ISA-12:
12th Joint ACL - ISO Workshop
on Interoperable Semantic Annotation

Saturday, May 28, 2016

ABSTRACTS

Editor:

Harry Bunt, Tilburg University

Workshop Programme

08.45 – 09:00 Registration
09:00 – 09:10 Opening

09:15 – 09:45 Claire Bonial, Susan Brown and Martha Palmer: *A Lexically-Informed Upper Level Event Ontology*
09:45 – 10:00 James Pustejovsky, Martha Palmer, Annie Zaenen, and Susan Brown: *Integrating VerbNet and GL Predicative Structures*
10:00 – 10:15 Elisabetta Jezek, Anna Feltracco, Lorenzo Gatti, Simone Magnolini and Bernardo Magnini: *Mapping Semantic Types onto WordNet Synsets*
10:15 – 10:30 Petya Osenova and Kiril Simov: *Cross-level Semantic Annotation of Bulgarian Treebank*

10:30 – 11:00 Coffee break

11:00 – 11:30 Ielka van der Sluis, Shadira Leito and Gisela Redeker: *Text-Picture Relations in Cooking Instructions*
11:30 – 12:00 Kiyong Lee: *An Abstract Syntax for ISOspace with its <moveLink> Reformulated*
12:00 – 12:15 James Pustejovsky, Kiyong Lee and Harry Bunt: *Proposed ISO Standard Amendment AMD 24617-7 ISOspace*

12:15 – 14:00 Lunch break, during which:
12:30 – 13:30 ISO TC 37/SC 4 Working groups 2 and 5 plenary meeting

14:00 – 14:30 Ludivine Crible: *Discourse Markers and Disfluencies: Integrating Functional and Formal Annotations*
14:30 – 15:00 Harry Bunt and Rashmi Prasad: *ISO DR-Core: Core Concepts for the Annotation of Discourse Relations*
15:00 – 15:15 Benjamin Weiss and Stefan Hillmann: *Feedback Matters: Applying Dialog Act Annotation to Study Social Attractiveness in Three-Party Conversations*
15:15 – 15:30 Andreea Macovei and Dan Cristea: *Time Frames: Rethinking the Way to Look at Texts*
15:30 – 15:45 Tuan Do, Nikhil Krishnaswamy, and James Pustejovsky: *ECAT: Event Capture Annotation Tool*

15:45 – 16:15 Tea break

16:15 – 16:45 Julia Lavid, Marta Carretero and Juan Rafael Zamorano: *Contrastive (English-Spanish) Annotation of Epistemicity in the MULTINOT Project: Preliminary Steps*
16:45 – 17:15 Elisa Ghia, Lennart Kloppenburg, Malvina Nissim and Paola Pietrandrea: *A Construction-Centered Approach to the Annotation of Modality*
17:15 – 17.30 Kyeongmin Rim: *MAE2: Portable Annotation Tool for General Natural Language Use*

17.30 –17.31 ISA-12 Workshop Closing, followed by discussion of proposed new ISO activities:

17:31 – 17.45 James Pustejovsky and Kiyong Lee: *Proposal for New ISO activity PWI 24617-x ISOspaceSem*
17:45 – 18:00 James Pustejovsky: *Proposal for New ISO activity PWI 24617-x VoxML*

Workshop Organizers

Harry Bunt   Tilburg University
Nancy Ide   Vassar College, Poughkeepsie, NY
Kiyong Lee   Korea University, Seoul
James Pustejovsky   Brandeis University, Waltham, MA
Laurent Romary   INRIA and Humboldt Universität Berlin

