be relevant for a future phase of the dialogue?

- Principles of a model of forgetting based on Morton’s Logogen Model
  - Tracking attentional scores, for every object, category, event, or property
    - the more the users refers to an object, the greater the object’s score
    - the more he evokes a property, the greater the score of its related concept
    - ...

History = Model of Forgetting? (cont.)

- **Illustration: MIAMM (access to a music database)**
  - Play this tune, please
  - Play this tune
  - This tune
  - This tune

- **Illustration: OZONE (transport information service)**
  - I want to go to Paris – here are the ways – how long does it take to go by this way? – what is this station? – are there a lot of trains that take this way? – where is Meudon station? – etc.
  - Where do you want to go?/Do you still want to go to Paris?

Multimodal dialogue system architecture
Illustration: MIAMM Architecture

Illustration: OZONE Architecture
Part II

Multimodal Meaning Representation Format

General Purpose

• Providing means to specify models for semantic content representation
  – Cf. work conducted on the Linguistic Annotation Framework (LAF) within ISO/TC 37/SC 4
  – Objective: independence vis-à-vis specific (XML) formats

• Main elements to consider
  – Meta-models
    • General, underlying information structure that informs current practice
      in the field
  – Data-categories
    • Elementary descriptors allowing one to implement a given meta-model
      – Obtained by sub-setting a Data Category Registry
**Metamodels and Data Categories**

**Catalogue of meta-models**

**Data category Registry**

- /inform/
- /bridging/
- /accept/
- /agent/
- /narration/
- /coreference/

**Application specific model**

**MMIL: Context**

- EU IST/MIAMM project (Multidimensional Information Access using Multiple Modalities - with DFKI, TNO, Sony, Canon)
  - Study of the design factors for a future haptic PDA like device
  - Underlying application: multidimensional access to a musical database

- Design and implementation of a unified representation language within the MIAMM demonstrator
  - MMIL: MultiModal Interface Language
  - “Blind” application of (Bunt & Romary 2002)
Main Characteristics of MMIL

- Basic ontology
  - Events and participants (organized as hierarchies)
  - Restrictions on events and participant
  - Relations: event-participant, event-event, participant-participant

- Additional mechanisms
  - Temporal anchoring of events
  - Ranges and alternatives

- Representation
  - Flat meta-model

MMIL Meta-Model (UML)
Example (Parser Output)

Spiel bitte das Lied
(Please play the song)

```
evtType=speak
dialogueAct=request

_evtType=play
_lex=vorspielen

 objectType=user

 objType=tune
 refType=definite
 refStatus=pending
```

XML Representation

```xml
<mmilComponent>
  <event id="e0">
    <evtType>speak</evtType>
    <dialogueAct>request</dialogueAct>
    <speaker target="p1"/>
  </event>
  <event id="e1">
    <evtType>play</evtType>
    <lex>vorspielen</lex>
  </event>
  <participant id="p1">
    <objType>user</objType>
  </participant>
  <participant id="p2">
    <objType>tune</objType>
    <refType>definite</refType>
    <refStatus>pending</refStatus>
  </participant>
  <relation source="e1" target="e0" type="propContent"/>
  <relation source="p1" target="e1" type="destination"/>
  <relation source="p2" target="e1" type="object"/>
</mmilComponent>
```
MMIL: Overview

- A highly specific meta-model
  - Events + participants (flat structure)
    - Relies on strong ontological assumptions
  - Subject to debate in the community

- Generic data categories
  - Dialogue acts (/opening/, /inform/, etc.), basic argument types (/agent/, /patient/)
    - Note: strong link to application specific concepts
  - Possibility to share expertise across applications

Conclusion and Future Work

- All parts of the dialogue history can be modeled with reference domains
  - Linguistic history = linguistic domains + ?
  - Visuo-gesture history = visual domains + gesture domains + ?
  - Task history = task domains + ?

- MMIL appears as an efficient framework for representing heterogeneous information in a homogeneous manner
  - Towards a stronger connection between reference domains and MMIL
  - Towards a normalization of MMIL
  - From specific dialogue systems to a generic architecture based on MMIL