Linguistic and multimodal annotation in ISLE/NIMM

The Pisa Group

International Standards for Language Engineering NSF-EC Project

Coord: A. Zampolli; M. Palmer

3 Working Groups:

- **Multilingual Computational Lexicons CL-WG**
  - Coord: N. Calzolari, R. Grishman, M.Palmer
- **Natural Interaction and Multimodality NIMM-WG**
  - Coord: O.N. Bersen, M. Liberman
- **Evaluation of HLT Systems E- WG**
  - Coord: M. King, E. Hovy

" several subgroups of experts, from academia and industry
- ↑ representing projects/groups/know-how to be conveyed in ISLE
- ↑ to build consensus around international workshops

ISLE standards & guidelines will be
- ↑ validated in RTD and National projects,
- ↑ disseminated widely, with exemplary data & prototype tool

Main Results in Lexicon & Corpus WGs: First Phase

(www.ilc.pi.cnr.it/EAGLES96/home.html)

- Standard for **morphosyntactic encoding of lexical entries**, in a multi-layered structure, with applications for **all** the EU languages
- Standard for encoding **subcategorisation in the lexicon**: a set of standardised basic notions using a frame-based structure
- Towards a proposal for a basic set of notions in **lexical semantics**: focus on requirements of Information Systems and MT

- **Corpus Encoding Standard (CES)** developed from TEI
- Standard for **morphosyntactic annotation** of corpora, to ensure compatibility/ interchangeability of concrete annotation schemata
- Preliminary recommendations for **syntactic annotation** of corpora
- **Dialogue annotation,** for integration of written and spoken annotation

Content vs. Format/Representation

Work on lexical description deals with two aspects:

- ***linguistic description*** of lexical items (***content***)
- ***formal representation*** of lexical descriptions (***format***)

⇩ EAGLES has concentrated on ***linguistic content***, not disregarding the formal representation of the proposal

MILE
Multilingual ISLE Lexical Entry

→**Objective**: definition of the **MILE**,

→as a ***meta-entry***

- → its basic notions
- → general architecture

- to be encoded in a DTD
- with a tool to support it

open to **task- & system-dependent parameterization**

(www.ilc.pi.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm)

Outline

- Focus on current practice in spoken and multimodal annotated dialogue corpora
- Recent standardization efforts
- to discuss: possible relationships with TEI...

Annotated corpora in HLT

- Annotated corpora as a critical component of research in the speech and language sciences.
- A wide variety of spoken and multimodal language resources are being created, together with:
  - coding schemes
  - coding tools
  - coding formats

- Multimedia extensions and rich annotation are making the resources more complex
- Multimedia extensions are making the resources much larger

Some terminology

- Language resource: a digital repository of structured information intended to document natural language and natural communicative interaction (S. Bird)
- Corpus: a specialized form of LR, where the primary content is either a sequence of characters or a sequence of samples taken from human output

- Annotation: any descriptive or analytic notations applied to raw language data
- Coding scheme: the inventory of descriptive categories used for annotation
- Coding format: SGML, XML

Speech corpora

- CHILDES
- CHRISTINE
- MAPTASK
- Switchboard
- Timit
- BAS Partitur
- Lacito
- LDC Callhome
- Trains

- TAL
- AVIP

Sound-only communication between humans (and machines)

An abstraction

Multimodal corpora

Common situated dialogue is natural interactive dialogue

To study actual human communication

To acquire knowledge of the different communicative components of natural human communication

- Some 26 resources surveyed by ISLE/NIMM

Spoken dialogue annotation

- Multi-level
- Multi-layered
- More than one level of annotation on the same data
- Queries including more than one level of annotation

Multi-level Annotation: MATE

- Transcription
- Prosody
- Morphosyntax (PoS)
- Syntax
- Co-reference
- Pragmatics
- Communication problems

Multi-level annotation: TAL

- Transcription
- Prosody
- Morphosyntax (PoS)
- Syntax
- Semantics
- Pragmatics

Multimodal Annotation

- A modality is a particular way of presenting information in some medium
- A medium is a physical channel for information encoding/decoding: visual, audio
- Multimodal representations are representations which can be decomposed into two or more unimodal modalities

Natural interactive dialogue modalities

- Speech
- Lip movements
- Facial expression
- Gaze
- Hand and arm gesture
- Bodily posture
- Cross-level
- Cross- modality

Types & Role of Gesture

Types of gestures (form)

- spontaneous co-speech gestures
  - iconic gestures
  - pointing gestures
  - emblematic gestures (stand alone)

Role of gestures (meaning)

- facilitating speech comprehension (disambiguation)
- facilitating speech production (planning)
- adding semantic, pragmatic, and discourse level information

Gesture annotation

Time Dimension

- general agreement on different phases

Movement Encoding

Inherent dimensions

- body parts
- static vs dynamic components
- path shape + direction, hand orientation + shape
- location relative to the body

FACIAL ACTION CODING SYSTEM (FACS)

To encode facial expression.

FACS has been developed by Paul Ekman and Wallace Friesen in 1976.

