

# The Mechanics of Semantic Annotation

Nancy Ide  
Vassar College

# Annotation Mechanics

- Annotations have to be represented some way
- GrAF: Graph Annotation Format
  - ISO TC37 SC4 WG1
  - XML serialization of Linguistic Annotation Framework (LAF) “pivot” format
  - Generic graph-based format
  - Enable merging of different annotations
  - Enable transduction among formats (“pivot”)

# Observations

- “Low-level” annotations
  - Close to text
  - Typically refer to one or more objects in linear order
  - Link signifies constituency
- Semantic++ annotations
  - Links among other annotations
  - Link signifies a wide variety of (named) relations

# MASC



- **Manually Annotated Sub-Corpus**
  - Drawn from free, unrestricted 15 million word **Open ANC**
  - 1<sup>st</sup> release (very soon) : 80K words
  - 2<sup>nd</sup> release (not too long): 140K words
  - **Validated annotations:**
    - Token, sentence, POS (Penn), shallow parse, named entities
    - WordNet senses, FrameNet frames
      - Feeding FN/WN harmonization effort
    - Penn Treebank syntactic annotation
    - (hopefully) others to be contributed for part or all:
      - PropBank, TimeML, Penn Discourse Treebank, HPSG annotation, co-reference (Poesio)...

<http://www.anc.org>

# MASC



- All in GrAF
- Tool to produce in many formats
  - Merge annotations of user's choice
  - XML in-line (XCES), UIMA format, GATE format, NLTK input, etc.
- Graph format itself very valuable as a structure for search and retrieval of complex patterns among multiple layers of annotation

**SILT**

# The Point

- If you annotate, consider some or all of the MASC corpus
  - Multiple genres : fiction, journal, blog, non-fiction, travel guides, ...
  - Completely unrestricted for any use
  - We will put your annotations into GrAF and make freely available
- Goal:
  - community-developed resource
  - a freely sharable, richly annotated corpus with inter-layer searchability

<http://www.anc.org>

# The Request

- What kinds of inter-layer relations are needed?
  - What kind of queries over graph structure?
- What kinds of links (generic) needed?
  - Dependent, two-way (equal), weighted, unweighted...
  - What should GrAF accommodate?