OntoLingAnnot's Ontologies – Facilitating Interoperable Linguistic Annotation up to the Pragmatic Level

Antonio Pareja-Lora

DSIC / ILSA – Universidad Complutense de Madrid (UCM)

ATLAS – Universidad Nacional de Educación a Distancia (UNED)

apareja@sip.ucm.es

Contents

- The OntoLingAnnot annotation model: background and foundations
- OntoLingAnnot's ontologies:
  - Development and overview
  - Annotation levels and layers
  - The pragmatic concepts/terms
    - The pragmatic (sub-)layers and (sub-)strata in the Linguistic Level Ontology (LLO)
    - The pragmatic units in the Linguistic Unit Ontology (LUO)
    - The pragmatic attributes in the Linguistic Attribute Ontology (LAO)
    - The pragmatic values in the Linguistic Value Ontology (LVO)
    - The pragmatic relations in the Linguistic Relationship Ontology (LRO)
    - Some statistics
- Conclusions
- Annex: Pragmatic annotation of a simple dialogue example

Contents

- The OntoLingAnnot annotation model: background and foundations
- OntoLingAnnot's ontologies:
  - Development and overview
  - Annotation levels and layers
  - The pragmatic concepts/terms
    - The pragmatic (sub-)layers and (sub-)strata in the Linguistic Level Ontology (LLO)
    - The pragmatic units in the Linguistic Unit Ontology (LUO)
    - The pragmatic attributes in the Linguistic Attribute Ontology (LAO)
    - The pragmatic values in the Linguistic Value Ontology (LVO)
    - The pragmatic relations in the Linguistic Relationship Ontology (LRO)
    - Some statistics
- Conclusions
- Annex: Pragmatic annotation of a simple dialogue example

Background (1)

LINGUISTIC ANNOTATION STANDARDISATION (ISO TC37 SC4)		
Linguistic Annotation Level	**Related ISO standard**	**Related terms/concepts**
Level-independent	LAF/GrAF (ISO, 2008a)	LINGUISTIC ANNOTATION, STAND-OFF ANNOTATION, ANNOTATION ENCODING, XML, RDF, OWL, etc.
	DCR/ISOCat (ISO, 2010)	DATA CATEGORY, ONTOLOGY, FEATURE, etc.
	LMF (ISO, 2008b)	NLP LEXICON, MACHINE READABLE DICTIONARY, INFLECTED ENTRY, MORPHOLOGICAL PATTERN, LEMMA, etc.
Morpho-syntactic annotation	MAF (ISO, 2008c)	NOUN, GENDER, MASCULINE, etc.
Syntactic annotation	SynAF (ISO, 2006)	SENTENCE, PHRASE, CONSTITUTION, DEPENDENCY, etc.
Semantic annotation	SemAF/Time (ISO, 2007)	EVENT, EVENT STRUCTURE, TEMPORAL RELATION, etc.
	SemAF-NE (ISO, 2010)	NAMED ENTITY, SEMANTIC TYPE, ORGANIZATION, etc.
	SemAF-DS (ISO, 2010)	DISCOURSE UNITS, COHERENCE RELATIONS (basically).
	SemAF-Dacts (ISO, 2009)	DIALOGUE, TURN, DIALOGUE ACT, etc.

- Level-dependent standards do not benefit from a comprehensive and holistic view of their annotations
- So it is difficult to define a common annotation scheme or meta-scheme.

Background (2)

OntoTag (Aguado de Cea; 2002; 2004)

Linguistic Knowledge & Terminology

ISO

Data Category Registry (DCR/ISOCat)

ISO

Linguistic Ontologies

Input text

Standardised Annotation Encoding

Standardised Annotation

LAF/GrAF (ISO, 2008a)

Subject | *Predicate* | *Object*

Triples

Standardised linguistic annotations

Languages

OWL

RDF(S)

XML

Semantic Web

What is an ontology?

An ONTOLOGY is:

a formal (and explicit) specification of a shared conceptualization

Computable

The type of concepts used, as well as their sense and application restrictions, are fully defined

Consisting of concepts, properties, relationships, funtions, constraints (rules) and axioms

Models consensual knowledge

An abstract model of a certain phenomenon that identifies its main components (concepts) and their inter-relations
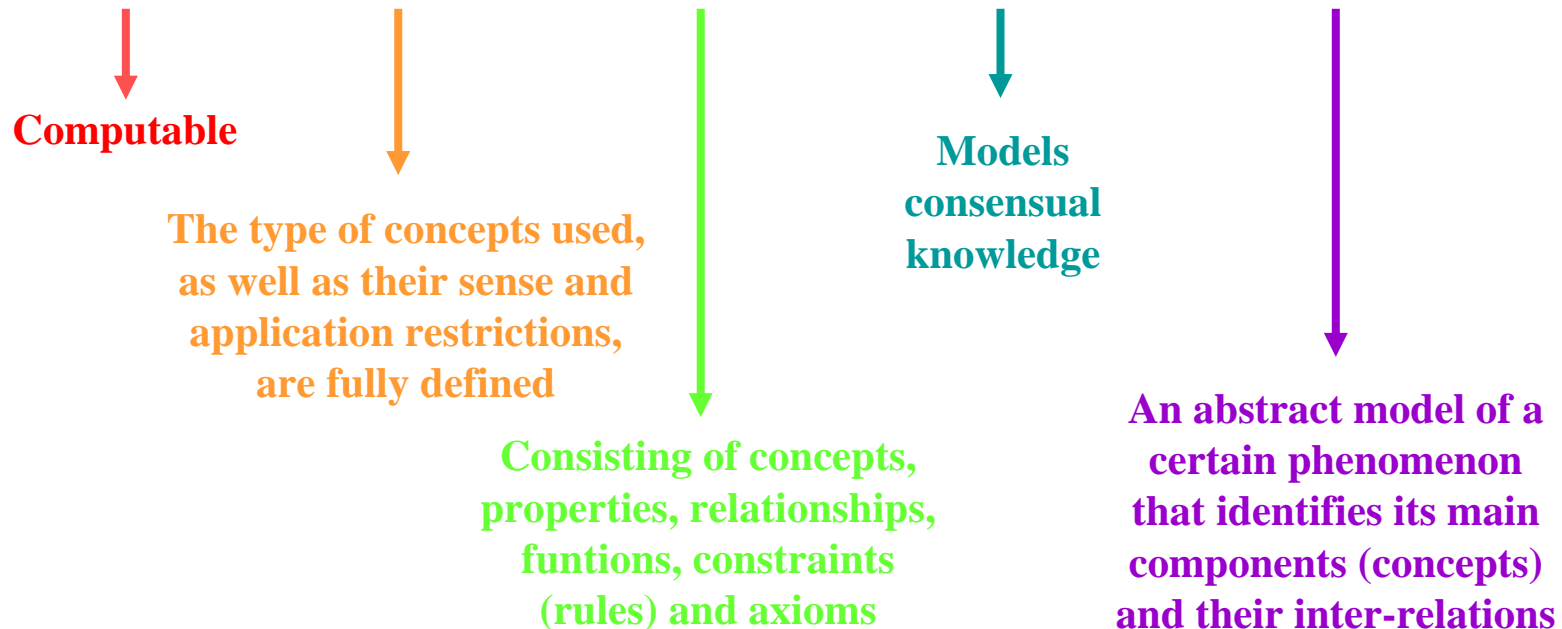
- Gruber, T. R. (1993) A Translation Approach to Portable Ontologies. *Journal on Knowledge Acquisition,* Vol. 5(2), 199-220