- describes visible facial movements
- anatomically based
- Action Unit (AU): action produced by one muscle or group of related muscles

any expression described as a set of Aus

combinations of Aus also allowed

FACS involves four operations:

- Determining which **AUs** are responsible for the observed move
- Scoring the intensity of the actions on a three-point scale: low (X), medium (Y), and high (Z)
- Deciding whether an action is asymmetrical or unilateral
- Determining the position of the head and the position of the eyes during a facial movement.

46 AUs are defined

ALPHABET OF THE EYES

- Created by Isabella Poggi, Nicoletta Pezzato and Catherine Pelachaud
- Annotation level: gaze;
  - eyebrow movements, eyelid openness, wrinkles, eye direction, eye reddening and humidity.

Description

- 1.eyebrows:
- right/left: inner part: up / down/ central
  medial part: up / down/ central
  outer part: up / down/ central
- 2.eyelids:
- right/left: upper: default / raised / lowered
  default / tense / corrugated
  blinking / winking / closed
  lower: default / raised / lowered
  default / tense / corrugated

EXAMPLE: FEAR

- Eyebrow: right/left: internal, central, external: up •
- Eyelid: right/left: upper: up, tense
- Eyelid: right/left: lower: down, tense
- Humidity: default
- Reddening: default
- Pupil dilation: no
- Focusing: yes
- Iris position: central
- Iris direction: right
- Face direction: right
- Head inclination: default
- Trunk direction: forward
- Interlocutor direction: right
- Duration: short

What is needed for LR

- Dissemination
- Reuse
- Reannotation

The problem

- Data: different types of data (text, audio/video streams) in different formats
- Metadata: necessary for ease and reliability in accessing, locating and evaluating LR
- Encoding formats
- Superimposed structure (annotation): phonetic features, discourse structure, grammar, syntax, semantic tagging, named entity identification, co-reference annotation, gesture, gaze, facial expression, etc.

Standards

- To facilitate interchange, processing and reuse of data among researchers, tools and application and research purposes

Initiatives on spoken and multimodal annotation

- EAGLES: Expert Advisory Group on Language Engineering Standards
- MATE: Multilevel Annotation, Tools Engineering
- Talkbank

Initiatives on multimodal annotation

- ISLE/NIMM: International Standards for Language Engineering.
- NITE: Natural Interactivity, Tools Engineering

EAGLES

- Special interest group on spoken language
- Handbook on Standard and Resources for Spoken Language Systems (see http://coral.lili.uni-bielefeld.de/~gibbon/EAGLES/)
- A set of recommendations
- what to encode

MATE

- Http://mate.nis.sdu.dk
- Multilevel annotation and tools engineering
- March 1998- Dec 1999
- Eight partners
- Five countries

MATE goals

- Facilitate use and re-use of spoken language resources
- Standard framework or architecture for spoken language corpora
- MATE tool
- MATE best practice annotation schemes

ISLE/NIMM

- http://lingue.ilc.pi.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm
- Natural Interaction and Multimodality working group
- 10 European sites
- three WP

ISLE/NIMM

Annotation Schemata and Systems

- A survey of annotation schemata in relation to gesture, facial expression, eye-gaze, sign language and cross-modality.
- Create, refine and validate guidelines and standards for annotation of spoken dialogue, gesture, face, sign language, cross-modality, and discourse and towards identifying a common framework for NIMM annotation tools and systems.

Meta-data description for browsing of large multimodal databases

- Define flexible schema which allows the creator of a multimodal resource to describe format and content.
- Define the type of search and browse tools necessary for navigating in the universe of meta-descriptions;

ISLE/NIMM

Tools for natural interaction and multimodal data annotation

- A survey and analysis of already existing tools and of current and emerging standards
- to create a basis for establishing guidelines and requirements for further development of such an environment.

NITE

- Http://nite.nis.sdu.dk
- extends MATE goals to multimodal natural interaction

- There is a need for a general framework for linguistic annotation that is flexible and extensible enough to accommodate different annotation types and different theoretical and practical approaches.

Emerging consensus

- The standardization of tagsets is necessarily an open-ended task, and is always subject to revision as the underlying domains change and the theories evolve.
- Yet the standardization of the annotation structures themselves is feasible goal in a relatively short timeframe. This is the primary step for the creation of general-purpose tools and formats.
- "Semantic" standards
- "Architectural" standards

Emerging consensus

- Flexibility and extensibility a must (allow a great variety of schemes to be represented and mutually substituted in a theory-independent way)
- core and periphery (modular approach)
  - "how to encode" vs. "what to encode"
- Need for architecturally-oriented standards
- Standoff annotation

Standoff Annotation

- Each layer of annotation is kept conceptually separated from others by establishing identifiers to which other layers are linked
- Layers can be kept physically separate in different documents
- There is no need to keep all data in one document
- Improves readability, maintenance and use of data

What about TEI?

- **Two issues:**
- Roots: humanities vs. speech and natural language processing
- Focus: representation of objective features of text vs. explicitation of information implicit in linguistic signal

Examples of Differences

- Prototypical types of data: text vs audio/visual streams
- Representational issues:
  - The distinction between representation and annotation has become commonplace
  - wide amount of levels of linguistic annotation
- Markup structure
  - standoff annotation vs. embedded annotation

Convergences

- Common goals
- similarly against a prescriptive approach and for a recommendation-oriented approach
- similarity between the TEI notion of interlingua and that of Eagles/ISLE or MATE meta-scheme.
- Markup language