

**LREC 2016 Workshop**

**ISA-12**

**12<sup>th</sup> Joint ACL - ISO Workshop on  
Interoperable Semantic Annotation**

**PROCEEDINGS**

Edited by

Harry Bunt

28 May 2016

Proceedings of the LREC 2016 Workshop  
ISA-12 – 12<sup>th</sup> Joint ACL - ISO Workshop on Interoperable Semantic Annotation

28 May 2016 – Portorož, Slovenia

Edited by Harry Bunt

## Organising Committee

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

## Programme Committee

- Jan Alexandersson, DFKI
- 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 of Tübingen



- Piek Vossen, Free University Amsterdam
- Annie Zaenen, Stanford University
- Sandrine Zufferey, Université de Fribourg



# Programme

09.00 – 09.10 Opening

**Session 1, 9.15 - 10.30**

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

**Session 2, 11.00 - 12.15**

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

**Session 3, 14.00 - 15.45**

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

**Session 4, 16.15 - 17.30**

- 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:  
*MAE 2: Portable Annotation Tool for General Natural Language Use*
- 17.30 –17.31 **ISA-12 Workshop Closing,**  
followed by discussion of proposed new ISO activities:

**ISO TC 37/SC 4/WG 2 Session, 17.31 - 18.00**

- 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*

# Table of Contents

<i>A Lexically-Informed Upper Level Event Ontology</i> Claire Bonial, Susan Windisch Brown and Martha Palmer .....	1
<i>Verb Meaning Operability: Keeping Complex Resources Alive</i> James Pustejovsky, Martha Palmer, Annie Zaenen, and Susan Windisch Brown .....	7
<i>Mapping Semantic Types onto WordNet Synsets</i> Elisabetta Jezeq, Anna Feltracco, Lorenzo Gatti, Simone Magnolini and Bernardo Magnini .....	11
<i>Cross-level Semantic Annotation of Bulgarian Treebank</i> Petya Osenova and Kiril Simov .....	16
<i>Text-Picture Relations in Cooking Instructions</i> Ilka van der Sluis, Shadira Leito and Gisela Redeker .....	22
<i>An Abstract Syntax for ISOSpace with its &lt;moveLink&gt; Reformulated</i> Kiyong Lee .....	28
<i>Discourse Markers and Disfluencies: Integrating Functional and Formal Annotations</i> Ludivine Crible .....	38
<i>ISO DR-Core: Core Concepts for the Annotation of Discourse Relations</i> Harry Bunt and Rashmi Prasad .....	45
<i>Feedback Matters: Applying Dialog Act Annotation to Study Social Attractiveness in Three-Party Conversations</i> Benjamin Weiss and Stefan Hillmann .....	55
<i>Time Frames: Rethinking the Way to Look at Texts</i> Andreea Macovei and Dan Cristea .....	59
<i>ECAT: Event Capture Annotation Tool</i> Tuan Do, Nikhil Krishnaswamy, and James Pustejovsky .....	63

<i>A Construction-Centered Approach to the Annotation of Modality</i> Elisa Ghia, Lennart Kloppenburg, Malvina Nissim and Paola Pietrandrea .....	67
<i>MAE 2: Portable Annotation Tool for General Natural Language Use</i> Kyeongmin Rim .....	75
<i>Contrastive (English-Spanish) Annotation of Epistemicity in the MULTINOT Project: Preliminary Steps</i> Julia Lavid, Marta Carretero and Juan Rafael Zamorano .....	81

# A Lexically-Informed Upper-Level Event Ontology

Claire Bonial,<sup>1</sup> Susan Windisch Brown,<sup>2</sup> Martha Palmer<sup>2</sup>

<sup>1</sup>U.S. Army Research Laboratory, <sup>2</sup>University of Colorado, Boulder

<sup>1</sup>2800 Powder Mill Rd, Adelphi, MD 20783

E-mail: Claire.N.Bonial.civ@mail.mil, Susan.Brown@colorado.edu, Martha.Palmer@colorado.edu

## Abstract

We describe the development of an upper-level event ontology that is informed by existing ontologies and draws its lexical sense distinctions from VerbNet, FrameNet, and the Rich Entities, Relations and Events project. This ontology is unique in that it is being developed within the theoretical framework of other existing upper-level ontologies, but also with the intention that the ontology provide coverage of the concepts represented in these computational lexical resources. As a result, the ontology allows for the combination and comparison of semantic representations from each of the resources, which vary in the level and type of semantic information provided. Furthermore, the ontology facilitates interoperability and the combination of annotations done for each independent resource. Additionally, the ontology reveals higher-order relationships between events, including temporal and causal relationships. This ontology is still under development. In this paper, we describe the lexical resources that are serving as the sense inventories for our upper-level event ontology, the foundational ontologies we have taken inspiration from in the organisation of our ontology, and explore a semantic domain of interest in more detail, considering possible use cases of the ontology.

**Keywords:** lexicon, semantic roles, ontology

## 1. Introduction & Background

A great deal of work has gone into creating the valuable computational lexical resources, VerbNet (Kipper et al., 2008), FrameNet (Fillmore et al., 2002), and the Rich Entities, Relations and Events (ERE) project (Song et al., 2015), each of which provide somewhat distinct information about which eventualities are related syntactically, semantically, or both, and which types of participants are involved in classes of eventualities. Furthermore, a great deal of time and resources have gone into constructing annotated corpora using the class and participant type labels set out in these resources, and these annotated corpora have proved to be useful sources of training data for a variety of Natural Language Processing (NLP) systems, including automatic semantic role labeling, word sense disambiguation, and question-answering systems.

Simultaneously, we have recently seen an explosion of efforts in the construction of ontologies as part of the Semantic Web (Berners-Lee, 1998). These ontologies facilitate machine reasoning over data that was previously only available for human consumption. Unfortunately, the progress to integrate computational lexical resources into the Semantic Web (e.g., Eckle-Kohler et al., 2014) has been somewhat slow and difficult, given that the conversion of resources like FrameNet, which include quite nuanced and complex ontological relations, into the minimalist Resource Descriptive Framework (RDF) schema used in the Semantic Web is not necessarily a trivial conversion and may involve some loss of information (e.g., Nuzzolese et al., 2011; Scheffczyk et al., 2006).

Our goal is to create an upper-level event ontology that provides conceptual coverage for the aforementioned lexical resources, and uses them as the sense inventories housing the linguistic realizations of those concepts in English. Our approach is somewhat distinct from that of the Semantic Web, wherein mapping distinct resources is common, but we are actually merging the individual resources under the umbrella of the upper-level event ontology. Individually, each lexical resource provides

valuable information about related eventualities and participant types, or semantic roles. However, each resource provides slightly different information as far as what type of “relatedness” is tapped into, and what level of semantic specificity participants are described with. The event ontology allows us to combine this information into one resource. Each resource has also been involved in some of the largest-scale, longest-running annotation projects to date in NLP. The event ontology allows us to combine annotations done for independent resources into one larger, more diverse training corpus. Furthermore, we aim to capture richer temporal and causal relationships between eventualities by basing some of the ontology’s relations on the fine-grained temporal and causal relations of the Richer Event Description annotation (RED) project (Ikuta et al., 2014).

In this paper, we describe each of the NLP resources we are drawing upon in constructing the upper-level event ontology, as well as the existing ontologies that we have taken inspiration from when deciding upon the most basic distinctions of the ontology. We also offer some insights into possible use cases of the ontology, focusing on one semantic domain of the ontology. We close with our next steps in developing the ontology, including plans for testing its utility in an end-to-end application.

## 2. Initial Design Decisions

To facilitate compatibility with the Semantic Web, our ontology was developed using the open-source ontology editor, Protégé (Noy et al., 2000) in OWL format. An early design decision we faced was how to incorporate the lexicons of interest into an ontology. Our initial decision was to include a top level class “Lexicon\_Features” within which there were separate sister classes representing the VerbNet, FrameNet, and ERE lexicons. However, in order to maintain the separation of concepts in the ontology, which represent generic sets of events, states and their participants, and the lexical denotation of those entities captured in a lexicon, we are currently shifting the ontology’s structure such that the VerbNet, FrameNet, and ERE lexicons are distinct, stand-alone ontologies that are imported into the upper-level event ontology. The

individual lexicons and the ontology are linked through the “has\_Sense” relation: conceptual nodes in the ontology have senses spelled out in the lexicons. (However, we are exploring adopting, or mapping this property to, the Lemon Uby relation “isReferenceOP” (Eckle-Kohler et al., 2014: 2).) This design also has a practical advantage: it allows for independent maintenance and updates of the lexical ontologies separate from the upper-level event ontology.

### 3. Sense Inventories of the Ontology

Currently, we have successfully implemented both VerbNet and ERE in OWL, since these lexicons have only very shallow hierarchical class structure. However, we are still developing the OWL-implementation of FrameNet, since, as mentioned previously, the ontological structure and inheritance types in FrameNet are quite complex. To this point, we have been developing the FrameNet lexical ontology on an as-needed basis, and including only basic inheritance links within the FrameNet ontology. However, we are exploring the feasibility of adopting an existing OWL-implementation of FrameNet (e.g., Scheffczyk et al., 2006), and importing this directly into the upper-level ontology. Here we describe each of the lexical resources included in the ontology, with some description of how the sense distinctions vary across each resource.

#### 3.1 VerbNet

VerbNet, based on the work of Levin (1993), groups verbs into “classes” based on compatibility with certain “diathesis alternations” or syntactic alternations (e.g., *She loaded the wagon with hay* vs. *She loaded hay into the wagon*). Although the groupings are primarily syntactic, the classes do share semantic features as well, since, as Levin posited, the syntactic behavior of a verb is largely determined by its meaning. Within each class, VerbNet lists the verbs of a class, the diathesis alternations characterizing a class (exemplified in typical usage examples), the syntactic constituents of each usage example, the semantic roles assigned to each constituent, and a semantic representation of each usage in the form of semantic predicates (e.g., CAUSE, MOVE, TRANSFER\_INFO). With this rich syntactic and semantic information, VerbNet serves as a resource for automatic semantic role labeling, word sense disambiguation, and question-answering systems.

Within the ontology, VerbNet will serve as a lexicon imported into the ontology, and it will therefore provide one set of sense distinctions for the English lexical items that denote concepts within the ontology. Because class membership in VerbNet is, in part, based on syntactic information, VerbNet captures the level of sense distinctions that are clearly evidenced by differences in syntactic behaviors (Brown et al., 2011). For example, the *creation* sense of *make* found in VerbNet’s Build class is often realized in a frame wherein the Product is a noun phrase direct object, while the Material is a prepositional phrase: *Martha carved a toy out of a piece of wood*. A quite distinct sense of *make* is found in the Reach class, exhibiting very different syntactic and semantic characteristics, i.e. “subcategorization frame.” In this sense, the noun phrase direct object indicates a Goal: *They made the finish line*.

#### 3.2 FrameNet

FrameNet, based on Fillmore’s frame semantics (Fillmore, 1976; Fillmore & Baker, 2001), groups verbs, nouns and adjectives into “frames” based on words or “frame elements” that evoke the same semantic scene or frame: a description of a type of event, relation, or entity and the participants in it. For example, the Apply\_heat frame includes the frame elements Cook, Food, Heating\_instrument, Temperature\_setting, etc. Thus, the frame evokes a real-world cooking scenario. Although these groupings are purely semantic, usage examples listing common syntactic realizations are given in each frame. FrameNet’s “net” of frames makes up a rather complex ontological network, including simple “is\_a” inheritance relations as well as more complex relations, such as Precedes and Perspective\_on. Like VerbNet, FrameNet has been used for a variety of different NLP tasks, including semantic role labeling and Natural Language Understanding.

FrameNet will serve as another lexicon within the ontology, providing a different set of sense distinctions for the lexical items denoting concepts. Since the classification of FrameNet is purely semantic and based on shared frame elements, the sense distinctions made in FrameNet are more fine-grained than VerbNet. Furthermore, the distinctions between participant types, or semantic roles, are much more fine-grained. For example, given the sentence *Sally fried an egg*, VerbNet would label *Sally* with the traditional semantic role label Agent, while FrameNet would label *Sally* with the more semantically specified label of Cook.

#### 3.3 Rich Entities, Relations, and Events

The Rich Entities, Relations, and Events (ERE) project is based on both the Automatic Content Extraction (ACE) project (Doddington et al., 2004) and the PropBank (Palmer et al., 2005) semantic role annotation schemas. The goal of the ERE project is to mark up the events (and other types of relations; i.e. “eventualities”) and the entities involved in them, and to mark coreference between these. This provides a somewhat shallow representation of the meaning of the text. Thus, ERE is somewhat different than VerbNet and FrameNet, which were created as lexical resources. ERE was instead created as an annotation schema used in the construction of training corpora for machine learning.