- Borst, W. N. (1997) *Construction of Engineering Ontologies*. PhD thesis, University of Twente, Enschede

- Studer, R.; Benjamins, R.; Fensel, D. (1998) Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering, (DKE)* Vol. 25, 1-2: 161-197
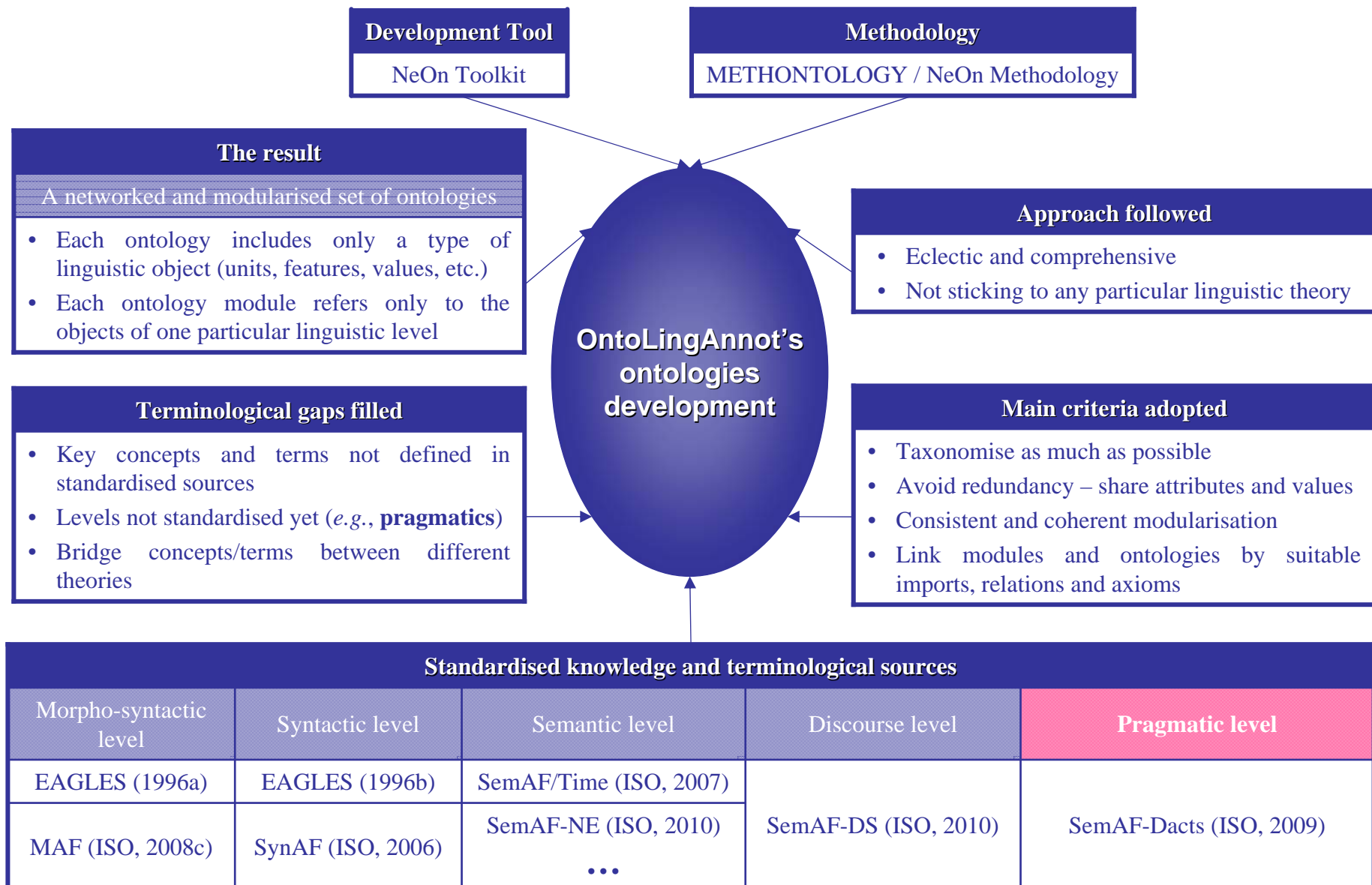
Foundations of the OntoLingAnnot annotation model
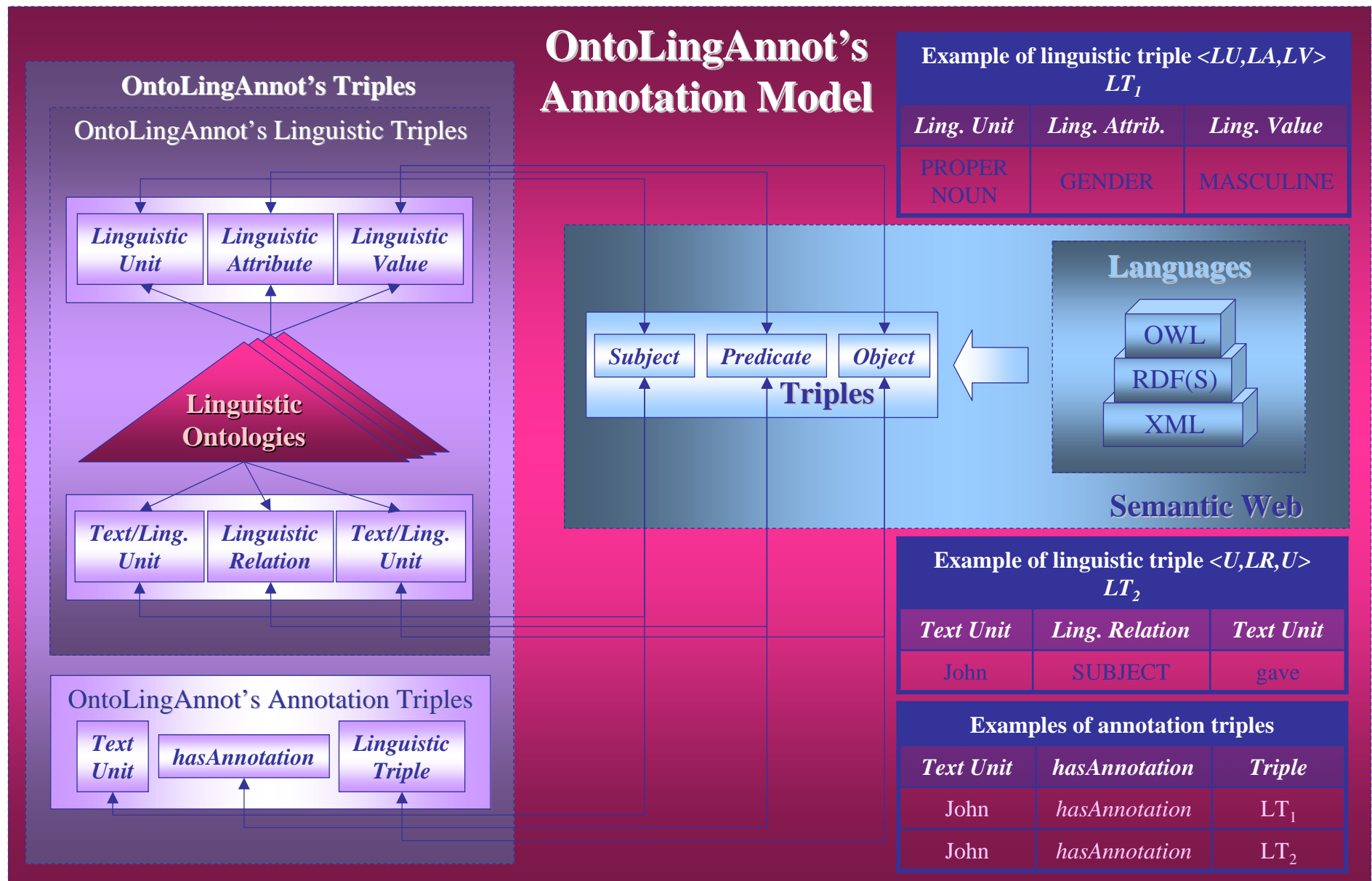
OntoLingAnnot's Triples

