Tools and Methods for Multimodal Annotation: the OTIM Project

Philippe Blache

Laboratoire Parole et Langage
CNRS & Université de Provence

Programme

Monday, May 23rd	
9h15-9h45	General presentation *Philippe Blache*
9h45-10h30	Primary data : Transcription, Alignment, Methodology *Robert Espesser, Brigitte Bigi*
11h-11h45	Automatic annotations: tools, results *Stéphane Rauzy et Daniel Hirst*
11h45-12h30	Manual annotations: overview *Roxane Bertrand, Béatrice Priego-Valverde*
13h45-14h45	Manual annotations: Disfluencies, Discourse, Gestures *Laurent Prévot, Ning Tan, Gaëlle Ferré, Marion Tellier*
14h45-15h15	Illustration: Backchannels, Reinforcing gestures Gaëlle Ferré, Roxane Bertrand
15h15-16h	Coding scheme and XML *Philippe Blache, Julien Seinturier*
16h15-17h	Querying multiple documents *Elisabeth Murisasco, Emmanuel Bruno, Julien Seinturier*

Programme

Tuesday, May 24th	
9h30-10h15	*Harry Bunt, Tilburg University* ISO/DIT dialogue annotation and its semantics
10h15-11h	*Nancy Ide, Vassar College* The Open American National Corpus : An Interoperable, Open Collaborative Annotation Project
11h30-12h15	*Christopher Cieri, UPenn, LDC* Language Resources for ? Linguists??? Adapting Corpora and Methods for Interdisciplinary Research
14h-14h45	*Michael Kipp, DFKI Saarbruecken* Sign language coding, 3D behavior data and ANVIL
14h45-15h30	*Nick Campbell, Trinity College Dublin* Some Corpora Illustrating our Approach to the Collection of Unstructured Social Speech (and ways to describe it)
16h00-16h45	*Jonathan Ginzburg, Paris 7* Integrating multimodal interaction into learning in dialogue
16h45-18h	*Daniel Hirst, LPL* Do we need explicit models of prosodic form to interpret spoken data?

Outline

1. The OTIM Project: general overview
2. Multimodal annotations: a life-size experiment
3. The formal background

Part I

The OTIM Project: General Overview

OTIM: *Outils pour le Traitement de l'Information Multimodale*

- Multimodal annotation
  - Different modalities
  - Different domains: phonetics, prosody, syntax, pragmatics, etc.
- Goals
  - Description of modalities and their interaction
  - Analysis of natural communication
- Questions
  - Generality: annotation reusability
  - Representation, encoding
  - Alignment vs. synchronization
  - Diversity of annotation tools and formats
  - Data manipulation, querying
- Method
  - Rich annotation for each domain
  - Stand-off, independent annotations
  - Homogeneous representation

Multimodal Corpora: a survey

- Switchboard in NXT project
  - 642 conversations, 830,000 words.
  - *Syntax, turns, disfluency, information status, coreference, phonemes, syllables, prosodic phrases, breaks, accents*
- LUNA (Spoken Language Understanding in Multilingual Communication Systems)
  - 8100 human-machine dialogues and 1000 human-human dialogues in Polish, Italian and French.
  - *Turns, POS, chunks, dialogue acts, reference*
- SAMMIE (Saarbrücken Multimodal MP3 Player Interaction Experiment)
  - Multimodal dialogue system, human-machine multimodal interaction (Wizard of Oz)
  - *Transcription, turns, clauses, discourse entities, dialogue acts*

Multimodal Corpora: a survey

- AMI (Augmented Multi-party Interaction)
  - 100h meeting, full manual transcription
  - *Dialogue acts, focus of attention, movement (hand, head, leg), named entities, topic segmentation*
- The ITC Corpus
  - 11 groups of 4 people (25 minutes each). Task: decision making scenario
  - *No transcription, functional role, socio emotional, speech activity, body activity*
- The ATR Corpus
  - 10 meetings, 1 hour each
  - *No transcription, speech activity, body movements, activity type*