The ERE schema will also serve as a sort of lexicon imported into the ontology, with its event type and subtype designations serving as links to the lexical items marked up with that designation. ERE annotated eventualities are limited to certain types and subtypes of special interest within the defense community, with top-level types referred to as *Life, Movement, Transaction, Business, Conflict, Manufacture, Contact, Personnel* and *Justice* events. Thus, the sense distinctions made by this resource are grounded in practical considerations of what event types are deemed to be of interest, and therefore offer very different insights and information into related events when compared with either FrameNet or VerbNet.

## 4. Ontology Structure & Relations

A goal of our upper-level event ontology is to provide unified conceptual coverage for, and interoperability



between, the lexical resources described in the preceding section. There are many critical decisions to be made when deciding what concepts and relations to represent in an ontology; thus, we draw upon established ontologies. Basic structural decisions in the ontology and the related work that influenced those decisions are described below.

#### 4.1 Temporal/Causal Relations

Within the ontology, we would like to capture richer information about the temporal and causal relations between events than any of the lexical resources described thus far are currently capturing independently. To ensure that the ontology captures temporal and causal relations of utility within NLP, we use relations from the Richer Event Description (RED) project. Like ERE, the RED project also aims to markup text with mentions of eventualities and entities, but the primary focus of RED is to represent the temporal and causal relationships between those eventualities. The final goal is to produce annotations rich enough that a computer, using complex inferencing, coreference, and domain-specific algorithms, would be able to construct an accurate event representation, including an accurate timeline of when the events in a given document occur relative to any fixed dates present and relative to one another (e.g., automatically constructed timelines of medical histories). RED builds on THYME (Styler et al., 2014), a temporal relationship annotation of clinical data that is based on TimeML (Pustejovsky et al., 2010). The temporal relations are quite fine-grained, including *Before*, *Before+overlap*, *Overlap*, *After+Overlap*, *After*. These labels are further distinguished with causal labels where appropriate: *Before/Causes*, *Before/Preconditions*. To anchor the events into a timeline, RED links the event to a document time or section time where applicable, and marks up explicit references to time in the document.

#### 4.2 Basic Class Distinctions

A variety of existing ontologies have been researched to serve as a “jumping off point” from which the upper-level ontology will be adapted to best suit its unique goals.

**WordNet** – WordNet (Fellbaum, 1998) is a large electronic database of English words, which was inspired by work in psycholinguistics investigating how and what type of information is stored in the human mental lexicon (Miller 1995). WordNet is divided firstly into syntactic categories: nouns, verbs, adjectives, and adverbs, and secondly by semantic relations. The semantic relations that organize WN are: synonymy (given in the form of ‘synsets’), antonymy, hyponymy (e.g. a Maple is a tree; therefore, tree is a hypernym of Maple), and meronymy (part-whole relations). These relations make up a complex network of associations that is both useful for computational linguistics and NLP, and also informative in situating a word’s meaning with respect to others. The highest-level semantic distinction made is between concrete entities and abstract entities, with events falling under abstract entity. The verb hypernym hierarchy is much more shallow, basically grouping verb synsets into one of 15 categories.

**SUMO** – The Suggested Upper Merged Ontology (Pease, 2002) is a formal ontology that maps to, and therefore reflects the ontological structure of, the WordNet lexicon. It serves as the upper-level ontology for a variety of domain

ontologies, varying in focus from emotions to weapons of mass destruction. As SUMO was based, in part, on WordNet, SUMO also makes a primary distinction between physical entities and abstract entities. However, SUMO’s next distinction for physical entities is between objects and processes, such that most events are represented as physical, as opposed to abstract, entities.

**DOLCE** – The Descriptive Ontology for Linguistic and Cognitive Engineering (Masolo et al., 2003) is the first module of the WonderWeb Foundational Ontologies effort, which aims to provide a library of ontologies that facilitate mutual understanding. The approach of DOLCE is “multiplicative,” meaning that entities can be co-located in the same space-time (e.g., a vase and the clay that constitutes the vase). This impacts the number of categories in the ontology, in the sense that it is preferred to introduce new categories despite their possible mutual reducibility. Within DOLCE, a primary distinction is made between endurants and perdurants (also sometimes called continuants and occurrents, respectively). Endurants are wholly present when present, while perdurants extend in time by accumulating different temporal parts, so they are only partially present when present.

**BFO** – The Basic Formal Ontology (Smith & Grenon, 2002) is intended to be an upper-level ontology to support the creation of lower-level domain ontologies; therefore, it is designed to be neutral with regard to the domains in which it is applied. BFO and DOLCE share many goals and properties, including an initial split between what BFO calls continuants (entities that can be sliced to yield parts only along spatial dimensions, e.g. table) and occurrents (entities that can be sliced along spatial and temporal dimensions to yield parts, e.g. events – childhood, throwing).

After considering the somewhat unique purposes and structures of each of these existing ontologies, we have selected a top level concept, *All Entities*, with an initial distinction between Endurant and Perdurant Entities. We define “Entity” as a unique object or set of objects in the world – for instance a specific person, place or organization – that typically functions as a participant. We define “Endurants” as those entities that can be observed/perceived as a complete concept, no matter which given snapshot of time – were we to freeze time, we would still be able to perceive the entire endurant. We define “Perdurants” as those entities for which only a part exists if we look at them at any given snapshot in time. Various events, processes, phenomena, activities and states, perdurants have temporal parts or spatial parts and participants. Beyond this primary distinction, our ontology makes secondary distinctions between physical and nonphysical endurants, as well as eventive and stative perdurants.

#### 4.3 Current Development Status

The construction of the ontology is still underway, and has involved a combination of bottom-up and top-down strategies. As ERE event types provide useful constraints on which events to focus on initially, our efforts generally begin with an examination of a particular ERE event type, a comparison of sense distinctions and lexical items made

in VerbNet and FrameNet, and a preliminary fleshing-out of one area of the ontology. At this point, we have situated the top level ERE event types *Life*, *Conflict*, *Contact*, *Justice* and *Personnel*. Thus, we have also situated most of the subtypes within these event types, although our approach involves some iterative refinement of the ontology’s class structure, so surely some of these preliminary placements will change as we examine the relationships and interactions between events in the ontology.

### 5. Life in the Ontology

To give a sense of what information is captured in the ontology, and how this information could be useful and valuable, the area of the ontology covering *Life* events is described in greater detail here. What are *Life* events? Perhaps the better question is, what are not *Life* events? Our first job is to constrain what concepts we are focusing on. We use ERE designations to determine where to focus our initial efforts. ERE *Life* events include the subtypes *Be-Born*, *Marry*, *Divorce*, *Injure* and *Die*. Within the ontology, we capture *Life\_Events* as a direct daughter class of *Eventive\_Perdurants*, which is a daughter class of *Perdurant\_Entities*, as mentioned previously. *Life\_Events* are a sister class to *Intentional\_Acts*, contrasting the non-volitional nature of many (but not all) of these events. Specifically in our ontology, the daughters of the *Life\_Event* class are *Birth*, *Death*, *Injury*, and *Life\_Sustaining\_Activity*. While *Birth* and *Death* are currently terminal nodes in the ontology, *Injury* splits into *Cause\_Injury* (which is currently underspecified as to volitionality) and *Experience\_Injury*, and *Life\_Sustaining\_Activity* remains under development. Given the dynamic and often ambiguous nature of events, we are still exploring how strictly disjoint to make classes like those described above.

As mentioned previously, each lexicon is constructed as a separate ontology that is imported into the main event ontology. Where appropriate, a concept in the ontology is currently linked to its English lexical realizations through the “has\_Sense” property. For example, the *Birth* concept has senses in the following event classes, listed by lexical resource: FrameNet *Giving\_birth*, *Birth\_scenario*, and *Being\_Born*, VerbNet *Birthing*, which is a meta-class encompassing the *Birth-28.2* and *Calve-28.1* sister classes, and the ERE *Life.Be-Born* subtype. The *Life\_Event* portion of the ontology is shown in Figure 1. This extract shows the sense mappings to VerbNet only, as the full visualization of the ontology and its lexical links can quickly become overwhelming to view.

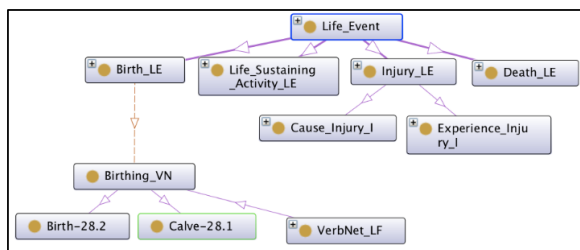


Figure 1: Life\_Event extract of event ontology.

Within each of the lexical classes, the individual lexical items denoting senses of a concept are listed within each resource. For example, the VerbNet *Birth* class lists the lexical items *bear*, *birth*, *deliver*, *father*, *mother*, *sire*, *spawn*, etc. The FrameNet *Being\_born* frame includes just *born* and the phrase *come into the world*. ERE realizations include any lexical item tagged as a trigger indicating this type of event during annotation, such as the verb *born* and the noun *birth*.

Events within the ontology are also related via temporal and causal relations, such as “has\_Result” and “has\_Precondition.” Here, for example, *Birth* life events are linked to the *Life\_State*, *Alive* (a daughter node to *Stative\_Perdurant*), through the “has\_Result” link: once something is born, it is alive. While some of these relations have been developed (and are still under development) specifically for the upper-level ontology, other relations stem directly from the RED project.

From this example, we can begin to see a variety of uses for the ontology. First, the ontology reveals the differences in coverage and semantic specificity of the classifications in each resource, and provides the ontological structure needed to understand relationships between events at higher levels of generalization. This could aid in overcoming data sparsity that can be problematic for a variety of different NLP systems relying on training data to learn verb behaviour. For example, if FrameNet’s specificity, which limits *Being\_born* events to two lexical realizations, is overly restrictive, then the ontology facilitates pinpointing other lexical items that are more generally related to the entries in *Being\_born*. Thus, the ontology facilitates some interoperability between the individual lexical resources and makes explicit some of the previously unseen relations between them.

Notably, the SemLink project (Palmer, 2009), which maps VerbNet, FrameNet, PropBank, and the OntoNotes sense groupings (Pradhan et al., 2007), can also be used to facilitate interoperability among most of these resources. The ontology, however, has value far beyond this, for the ontology not only captures similarity of events at a higher order than the individual classifications, but also provides information about causal and temporal relationships between events. Although these relations are relatively simple, they allow for rather powerful reasoning about preconditions and resulting states, as demonstrated by related work on the Event Situation Ontology (ESO) (Segers et al., 2015), which we also drew inspiration from. When we consider how the ontology combines information about related events and how those events tend to occur in larger temporal and causal chains, we see the potential for the ontology to provide information about the ways in which different events combine to form common higher-order scenarios. For example, within the *Justice* area of the ontology, it is clear that the charge of a crime is commonly followed by trial and a verdict. This carries both explanatory and potentially predictive power.

### 6. Conclusions & Future Work

As we continue to define and refine the ontology, we will focus on the other ERE event types that we have not yet examined, including *Movement*, *Business*, and *Transaction* events. Once we have made the preliminary decisions

regarding the placement of these events within the ontology, we will continue to refine the ontology with richer relations between events. Another challenging task to tackle will be the specificity with which event participants are described. Currently, the ontology includes the relation “has\_Participant,” which provides a general link between eventualities and their participants. In the future, we will be making this label more fine-grained by subdividing it into relations specific to a particular semantic role, e.g., “has\_Agent,” “has\_Patient,” etc. Although we will explore using the VerbNet thematic roleset, research is always needed into what roleset is the most appropriate for a particular application, and here we must determine the roleset that would be the most informative, but also general enough to apply across many different event types. Thus, we are also considering the LIRICS thematic roleset (Petukhova et al., 2008), which is closely related, but somewhat more general than the VerbNet thematic roleset (Bonial et al., 2011), as well as the FrameNet frame elements.

Additionally, we will begin to explore the utility of the ontology in NLP and multi-modal, text/video processing tasks. One simple set of experiments will examine the utility of the ontology in overcoming data sparsity, for example in training a Word Sense Disambiguation system. Another planned set of experimentation will evaluate the utility of the ontology in human activity recognition in videos (Tahmoush, 2015), examining whether the ontological relations between smaller actions and the larger events and scenarios that they are involved with can be valuable in improving the precision of human activity recognition. In general, we hope that in raising awareness of the development of this resource, we can receive feedback from the community on other potential use cases and users, and develop the ontology with these users in mind.

## 7. Acknowledgements

We gratefully acknowledge the support of DARPA DEFT - FA-8750-13-2-0045 and DTRA HDTRA1 -16-1-0002/Project # 1553695, eTASC - Empirical Evidence for a Theoretical Approach to Semantic Components. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA or the US government.

