# WG2: Scope of Standards and Standardization

Nicoletta Calzolari, Nancy Ide,
Rashmi Prasad, James Pustejovsky, Janyce Wiebe

November 7, 2009
CLASP, NYU

# Which classes of content categories are ready for standardization?

- POS
- Chunk
- Time
- Sentiment
- Discourse Markers
- Semantic Roles
- Dialogue Acts
- Grammatical Functions and Relations

# Which classes of content categories are ready for standardization?

- POS
- Chunk
- Time
- Sentiment
- Discourse Markers
- Semantic Roles
- Dialogue Acts
- Grammatical Functions and Relations

# Which classes of content categories are ready for standardization?

- Anything that's reached maturity within a sub-community of interest.

# Repository Use Case

- You have an idea for something to annotate/markup
- Visit repository searching for:
  - The same annotation ; reuse
  - Similar annotations: identify differences
- In either case, you are leveraging on existing annotations
  - Links to other annotations
  - Community
  - Corpora
  - Software, etc.

# Defining our Terms

1. **Interoperability**: Reuse and exchange and integration of data, resources, and tools.
2. **Harmonization**: Processes by which interoperability can be achieved.
3. **Standards**: Specifications which are required to achieve interoperability
4. **Best Practice**: Categories or procedures adopted by a community based on collective positive experience

# Motivations

1. Interoperability is critical for
   - Reuse of resources and tools
   - Distributed research and deployment
   - Multi-lingual, multi-modal application
1. Harmonization and standards for resources and tools enable interoperability:
   - But we have not so far achieved universally accepted standards