Multimodal Corpora

Part II

Multimodal Annotation: a Life-size Experiment

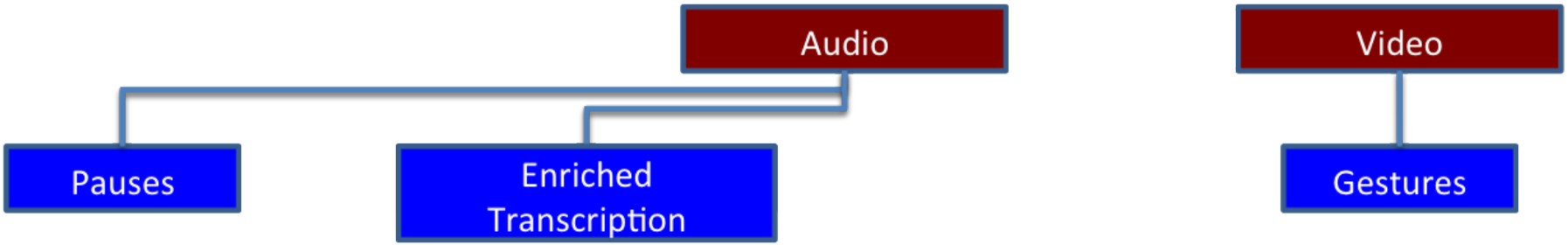
A life-size experiment: CID (Corpus of Interactional Data)

- 8 dialogs, 1 hour each (4 male/male; 4 female/female)
- Task: *tell something unusual which happened to you*
  tell about professional conflicts you may have met
- Setting
  - Anechoic room
  - 1 camcorder / 2 microphones
- Annotations (aligned on the signal)
  - Phonetic and orthographic transcription
  - Prosody (units, intonation, contours)
  - Morphosyntax, syntax
  - Discourse (markers, turns, etc.)
  - Gestures