## 8. Bibliographical References

- Berners-Lee, T. (1998). Semantic web road map.
- Bonial, C., Corvey, W., Palmer, M., Petukhova, V. V., & Bunt, H. (2011, September). A hierarchical unification of LIRICS and VerbNet semantic roles. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on* (pp. 483-489). IEEE.
- Brown, S. W., Dligach, D., and Palmer, M. (2011). VerbNet class assignment as a WSD task. *IWCS*, Oxford, UK, January.
- Doddington, G. R., Mitchell, A., Przybocski, M. A., Ramshaw, L. A., Strassel, S., & Weischedel, R. M. (2004, May). The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In *LREC* (Vol. 2, p. 1).
- Eckle-Kohler, J., McCrae, J., & Chiarcos, C. (2014). lemonUby-a large, interlinked, syntactically-rich resource for ontologies. *Semantic Web Journal*, submitted. special issue on Multilingual Linked Open Data.
- Fellbaum, C. (Ed.) (1998.) *WordNet: An Electronic Lexical Database*. MIT Press.
- Fillmore, C. J. (1976). Frame semantics and the nature of language\*. *Annals of the New York Academy of Sciences*, 280(1), 20-32.
- Fillmore, C. J., & Baker, C. F. (2001, June). Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*.
- Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck. (2002.) Background to FrameNet. *International Journal of Lexicography*, 16(3):235-250.
- Ikuta, R., Styler IV, W. F., Hamang, M., O’Gorman, T., & Palmer, M. (2014). Challenges of Adding Causation to Richer Event Descriptions. *ACL 2014*, 12.
- Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer. (2008.) A large-scale classification of English verbs. *Language Resources and Evaluation Journal*, 42: pp. 21– 40.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., & Schneider, L. (2003). Dolce: a descriptive ontology for linguistic and cognitive engineering. *WonderWeb Project, Deliverable D, 17*.
- Miller, George A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.
- Noy, N. F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R. W., & Musen, M. A. (2001). Creating semantic web contents with protege-2000. *IEEE intelligent systems*, (2), 60-71.
- Nuzzolese, A. G., Gangemi, A., & Presutti, V. (2011, June). Gathering lexical linked data and knowledge patterns from FrameNet. In *Proceedings of the sixth international conference on Knowledge capture* (pp. 41-48). ACM.
- Palmer, M. (2009, September). Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference* (pp. 9-15).
- Palmer, Martha, Daniel Gildea, and Paul Kingsbury. (2005.) The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71– 106.
- Pease, A., Niles, I., & Li, J. (2002, July). The suggested upper merged ontology: A large ontology for the semantic web and its applications. In *Working notes of the AAAI-2002 workshop on ontologies and the semantic web* (Vol. 28).
- Petukhova, V., & Bunt, H. (2008). LIRICS Semantic Role Annotation: Design and Evaluation of a Set of Data Categories. In *LREC*.
- Pradhan, S. S., Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., & Weischedel, R. (2007). Ontonotes: A unified relational semantic representation. *International Journal of Semantic Computing*, 1(04), 405-419.

- Pustejovsky, J., Lee, K., Bunt, H. and Romary, L., (2010). ISO-TimeML: An International Standard for Semantic Annotation. In *LREC*.
- Scheffczyk, J., Baker, C. F., & Narayanan, S. (2006). Ontology-based reasoning about lexical resources. In *Proc. of OntoLex* (pp. 1-8).
- Segers, R., Vossen, P., Rospocher, M., Serafini, L., Laparra, E., & Rigau, G. (2015). Eso: A frame based ontology for events and implied situations. *Proceedings of Maplex2015*.
- Smith, B., & Grenon, P. (2002). Basic formal ontology. *Draft. Downloadable at <http://ontology.buffalo.edu/bfo>.*
- Song, Z., Bies, A., Strassel, S., Riese, T., Mott, J., Ellis, J., ... & Ma, X. (2015, June). From light to rich ERE: annotation of entities, relations, and events. In *Proceedings of the 3rd Workshop on EVENTS at the NAACL-HLT*(pp. 89-98).
- Styler, William, F., Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova and James Pustejovsky. (2014). Temporal Annotation in the Clinical Domain. *Transactions of the Association of Computational Linguistics*, 2, pp. 143-154.
- Tahmoush, D. (2015). Applying Action Attribute Class Validation to Improve Human Activity Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 15-21).

## Verb Meaning in Context: Integrating VerbNet and GL Predicative Structures

James Pustejovsky<sup>†</sup>, Martha Palmer<sup>‡</sup>, Annie Zaenen<sup>§</sup>, and Susan Brown<sup>‡</sup>

<sup>†</sup>Brandeis University <sup>‡</sup>University of Colorado <sup>§</sup>Stanford University  
jamesp@cs.brandeis.edu, {martha.palmer,susan.brown}@colorado.edu, azaenen@stanford.edu

### Abstract

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.

**Keywords:** Event Semantics, Event Structure, VerbNet, Generative Lexicon

### 1. Introduction

In this research note, we report on a newly funded effort towards integrating VerbNet’s lexical structure and Generative Lexicon’s (GL) semantic representation.<sup>1</sup> Our overall goal is to address two of the major problems in the representation and annotation of verb meaning in natural language: (i) how to encode the context-dependence of the meaning of a verb; and (ii) how to adequately represent the subeventual predication that inheres in complex verb meanings and is associated with polysemy arising in distinct contexts. Specifically, we propose integrating GL’s compositional approach to event semantics with the predicative representations in VerbNet. This includes making explicit reference to the conditions that hold before, during, and as a result of an activity or event. Here we focus primarily on the predicative content of these conditions and how this technique might contribute additional structural distinctions within VerbNet classes.

It is well known that verbs can be notoriously polysemous. Sometimes this occurs with overt syntactic markers that are relatively easy to identify, as when a “moved” argument alternation signals both a new subcategorization frame as well as a shift in meaning, as illustrated in (1) below. In fact, there is controversy over whether such meaning preserving diathesis alternations actually constitute true polysemy or not (Levin, 1993).<sup>2</sup>

- (1) a. *The wind broke the glass.*  
break-45.1, [NP V NP]  
b. *The glass broke suddenly.*  
break-45.1, [NP.patient V]

But just as often, polysemy emerges not from argument al-

ternation, but from PP or other forms of predicative adjunction, cf. (2).

- (2) a. *The books slid.*  
slide-11.2, [NP V]  
b. *The books slid from the table.*  
slide-11.2, [NP V PP.init\_loc]  
c. *The books slid to the floor.*  
slide-11.2, [NP V PP.dest]

Here we see a manner-of-motion verb lexically typed as a process in (2a), and in (2b) and (2c) as a telic event. The semantics for each of these senses is illustrated in (3):<sup>3</sup>

- (3) a. [NP V]: motion(during(E), Theme)  
b. [NP V PP.init\_loc]: motion(during(E), Theme)  
path\_rel(start(E), Theme, Init\_Loc, ch.of\_loc, prep)  
c. [NP V PP.dest]: motion(during(E), Theme)  
path\_rel(end(E), Theme, Dest, ch.of\_loc, prep)

Other examples can be seen with the verbs *yank* and *push*.

- (4) a. *Nora yanked the button loose.*  
push-12-1, [NP V NP ADJP-Result]  
b. *Nora pushed the tables apart.*  
push-12-1, [NP V NP ADJP-Result]

These are typically analyzed as cases of constructional meaning (Goldberg and Jackendoff, 2004; Croft, 2001) or co-composition (Pustejovsky, 1995b; Pustejovsky and Busa, 1995), where the construction associated with these examples reflects a contextualized interpretation of the verb meaning. For example, the semantic representation for *yank* in (4a) given in VerbNet is shown in (5).

- (5) cause(Agent, E) contact(during(E), Agent, Theme)  
exert\_force(during(E), Agent, Theme) Pred(result(E), Theme)

Further, notice that the two verbs in (4) are currently annotated as members of the same VerbNet class, since the goal

<sup>3</sup>See (Hwang et al., 2014) for discussion of path\_rel in the context of motion and caused motion constructions.

<sup>1</sup>This work is being carried out in the context of two grants: CwC, a DARPA effort to identify and construct computational semantic elements, for the purpose of carrying out joint plans between a human and computer through NL discourse; and eTASC, a DTRA effort to identify and build semantic components in natural language.

<sup>2</sup>In the examples below, we annotate verb uses with VerbNet class identifiers and the specific construction invoked (Kipper et al., 2006; Brown et al., 2014).

of VerbNet is to capture commonalities of syntax-semantics interaction across members of a class. However, this leaves within-class semantic distinctions still needing further clarification.

The representations make no mention of predicative content differentiating them from other within-class verb members (such as *bounce* and *float*). In addition, the approach VerbNet currently uses for capturing event structure, which distinguishes between the start, end and middle (during) of an event, does not always provide a consistent, detailed representation of different event structures for different types of events.

From such observations, we have begun exploring how the Qualia and Event Structures from Generative Lexicon Theory (GL) can help overcome some of these problems. First, by incorporating a richer subeventual predicative structure within VerbNet's representation, we will be better able to distinguish within-class coherent semantic groupings. Secondly, a more structured and compositional approach to subeventual semantics will help explain the semantics encoded in cases of VerbNet constructional polysemy. In the remainder of this short note, we focus on how to enrich VerbNet's predicative structure, while deferring discussion of changes to the event structure for a later venue.

## 2. Review of VerbNet

VerbNet is a lexicon of around 5,200 English verbs, organized primarily around Levin's (1993) verb classification. Classes in VerbNet are structured according to the verb's syntactic behavior. As described in (Kipper et al., 2006; Palmer, 2009; Bonial et al., 2011), VerbNet describes the sets of diathesis alternations that are compatible with each verb in the lexicon. For example, the verb *break* expresses both an inchoative form as well as a causative form, as already encountered in (1) above. Verbs such as *appear*, however, are compatible with an inchoative form (*A cloud appeared.*), but not in a causative construction. Classes are arranged hierarchically, with subclasses of verbs inheriting all the characteristics and frames of the parent class but exhibiting additional syntactic alternations.

Although the basis of the classification is largely syntactic, the verbs of a given class do share semantic regularities as well because, as Levin hypothesized, the syntactic behavior of a verb is largely determined by its meaning. Each class contains semantic predicates that are compatible with the member verbs and the class's syntactic frames. The semantic representations describe the participants at various stages of the event. For example, the representations for the **break** class, which includes such verbs as *shatter*, *snap*, and *tear*, describe a general Initial\_state at the start of the event and a general Result at the end of the event.

- (6) a. **break**: [NP V NP]  
 b. **example**: "Tony broke the window."  
 c. **syntax**: Agent V Patient  
 d. **semantics**: path\_rel(start(E), Initial\_State, Patient, change\_of\_state) & path\_rel(end(E), Result, Patient, change\_of\_state) & cause(Agent, E) & contact(during(E), Instrument, Patient) & degradation\_material\_integrity(result(E), Patient) & physical\_form(result(E), form, Patient)

The class does not refer to the type of contact that occurs or the specific form that results, although such distinctions could be made for subgroups of the class's verbs.

The related class **calibratable change of state**, covers events of change along a scale, such as *rise*, *fluctuate*, and *dwindle*. Its semantic representation makes no mention of contact or a degradation of material integrity. However, it also uses the path\_rel *start(E)* and *end(E)* predicates, but substitutes *change\_on\_scale* for *change\_of\_state* and adds the predicate *change\_value(during(E), Patient, Direction)*. The direction is left underspecified, and no reference is made to any manner of the change, such as its speed.

## 3. VerbNet Predicative Structure

The first proposed change to VerbNet's semantic representation involves an enrichment to the predicative content associated with subevents that will help differentiate the meaning of within-class verbs. We believe that GL provides a framework with which to perform this kind of semantic componential analysis of word classes. To this end, there are two aspects of GL's semantic structure that will prove useful: predicate opposition structure and subeventual componential analysis. In addition, recent work on scalarity provides useful insights into how to distinguish verb classes involving incremental change (Kennedy and Levin, 2008).

Without an explicit representation of change of state, the lexical structure for a verb does not adequately model change dynamically. For this reason, the concept of *opposition structure* was introduced in GL as an enrichment to event structure (Pustejovsky, 2000). This makes explicit which predicate opposition is lexically encoded in a verb. For example, the verbs *die* and *kill* are both encoded with the opposition structure  $[-dead(x), dead(x)]$ . A binary opposition such as this can have distinct grammatical consequences, and this is reflected in VerbNet by membership in a specific class of *change\_of\_state* (COS), i.e., class 45.<sup>4</sup> In fact, identification of the mode of change and the scale associated with that change goes a long way towards explaining much of the grammatical behavior of such verbs (Hay et al., 1999; Kennedy and Levin, 2008).