OntoLingAnnot's Linguistic Triples

- Linguistic Unit
- Linguistic Attribute
- Linguistic Value

Linguistic Ontologies

- Text/Ling. Unit
- Linguistic Relation
- Text/Ling. Unit

OntoLingAnnot's Annotation Triples

- Text Unit
- hasAnnotation
- Linguistic Triple

OntoLingAnnot's Annotation Model

- Subject
- Predicate
- Object

Triples

Languages

- OWL
- RDF(S)
- XML

Semantic Web

Example of linguistic triple <LU,LA,LV> LT_1

Ling. Unit	Ling. Attrib.	Ling. Value
PROPER NOUN	GENDER	MASCULINE

Example of linguistic triple <U,LR,U> LT_2

Text Unit	Ling. Relation	Text Unit
John	SUBJECT	gave

Examples of annotation triples

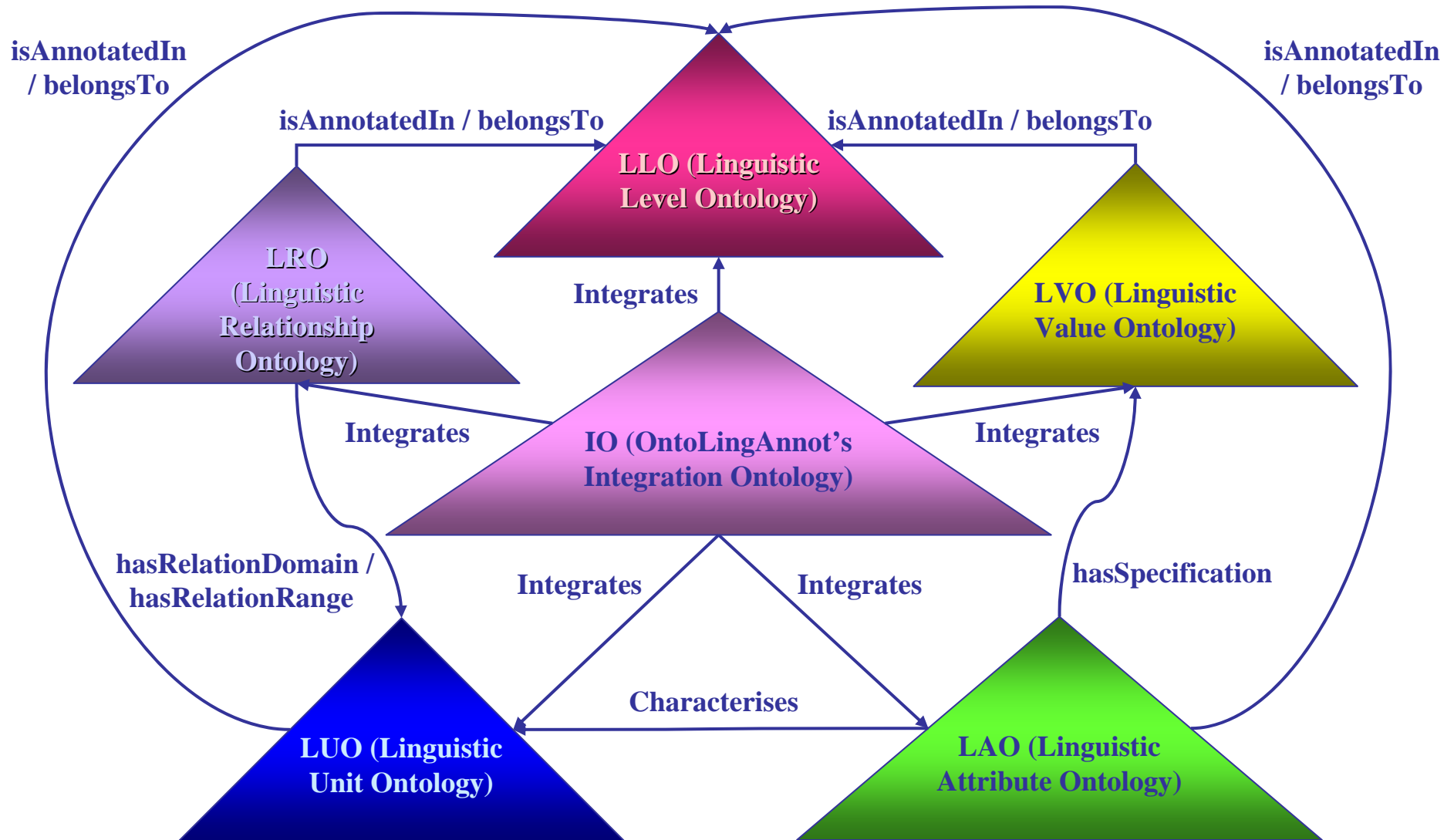| Text Unit	hasAnnotation	Triple
John	hasAnnotation	LT_1
John	hasAnnotation	LT_2

Contents

- The OntoLingAnnot annotation model: background and foundations
- OntoLingAnnot's ontologies:
  - Development and overview
  - Annotation levels and layers
  - The pragmatic concepts/terms
    - The pragmatic (sub-)layers and (sub-)strata in the Linguistic Level Ontology (LLO)
    - The pragmatic units in the Linguistic Unit Ontology (LUO)
    - The pragmatic attributes in the Linguistic Attribute Ontology (LAO)
    - The pragmatic values in the Linguistic Value Ontology (LVO)
    - The pragmatic relations in the Linguistic Relationship Ontology (LRO)
    - Some statistics
- Conclusions
- Annex: Pragmatic annotation of a simple dialogue example

OntoLingAnnot's ontologies: development

Development Tool

NeOn Toolkit

Methodology

METHONTOLOGY / NeOn Methodology

The result

A networked and modularised set of ontologies

- Each ontology includes only a type of linguistic object (units, features, values, etc.)
- Each ontology module refers only to the objects of one particular linguistic level

OntoLingAnnot's ontologies development

Approach followed

- Eclectic and comprehensive
- Not sticking to any particular linguistic theory

Terminological gaps filled

- Key concepts and terms not defined in standardised sources
- Levels not standardised yet (*e.g.*, **pragmatics**)
- Bridge concepts/terms between different theories

Main criteria adopted

- Taxonomise as much as possible
- Avoid redundancy – share attributes and values
- Consistent and coherent modularisation
- Link modules and ontologies by suitable imports, relations and axioms

Standardised knowledge and terminological sources				
Morpho-syntactic level	Syntactic level	Semantic level	Discourse level	**Pragmatic level**
EAGLES (1996a)	EAGLES (1996b)	SemAF/Time (ISO, 2007)		
MAF (ISO, 2008c)	SynAF (ISO, 2006)	SemAF-NE (ISO, 2010) •••	SemAF-DS (ISO, 2010)	SemAF-Dacts (ISO, 2009)

OntoLingAnnot's ontologies: an overview

isAnnotatedIn / belongsTo

isAnnotatedIn / belongsTo

isAnnotatedIn / belongsTo

isAnnotatedIn / belongsTo

LLO (Linguistic Level Ontology)

LRO (Linguistic Relationship Ontology)