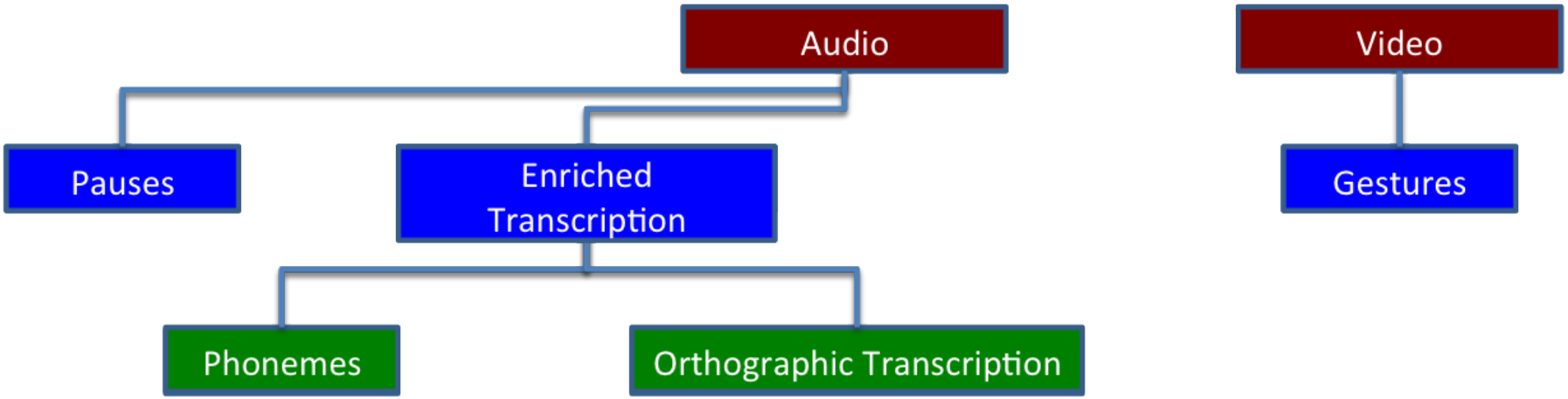
Illustration

Illustration

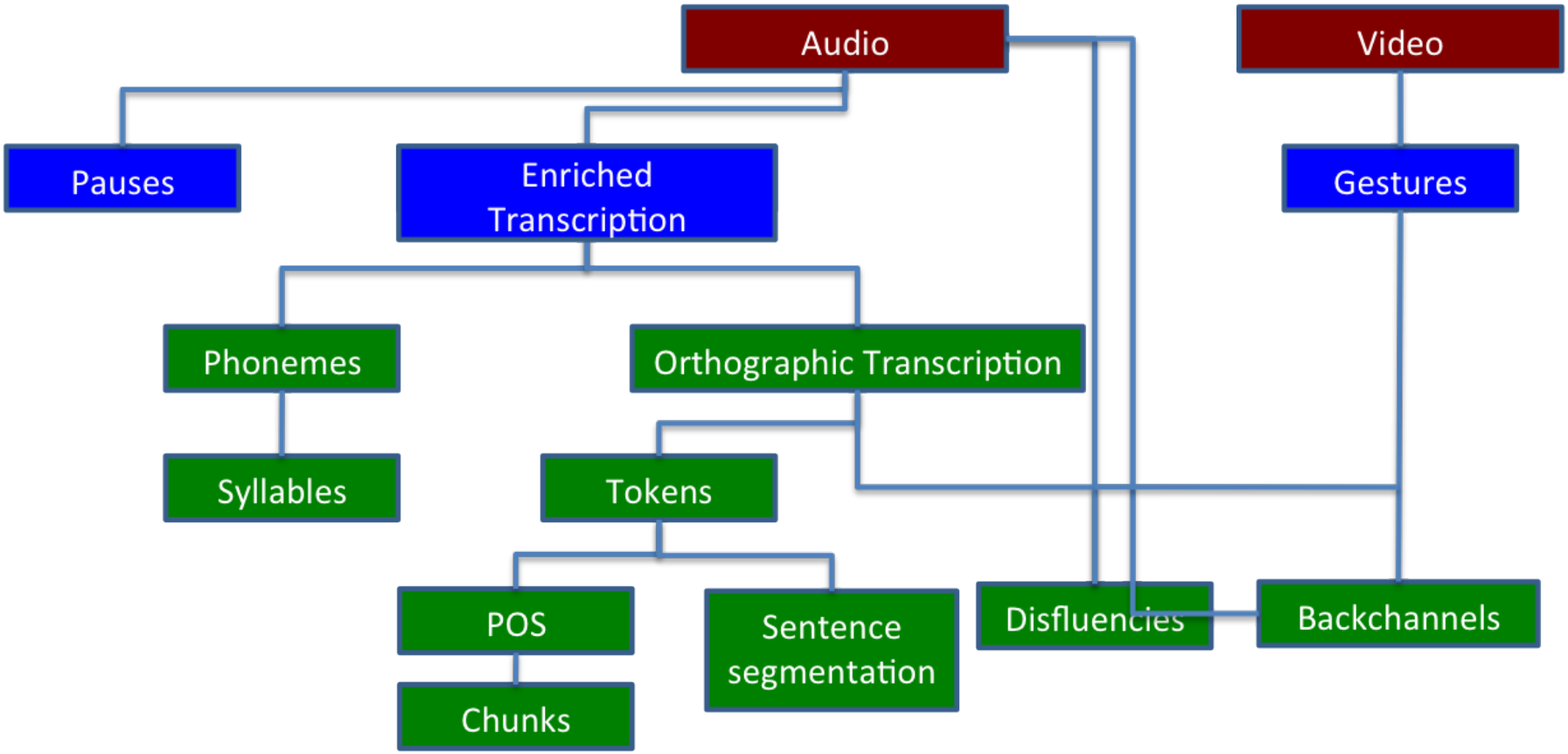
The Annotation Architecture



The Annotation Architecture



The Annotation Architecture



Main steps and contributions

1. Primary Data Preparation
   - Transcription: Enriched Transcription Convention <<< OTIM
   - Generation of orthographic and phonetic transcriptions
   - Aligning transcriptions with the signal <<< OTIM
2. Automatic Annotation
   - Syllabification <<< OTIM
   - Intonation
   - Sentence segmentation <<< OTIM
   - POS-tagger
   - Chunker

Main steps and contributions

3. Manual Annotation <<< OTIM
   - Gestures: hands, head, arms
   - Prosody: phrasing, contours, intonation
   - Disfluences
   - Discourse: turns, backchannels, reported speech, information structure
   - Syntax: detachments

4. Formal representation, data
   - Abstract schema: Typed Feature Structures <<< OTIM
   - Generation of the XML schema <<< OTIM
   - Formatting data
   - Querying