VerbNet classes are motivated on the basis of syntactic and alternation-based behavior. We believe that it is possible to also identify semantically coherent clusters of verbs within these classes. A few examples will suggest our approach. Using GL-inspired componential analysis applied to the *run*-class (Verbnet 51.3.2), six distinct semantic dimensions emerge, which provide clear differentiations in meaning within this class. They are: 1) SPEED: *amble*, *bolt*, *sprint*, *streak*, *tear*, *chunter*, *flit*, *zoom*; PATH SHAPE: *cavort*, *hopsotch*, *meander*, *seesaw*, *slither*, *swerve*, *zigzag*; PURPOSE: *creep*, *pounce*; BODILY MANNER: *amble*, *ambulate*, *backpack*, *clump*, *clamber*, *shuffle*; ATTITUDE: *frolic*, *lumber*, *lurch*, *gallivant*; ORIENTATION: *slither*, *crawl*, *walk*, *backpack*. The benefit that such component-based analysis provides, as pointed out above, is that within-class semantic distinctions can be identified and also associated with behavior.

<sup>4</sup>The verb *die* is not formally marked as COS in VerbNet, but we can ignore this for the present discussion.

Theoretically inspired distinctions in meaning (e.g., the two motion verb classes of *path* and *manner*), can be systematically associated with (and hence identified with) specific grammatical realizations in the language. That is, given the right semantic vocabulary, linking components to syntactic behavior in the language can be annotated and then used for training classifiers and clustering algorithms. What is interesting about the class distinctions above, is that each dimension links to (associates with) clusters of syntactic clues and constructions. For example, PURPOSE associates with rationale and purpose clauses; the SPEED, ORIENTATION, and ATTITUDE dimensions select for adverbials for those attributes, respectively. This is the underlying benefit of deep semantic modeling: revealing underlying aspects of the event that are expressed syntactically, given a rich enough description, and an annotation strategy over datasets.

Let us return briefly to the examples mentioned in Section 1.0., where the verbs *slide*, *float*, and *roll* are all annotated as the same class, *slide-11.2*. We wish to identify those predicative forms that will sufficiently distinguish the meanings of these verbs. As pointed out in (Mani and Pustejovsky, 2012), the manner introduced by a verb such as *slide* is a *mereo-topological* specialization in meaning of a generic directed motion verb. This means that the nature of the movement is definable in terms referring to spatial configurations between an object (the ground) and the mover — or part of the mover (the figure). For example, *slide* and *roll* presuppose different modes of contact of the figure's surface to the ground, as well as presupposing a component of rotational symmetry for the figure. The VerbNet entries should reflect this distinction, which will entail reference to a Ground (G) role that is not currently part of the role inventory. Assuming such a participant (or something similar) is added to the inventory of roles in VerbNet, we can introduce the relation of “contact” between the mover and the ground to account for the first distinction, and a predicate for “rotational symmetry” to distinguish the second.

Using these features, we can distinguish several of Levin's classes of manner (including the members of **slide-11.2-1**), where a class is defined by certain constraints that hold throughout the event, E. For designating contact, we adopt RCC8's relations of “externally connected” (EC) and “disconnected” (DC) (Randell et al., 1992). To account for rotational symmetry, we introduce a relation between the moving object and its surface, which is in contact with the ground, i.e., “rot-surface”. These predicates facilitate three basic distinctions within this class: whether the mover is in touching the relative ground (*slide* vs. *fly*), when it is touching it (*slide* vs. *bounce*), and how it is touching it (*slide* vs. *roll*). Consider the definitions in (7).

(7) **Mereo-topological Distinctions:**

For Figure (F) relative to Ground (G):

- a. EC(F,G), throughout E;
- b. DC(F,G), throughout E;
- c. (EC(F',G), throughout E, where rot-surface(F',F):
- d. (EC(F,G), DC(F,G))\* , throughout E.

For example, (7d) expresses the iterating step-wise motion

involved in bouncing or hopping, where contact is followed by no contact, iterated throughout the event. That in (7c) expresses the condition present for a rotating surface in contact with the ground, i.e., *roll*. Finally, (7a) holds for motion of an object, F, involving continuous contact with the surface of the ground, G, while (7b) holds for motion with no contact between F and G. This distinguishes the verbs *slide* and *roll* from *float* and *fly*. The VerbNet representations with these distinction, for *slide* and *roll* might look like the following:

- (8) a. [NP V]: motion(during(E), Figure) & while(E, EC(Figure,Ground))
- b. [NP V]: motion(during(E), Figure) & while(E, EC(F',Ground)) & rot-surface(F',Figure)

This helps clarify the distinction between continuous contact verbs, such as *roll*, *drive*, and *walk*, from *float* and *fly*. This also has consequences when these verb classes each compose with orientational prepositions such as *over*, as illustrated in (9).

- (9) a. The ball rolled over the grass.  
(contact with the grass)
- b. The balloon floated over the grass.  
(no contact with the grass)

This illustrates that, while the orientation introduced by *over* is preserved in both classes, the semantics of contact is conveyed by the motion verb itself.

Finally, consider briefly the distinctions in VerbNet between the *change\_of\_state* classes, two of which were discussed in Section 2 above.

- (10) a. 45.1: break-45.1
- b. 45.2: bend-45.2
- c. 45.3: cooking-45.3
- d. 45.4: other\_cos-45.4
- e. 45.5: entity\_specific\_cos-45.5
- f. 45.6.1: calibratable\_cos-45.6.1
- g. 45.6.2: caused\_calibratable\_cos-45.6.2
- h. 45.7: remedy-45.7
- i. 45.8: break\_down-45.8

For each class, we propose that the opposition structure be explicitly encoded. Further, the nature of the scale structure should be identified, differentiating the following: what scale theory is assumed (nominal, binary, ordinal, interval, ratio); the attribute undergoing change; and whether the predicate denoting the attribute is associated with an *open* or *closed* scale. Through a similar strategy of differential semantic analysis applied across these classes, the nature of the change can be characterized using the vocabulary of GL qualia structure and types. For example, 45.1 involves an opposition structure over the FORMAL qualia role (denoting material integrity), while 45.2 refers to an aspect of the FORMAL, i.e., its “shape”. The calibratable change verbs of 45.6.1 are incremental change predicates that are identified as changing along a specific attribute, whether the scale is open or closed, and the nature of the scale theory.

#### 4. Conclusion

In this brief note, we have reported on some aspects of a new research project aimed at enriching VerbNet's predicative representations. This involved the introduction of systematic predicative enrichment to the verb's predicate structure. One part of this is an explicit identification of the mode of opposition structure inherent in the predicate. Another strategy involved GL-inspired semantic componential analysis over VerbNet classes.

We are also currently investigating a second modification concerning VerbNet's event representation, where we are studying how to integrate aspects of the event structure from GL (Pustejovsky, 1995a), specifically the notion of Dynamic Event Models (Pustejovsky and Moszkowicz, 2011; Pustejovsky, 2013) and Dynamic Argument Structure (Jezek and Pustejovsky, 2016). This is a significant issue, since VerbNet aims to represent the subeventual properties of the event as it unfolds, and it is important to ensure that the representation is both systematic and compositional in nature. This is a topic for ongoing research within this effort.

#### Acknowledgements

This work was supported to Brandeis by Contract W911NF-15-C-0238 with the US Defense Advanced Research Projects Agency (DARPA), to University of Colorado by Subaward 2015-06166-02 from UIUC for DARPA, and to both Brandeis and CU by the Defense Threat Reduction Agency (DTRA). Approved for Public Release, Distribution Unlimited. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. All errors and mistakes are, of course, the responsibilities of the authors.

Claire Bonial, William Corvey, Martha Palmer, Volha V Petukhova, and Harry Bunt. 2011. A hierarchical unification of lyrics and verbnet semantic roles. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on*, pages 483–489. IEEE.

Susan Windisch Brown, Dmitriy Dligach, and Martha Palmer. 2014. Verbnet class assignment as a wsd task. In *Computing Meaning*, pages 203–216. Springer.

William Croft. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

Adele E Goldberg and Ray Jackendoff. 2004. The english resultative as a family of constructions. *Language*, pages 532–568.

Jennifer Hay, Christopher Kennedy, and Beth Levin. 1999. Scalar structure underlies telicity in "degree achievements". In *Semantics and linguistic theory*, volume 9, pages 127–144.

Jena D Hwang, Annie Zaenen, and Martha Palmer. 2014. Criteria for identifying and annotating caused motion constructions in corpus data. In *LREC*, pages 1297–1304.

Elisabetta Jezek and James Pustejovsky. 2016. Dynamic argument structure.

Christopher Kennedy and Beth Levin. 2008. Measure of change: The adjectival core of degree achievements. *Adjectives and adverbs: Syntax, semantics and discourse*, pages 156–182.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending verbnet with novel verb classes. In *Proceedings of LREC*, volume 2006, page 1. Citeseer.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

Inderjeet Mani and James Pustejovsky. 2012. *Interpreting Motion: Grounded Representations for Spatial Language*. Oxford University Press.

Martha Palmer. 2009. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference*, pages 9–15.

James Pustejovsky and Federica Busa. 1995. Unaccusativity and event composition. *Temporal reference, aspect, and actionality*, 1:159–177.

James Pustejovsky and Jessica Moszkowicz. 2011. The qualitative spatial dynamics of motion. *The Journal of Spatial Cognition and Computation*.

J. Pustejovsky. 1995a. *The Generative Lexicon*. Bradford Book. MIT Press.

James Pustejovsky. 1995b. *The Generative Lexicon*. MIT Press, Cambridge, MA.

J. Pustejovsky. 2000. Events and the semantics of opposition. In C. Tenny and J. Pustejovsky, editors, *Events as Grammatical Objects*, pages 445–482. Center for the Study of Language and Information (CSLI), Stanford, CA.

James Pustejovsky. 2013. Dynamic event structure and habitat theory. In *Proceedings of the 6th International Conference on Generative Approaches to the Lexicon (GL2013)*, pages 1–10. ACL.

D.A. Randell, Z. Cui, A. Cohn, B. Nebel, C. Rich, and W. Swartout. 1992. A spatial logic based on regions and connection. In *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176, San Mateo. Morgan Kaufmann.



## Mapping Semantic Types to WordNet Synsets

Elisabetta Jezeq<sup>1</sup>, Anna Feltracco<sup>1,2</sup>, Lorenzo Gatti<sup>2,3</sup>, Simone Magnolini<sup>2,4</sup>, Bernardo Magnini<sup>2</sup>

<sup>1</sup>University of Pavia, Strada Nuova 65, 27100 Pavia, Italy

<sup>2</sup>Fondazione Bruno Kessler, Via Sommarive 18, 38100 Povo-Trento, Italy

<sup>3</sup>University of Trento, Via Calepina 14, 38122 Trento, Italy

<sup>4</sup>University of Brescia, Piazza del Mercato 15, 25121 Brescia, Italy

jezeq@unipv.it, feltracco@fbk.eu, l.gatti@fbk.eu, magnolini@fbk.eu, magnini@fbk.eu

### Abstract

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.

**Keywords:** semantic type, synset, lexical resource, mapping, semantic annotation, taxonomy, ontology

### 1. Introduction

It is common practice in computational linguistics to use conceptual categories organized in a hierarchy (i.e. ontologies) as primary knowledge resources to perform several tasks. For example, for word sense disambiguation (WSD) tasks applied to verbs, a widely adopted methodology is to use a given inventory of categories (Human, Location, Artifact, etc.) to encode the combinatorial constraints a verb places on its arguments, and employ this feature to guide the discrimination of different senses for verbs in context. “Categories” are often equated with “senses”, and structured sense repositories such as WordNet (henceforth WN) (Fellbaum, 1998) are treated as ontologies and widely used for WSD tasks (McCarthy, 2006).

In this paper, we report the results of an experiment of mapping categories reflecting verb selectional constraints, acquired from distributional analysis and clustering of corpus data (corpus-derived Semantic Types, henceforth STs) to WN synsets. The mapping is performed as the first step of a broader experiment aiming at populating STs with argument fillers extracted from corpora, using WN (Feltracco et al, 2016; see also section 4.1). Here, we are interested in examining to what extent linguistic objects which share common properties but are defined based on different criteria (STs and their extensions on the one hand, WN synsets and their hyponyms on the other hand) can be linked.

The paper is structured as follows. Section 2 briefly introduces the context in which the list of STs object of the mapping has been compiled. Section 3 focuses on the comparison between STs, conceptual categories and WN synsets, and provides the theoretical background of the experiment. Section 4 illustrates the various steps of the mapping, together with its evaluation. Section 5 reports our concluding observations and highlights issues for further work.