Workshop Programme Committee

Jan Alexandersson   DFKI, Saarbrücken
Harry Bunt   Tilburg University
Nicoletta Calzolari   CNR-ILC, Pisa
Thierry Declerck   DFKI, Saarbrücken
Liesbeth Degand   UCL, Louvain-la-Neuve
David DeVault   USC, Playa Vista, CA
Alex Chengyu Fang   City University of Hong Kong
Robert Gaizauskas   University of Sheffield
Daniel Hardt   Copenhagen Business School
Koiti Hasida   Tokyo University
Elisabetta Jezek   University of Pavia
Michael Kipp   Augsburg University of Applied Sciences
Kiyong Lee   Korea University, Seoul
Philippe Muller   Université Paul Sabatier, Toulouse
Malvina Nissim   CLCG, University of Groningen
Volha Petukhova   Universität des Saarlandes, Saarbrücken
Paola Pietrandrea   Université de Tours and CNRS LLL
Andrei Popescu-Belis   IDIAP, Martigny
Laurent Prévot   Université Aix-Marseille
James Pustejovsky   Brandeis University, Waltham, MA
Laurent Romary   INRIA and Humboldt Universität Berlin
Ted Sanders   University of Utrecht
Manfred Stede   Universität Potsdam
Thorsten Trippel   University og Tübingen
Piek Vossen   Free University Amsterdam
Annie Zaenen   Stanford University
Sandrine Zufferey   Université de Fribourg

A Lexically-Informed Upper Level Event Ontology

Claire Bonial, Susan Windisch Brown and Martha Palmer

This paper summarizes ISO 24617-8 (ISO DR-Core), a new part of the ISO SemAF framework for semantic annotation. Within this framework a range of standards is developed to support the interoperable annotation of semantic phenomena. The effort to develop a standard for the annotation of semantic relations in discourse is split into two parts, of which ISO 24617-8 concerns the first part, formulating desiderata for the annotation of discourse relations and providing clear definitions for a set of 'core' discourse relations, based on an analysis of a range of theoretical approaches and annotation efforts. Following the ISO principles for semantic annotation, an abstract syntax as well as a concrete XML-based syntax for annotations were defined, together with a formal semantics. Mappings are provided between the ISO core relations and various other annotation schemes.

Verb Meaning in Context: Integrating VerbNet and GL Predicative Structures

James Pustejovsky, Martha Palmer, Annie Zaenen, and Susan Brown

This paper reports on aspects of a new research project aimed at enriching VerbNet's predicative structures with representations and mechanisms from Generative Lexicon Theory. This involves the introduction of systematic predicative enrichment to the verb's predicate structure, including an explicit identification of the mode of opposition structure inherent in the predicate. In addition, we explore a GL-inspired semantic componential analysis over VerbNet classes, in order to identify coherent semantic cohorts within the classes.

Mapping Semantic Types onto WordNet Synsets

Elisabetta Jezek, Anna Feltracco, Lorenzo Gatti, Simone Magnolini and Bernardo Magnini

In this paper, we report the results of an experiment aimed at automatically mapping corpus-derived Semantic Types to WordNet synsets. The algorithm for the automatic alignment of Semantic Types with WordNet synsets relies on lexical correspondence, i.e. it performs an automatic alignment of Semantic Types labels with the corresponding WordNet entry nouns, when present (for example, the Semantic Type [[Activity]] is mapped to synsets containing the entry noun *activity#n*). In this way, 150 Types out of 180 are mapped automatically, while 30 gaps have to be resolved manually. Automatic mapping based on lexical correspondence, however, does not guarantee that the mapping is good, i.e. that the items which make up the extension of a certain Semantic Types match the set of hyponyms of the corresponding synset(s). An evaluation of 43 Semantic Types against a gold standard reveals that, for 30% of them, a manual revision is needed
.

Cross-level Semantic Annotation of Bulgarian Treebank
Petya Osenova and Kiril Simov

The paper focuses on the cross-level semantic annotation of BulTreeBank. It discusses the annotation of distinct lexemes as well as MultiWord Expressions with senses from the BTB Wordnet, valency frames dictionary, and DBPedia URIs and classes. Also, one important application of the semantically annotated treebank is discussed – namely, for improving the Knowledge-based Word Sense Disambiguation task via extraction of new semantic relations.