Some descriptions

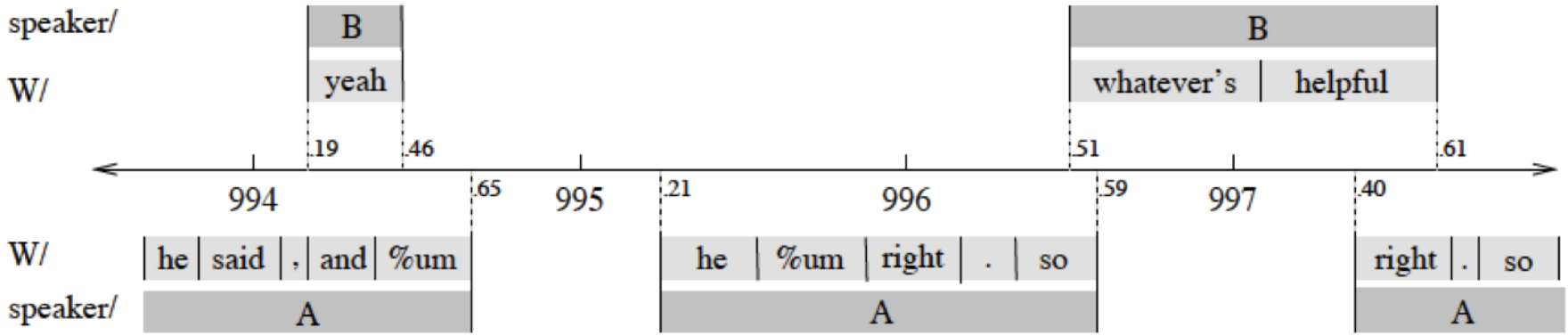
1. Backchannels <<< OTIM
   - Vocal and gestural
   - Description in terms of prosody, discourse, morpho-syntax

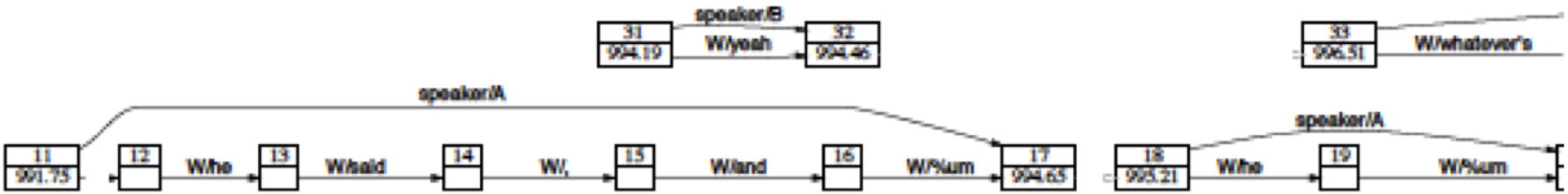
2. Detachments <<< OTIM
   - Dislocation, cleft, topicalization
   - Annotation of the detachment type, the category, the function, the anaphor