LVO (Linguistic Value Ontology)

Integrates

IO (OntoLingAnnot's Integration Ontology)

Integrates

Integrates

Integrates

Integrates

hasRelationDomain / hasRelationRange

hasSpecification

Characterises

LUO (Linguistic Unit Ontology)

LAO (Linguistic Attribute Ontology)

Contents

- The OntoLingAnnot annotation model: background and foundations
- OntoLingAnnot's ontologies:
  - Development and overview
  - Annotation levels and layers
  - The pragmatic concepts/terms
    - The pragmatic (sub-)layers and (sub-)strata in the Linguistic Level Ontology (LLO)
    - The pragmatic units in the Linguistic Unit Ontology (LUO)
    - The pragmatic attributes in the Linguistic Attribute Ontology (LAO)
    - The pragmatic values in the Linguistic Value Ontology (LVO)
    - The pragmatic relations in the Linguistic Relationship Ontology (LRO)
    - Some statistics
- Conclusions
- Annex: Pragmatic annotation of a simple dialogue example

The Linguistic Level Ontology (LLO): main **annotation levels**

LEVEL	ANNOTATION SCOPE	UNITS
Morphology	Up to word structure and meaning (including morph-related annotation).	• Morph • Word
Syntax	Up to sentential structure (including multiword token, phrasal and clausal structure).	• Token • Phrase • Clause • Sentence
Semantics	Up to propositional structure and sentence meaning (including phrase and clause meaning), that is, propositional meaning.	• Sense • Proposition, etc.
Discourse	Up to discourse structure (including coherence relation-based structure) and the supra-sentential and locutionary meaning of texts and speech (including anaphora resolution).	• Macroproposition • Turn, etc.
Pragmatics	Up to the illocutionary and perlocutionary structure and meaning of texts and speech (including deictic resolution and other pragmatic relation annotation).	• Macroproposition • Pragmateme, etc.

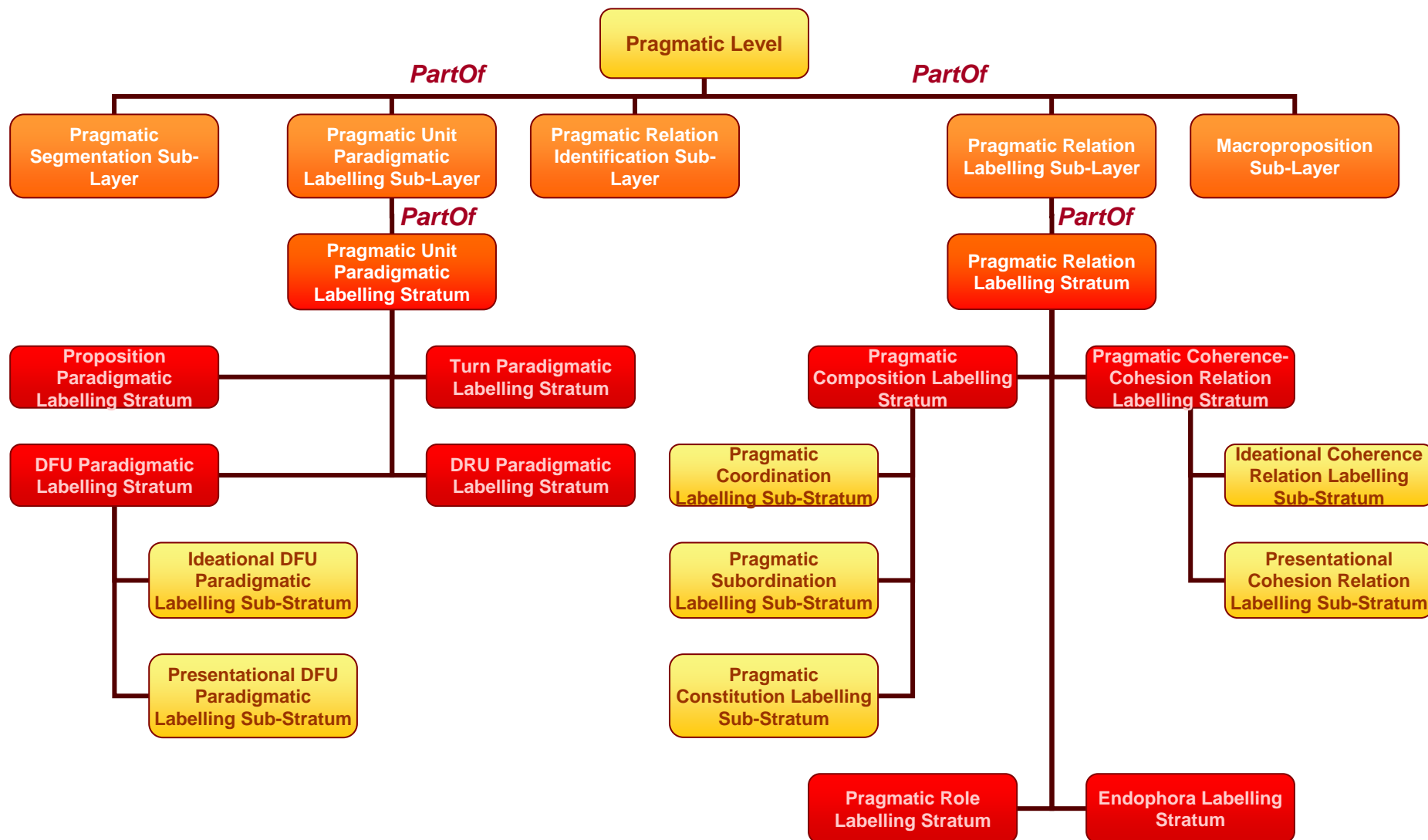# The LLO: common **processes** (the **annotation layers**)

Segmentation (into its constituent units)