### 2. The resource

The Semantic Types used in our mapping experiment are derived from the T-PAS resource. The T-PAS resource (Jezeq et al, 2014)<sup>1</sup> is an inventory of Typed Predicate Argument Structures for Italian verbs manually acquired from corpora following the Corpus Pattern Analysis (CPA) methodology (Hanks, 2013).<sup>2</sup>

T-PASs are semantically motivated and are identified through inspection and annotation of actual uses of the analyzed verbs in a corpus of sentences extracted from a reduced version of the ItWAC corpus (Baroni and Kilgarriff, 2006). An example of T-PAS for the Italian verb *divorare* (Engl. to devour) is given in (1):

- (1) T-PAS#2 of the verb *divorare* (Eng. to devour):  
[[Human]] divorare [[Document]]<sup>3</sup>

According to the CPA procedure, after analyzing a random sample of 250 concordances of the verb in the corpus, the lexicographer defines each T-PAS recognizing its relevant structure and identifying the STs for each argument slot by generalizing over the argument fillers (“lexical set”) observed in the concordances. For instance, in (1) [[Document]] generalizes over the lexical set *libro, romanzo, saggio*, etc. (Eng. book, newspaper, essay). Then, the lexicographer associates the instances in the corpus to the corresponding T-PAS; these sentences in the corpus correspond to a list of examples of the particular sense of the verb. The latter is described and encoded in the resource by the lexicographer in the form of an implicature anchored to the STs of the T-PAS, i.e. the implicature for the T-PAS in (1) is: [[Human]] legge [[Document]] con grande interesse.

<sup>1</sup><http://tpas.fbk.eu/>

<sup>2</sup>The first release contains 1,000 analyzed average polysemy verbs.

<sup>3</sup>Names of Semantic Types are conventionally written in double square brackets with capital initial letters.

### 3. Semantic types, conceptual categories and WordNet synsets

The list of STs used in the T-PAS resource was acquired by applying the procedure described in Section 2 to the analysis of concordances for ca 1,500 English, Italian and Spanish verbs (cf. Hanks and Pustejovsky, 2005 for the English project). Specifically, a list of 224 Semantic Types was obtained by manual clustering and generalization over sets of lexical items found in the argument positions in the corpus. The type list was organized into a hierarchy to capture the appropriate level of selection of verbs.<sup>4</sup> The main relation in the taxonomic structure is the “IS\_A” relation (subsumption), i.e. [[Plane]] is a type of [[Vehicle]] which is a type of [[Artifact]] (see Table1), and so forth.

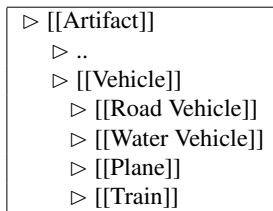


Table 1: Section of the STs Taxonomy.

These types look very much like conceptual / ontological categories for nouns but should instead be conceived as linguistic objects. For example, [[Horse]] is included in the type list because people *cavalca* ‘ride’ horses, *ferra* ‘shoe’ horses, and *sella* ‘saddle’ horses; horses *si imbizzarriscono* ‘bolt’; horses *galoppiano* ‘gallop’; and horses *trottano* or *vanno al trotto* ‘trot’; for all of these verbs, the concept <horse> serves to distinguish a particular event type from events involving other animals (Hanks, 2000). On the other hand, the category <chordate> is not included in the type list, because there are no events encoded in natural language in which <chordate> makes a useful distinction.

STs differ from categories of entities defined on the basis of ontological axioms, such as those of DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), which, despite “aiming at capturing the ontological categories underlying natural language and human common sense” (Gangemi et al, 2002), does not base category distinctions on systematic observation and clustering of language data.

STs also differ from WN synsets - which are sets of cognitive synonyms - each expressing a distinct concept, compiled on the base of psychological assumptions regarding the semantic relations holding among words in the mental lexicon (Fellbaum, 2008). In WN, membership of a word-meaning pair to a synset is not grounded on the systematic analysis of its distribution in the corpus; as a result, synsets can but do not necessarily reflect selectional constraints, as is the case of the STs and their extensions (lexical sets)

<sup>4</sup>Taxonomic structure is mostly based on *prima facie* decisions reflecting the intuition of the lexicographer about the meaning ascribed to the terms used and by manually comparing the paradigmatic set of words that fill the argument positions of different verbs.

identified with the CPA methodology.<sup>5</sup>

In our mapping exercise, we are interested to examine to what extent corpus-derived STs and their extensions in corpora can be mapped onto WN synsets and their hyponyms.

### 4. Types - Synsets Mapping: a Case Study

In this Section we describe the task we performed in Feltracco et al. (2016), for which a mapping between corpus-derived STs from the T-PAS resource and the corresponding WN 1.6 synsets was required (Section 4.1); Section 4.2 discusses in details the experimental mapping we carried out in two phases, i.e. an automatic and a manual one. Finally, we present the evaluation of this mapping, conducted observing the results of the task (Section 4.3).

#### 4.1. Case Study Description

In Feltracco et al. (2016), we aimed at automatically acquiring the paradigmatic sets of words (i.e. lexical sets) corresponding to the STs of specific T-PASs, from the sentences of the corpus already linked to those T-PASs (see Section 2), by using WN synsets and benefitting from its structure.

The task was defined as follows. The system receives as input (i) one T-PAS of a certain verb and (ii) one of the corpus sentences already associated to that T-PAS in the resource. The system should correctly mark, where present, the lexical items corresponding to the STs of each argument position specified by the T-PAS in question. By replicating this step for all the sentences of a T-PAS, the system should build the lexical set for each ST of the T-PAS (i.e. lexical set population).

For instance, Example (2a) shows the T-PAS#1 of the verb *preparare* (Eng. to prepare), and (2b) a sentence associated to it. In this case, the system should identify *nonna* (Eng. grandmother) as a lexical item for [[Human]]-subj. If this annotation is repeated for all the sentences of the T-PAS#1 of the verb, the system will build the lexical set for the ST [[Human]] in subject position in the T-PAS, such as {*nonna, chef, Gino, bambina,...*}.

- (2) a. [[Human]] **preparare** [[Food | Drug]]  
 b. “La *nonna*, prima di informare le patate, **prepara** una torta”  
 (Eng. “the *grandmother*, before baking the potatoes, **prepares** a cake”)

In order to identify possible candidate items for a ST, the system uses MultiWordNet (henceforth, MWN) synsets (Pianta et al, 2002)<sup>6</sup>, which we have mapped on T-PAS STs.

<sup>5</sup>The relevance of the link between synset membership, word sense distinctions and the context of use for both linguistic and NLP purposes has been noted at least since Miller (1995): “WordNet would be much more useful if it incorporated the means for determining appropriate senses, allowing the program to evaluate the contexts in which words are used.” This has led to several initiatives aiming at reducing the high granularity of WN senses to attain higher automatic tagging performance in WSD tasks. A major contribution in this direction is the OntoNotes project (Hovy et al, 2006; Weischedel et al, 2011).

<sup>6</sup>In MWN Italian synsets are aligned with the English synsets of WN 1.6. MWN is one of the features the system uses.

The starting assumption is that the candidate lexical items that the system has to label are hyponyms of those synsets. In Example (2), the system has to identify possible candidate items for the three STs in the sentence, i.e. [[Human]], [[Food]] or [[Drug]]. We consider the labels of these STs (i.e. *human*, *food*, *drug*) as entry point nouns in WN 1.6 (thus obtaining *human#n*, *food#n*, *drug#n*). Finally, we associate the STs with the synsets containing these entry points as in Table 2.

ST	entry point	→ SYNSETS
[[Human]]	→ <i>human#n</i>	→ { <i>human#n#1</i> ; ...}; { <i>human#n#2</i> ; ...}
[[Food]]	→ <i>food#n</i>	→ { <i>food#n#1</i> }
[[Drug]]	→ <i>drug#n</i>	→ { <i>drug#n#1</i> }

Table 2: Example of Automatic Semantic Type-Synsets mapping.

Once initialized, the system takes each Italian lemma in the sentence, looks for all the synsets in MWN that contain it, and retrieves the English synsets aligned with them. Then, it determines whether one of these retrieved synsets is actually one of those previously mapped to the ST in the T-PAS under consideration, or one of its hyponyms.

In Example (2), the Italian lemma *nonna* is searched in MWN and the equivalent English synset {*grandma#n#1*, *grandmother#n#1*, *granny#n#1*, *grannie#n#1*} is found. None of these synset members match with any synset members of *human#n* (i.e. {*human#n#1*, ...} and {*human#n#2*, ...}), *food#n*, or *drug#n*; thus, the MWN hierarchy is traversed until *human#n#1* is found (see a representation in Figure 1).<sup>7</sup>

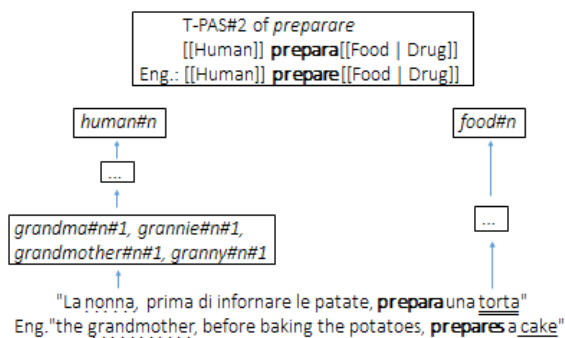


Figure 1: Lexical Set identification for T-PAS#2 for the verb *preparare* as in Feltracco et al. (2016).

We run the experiment for 500 sentences selected by extracting 10 sentences for 10 different STs in 5 different T-PASs (for a total of 50 different T-PASs belonging to 47 verbs). In particular, we chose 10 types that are used in at least 5 different T-PASs, each of them having at least 10 (potential) sentences associated in the reference corpus<sup>8</sup>.

<sup>7</sup>The system is composed by two algorithm: the Baseline and the LEA algorithm. The former lemmatizes each sentence using TextPro (Pianta et al, 2008) and considers the part of speech of the candidate items; the latter also takes into account the dependency tree of the sentence, named entities and multiword expressions.

<sup>8</sup>This is a selection criteria. Considering that we analyzed a limited number of examples for each verb, and that more than one

[[Inanimate]]
▷ [[Artifact]]
▷ [[Garment]]
▷ [[Device]]
▷ [[Food]]
▷ [[Building]]
▷ [[Machine]]
▷ [[Vehicle]]
▷ [[Artwork]]
▷ [[Document]]
▷ [[Drug]]

Table 3: The 10 chosen STs within the [[Inanimate]] branch.

This selection of few STs was intended to better compare performances of the algorithms for different lexical sets. We selected the 10 STs among all the STs within the [[Inanimate]] branch (see Table 3), a branch that displays a high level of granularity and appears to be very representative of the structure of the taxonomy of STs.

We created a gold standard for the task by manually annotating the 500 sentences, i.e. three annotators marked the lexical items that correspond to the members of the lexical set of the STs (for a total of 43 STs).

## 4.2. The mapping

The mapping between the T-PAS STs and the corresponding WN 1.6 synsets has been carried out for the 180 STs that are actually used in the released version of the T-PAS resource (out of the 224 at disposal, see Table 4). The mapping used in the task is created in two steps: in the first step (the *automatic mapping step*) we map the STs to the synsets by automatically associating the STs to MWN entry nouns; in the second step (the *manual resolution step*) we solve the gaps in which there is no exact match between the ST and at least one entry point in MWN.

### The automatic mapping step.

The mapping consists in searching the ST (e.g. [[Human]]) in MWN and matching it with all the synsets of the found entry noun (*human#n#1*; *human#n#2*). The mapping is done automatically, this strategy being less time consuming than manual inspection of all the STs. We also want to test the accuracy of a system that does not require human intervention.

This approach leads to the mapping of 150 STs, such as [[Action]] → *action#n* → *action#n#1-9*, and [[Hair]] → *hair#n* → *hair#n#1-5*.

Since this strategy is only based on a lexical correspondence between the label for the ST and the entry noun in WN 1.6, there is no guarantee that the mapping is always correct. In fact, even when there is a total lexical correspondence, such as between [[Animal]] and *animal#n*, this does not mean that the two terms indicate the same concept, and consequently that their semantic relations are

ST can be specified for each argument slot, it is also possible that none of the sentences extracted for a ST for a verb instantiate one of the 10 selected STs.

consistent (e.g. is [[Human]] an [[Animal]] in the T-PAS taxonomy? Is *human#n* an hyponym of *animal#n* in WN 1.6?). This problem calls for an evaluation of the automatic mapping, which is discussed in Section 4.3.

#### The manual resolution step.

In 30 cases there was no exact match between a ST and at least one entry noun in MWN. These gaps were manually inspected.

First, we found a candidate entry noun that could map the ST on a purely lexical level (e.g. *alcoholic\_beverage#n* was considered a candidate for [[Alcoholic\_Drink]] due to the synonymy between the entry noun and the label); then, we checked this candidate by looking at the ST guidelines (Bradbury et al, 2014), and the usage of the ST in the T-PAS resource. We also evaluated whether these examples of usage correspond to hyponyms of the synsets of the candidate (e.g. whether *champagne*, *beer*, *vodka* that are examples of [[Alcoholic\_Drink]] in the ST guidelines are also hyponyms of the synsets of *alcoholic\_beverage#n*).