Part III

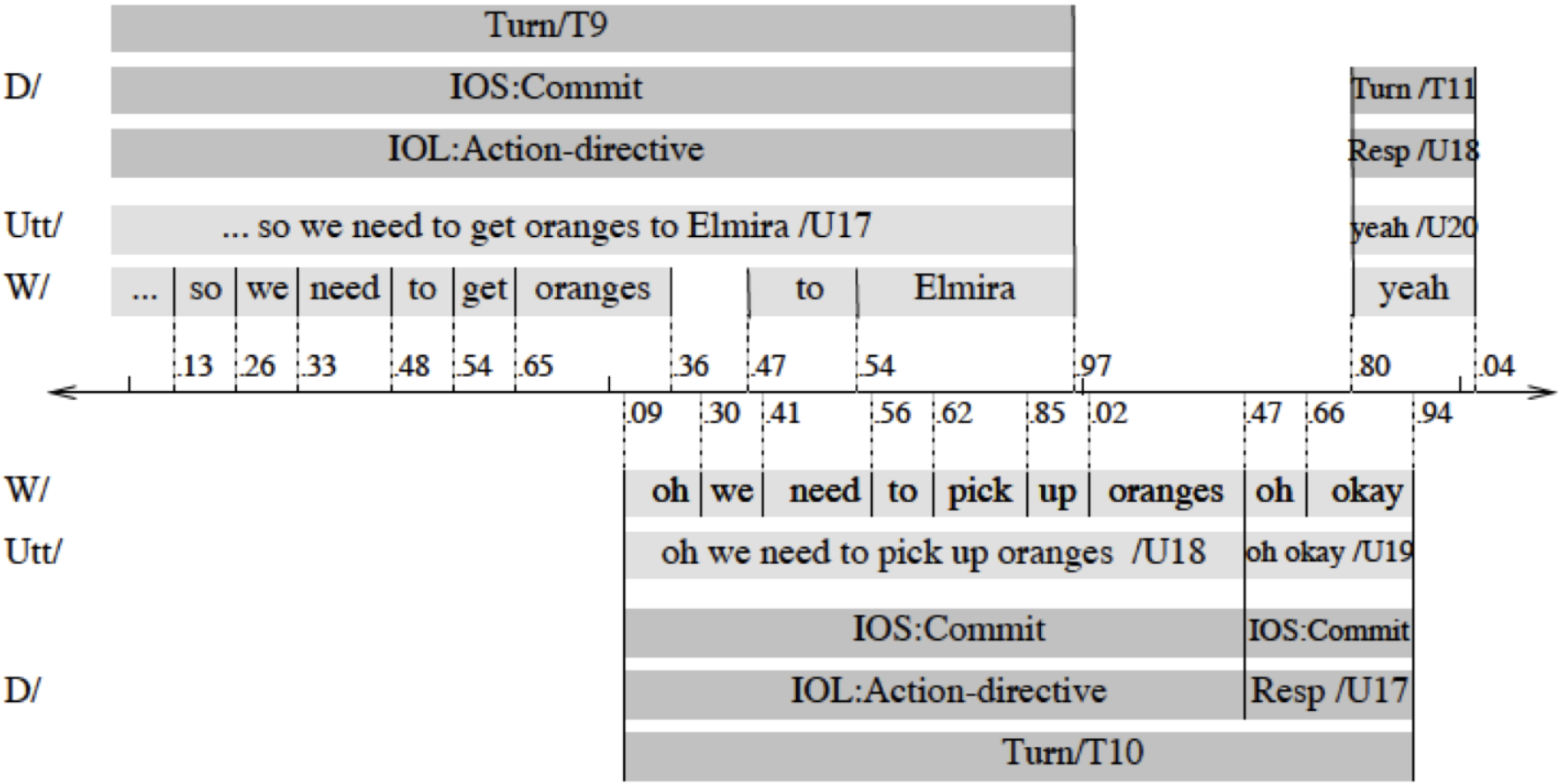
The Formal Background

Annotation Graphs

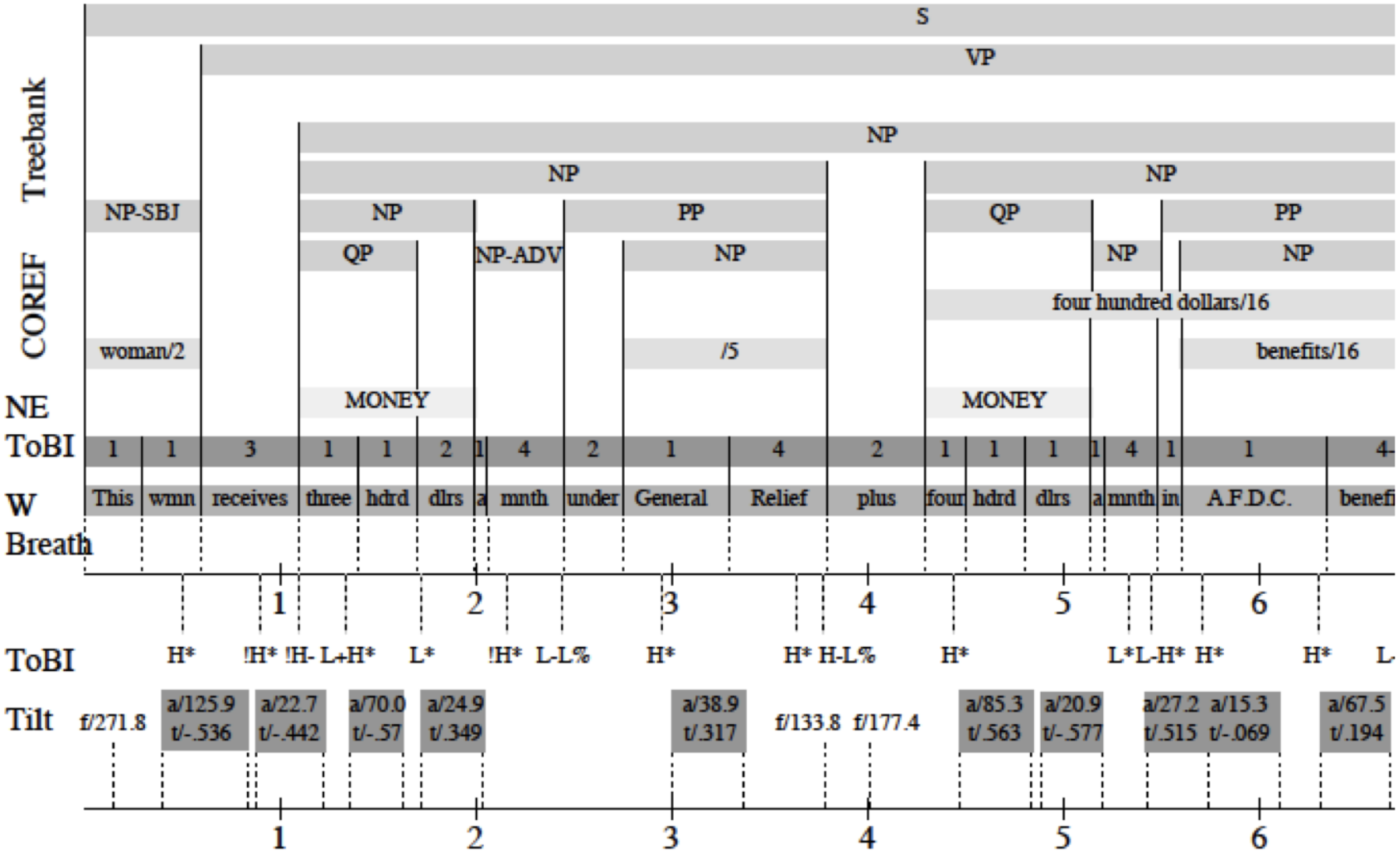




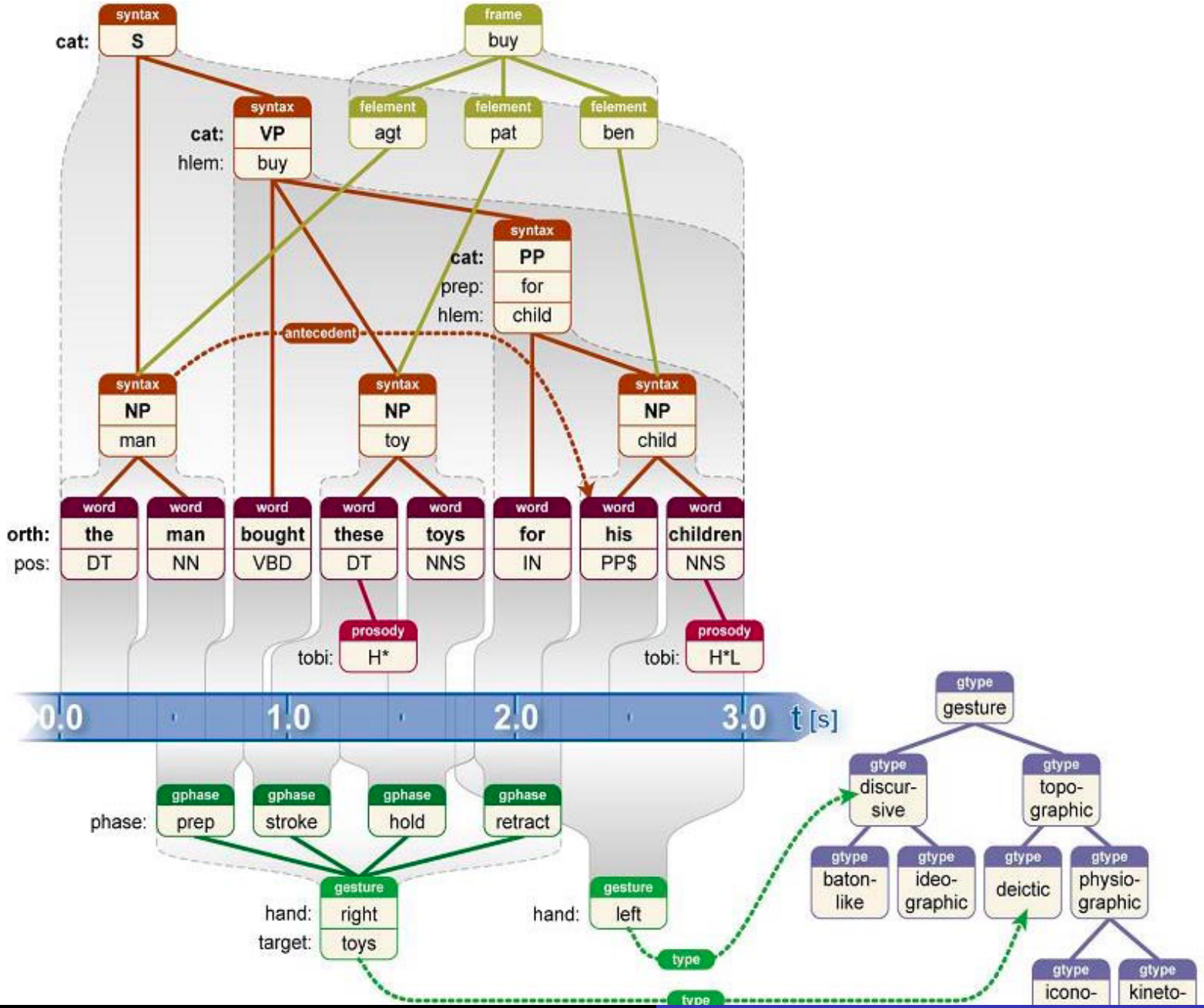
Annotation Graphs



Annotation Graphs

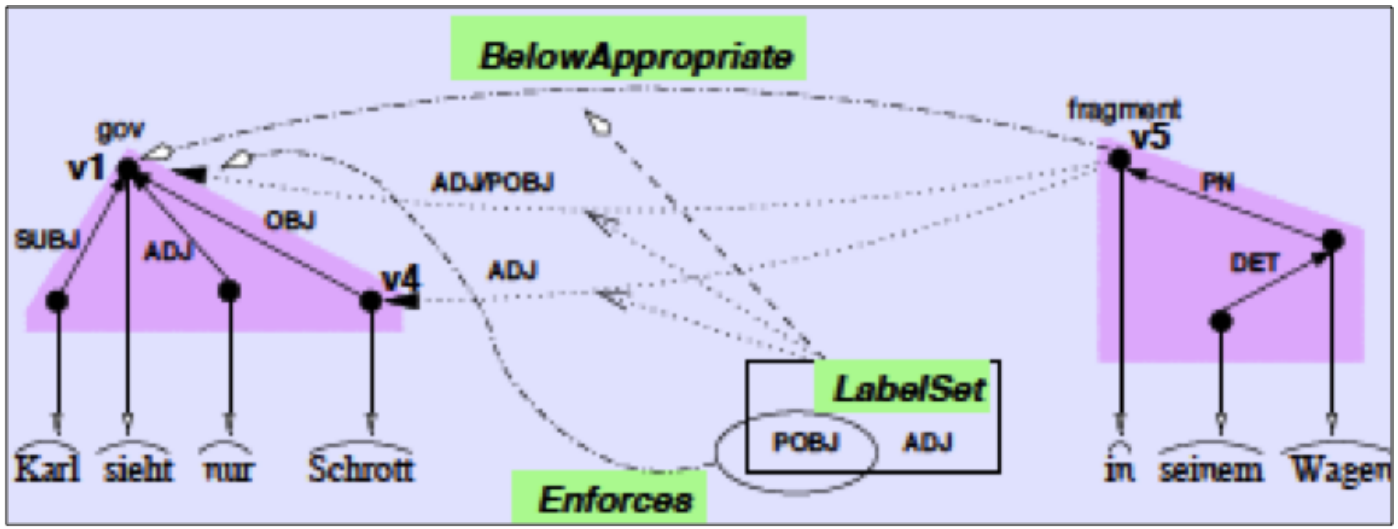


NITE XML Toolkit (NXT)



Graph Annotation Format (GrAF)

- GrAF: nodes and edges, decorated with feature structures
- Annotations label nodes (rather than edges as in AG)
- Nodes may be linked to:
  - Primary data
  - Other nodes in the graph



Graph Annotation Format (GrAF)

- Base segmentation:

```
<seg:sink seg:id="42" seg:start="24" seg:end="35"/>
```

- Annotation over the base segmentation:

```
<msd:node msd:id="16">
    <msd:f name="cat" value="NN"/>
</msd:node>
<msd:edge from="msd:16" to="seg:42"/>
```