Session 2
11:00 – 12:15

Text-Picture Relations in Cooking Instructions
Ielka van der Sluis, Shadira Leito and Gisela Redeker

Like many other instructions, recipes on packages with ready-to-use ingredients for a dish combine a series of pictures with short text paragraphs. The information presentation in such multimodal instructions can be compact (either text or picture) and/or cohesive (text and picture). In an exploratory corpus study, 30 instructions for semi-prefabricated meals were annotated for text-picture relations. A slight majority of the 452 actions in the corpus were presented only textually. A third were presented in text and picture, indicating a moderate amount of cohesion. A minority of 31 actions (7%) were presented only pictorially, suggesting that the potential for compact multimodal presentation may be rather limited in these instructions.

An Abstract Syntax for ISOspace with its <moveLink> Reformulated
Kiyong Lee

ISOspace introduces the movement link, tagged `<moveLink>`, to annotate how motions are related to spatial entities in language. As pointed out in SemAF principles 9iso 24617-6), ISOspace overlaps SemAF-SR (ISO 24617-5), which treats semantic roles in general. It also fails to conform to the *link* structure $<\eta, E, \rho>$ as formulated in Bunt et al. (2016). To resolve these problems, we first construct the general abstract syntax $\mathscr{A}\mathscr{S}yn$ of annotation structures on which the abstract syntax $\mathscr{A}\mathscr{S}yn_{isoSpace}$ of ISOspace is instantiated. Following the two axioms on motion-events and event-paths, discussed by Pustejovsky and Yocum (2013), we then propose to restore the event-path, introduced earlier by Pustejovsky et al. (l2010), as a genuine basic entity in the abstract syntax, while implementing it as such into a concrete syntax. We finally reformulate the movement link as relating the mover of a motion-event to an event-path, as triggered by that motion-event. We also illustrate how the newly formulated movement link (`<moveLink>`) interacts with the other links in ISOspace.

Session 3
14:00 – 15:45

Discourse Markers and Disfluencies: Integrating Functional and Formal Annotations
Ludivine Crible

While discourse markers (DMs) and (dis)fluency have been extensively studied in the past as independent phenomena, combining DM-level and disfluency-level annotations however has never been carried out before at a fine-grained level of informativeness. It is argued that the integration of formal and functional annotations, while facing a number of methodological and theoretical challenges, is not only possible and innovative (addressing the lack of consensus in the field) but also highly relevant to the investigation of form-meaning patterns. This paper reports the methodological aspects of an annotation protocol which integrates formal identification of (dis)fluency markers and a multi-layered description of discourse markers featuring, among others, semantic-pragmatic variables such as their domain and function in context. The challenges and merits of this integration are illustrated by a comparison of clustering tendencies between different functions of DMs in DisFrEn, a French-English comparable dataset. Quantitative results allow us to generate tentative interpretations of the relative fluency of some DM functions based on co-occurrence patterns, in line with a cognitive-functional approach to spoken language.

ISO DR-Core: Core Concepts for the Annotation of Disccourse Relations
Harry Bunt and Rashmi Prasad

This paper summarizes ISO 24617-8 (ISO DR-Core), a new part of the ISO SemAF framework for semantic annotation. Within this framework a range of standards is developed to support the interoperable annotation of semantic phenomena. The effort to develop a standard for the annotation of semantic relations in discourse is split into two parts, of which ISO 24617-8 concerns the first part, formulating desiderata for the annotation of discourse relations and providing clear definitions for a set of 'core' discourse relations, based on an analysis of a range of theoretical approaches and annotation efforts. Following the ISO principles for semantic annotation, an abstract syntax as well as a concrete XML-based syntax for annotations were defined, together with a formal semantics. Mappings are provided between the ISO core relations and various other annotation schemes.

Feedback Matters: Applying Dialog Act Annotation to Study Social Attractiveness in Three-Party Conversations
Benjamin Weiss and Stefan Hillmann

The relationship between verbal behavior and social attractiveness ratings are studied based on three-party conversation scenarios. The recorded conversations are annotated according to ISO 24617-2:2012, applying 11 classes. Intra-group likability ratings given by each interlocutor are correlated with frequencies of each dialog-act class. A linear model shows significant relations between likability ratings given to interlocutors and frequencies of three dialog act classes uttered by the rater. Two classes "positive" and "negative allo-feedback" are negatively related to likability, whereas "positive auto-feedback" shows a positive relation. An effect for the receivers side was not found. All participants met briefly before starting the experiment and also conducted a training conversation, which is why no assumption on cause and effect have been made. This exploratory study motivates to look deeper into the interdependence between verbal behavior and social relationships than just on surface features as speaking time and number of turns.