In most cases, a synonym of the type label was chosen as candidate, e.g. [[Alcoholic\_Drink]] was mapped to *alcoholic\_beverage#n*, [[Weather\_Event]] to *atmospheric\_phenomenon#n*, [[Animate]] to *living\_thing#n*. In the case of [[Road\_Vehicle]] an hypernym was selected, as no lexical synonym was found; the ST was mapped to *vehicle#n*.<sup>9</sup>

We also found that some of candidates were not satisfactory solutions for the mapping. This is e.g. the case of [[Abstract\_Entity]]. Initially, we mapped this ST to *abstraction#n*, but this entry alone was considered inadequate for covering the whole type. After examining the ST guidelines, the examples in the corpus, and the hyponyms of the synsets of the candidate entry noun we chose to include also a second synset in the mapping, i.e. *psychological\_feature#n*.<sup>10</sup>

### 4.3. Evaluation of the mapping

As described in Section 4.1, the experiment reported in Feltracco et al. (2016) required a mapping between the two resources based on the correspondence between STs and synsets. This mapping was mainly carried out automatically, following the procedure described in Section 4.2. To

<sup>9</sup>The presence of the ST [[Road\_Vehicle]], for which a corresponding synset has not been found, gives us the opportunity for reasoning on the different criteria followed for creating the two resources. In particular, the need for distinguishing [[Road\_Vehicle]] from other vehicles -and in general the level of granularity of the entire taxonomy- in T-PAS is justified by the presence in the language of specific senses of verbs that requires specific STs (rather than a more general [[Vehicle]]) to be disambiguated. For example, the T-PAS#1 of *guidare* (Eng. to drive) selects [[Road\_Vehicle]] as an object, while for the T-PAS#2 of *viaggiare* (Eng. to move, to travel) the most suitable ST that generalizes over the items in the corpus is the more general [[Vehicle]]. On the other hand, “Road\_Vehicle” do not appear as an entry in WN as this resource does not follow the same criteria; moreover, WN appears to favour single entries over multi-word expressions as entry points.

<sup>10</sup>In the ST taxonomy, [[Psych]] IS A [[Abstract\_Entity]], but in WordNet 1.6 *psychological\_feature#n* is not an hyponym of *abstraction#n*.

evaluate the quality of such a mapping, the ideal criterion would be taking all the items over which a ST generalized and identify a WN synset (or a set of synsets) that is a hypernym of most of these items without being too broad. Ideally, the perfect synset should thus match all of the lexical set for a ST, without matching anything outside it (for example, the best mapping for [[Drug]] is *drug#n#1* an hypernym of all the items generalized by [[Drug]] such as *drug*, *medicine*, *anesthetic*, *etc.* without being an hypernym of e.g. *cake*, *bottle*, *dust* that are obviously not [[Drug]]). The fundamental prerequisite for this evaluation method, however, is having already the entire set of items for each ST, thus it was not applicable in our experiment (as we actually used the mapping for extracting the lexical set).

Still, a partial evaluation can be obtained by comparing the results of the system with our gold standard. This evaluation is “subtractive”, in the sense that it assesses if the mapped synsets do not exhaustively cover the ST fillers, but does not assess whether the mapping is too broad. Also, mapping each ST to the set of synsets that contains the corresponding entry noun (e.g. [[Human]] → *human#n* → *human#n#1-2*), does not allow us to match all and only the senses of the entry noun that actually satisfy the criteria (e.g. is the best mapping for [[Human]] the synset of *human#n#1*, the synset of *human#n#2* or both? This is a point for future work.

Through the evaluation process we carried out, we discovered examples for which the system failed to recognize an item due to a lack of coverage of the mapping, that thus needed to be revised. For instance, in one of the gold standard sentences *elicottero* (Eng.: helicopter) was annotated as the item for the ST [[Machine]] in the obj. context of the verb *manovrare* (Eng. to operate) in T-PAS#1. In MWN the same lemma *elicottero* is an hyponym of *transport#n* and not of *machine#n* and the two entries do not have direct relation in any of their synsets. As a consequence, in the sentences in which *elicottero* (or other vehicles) are considered members of the lexical set corresponding to [[Machine]], even traversing the MWN hierarchy the system can not consider these items as valid candidates for the ST [[Machine]]. Thus, the mapping between [[Machine]] and the automatic extracted entry noun *machine#n* was revised by mapping [[Machine]] to both *machine#n* and *transport#n*.

We conducted this evaluation by observing the systems results for the 43 STs involved in the task (see Table 4). For 36 of these STs, one entry noun in MWN was found automatically through the *automatic mapping step*, while the remaining 7 underwent the *manual resolution step*. We found that a review was necessary for 10 of the STs mapped automatically (e.g., for [[Machine]] as explained above), while none of the manually mapped STs required further intervention.<sup>11</sup>

Considering that the algorithm for automatically aligning STs and WN synsets relies only on lexical correspondence, obtaining a reasonable mapping for more than 70% of the entries is a good starting point.

This manual inspection was triggered by a lack of precision

<sup>11</sup>This enhancement led to a significant improvement of the results of the mapping; for details see Feltracco et al. (2016).

of the system for some STs, which was detectable thanks to our gold standard. In order to obtain a more accurate mapping between the two resources, further work includes the evaluation of the other STs as possibly they would also require a manual revision.

Total STs	224
with multiple inheritance	17
▷ STs used in T-PAS	180
▷ <i>Automatically mapped</i>	150
▷ <i>Manually map - gap resolution</i>	30
▷ STs used in the task and evaluated	43
▷ <i>Automatically mapped</i>	36
▷ Manual revision	10
▷ <i>Manually map - gap resolution</i>	7
▷ Manual revision	0
▷ STs not yet used in T-PAS	44

Table 4: Quantitative Data on STs

## 5. Conclusion

In this paper, we report the results of an experiment that establishes a set of mappings between corpus-derived STs and WN synsets. Results show that automatic mappings perform well in terms of matching ST labels to WN entry points (150 over 180 total, 83%). Evaluation of the results of the mappings for 43 STs, based on manual inspection of the fillers of the STs in a gold standard, and of the hyponyms of the corresponding synsets in WN, achieves an overall performance of 72%. In order to improve the accuracy of the mapping between the two resources, we plan to examine the STs that were not considered in the population task through the use of a larger gold standard. The best mappings will be obtained by linking the STs with all and only the senses of the corresponding entry noun(s) in WN, whose hyperonyms match the extensions of STs to the highest degree.

## References

- Marco Baroni and Adam Kilgarriff. 2006. Large Linguistically-Processed Web Corpora for Multiple Languages. In *Proceedings of the 11<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, 87-90.
- Jane Bradbury, Elisabetta Jezeq and Patrick Hanks. 2014. *Semantic Types Annotation Guidelines*. University of Wolverhampton, Università di Pavia.
- Christiane Fellbaum (ed) 1998. *WordNet: An Electronic Lexical Database*. Cambridge MA, The MIT Press.
- Anna Feltracco, Lorenzo Gatti, Simone Magnolini, Bernardo Magnini and Elisabetta Jezeq. 2016. Using WordNet to Build Lexical Sets for Italian Verbs. In *Proceedings of the 8<sup>th</sup> International Global WordNet Conference 2016*. Bucharest, Romania, January 27-30, 100-104.
- Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, Luc Schneider. 2002. Sweetening Ontologies with DOLCE. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management, Ontologies and the Semantic Web*, Berlin, Springer, 166-181.
- Patrick Hanks. 2000. Contributions of Lexicography and Corpus Linguistics to a Theory of Language Performance. *Proceedings of the 9<sup>th</sup> Euralex International Conference*, Stuttgart, Germany, 3-13.
- Patrick Hanks. 2004. Corpus Pattern Analysis. In *Proceedings of the 11<sup>th</sup> EURALEX International Congress*. Lorient, France (July 6-10, 2004), 87-98.
- Patrick Hanks. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge MA, The MIT Press.
- Patrick Hanks and James Pustejovsky. 2005. A Pattern Dictionary for Natural Language Processing. In *Revue Française de linguistique appliquée*, 10.2, 63-82.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw and Ralph Weischedel. 2006. OntoNotes: the 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, ACL*, 57-60.
- Elisabetta Jezeq and Patrick Hanks. 2010. What lexical sets tell us about conceptual categories. In *Corpus Linguistics and the Lexicon*, special issue of *LEXIS*, vol. 4, 7-22.
- Jezeq, Elisabetta, Bernardo Magnini, Anna Feltracco, Alessia Bianchini and Octavian Popescu. 2014. T-PAS: A resource of corpus-derived Types Predicate-Argument Structures for linguistic analysis and semantic processing. In *Proceedings of the 9<sup>th</sup> International Conference on Language Resources and Evaluation (LREC'14)*, May 26-31, Reykjavik, Iceland, ELRA.
- Diana McCarthy. 2006. Relating WordNet senses for word sense disambiguation. In *Proceedings of the Workshop Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together, (EACL 2006)* Trento, Italy, April 4, 17-24.
- George A. Miller. 1995. WordNet: A Lexical Database for English. In *Communications of the ACM*, Vol. 38, No. 11, 39-41.
- Emanuele Pianta, Luisa Bentivogli and Christian Girardi. 2002. Multiwordnet: developing an aligned multilingual database. In *Proceedings of the 1<sup>st</sup> International Conference on Global WordNet*, Mysore, India, 55-63.
- Emanuele Pianta, Christian Girardi and Roberto Zanolli. 2008. The *TextPro* Tool Suite. In *Proceedings of the 6<sup>th</sup> International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.
- Ralph Weischedel, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin, Sameer Pradan, Lance Ramshaw and Nianwen Xue. 2011. OntoNotes: A Large Training Corpus for Enhanced Processing. In *Handbook of Natural Language Processing and Machine Translation: Global Automatic Language Exploitation*, Berlin, Springer, 54-63.

# Cross-level Semantic Annotation of Bulgarian Treebank

Petya Osenova, Kiril Simov

Linguistic Modelling and Knowledge Processing Department  
Institute of Information and Communication Technologies, BAS  
Acad. G. Bonchev 25A, 1113 Sofia, Bulgaria  
petya@bultreebank.org, kivs@bultreebank.org

## Abstract

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.

**Keywords:** Sense annotation; Semantic annotation of Named Entities; Extraction of Syntagmatic Relations

## 1. Introduction

In this paper we present our approach to sense annotation of open class words in the Bulgarian treebank — BulTreeBank. In the semantic annotation we selected three groups of items for annotation: common words, MultiWord Expressions (MWE) and Named Entities (NE). The semantic information was selected from three sources: BTB-WordNet<sup>1</sup>, Bulgarian DBpedia instances (Wikipedia pages) and DBpedia ontology. The three resources are interlinked in the following way: DBpedia instances are related to one or more classes in DBpedia ontology. DBpedia classes have been manually mapped to the WordNet synsets. The information from the Bulgarian DBpedia instances (Wikipedia pages) and DBpedia ontology classes has been used for Named Entity annotation. Our goal was to assign the most specific meaning for each item (being either a common word, a MWE or NE).

The syntactic structure of the treebank was used for extracting of new relations between semantic units (synsets, DBpedia instances and classes). These relations are mainly syntagmatic in their nature and thus might be exploited in many applications. We use them to check the semantic restrictions over the valency frames in a valency lexicon for Bulgarian. Another important application is the Knowledge-based Word Sense disambiguation task. This knowledge is added in the form of arcs in the current knowledge graph.

The paper is structured as follows: in Sect. 2. the related works are mentioned; Sect. 3. describes the Bulgarian treebank; in Sect. 4. the methodology of the sense annotation is described; Sect. 5. compares our resource with SemCor; an application of the created semantic resource for improving knowledge-based Word Sense Disambiguation task is presented in Sect. 6.; Sect. 7. concludes the paper.

## 2. Related Work

There are a number of resources which are sense annotated. Most of them rely on WordNets and/or other lexical resources that provide sense differentiation, such as

<sup>1</sup>BTB-Wordnet is a WordNet of Bulgarian in a process of creation.

language-specific lexicons. Sense annotated corpora take their origins from seminal corpora, such as SemCor, and are realized as particular variants of them in other languages, such as Dutch, Basque, Bulgarian, etc.<sup>2</sup> Unfortunately, most of them are not freely available in their full capacity and for further third-party research.