Constituent Unit Paradigmatic Labelling

Syntagmatic Relation Identification

Syntagmatic Relation Labelling

Resulting Unit Determination

Morphology

Syntax

Pragmatics

Semantics

Discourse
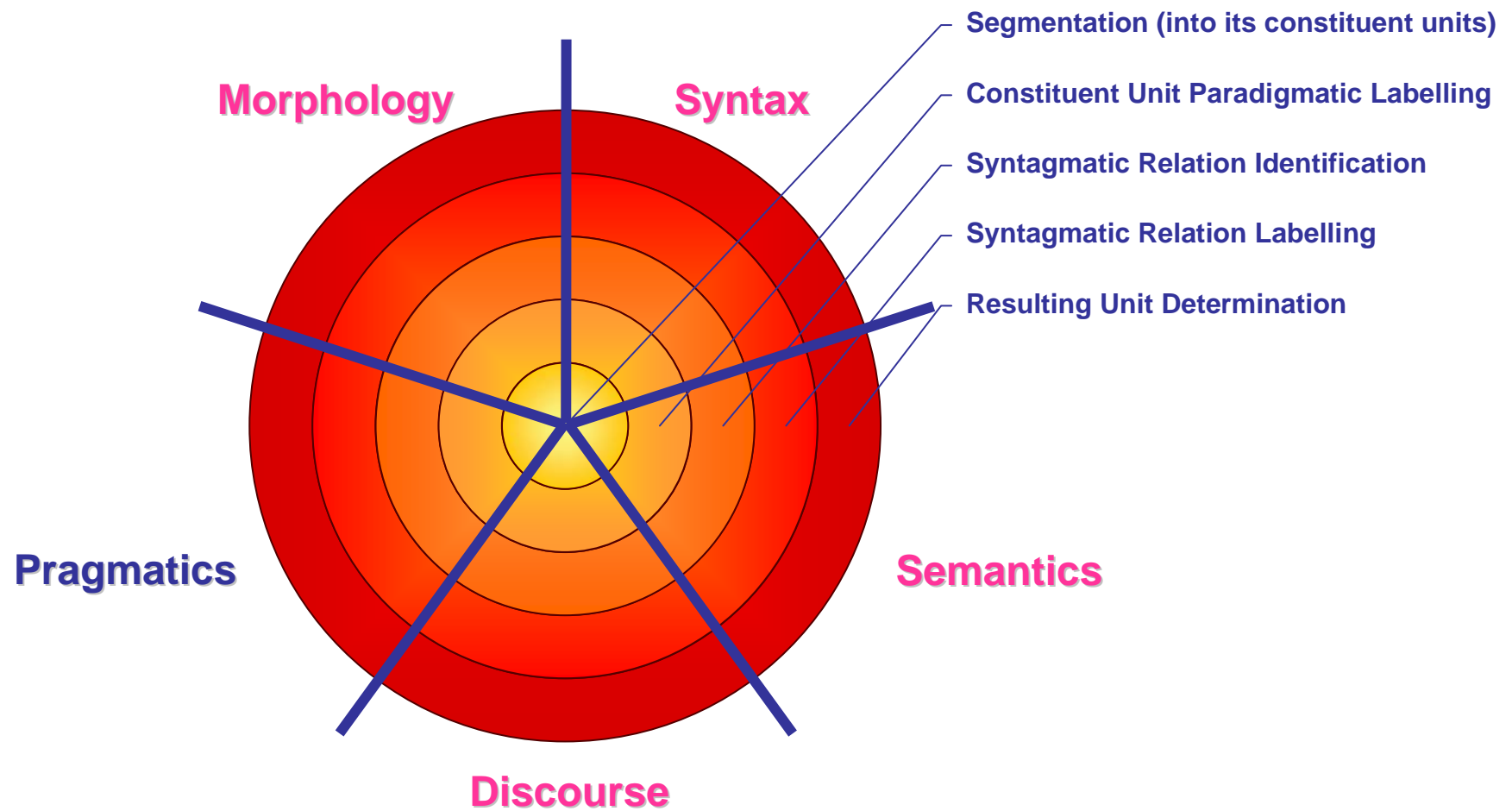
The LLO: annotation sub-layers

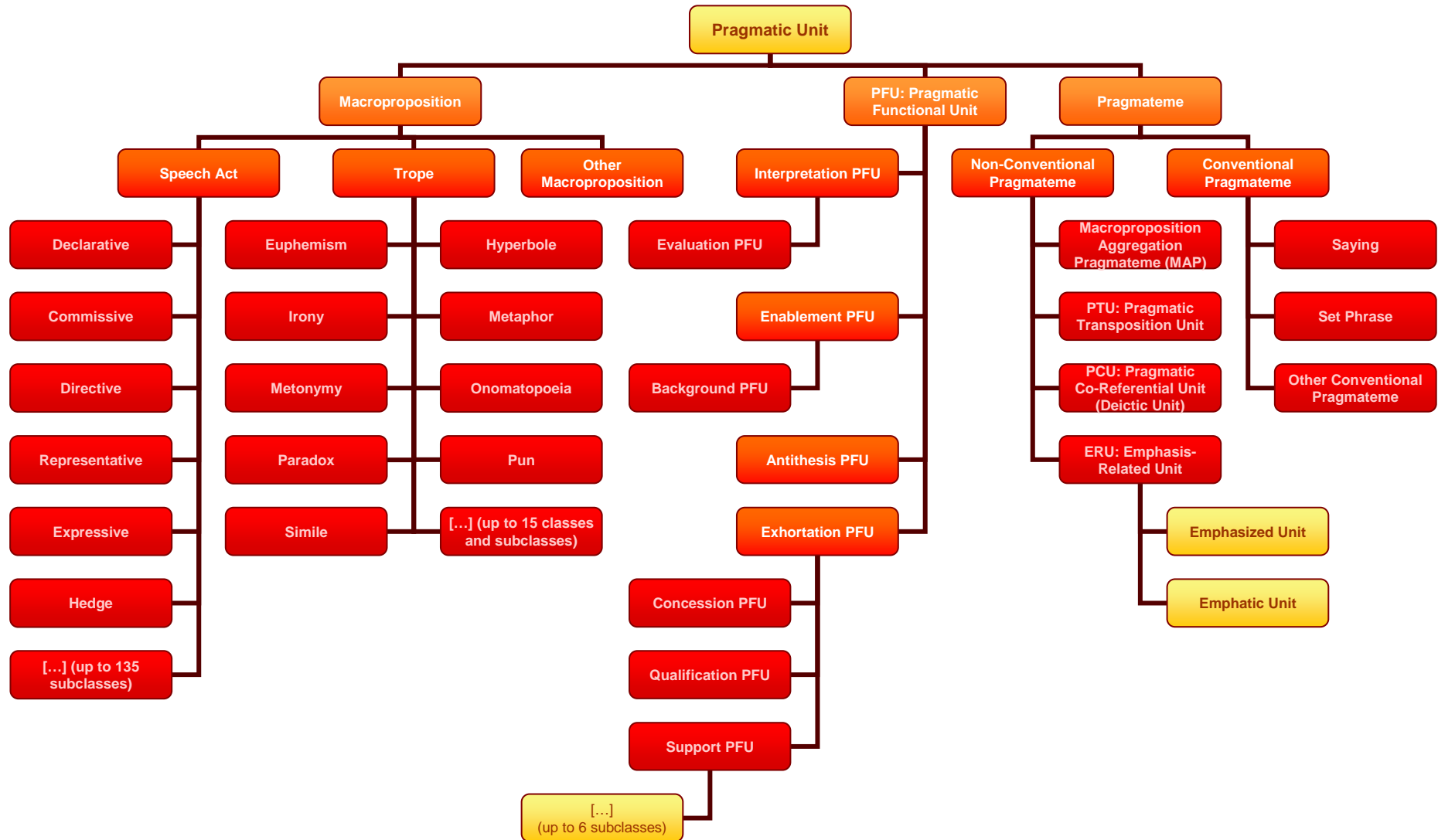| LAYER → / LEVEL ↓	SEGMENTATION (CONSTITUENT UNIT RECOGNITION)	CONSTITUENT UNIT PARADIGMATIC LABELLING	SYNTAGMATIC RELATION IDENTIFICATION	SYNTAGMATIC RELATION LABELLING	RESULTING UNIT DETERMINATION
MORPHOLOGY	Sub-Layer 1.1: Morph Recognition (Morphological Segmentation)	Sub-Layer 1.2: Morph Paradigmatic Labelling	Sub-Layer 1.3: Word Formation Relation Identification	Sub-Layer 1.4: Word Formation Relation Labelling	Sub-Layer 1.5: Word Sub-Layer
SYNTAX	Sub-Layer 2.1: Syntactic Segmentation	Sub-Layer 2.2: Syntactic Unit Paradigmatic Labelling	Sub-Layer 2.3: Syntactic Relation Identification	Sub-Layer 2.4: Syntactic Relation Labelling	Sub-Layer 2.5: Aggregated Syntactic Unit Sub-Layer
SEMANTICS	Sub-Layer 3.1: Semantic Segmentation	Sub-Layer 3.2: Semantic Unit Paradigmatic Labelling	Sub-Layer 3.3: Proposition Formation Relation Identification (Predication Identification)	Sub-Layer 3.4: Proposition Formation Relation Labelling (Predication Labelling)	Sub-Layer 3.5: Proposition Sub-Layer
DISCOURSE	Sub-Layer 4.1: Discourse Segmentation	Sub-Layer 4.2: Discourse Unit Paradigmatic Labelling	Sub-Layer 4.3: Discourse Relation Identification	Sub-Layer 4.4: Discourse Relation Labelling	Sub-Layer 4.5: Macroproposition Sub-Layer
PRAGMATICS	Sub-Layer 5.1: Pragmatic Segmentation	Sub-Layer 5.2: Pragmatic Unit Paradigmatic Labelling	Sub-Layer 5.3: Pragmatic Relation Identification	Sub-Layer 5.4: Pragmatic Relation Labelling	Sub-Layer 5.5: Aggregated Pragmatic Unit Sub-Layer