Time Frames: Rethinking the Way to Look at Texts
Andreea Macovei and Dan Cristea

In this paper, we discuss the challenging task of identifying time frames that may intersect in a text (a novel, a news article, a Facebook post, etc.), in a form more or less visible for the reader. By time frame, we mean a sequence of events or statements that an author exposes voluntarily; these time frames can be considered specific writing techniques where diverse narrative threads are used for the purpose of capturing the reader's attention regarding the story as it develops. A particularity of time frames is the fact that the transition from one time frame to another one seems to be rather difficult to discern and put in evidence by a forewarned annotator, while the consequences of the temporal discontinuities are understood naturally by a casual reader of the text. We are going to explain this notion and to determine if it is necessary to propose a remodelled temporal annotation for this issue.

Session 4
Date / Time *16:15 – 17:30*

ECAT: Event Capture Annotation Tool
Tuan Do, Nikhil Krishnaswamy, and James Pustejovsky

This paper introduces the Event Capture Annotation Tool (ECAT), a user-friendly interface tool for annotating events and their participants in video, capable of extracting 3D positional and orientational data about objects in video captured by Microsoft's Kinect hardware. The modeling language VoxML (Pustejovsky and Krsihnaswamy, 2016) underlies ECAT's object, program, and attribute representations, although ECAT uses its own spec for explicit labeling of motion instances. The demonstration will show the toool's workflow and the options available for capturing event-participant relations and browsing visual data. Mapping ECAT's output to VoxML will also be addressed.

Contrastive (English- Spanish) Annotation of Epistemicity in the MULTINOT Project: Preliminary Steps
Julia Lavid, Marta Carretero and Juan Rafael Zamorano

In this paper we describe the preliminary steps undertaken for the annotation of the conceptual domain of epistemicity in English and Spanish, as part of a larger annotation effort of modal meanings in the context of the MULTINOT project. These steps focus on: a) the instantiation of existing linguistic theories in the area of epistemicity, identifying and defining the categories to be used as tags for annotation; b) the design of an annotation scheme which captures both the functional-semantic dimension of epistemicity, on the one hand, and the language-specific realisations of epistemic meanings in both languages, on the other. These two dimensions are shown to be necessary for investigating relevant contrasts between English and Spanish in the area of epistemicity and for the large-scale annotation of comparable and parallel texts belonging to different registers in English and Spanish.

A Construction-Centered Approach to the Annotation of Modality

Elisa Ghia, Lennart Kloppenburg, Malvina Nissim and Paola Pietrandrea

We propose a comprehensive annotation framework for modality, which encompasses and supports existing annotation schemes, by adopting a construction-centered view. Rather than seeing modality as a feature of a trigger or of a target, we view it as a feature of the triad "trigger-target-relation", which we name *construction*. We motivate the need for such an approach from a theoretical perspective, and we also show that a construction-centered annotation scheme is operationally valid. We evaluate inter-annotator agreement via a pilot study, and find that modalised constructions identified by different annotators can be successfully aligned, as a first crucial step towards further agreement evaluations.

Kyeongmin Rim: MAE2: Portable Annotation Tool for General Natural Language Use

A pair of general-purpose annotation/adjudication tools, MAE and MAI, has been available for years and has successfully proven itself useful in many semantic annotation projects. We are releasing a newer version, MAE2, that inherits the original pair's strengths of being adaptable, flexible, and portable. The new version is enhanced with new features to help rapid prototyping of the design of natural language annotation tasks, naturally modeling complex semantic structures, setting up a more focusable and consistent annotation work-flow, and assuring the quality of annotations. Also, as an open source project, to make it easier to modify the software for specialized features for a specific annotation task, the MAE2 has adopted common software design patterns.