At the same time, there are not so many treebanks available that have been annotated with senses. Here the following ones need to be mentioned, among others: for English, the sense annotated developments of Penn Treebank — PropBank (Palmer et al., 2005) and NomBank (Meyers et al., 2004) — as well as OntoNotes, which combines sense information from several resources; for German, the TÜBa-D/Z sense annotated treebank (Henrich and Hinrichs, 2013); for Italian, the syntactic-semantic treebank (Montemagni et al., 2000). In OntoNotes an ontology was used for mapping the WordNet senses. This is the Omega Ontology (Philpot et al., 2005).

Our resource differs from PropBank in that it does not provide detailed semantic role labels. We expect this information to come from the ontological labels in valency frames over the grammatical roles (subject, complement, adjunct). The sense annotated BulTreeBank remains closer to the OntoNotes strategy of combining syntactic analysis with sense annotations.

The novelty in our sense annotation endeavour lies, as far as we are aware, in the combination of assigned valencies, lexical senses and DBpedia URIs into a syntactic resource.

## 3. Bulgarian Treebank

The original BulTreeBank (Simov et al., 2004; Simov and Osenova, 2003) that has been used in the conversion to the universal format comprises 256,331 tokens, which form a little more than 15,000 sentences. Each token has been annotated with elaborate morphosyntactic information. The original XML format of the BulTreeBank is based on HPSG. The syntactic structure is presented through a set of constituents with head-dependant markings. The phrasal constituents contain two types of information: the domain of the constituent (*NP*, *VP* etc.) and the type of the phrase

<sup>2</sup><http://globalwordnet.org/wordnet-annotated-corpora/>

(head-complement (*NPC*, *VPC* etc.), head-subject (*VPS*), head-adjunct (*NPA*, *VPA* etc.). The treebank provides also functional nodes, such as clausal ones – *CLDA* (subordinate clause introduced by the auxiliary particle “*da*” *to*), *CLCHE* (subordinate clause introduced by the subordinator “*che*” *that*), etc.

Tracing back to the developments of BulTreeBank, its first conversion happened in 2006, when it was transferred into the shared CoNLL dependency format – (Chanev et al., 2006), (Chanev et al., 2007). The rich structure was flattened to a set of 18 relations.<sup>3</sup> This part consists of 196 000 tokens, because the sentences with ellipses were not considered.

Alternative versions of BulTreeBank exist in two other popular formats: PennTreebank (Ghahyoomi et al., 2014) and Stanford Dependencies (Rosa et al., 2014). The former was used for constituent parsing of Bulgarian, while the latter was part of a bigger endeavour towards universalizing syntactic annotation schemes of many languages.

Recently, BulTreeBank has become part of the common efforts that evolved from the previous initiatives towards the creation of comparable syntactically annotated multilingual datasets — Universal Dependency<sup>4</sup>. For the Universal Dependencies initiative we defined the dependency structure over the original BulTreeBank constituent-based format. In this way we managed both types of analyses simultaneously.

#### 4. Sense Annotation

As it was mentioned in the introduction, we exploit three interconnected sources of semantic information for the annotation: BTB-WordNet (BTB-WN), Bulgarian DBpedia instances / Wikipedia pages and DBpedia ontology.

The BTB-WN has been compiled in several steps. Initially, the Core WordNet<sup>5</sup> was created for Bulgarian, which covered 4,999 synsets. Then, nearly the same number of new synsets were added to the WordNet (now we have more than 11,000 synsets) on the basis the semantic annotation within the treebank. The additional definitions were taken from a machine-readable dictionary (MRD) if appropriate, or they were formulated by the annotators and later checked by a professional lexicographer.

The annotation of the common words in the treebank was done in the following order:

- The lemma of the existing synsets or definitions in MRD were mapped to the lemma of the common word in the treebank. The sentences were grouped by the lemmas of the common words. In this way the annotator has access to all occurrences of the same lemma in the treebank. If necessary, they were able to consult the whole text in the treebank.
- The annotator could select one of the synsets (if there is appropriate) or one of the definitions (if there is appropriate).

<sup>3</sup><http://www.bultreebank.org/dpbtb/>

<sup>4</sup><http://universaldependencies.org/>

<sup>5</sup><http://wordnetcode.princeton.edu/standoff-files/core-wordnet.txt>

- If there was no appropriate synset or definition from the MRD to describe the common word in the treebank, the annotator introduced a new definition.
- In case when a common word was annotated with a definition that has not been included into a synset yet, the annotator defined a mapping to the Princeton WordNet (see below for an explanation of the mapping relations).
- The annotator detected the other usages of the same common word in the treebank, and annotated them, too.
- The resulting annotations, together with the definitions and mappings to the Princeton WordNet have been transferred to the lexicographer who formed a new synset in the BTB-WN.
- When the new synset was added to the BTB-WN, its internal identifier was returned to the sense annotation of the corresponding common words in the treebank. In this way, any further editing of the synset definition would not destroy the annotation.

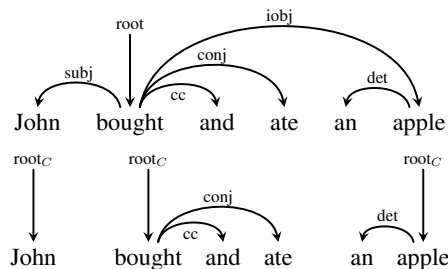


Figure 1: A complete dependency tree and some of its catenae.

In the process of annotating the common words in the treebank, the annotators analyzed also the MWEs. We rely on the catena approach for the description of MWEs in the lexicon and the treebank — (Osenova and Simov, 2015). The catena could be a word or an arbitrary subtree. One example is given in Fig. 1. Thus, the annotators determined the usage of a given MWEs by selecting the nodes of the corresponding catena. In the treebank we assume that there are two types of MWEs: those that obey the rules of Bulgarian syntax, so they were represented as catena in the syntactic trees; and those that do not obey the rules of syntax, so they were represented as complex lexical items. For example, “in order to” is considered a MWE from the second type and thus it was represented as a single lexical node in the tree; “kick the bucket” was presented as a syntactic (sub)tree. The nodes for “kick”, “the” and “bucket” were marked as belonging to the same catena. The idiomatic sense was annotated for the head node of the catena. Additionally, all the words in the catena were annotated with their literal meaning. In this way, we allowed for references to the literal meaning of the elements of the MWEs.

The annotation with DBpedia instances was performed as a separate activity. It covered 10 885 named entities —

2877 organizations, 2938 locations, 4195 people, the rest were from other different categories: events, books, others. Unfortunately, the coverage of the Bulgarian DBpedia is rather small. For that reason, the Bulgarian Wikipedia was used for adding the respective links into the data. The Named Entities in the treebank have been already annotated in the original treebank, and also classified as “person”, “location”, “organization”, and “other”. This information was used during the annotation with semantic information. The annotation of the Named Entities in the treebank was done in the following order:

- Gazetteers for the DBpedia instances were created automatically from a dump of Bulgarian DBpedia. We have extracted them by examination of the triples in dump and looking for the corresponding classes from DBpedia ontology. One alternative approach can be to load the DBpedia dump to a triple store and using SPARQL queries to extract the corresponding names.
- All the Named Entities in the treebank that matched gazetteers items were annotated with all possible URIs for DBpedia instances. The annotation was done by a regular grammar constructed from the DBpedia gazetteers. In some cases we applied some rules for partial matching of the names. For example, if the text mentions “Washington” annotated already in the treebank as “*person*” then it was annotated with all URIs for people with this name, including “George Washington” and “Denzel Washington”. We do not try automatically to solve some cases on the basis of the whole document, but we rely on the human annotator. Later we have used the idea of one reference by context to check possible mistakes in the annotation.
- The annotator selected an appropriate URI for the given NE. In this selection the annotator could consult Wikipedia page for the corresponding DBpedia instance. Additionally to the URI, we stored a list of all possible classes from the DBpedia ontology. The URI was presented only by its suffix. In such a case, it was relatively easy to access not only the instance from DBpedia, but also the Wikipedia page. For example, Barack Obama was annotated with the URI suffix `Barack.Obama` and the classes `dbo:Person`, `dbo:Politician`, and `dbo:President`.
- In case when there was no appropriate DBpedia instance, the annotator tried to find an appropriate Wikipedia page. If such page existed, then the suffix of its URI was stored in the annotation similarly to the suffix of the DBpedia URI. Additionally, the annotator performed a classification of the NE with respect to the DBpedia ontology. In cases when the annotator could not find a Wikipedia page, the annotation for the URI suffix remained empty, but still the classification with respect to DBpedia was required.

As it was mentioned above, the classes from DBpedia ontology used in the annotation were mapped to the appropriate synsets in the Princeton WordNet manually. Thus, we automatically annotated each NE with one or more synsets

from the WordNet. There exist corresponding Bulgarian synsets in BTB-WN, thus, the annotation is also related to BTB-WN.

During the DBpedia annotation process, the URIs pointed to the full names of the entities, while the text box kept the specific occurrences of the names in the text. Several kinds of challenging situations were encountered: the text provides a metaphoric name for the entity, while the DBpedia link uses its real name (for example, the politician Ahmed Dogan is referred to as Sokola (the Falcon) in many texts, but in DBpedia the link is constructed from his actual name); there is insufficient context for the selection of the correct URI; there is no matching URI in the Bulgarian DBpedia; there is no direct URI mapping to the name, available only under another URI.

There is an additional layer of annotation. It is with valency frames from a verb valency lexicon of Bulgarian — (Osenova et al., 2012). Each verb was first annotated with the appropriate sense in the context. The valency frames were assigned to the senses of the verbs. For some senses there are more than one valency frame. In this case semantic restrictions over the frame elements were used for the selection of the correct frame.

## 5. Comparison of BTB corpus to SemCor, Core WordNet, and GWA Base Concepts

The annotation of a new corpus with senses raises the question how good it is with respect to other similar resources. Here we provide a comparison of BulTreeBank with the English SemCor corpus<sup>6</sup> from the perspective of the overlapped senses. We also compare the representation of two special sets of senses – CoreWordNet<sup>7</sup> and GWA Base Concepts<sup>8</sup>. Table 1 below shows the figures:

<i>Items</i>	<i>SemCor</i>	<i>BTB</i>
Tokens	414,288	256,331
Token Senses	183,913	116,305
Type Senses	24,647	9,492/5,090
CWN	2,970	1,873
GWA-BC	3,529	1,562
CWN and GWA-BC	1,023	682

Table 1: Statistics over the two corpora. The **Tokens** are all tokens in the corpora; **Token senses** are the number of tokens that are annotated with senses; **Type Senses** are the different synsets used in the corpora. In the Bulgarian corpus the second statistics number indicates 5090 synsets mapped to the Princeton WN; **CWN** are the synsets from Core WordNet used in the annotation; **GWA-BC** are the synsets from GWA Base Concepts used in the annotation; **CWN and GWA-BC** are the common synsets from Core WordNet and GWA Base Concept used in the annotation.

We used a set of 4689 synsets from GWA Base Concepts

<sup>6</sup><http://web.eecs.umich.edu/~mihalcea/downloads.html#semcor>

<sup>7</sup><http://wordnetcode.princeton.edu/standoff-files/core-wordnet.txt>

<sup>8</sup><http://globalwordnet.org/gwa-base-concepts/>



and a set of 4997 synsets of most frequently used word senses from Core WordNet. The intersection of both sets is 1502 synsets. We assume that GWA Base Concepts are more representative for the structure of WordNet because of their role in the creation of WordNets. From the statistics above, we could conclude that SemCor corpus is more representative with respect to the structure of the WordNet and less representative with respect to the frequent senses. On the other hand, the BTB corpus is more representative with respect to the frequent senses than with respect to the WordNet structure. The common synsets of SemCor and BTB are 4076. From them 1637 are from CWN and 1465 from GWA-BC. This intersection statistics in our view demonstrates that the BTB corpus is more suitable for training WSD tools in the news domain than SemCor, because it represents in a better way the frequent senses. It however is less appropriate for supporting lexicographical work.

## 6. Application to Knowledge-based Word Sense Disambiguation

An initial application that we addressed using the semantically annotated treebank of Bulgarian, was the area of Knowledge-based Word Sense Disambiguation (WSD). Knowledge-based systems for WSD have proven to be a good alternative to supervised systems, which require large amounts of manually annotated data. Knowledge-based systems require only a knowledge base and no additional corpus-dependent information. An especially popular knowledge-based disambiguation approach has been the use of popular graph-based algorithms known under the name of “Random Walk on Graph” (Agirre et al., 2014). Most approaches exploit variants of the PageRank algorithm (Brin and Page, 2012). Agirre and Soroa (2009) (Agirre and Soroa, 2009) apply a variant of the algorithm to WSD by representing WordNet as a graph in which the synsets are represented as nodes and the relations between them are represented as arcs. The resulting graph is called a *knowledge graph*.