Contents

- The OntoLingAnnot annotation model: background and foundations
- OntoLingAnnot's ontologies:
  - Development and overview
  - Annotation levels and layers
  - The pragmatic concepts/terms
    - The pragmatic (sub-)layers and (sub-)strata in the Linguistic Level Ontology (LLO)
    - The pragmatic units in the Linguistic Unit Ontology (LUO)
    - The pragmatic attributes in the Linguistic Attribute Ontology (LAO)
    - The pragmatic values in the Linguistic Value Ontology (LVO)
    - The pragmatic relations in the Linguistic Relationship Ontology (LRO)
    - Some statistics
- Conclusions
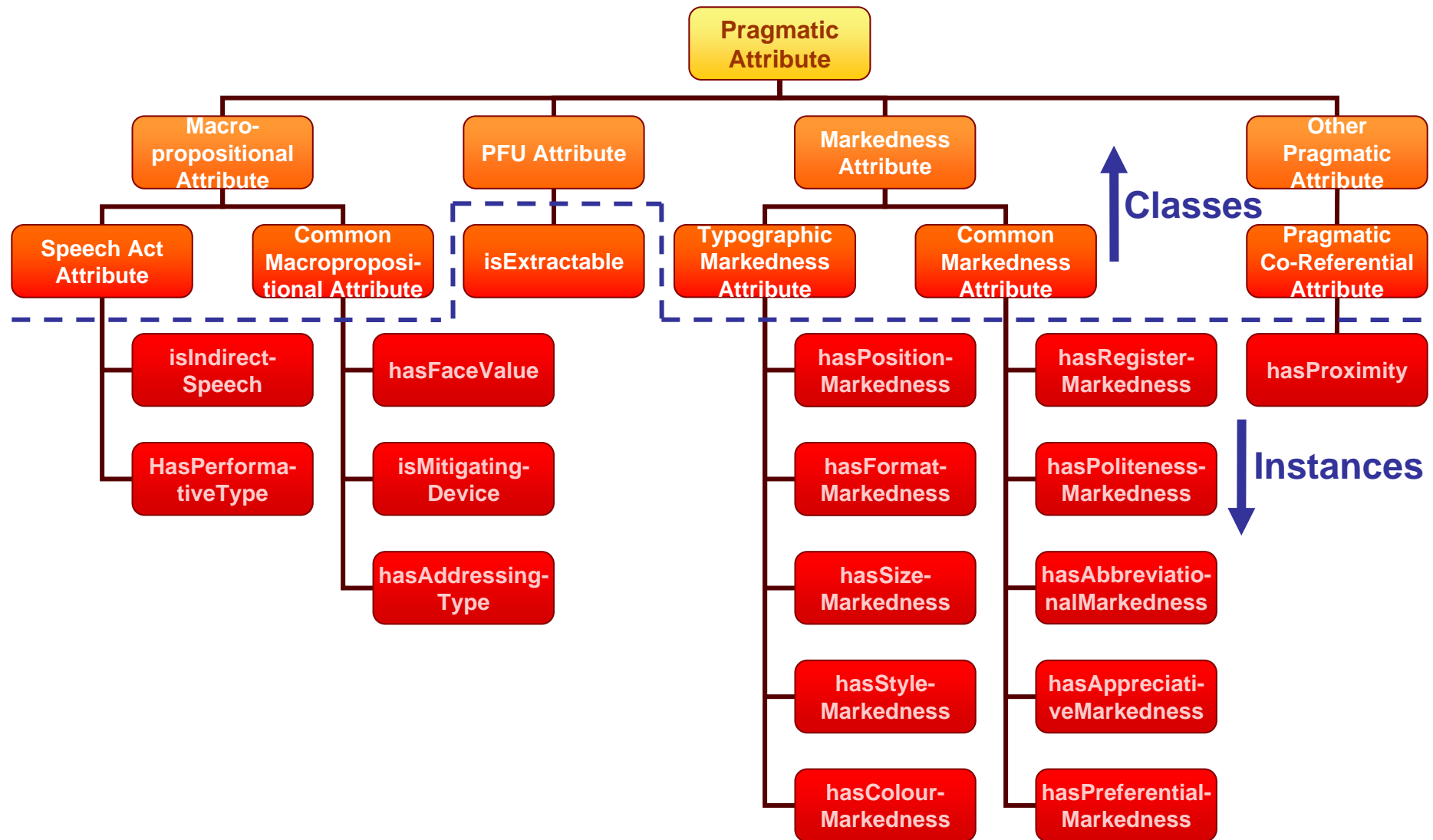- Annex: Pragmatic annotation of a simple dialogue example

The pragmatic layers, sublayers, strata and substrata in the LLO

The pragmatic units in the Linguistic Unit Ontology (LUO)

The pragmatic attributes in the Linguistic Attribute Ontology (LAO)

Pragmatic Attribute

- **Macro-propositional Attribute**
  - **Speech Act Attribute**
    - isIndirect-Speech
    - HasPerforma-tiveType
  - **Common Macroproposi-tional Attribute**
    - hasFaceValue
    - isMitigating-Device
    - hasAddressing-Type
- **PFU Attribute**
  - **isExtractable**
- **Markedness Attribute**
  - **Typographic Markedness Attribute**
    - hasPosition-Markedness
    - hasFormat-Markedness
    - hasSize-Markedness
    - hasStyle-Markedness
    - hasColour-Markedness
  - **Common Markedness Attribute**
    - hasRegister-Markedness
    - hasPoliteness-Markedness
    - hasAbbreviatio-nalMarkedness
    - hasAppreciati-veMarkedness
    - hasPreferential-Markedness
- **Other Pragmatic Attribute**
  - **Pragmatic Co-Referential Attribute**
    - hasProximity

Classes ↑

Instances ↓

The correspondences between pragmatic units and attributes in OntoLingAnnot

Pragmatic Unit

LUO

- Macroproposition
  - Speech Act
  - Trope
  - Other Macroproposition
- Pragmatic Functional Unit (PFU)
- Pragmateme
  - Non-Conventional Pragmateme
    - PCU: Pragmatic Co-Referential Unit (Deictic Unit)
  - Conventional Pragmateme

LAO

- isIndirectSpeech
- HasPerformativeType
- hasProximity
- hasPositionMarkedness
- hasFormatMarkedness
- hasSizeMarkedness
- hasStyleMarkedness
- hasColourMarkedness
- hasRegisterMarkedness
- hasPolitenessMarkedness
- hasAbbreviationalMarkedness
- hasAppreciativeMarkedness
- hasPreferentialMarkedness

- hasFaceValue
- isMitigatingDevice
- hasAddressingType
- isExtractable

- Common Macropropositional Attribute
- Speech Act Attribute
- Pragmatic Co-Referential Attribute
- Typographic Markedness Attribute
- Common Markedness Attribute

- Macropropositional Attribute
- PFU Attribute
- Other Pragmatic Attribute
- Markedness Attribute

