A Text-based Search Interface for Multimedia Dialectics

Katerina Pastra
Institute for Language & Speech Processing
Athens, Greece
kpastra@ilsp.athena-innovation.gr

Eirini Balta
Institute for Language & Speech Processing
Athens, Greece
ebalta@ilsp.athena-innovation.gr
COSMOROE & COSMOROE Search Engine

• COSMOROE (CrOSs-Media inteRactiOn rElations)
  – A framework for modelling multimedia dialectics (semantic interplay between images, language and human actions)
  – Defines a number of semantic relations between different modalities, for formulating multimedia messages

• COSMOROE Search Engine
  – A text-based search engine for multimedia documents
  – A support tool for COSMOROE framework
  – Indexes and retrieves information from audiovisual files, manually annotated
  – Facilitates the active exploration and presentation of the semantic interplay between different modalities
  – Allows interfacing with semantic lexical resources for automatically expanding and generalising the semantic relations
COSMOROE Framework

- Equivalence: the information expressed by the different media is semantically equivalent (paradigmatic relations between modalities)
- Complementarity: the information expressed in one medium is (an essential or not) complement of the information expressed in another medium (syntagmatic relations between modalities)
- Independence: each medium carries an independent message, the combination of which creates the multimedia message
**COSMOROE Examples**

**Equivalence**

**Literal**

Token-Token: the media refer exactly to the same entity, uniquely identified as such

**Figurative**

Metonymy: the modalities refer to different entities but the user considers them semantically equivalent, the referents come from the same domain, have same array of associations, there is no transfer of qualities from one to another

Metaphor: the modalities refer to different entities of different domains, the user considers them semantically equivalent, there is a transfer of qualities

Type-Token: one modality refers to a class of entities and the other to one or more representative members of the class

EACL 2009
## COSMOROE Examples

### Complementarity

<table>
<thead>
<tr>
<th>Essential</th>
<th>Non Essential</th>
</tr>
</thead>
<tbody>
<tr>
<td>the meaning of what is communicated is clearly comprehended only when information by all participating media is combined</td>
<td>one medium provides extra information to what the other expresses, which is not vital for the comprehension of the message</td>
</tr>
</tbody>
</table>

### Exophora

<table>
<thead>
<tr>
<th>Cases</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exophora: cases of &quot;anaphora&quot;, in which one medium resolves the reference made by another, signals of resolution are present in discourse</td>
<td>Exophora: cases of &quot;anaphora&quot;, in which the entity referred to in one medium is revealed in another, though this is not vital for understanding the message</td>
</tr>
</tbody>
</table>

### Agent-Object

<table>
<thead>
<tr>
<th>Cases</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent-Object: cases of &quot;ellipsis&quot;, in which one medium reveals the agent or object of an action, event or state, expressed by another</td>
<td>Agent-Object: one modality reveals the agent or the object of the action, state or event expressed by another, but this information is also mentioned in the close context or easily inferred</td>
</tr>
</tbody>
</table>

### Defining Apposition

<table>
<thead>
<tr>
<th>Cases</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defining Apposition: one medium provides extra information to another, which identifies or describes something or someone (the relation is tied to the specific context and should not be considered as generally valid)</td>
<td>Non Defining Apposition: one medium reveals a generic property/characteristic of the very concrete entity mentioned by another medium, which is not vital for the clear comprehension of the message</td>
</tr>
</tbody>
</table>

### Adjunct

<table>
<thead>
<tr>
<th>Cases</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjunct: one modality functions as an adjunct to the other (place-position, place-direction, manner)</td>
<td></td>
</tr>
</tbody>
</table>

---

© EACL 2009
COSMOROE Examples

Independence

Contradiction: one modality expresses the exact opposite information to what the other expresses (irony, humour, editorial mistake), one may distinguish all subtypes that Equivalence has

Symbiosis: cases of different pieces of information being expressed by the media, the conjunction of which (conjunction in time or space) serves "phatic" communication purposes, i.e. one medium provides some information and the other shows something that is thematically related, but does not refer or complement that information in any way

Meta-Information: one medium reveals extra information on what the other expresses, through its own means of realisation (e.g. camera movement) or through non-propositional aspects of its realisation (e.g. beats, prosody etc.), the medium expresses information that comments on aspects of what the other expresses, going beyond the message communicated to creation-related comments (who created the message, when, why, how – cf. typical archival metadata)
A more...challenging example

«απολαμβάνουμε μελωδίες από επαγγελματίες μουσικούς...»
“we enjoy melodies by professional musicians...”

Textual metonymy: logical metonymy / perception event for its trigger

Multimedia Metonymy: Perception event for its trigger

Body Movement: Play music
COSMOROE Search Engine

The engine allows one to:

- perform a simple text-based search, taking advantage of the COSMOROE relations -behind the scenes- for more precise and intelligent retrieval within multimedia archives, and
- perform an advanced text-based search, filtering the query with criteria that are directly related to multimedia semantics and to modality specific information

Search is done:

- using single or multiword query terms (keywords) that are mentioned in the transcribed speech (or other text) of the video or in the visual labels set of its visual-units, in order to find instantiations of different semantic relations in which they participate
- using a pair of single or multiword query terms (keywords) that are related through (un)specified semantic relations
- for specific types of relations and find out how these are realized through actual instances in a certain multimedia context
- for specific modality types (e.g. specific types of gestures, image-regions etc.) and find out all the different relations in which they appear

Simple Search: targets the general users, queries the audiovisual file in a straightforward manner, like in any text-based search engine

Advanced Search: is addressed to people with a special interest in multimedia semantics and/or multimedia system development, multimedia relations act as semantic filters
**COSMOROE Search Engine Results**

- The search results are the multimedia files that matched the given query
- They are presented in a list, accompanied by the transcribed text
- Simple Search results are centered around the video clips presenting the relation files, while Advanced Search results are centered around the relation files themselves
- Simple Search results are sorted according to a modified tf/idf score, tuned to favour equivalence and apposition relations, over contradiction, complementarity and symbiosis relations
- The option of saving the content of the retrieved relation files is also provided
- Statistics for the various types and subtypes of relations, as well as for the modalities participating in relations are provided
COSMOROE Search Engine Relation File

- Individual retrieved relation files are accessible, in which the complete structure of the multimedia relation is presented, exposing the semantic interplay between different modalities.
- Depicted here are textual and visual components of the relation, information and comments about the type, as well as highlighted indications about the query matches. The user can also watch the video clip "containing" the whole relation, or the video clips of the modalities that participate in it and see static images of specific objects, with the objects' contour highlighted.
- Each individual relation can also be saved.