- Annotation over another annotation:

```
<ptb:node ptb:id="23">
    <ptb:f name="type" value="NP"/>
    <ptb:f name="role" value="SBJ"/>
</ptb:node>
<ptb:edge from="ptb:23"to="msd:16"/>
```

Discussion

Problems

- Some referential objects are not aligned with the signal
  - Concrete objects (the objects of the scene)
  - Abstracts referents (ideas, concepts, etc.)
- Some phenomena are not synchronized (e.g. deictic gestures)

Elements of answer

- Different anchors: time, spatial, index
- An annotation is described by its properties and its anchor
- A construction is a set of annotations, not necessarily ordered

Formally

- A construction is an hypergraph
- Nodes are annotations
- Nodes and edges are labelled with feature structures

Discussion

Problems

- Some referential objects are not aligned with the signal
  - Concrete objects (the objects of the scene)
  - Abstracts referents (ideas, concepts, etc.)
- Some phenomena are not synchronized (e.g. deictic gestures)

Elements of answer

- Different anchors: time, spatial, index
- An annotation is described by its properties and its anchor
- A construction is a set of annotations, not necessarily ordered

Formally

- A construction is an hypergraph
- Nodes are annotations
- Nodes and edges are labelled with feature structures

Discussion

Problems

- Some referential objects are not aligned with the signal
  - Concrete objects (the objects of the scene)
  - Abstracts referents (ideas, concepts, etc.)
- Some phenomena are not synchronized (e.g. deictic gestures)

Elements of answer

- Different anchors: time, spatial, index
- An annotation is described by its properties and its anchor
- A construction is a set of annotations, not necessarily ordered

Formally

- A construction is an hypergraph
- Nodes are annotations
- Nodes and edges are labelled with feature structures

Nodes

- **Index**: absolute reference to the node
- **Reference**: description of the general characteristics
- **Features**: description of the linguistic properties

$$
\text{NODE}\left[\begin{array}{ll}
\text{INDEX} & \\
\text{LABEL} & \\
\text{REFERENCE} & \left[\begin{array}{l}
\text{LOCATION}\left[\begin{array}{ll}
\text{TEMPORAL} & \left[\begin{array}{l}\text{START } \textit{value}\\ \text{END } \textit{value}\end{array}\right]\\
\text{SPATIAL } \textit{coord} &
\end{array}\right]\\
\text{REALIZATION}\left\{\textit{concrete}, \textit{abstract}\right\}\\
\text{MEDIUM}\left\{\textit{acoustic}, \textit{tactile}, \textit{visual}\right\}\\
\text{PRODUCTION}\left\{\textit{produced}, \textit{accessible}\right\}\\
\text{DOMAIN}\left\{\textit{prosody}, \textit{syntax}, \textit{pragmatics}, \ldots\right\}
\end{array}\right]\\
\text{FEATURES} &
\end{array}\right]
$$

Edges

$$\text{RELATION}\begin{bmatrix}\text{INDEX}\\ \text{LABEL}\\ \text{DOMAIN}\left\{\textit{prosody},\ \textit{syntax},\ \textit{pragmatics},\ \ldots\right\}\\ \text{REL\_TYPE}\left[\left\{\begin{array}{l}\text{ORIENTED\_REL}\begin{bmatrix}\text{FROM}\ \textit{index}\\ \text{TO}\ \textit{index}\end{bmatrix}\\ \text{SET\_REL}\left\langle\textit{node list}\right\rangle\end{array}\right\}\right]\\ \text{ALIGNMENT}\left\{\textit{strict},\ \textit{fuzzy}\right\}\end{bmatrix}$$

Part IV

Results Presentation

Results Presentation

1. Transcription, phonetization, alignment

   Robert Espesser and Brigitte Bigi

2. Automatic annotations: tools, results

   Stéphane Rauzy and Daniel Hirst

3. Manual annotations: overview

   Roxane Bertrand and Béatrice Priego-Valverde

4. Manual annotations: disfluencies, discourse, gestures

   Laurent Prévot, Ning Tan, Gaëlle Ferré, Marion Tellier

5. Illustration: backchannels, beinforcing gestures

   Gaëlle Ferré and Roxane Bertrand

6. Coding scheme and XML

   Philippe Blache and Julien Seinturier

7. Querying multiple documents

   Elisabeth Murisasco, Emmanuel Bruno, Julien Seinturier