Pragmatic Attribute

The pragmatic values in the Linguistic Value Ontology (LVO) [TAXONOMY & CLASSES]
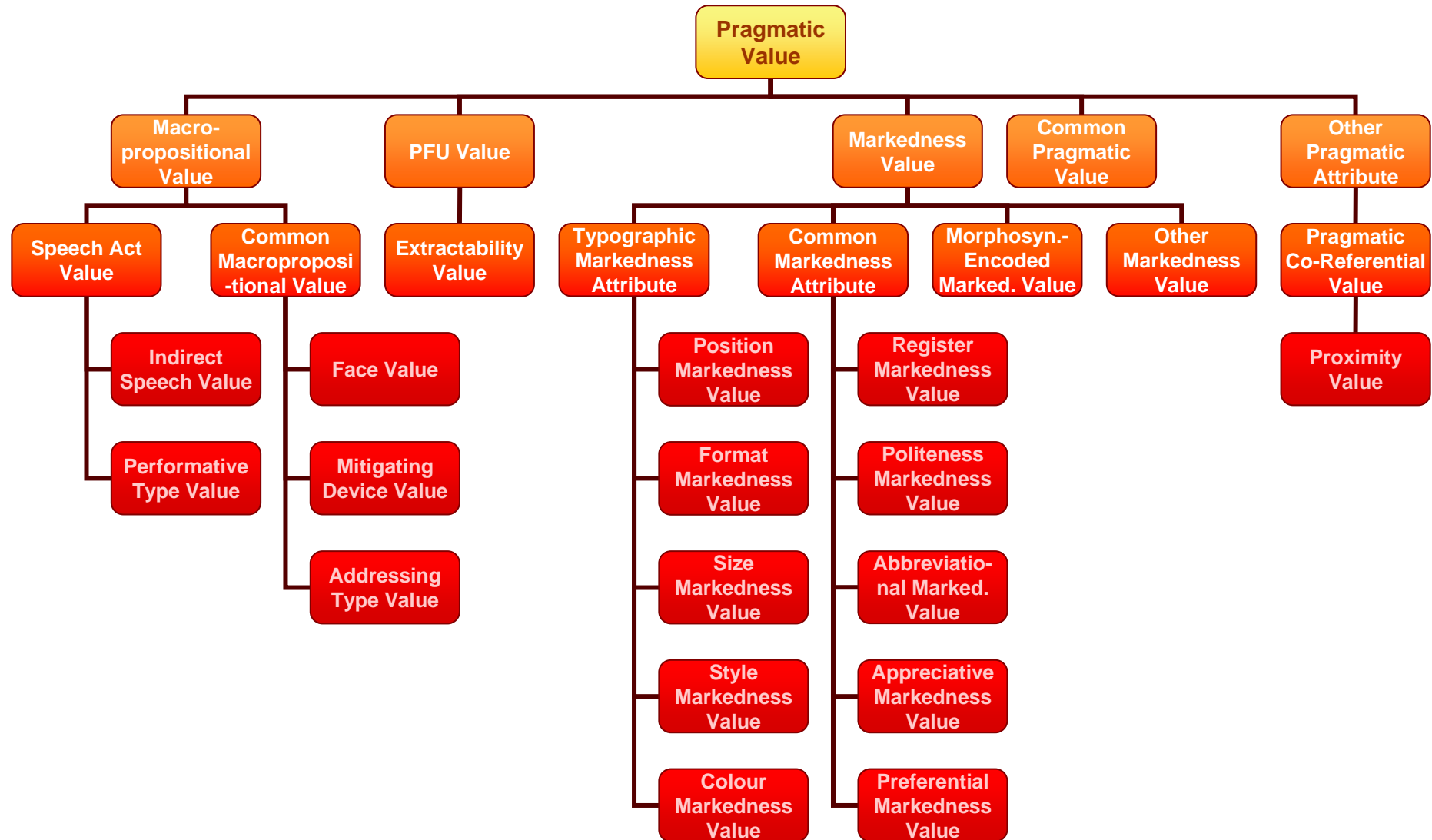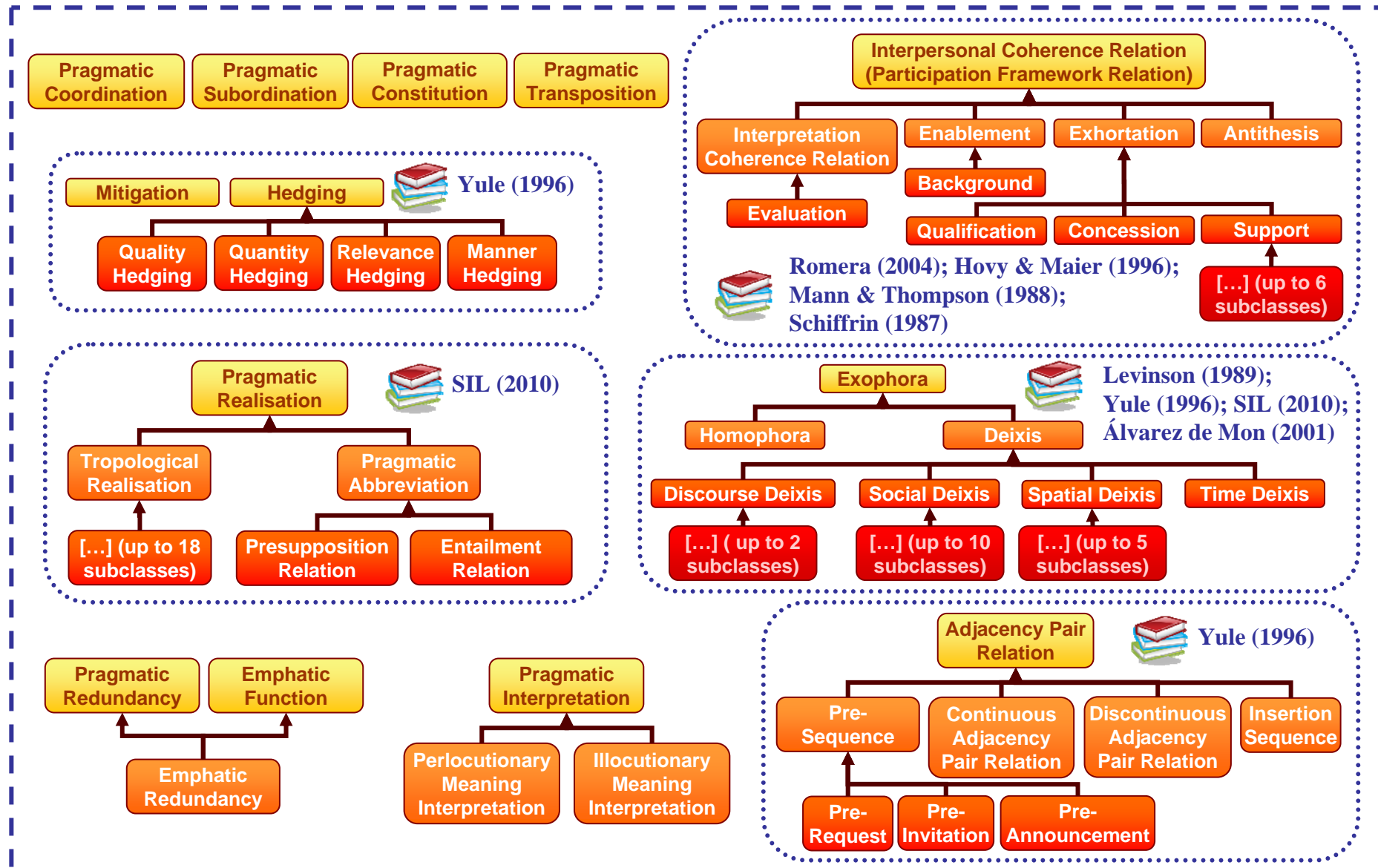
The pragmatic values in the Linguistic Value Ontology (LVO) [INSTANCES]