# What are the operational definitions for interoperability?

**two or more X, exchanging Y, via W, in order to Z**

1. Metadata
2. Data categories
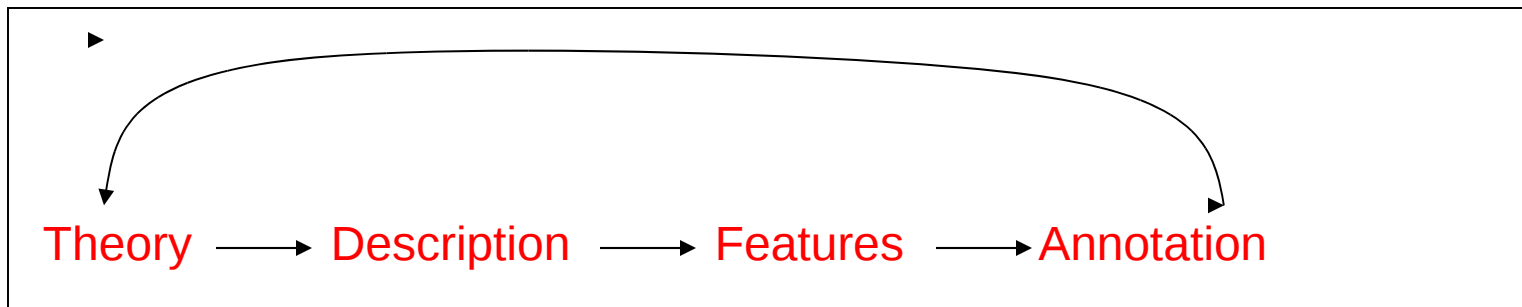3. Publication requirements
4. Software sharing

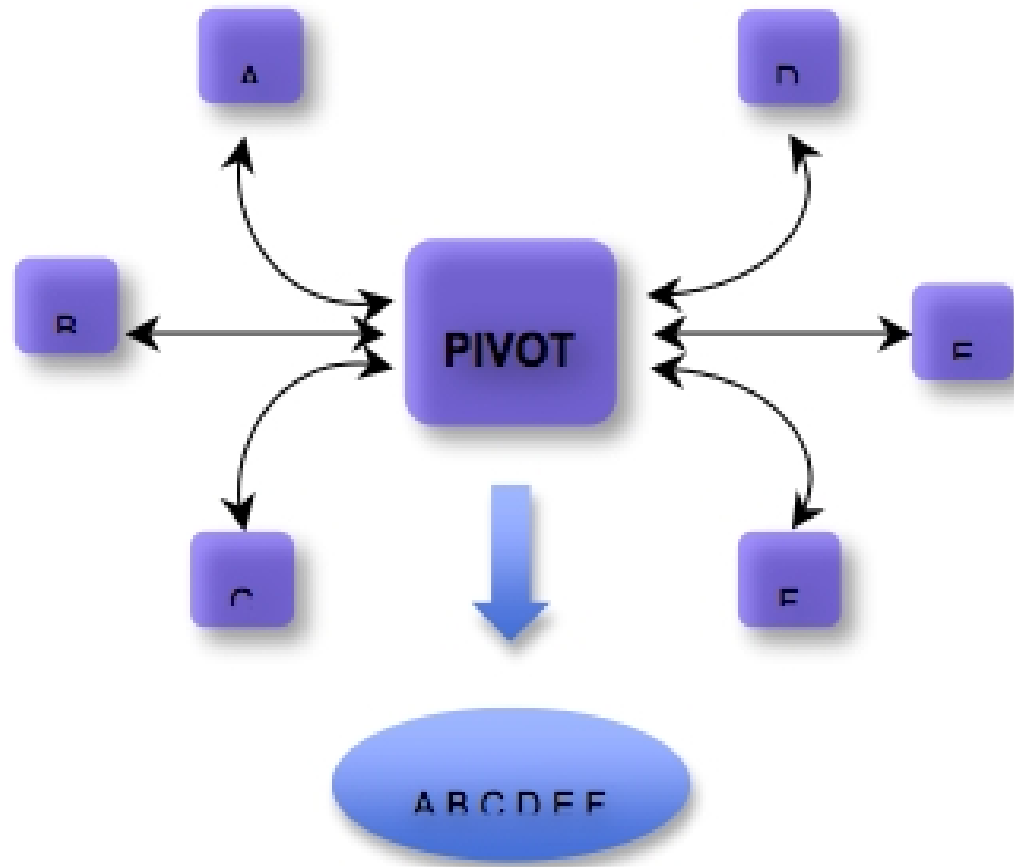**Find X, Y, W, Z**

# Consensus Building towards Standards

- Create abstract data model for a phenomenon
  - Define the set of basic notions
- Work to see if existing models can be expressed in terms of the model elements
- Iterate

# Adopting an abstract model

- Annotation scheme:
  - □ assumes a given feature set
- Feature set:
  - □ encodes specific structural descriptions and properties of the input data
- Structural descriptions:
  - □ theoretically-informed attributes derived from empirical observations over the data
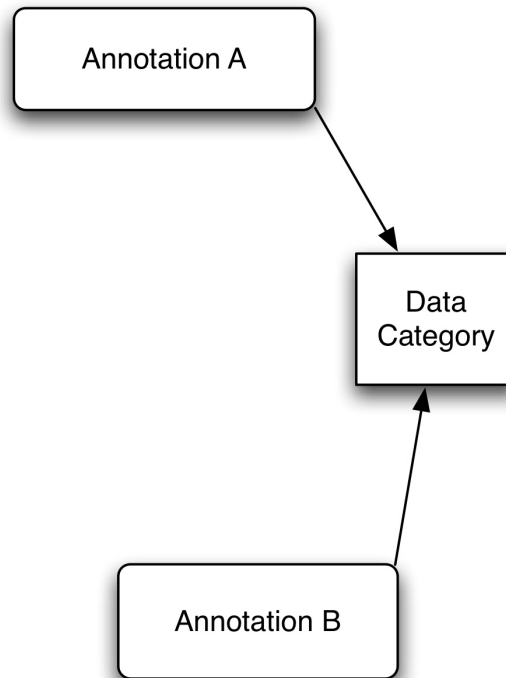
Theory → Description → Features → Annotation

# Harmonization via Pivot: LAF/GrAF

# Harmonization via Mapping

• Categories can be defined in terms of simpler, primitive features
• Different approaches bundle different subsets in different configurations

# Which classes of content categories are ready for standardization?

- The goal should be to promote Interoperability.
- To have interoperability, resources, data, and software should be harmonized.
- Therefore, standardization should be discussed in terms of harmonization

- So…. Anything that's reached maturity with a sub-community of interest.

# The Proper Role of Standards

- To promote harmonization
- Not to be prescriptive

# How should CLASP interact with prior standards?

- As much as possible. Where something is usable take it as a starting point, build from there.

# Should CLASP guidelines be anglo-centric?

No.

# Are there any more slides?

No.