CONCEPTS	INSTANCES
Common Pragmatic Value	{UNMARKED, MARKED}
Indirect Speech Value	*BOOLEAN* = {TRUE, FALSE}
Performative Type Value	{EXPLICIT, IMPLICIT}
Face Value	{SAVING, THREATENING, UNMARKED}
Mitigating Device Value	*BOOLEAN* = {TRUE, FALSE}
Addressing Type Value	{UNMARKED, OFF_RECORD, ON_RECORD, BALD_ON_RECORD}
Position Markedness Value	{UNMARKED, SUBINDEX, SUPERINDEX}
Format Markedness Value	{UNMARKED, UNDERLINED, ITALICS, BOLD, CAPITALISED, SMALL_CAPITALISED}
Size Markedness Value	{UNMARKED, BIGGER, SMALLER}
Style Markedness Value	{UNMARKED, SUBTITLE, TITLE}
Colour Markedness Value	{UNMARKED, MARKED}

CONCEPTS	INSTANCES
Register Markedness Value	{UNMARKED, FAMILIAR, SLANG, FORMAL, JARGON, TECHNICAL, OFFICIAL}
Politeness Markedness Value	{UNMARKED, POLITE, UNPOLITE}
Abbreviational Markedness Value	{UNMARKED, EXPECTEDLY_ABBREVIATED, UNEXPECTEDLY_ABBREVIATED, EXPECTEDLY_NON_ABBREVIATED, UNEXPECTEDLY_NON_ABBREVIATED}
Appreciative Markedness Value	{UNMARKED, AUGMENTATIVE, DIMINUTIVE, PEJORATIVE}
Preferential Markedness Value	{UNMARKED, PREFERRED, DISPREFERRED}
Proximity Value	{CLOSE, INTERMEDIATE, DISTANT, UNMARKED}

Pragmatic relations in the Linguistic Relationship Ontology (LRO)

Pragmatic Coordination

Pragmatic Subordination

Pragmatic Constitution

Pragmatic Transposition

Yule (1996)

Mitigation — Hedging
- Quality Hedging
- Quantity Hedging
- Relevance Hedging
- Manner Hedging

Interpersonal Coherence Relation (Participation Framework Relation)
- Interpretation Coherence Relation
  - Evaluation
- Enablement
  - Background
    - Qualification
    - Concession
    - Support
      - […] (up to 6 subclasses)
- Exhortation
- Antithesis

Romera (2004); Hovy & Maier (1996); Mann & Thompson (1988); Schiffrin (1987)

SIL (2010)

Pragmatic Realisation
- Tropological Realisation
  - […] (up to 18 subclasses)
- Pragmatic Abbreviation
  - Presupposition Relation
  - Entailment Relation

Levinson (1989); Yule (1996); SIL (2010); Álvarez de Mon (2001)

Exophora
- Homophora
- Deixis
  - Discourse Deixis
    - […] (up to 2 subclasses)
  - Social Deixis
    - […] (up to 10 subclasses)
  - Spatial Deixis
    - […] (up to 5 subclasses)
  - Time Deixis

Pragmatic Redundancy — Emphatic Function
- Emphatic Redundancy

Pragmatic Interpretation
- Perlocutionary Meaning Interpretation
- Illocutionary Meaning Interpretation

Yule (1996)

Adjacency Pair Relation
- Pre-Sequence
  - Pre-Request
  - Pre-Invitation
  - Pre-Announcement
- Continuous Adjacency Pair Relation
- Discontinuous Adjacency Pair Relation
- Insertion Sequence

Some Pragmatics-related statistics of OntoLingAnnot's ontologies

PRAGMATIC TERMS IN ONTOLINGANNOT'S ONTOLOGIES						
	UNITS (LUO)	ATTRIBUTES (LAO)	VALUES (LVO)	RELATIONS (LRO)	LEVEL, LAYERS, etc. (LLO)	TOTAL PRAGMATIC TERMS
CONCEPTS	192	10	27	86	24	339
INSTANCES	0	16	54	0	0	70
TOTAL TERMS	192	26	81	86	24	409

- There are also several other ontological terms concerning pragmatic (*attributes*, *SubclassOf*, *PartOf* and *ad hoc relations*, *rules* and *axioms*) not mentioned here for the sake of brevity.

Contents

- The OntoLingAnnot annotation model: background and foundations
- OntoLingAnnot's ontologies:
  - Development and overview
  - Annotation levels and layers
  - The pragmatic concepts/terms
    - The pragmatic (sub-)layers and (sub-)strata in the Linguistic Level Ontology (LLO)
    - The pragmatic units in the Linguistic Unit Ontology (LUO)
    - The pragmatic attributes in the Linguistic Attribute Ontology (LAO)
    - The pragmatic values in the Linguistic Value Ontology (LVO)
    - The pragmatic relations in the Linguistic Relationship Ontology (LRO)
    - Some statistics
- Conclusions
- Annex: Pragmatic annotation of a simple dialogue example

Conclusions (1)

CONTRIBUTIONS REGARDING LINGUISTIC ANNOTATION AND STANDARDISATION										
	LAF/ GrAF (ISO, 2008a)	DCR/ ISOCat (2010)	LMF (ISO, 2008b)	MAF (ISO, 2008c)	SynAF (ISO, 2006)	SemAF				Onto Ling Annot
						Time (ISO, 2007)	NE (ISO, 2010)	DS (ISO, 2010)	Dacts (ISO, 2009)	
Dialogue/discourse segmentation	-	MINIM.	-	-	-	-	-	PARTIAL.	PARTIAL.	YES
Turn definition and management	-	MINIM.	-	-	-	-	-	-	YES	YES
Speech acts	-	MINIM.	-	-	-	-	-	-	YES	YES
Other pragmatic units	-	-	-	-	-	-	-	-	-	YES
Pragmatic constitution, adjacency, etc.	-	-	-	-	-	-	-	-	-	YES
Mitigation / hedging	-	-	-	-	-	-	-	-	-	YES
Pragmatic cohesion/coherence	-	-	-	-	-	-	-	-	-	YES
Exophora / deixis	-	-	-	-	-	-	-	-	-	YES
Other pragm. relations	-	-	-	-	-	-	-	-	-	YES
PRAGMATIC ANNOT.	-	-	-	-	-	-	-	-	PARTIAL.	YES

- **OntoLingAnnot**,
  - gives a **holistic and uniform view** of linguistic annotation levels;
  - could be regarded as a perfect **complement to LAF/GrAF** (ISO, 2008a).

Conclusions (2): other relevant contributions

- We have presented the ontology modules dealing with the pragmatic level of the OntoLingAnnot (linguistic) annotation model.
- OntoLingAnnot's pragmatic ontology modules
  - formalise the different pragmatic units, features and relationships identified in the literature so far.
  - constitute a coherent distribution and structuring of these pragmatic elements.
- Up to date, this is the first global ontological conceptualization of linguistic annotation in general and of Pragmatics and pragmatic annotation in particular.
  - No other pragmatic model accounts globally and coherently for such a number of pragmatic concepts and/or terms.
- Hence, it is an important contribution *per se* to the areas of Ontological Engineering, Pragmatics and Linguistic Annotation.

Thank you!