For the experiments with Bulgarian data we have used the mapping from BTB Wordnet for Bulgarian to Princeton Wordnet for English. Several types of correspondences have been attested during the mapping process: full correspondence (one-to-one); partial correspondence (one-to-many or many-to-one); forced connectivity (re-design of Bulgarian definition); common general meaning; resolving metonymies; incorrect and extended correspondences. Here we present only the full and partial correspondences. These are the main mapping relations:

**Full Correspondence.** The ideal case in mapping is when equal concepts are encountered, i.e. the concepts in the two languages map one-to-one. That is, the Bulgarian concept matches the one in Princeton Wordnet.

For example, the Bulgarian “sigurnost” and English “certainty” that both mean *lack of danger*.

If a Bulgarian definition corresponds equally well to more than one definition in Wordnet, then all these definitions are mapped to the Bulgarian one, using a special separator. For example, English “answer” and “response” map to Bulgarian “otgovor”.

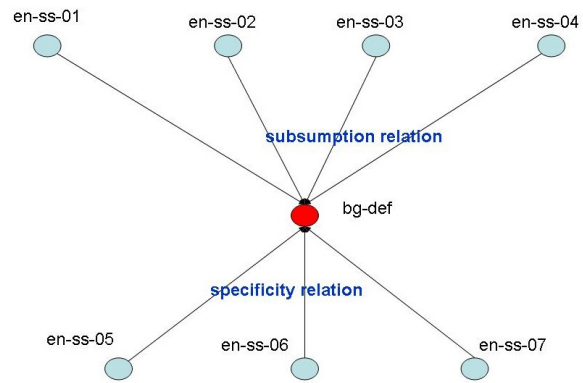


Figure 2: Classification of a Bulgarian definition with respect to English synsets in WordNet hierarchy. We use the relation *subsumption* to map Bulgarian concept (definition) to more general synsets in English WordNet, and the relation *specificity* to map it to more specific English synsets.

**Partial Correspondence.** In many cases, however, the concepts differ in terms of specificity in both language directions.

In the first case, the Bulgarian definition is more specific than the English one. In this case, it is mapped to a more general English one, but it is also marked with a specificity label. The most frequent cases here are the following ones: (i) regular polysemy — for example, in Bulgarian “prokuratura” (prosecutor’s office) is given as an institution and a building, while in English it is the institution, the group of people and the act; (ii) restrictive in Bulgarian vs. general in English definitions — for example, ‘direktsiya’ in the meaning of ‘director’s office’ is mapped to the more general concept ‘office’ with the meaning of ‘place of business’.

A second scenario is possible, where the Bulgarian definition is more general and subsumes one or more synonym sets from Wordnet. In this case, the following approach is taken – the common definition in Bulgarian is mapped once to the more specific English definitions (with relation *specificity*) and a second time to their hypernyms (with relation *subsumption*).

For instance, in Bulgarian ‘rezhisyor’ (director) has only one definition: The lead person in the making of a theater play, film, TV program, etc. However, in Wordnet there exist two synsets that can be related to it: director as someone who supervises the actors and directs the action in the production of a show (with a hypernym ‘supervisor’ as one who supervises or has charge and direction of) and director as the person who directs the making of a film (with a hypernym ‘film maker’ as a producer of motion pictures). In order to preserve both — the more abstract concept in Bulgarian as well as the hierarchical structure of WordNet — the Bulgarian definition is mapped to all four of these synsets with relation *specificity* to the specific ones, and with relation *subsumption* to their hypernyms. These mappings are presented in Fig. 2. Some more considerations are presented below.

**Ensuring a One-to-One Mapping.** In some cases of mismatch the one-to-one mapping can be achieved through

re-working the Bulgarian definitions. This often means dividing the Bulgarian definition into two separate ones. For example the word “sedmitsa” (week) has the following definition: seven consecutive days, usually counted from Monday to Sunday. All examples correspond to this definition. There are two synsets in English: “week” as any period of seven consecutive days, and “week” as a period of seven consecutive days starting on Sunday.

Such a division in nouns referring to the passing of time has been done in Bulgarian for the concept of ‘month’, so it can be implemented for the ‘week’ as well. Because the Bulgarian definition has been mapped to the second synset in English, it can remain as it is, while a second definition is introduced (Seven consecutive days), which is mapped to the first synset; the examples are correspondingly divided between the two definitions.

We are able to use the mapping between the two Wordnets as a knowledge graph for UKB system<sup>9</sup>. In the knowledge graph the nodes come from both Wordnets — nodes corresponding to synsets in Princeton Wordnet as well as nodes corresponding to synsets in BTB Wordnet. The arcs between nodes in the graph correspond to the relations in Princeton Wordnet and the mapping relations between the two wordnets. In this way, we are using the larger knowledge graph for English extended with Bulgarian nodes and relations for application to Bulgarian WSD.

Knowledge graph constructed on the basis of WordNet represents predominantly paradigmatic relations between the nodes. The intuition exploited in the current experiments on Bulgarian is that adding syntagmatic relations improves the disambiguation task.

We consider the semantically annotated treebank as a possible source of syntagmatic relations. In our experiments we exploited dependency relations from the Universal Dependency representation of the treebank. The main relations that we used in the experiments are *nsubj*, *nmod*, *amod*, *iobj*, *dojb* like in the examples: “the boy read the book”; “the man broke the vase with the hammer”; “the tall person”; “the woman with the hat”. From such examples we have extracted the following relations: [boy]-[read]<sup>10</sup>; [read]-[book]; [man]-[break]; [break]-[vase]; [break]-[hammer]; [tall]-[person]; [woman]-[hat]. Thus, using the combination of syntactic and semantic annotation in the treebank we added syntagmatic relations to the knowledge graph.

The next step to extend the set of syntagmatic relations was to apply inference using hyperonymy relations in WordNet. The inference is motivated by the fact that the syntagmatic relations represent mainly the relation between a participant and an event (state) or between two participants in the same event (state). Then if a noun synset represents a participant in an event, then each of its hyponymy synsets also represents a participant in the event (state). Similarly, a verbal synset can be substituted by its hyperonymy synsets. Using such inference we have inferred relations like the following from the examples above: [read]-[textbook]; [woman]-[bonnet], etc.

<sup>9</sup><http://ixa2.si.ehu.es/ukb/>

<sup>10</sup>With [] we denote the node in the knowledge graph that corresponds to the appropriate synsets.

In our experiments (Simov et al., 2015) with the system UKB for Knowledge-based WSD we have achieved more than 10% improvement of the accuracy. The knowledge graph used in these experiments is the knowledge graph constructed on the basis of Princeton WordNet and extended WordNet which is distributed with the UKB system and BTB Wordnet as described above. It was expanded with relations extracted from semantically annotated treebank.

This application demonstrates the usefulness of the semantic annotation over treebanks. Our future plans are to extend the set of extracted syntagmatic relations using paths in the dependency trees longer than one arc. For example, we can exploit path of *nsubj* and *dojb* to extract a relation such as [boy]-[book]. Of course there exist also other applications of such semantically and syntactically annotated resources.

## 7. Conclusions

In this paper we presented the methodology behind cross-level semantic annotation in BulTreeBank. The scheme relies on the annotation of verb valency frames, sense annotation of four parts-of-speech (nouns, verbs, adjectives, adverbs) as well as DBpedia URIs and classes.

Our efforts reported in the paper are similar to the sense annotation task performed by (Fellbaum et al., 1998) for English, but with some differences, such as: the annotation of the corpus and the creation of the BTB-WordNet have been performed simultaneously; the annotation was performed on a treebank, which provided the facility of using a derived valence lexicon; no confidence markers have been used by the human annotators – the superannotation technique and cross-resource mappings were adopted as quality assurance strategies instead; DBpedia data have been added.

In future, we also envisage the following extensions of the annotation: (1) Coreference annotation: currently the treebank is annotated with coreference chains within sentences. We would like to extend them on intersentential level; (2) Addition of semantic role labels: using the verbal valency frames we aim at mappings from the valency lexicon to the treebank; (3) Annotation of the internal structure of the NEs. Currently, the whole named entities are annotated, but many of them have internal structure which can be annotated. For example, in “The Bank of England” we could annotate ‘Bank’ with the appropriate synset from WordNet and ‘England’ with DBpedia information; (4) Logical structure of the sentences. We plan to add Minimal Recursive Semantic annotation over the syntactic annotation and the valency frames.

## 8. Acknowledgements

This research has received partial funding from the EC’s FP7 (FP7/2007-2013) under grant agreement number 610516: “QTLeap: Quality Translation by Deep Language Engineering Approaches”.

We are grateful to the three anonymous reviewers, whose remarks, comments, suggestions and encouragement helped us to improve the initial variant of the paper. All errors remain our own responsibility.

## 9. Bibliographical References

- Agirre, E. and Soroa, A. (2009). Personalizing PageRank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 33–41, Athens, Greece, March. Association for Computational Linguistics.
- Agirre, E., López de Lacalle, O., and Soroa, A. (2014). Random walks for knowledge-based word sense disambiguation. *Comput. Linguist.*, 40(1):57–84, March.
- Brin, S. and Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833.
- Chaney, A., Simov, K., Osenova, P., and Marinov, S. (2006). Dependency conversion and parsing of the BulTreeBank. In *Proceedings of the LREC workshop Merging and Layering Linguistic Information*, pages 16–23.
- Chaney, A., Simov, K., Osenova, P., and Marinov, S. (2007). The BulTreeBank: Parsing and Conversion. In *Proceedings of the Recent Advances in Natural Language Processing Conference*, pages 114–120.
- Fellbaum, C., Grabowski, J., and Landes, S., (1998). *Performance and Confidence in a Semantic Annotation Task*. MIT Press.
- Ghayoomi, M., Simov, K., and Osenova, P. (2014). Constituency parsing of Bulgarian: Word- vs class-based parsing. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- Henrich, V. and Hinrichs, E., (2013). *Language Processing and Knowledge in the Web: 25th International Conference, GSCL 2013, Darmstadt, Germany, September 25-27, 2013. Proceedings*, chapter Extending the TüBa-D/Z Treebank with GermaNet Sense Annotation, pages 89–96. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., and Grishman, R. (2004). The nombank project: An interim report. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, Massachusetts, USA, May 2 - May 7. Association for Computational Linguistics.
- Montemagni, S., Barsotti, F., Battista, M., Calzolari, N., O. Corazzari, Zampolli, A., Fanciulli, F., Massetani, M., Raffaelli, R., Basili, R., Pazienza, M. T., Saracino, D., Zanzotto, F., Mana, N., Pianesi, F., and Delmonte, R. (2000). The italian syntactic-semantic treebank: Architecture, annotation, tools and evaluation. In *Proceedings of the COLING-2000 Workshop on Linguistically Interpreted Corpora*, pages 18–27, Centre Universitaire, Luxembourg, August. International Committee on Computational Linguistics.
- Osenova, P. and Simov, K. (2015). Modeling lexicon-syntax interaction with catenae. 16(3):287–322.
- Osenova, P., Simov, K., Laskova, L., and Kancheva, S. (2012). A treebank-driven creation of an ontovallence verb lexicon for Bulgarian. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 2636–2640, Istanbul, Turkey. LREC 2012.
- Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106, March.
- Philpot, A., Hovy, E., and Pantel, P. (2005). The omega ontology. In *IJCNLP workshop on Ontologies and Lexical Resources*, pages 59–66.
- Rosa, R., Mašek, J., Mareček, D., Popel, M., Zeman, D., and Žabokrtský, Z. (2014). Hamledt 2.0: Thirty Dependency Treebanks Stanfordized. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- Simov, K. and Osenova, P. (2003). Practical annotation scheme for an HPSG treebank of Bulgarian. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-2003)*, Budapest, Hungary.
- Simov, K., Osenova, P., Simov, A., and Kouylekov, M. (2004). Design and implementation of the Bulgarian HPSG-based treebank. In *Journal of Research on Language and Computation, Special Issue*, pages 495–522, Kluwer Academic Publishers.
- Simov, K., Popov, A., and Osenova, P. (2015). Improving word sense disambiguation with linguistic knowledge from a sense annotated treebank. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 596–603.

# Text-Picture Relations in Cooking Instructions

I. van der Sluis, S. Leito and G. Redeker

Center for Language and Cognition  
University of Groningen

E-mail: I.F.van.der.Sluis@rug.nl, S.L.Leito@alumnus.rug.nl, G. Redeker@rug.nl

## Abstract

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.

**Keywords:** multimodal instruction, annotation, evaluation

## 1. Introduction, Background and Research Questions

Multimodal information presentation is ubiquitous in modern life. On any given day, we come across newspapers, advertisements, webpages, instructions etc. in which pictures and text are combined to send us a message. Making documents multimodal, that is, using pictures (photographs, diagrams etc.) as well as text, is one significant way to improve clarity (Bateman, 2015; Mayer, 2005; Schriver, 1997). However, the effectiveness of multimodal presentations critically depends on how well the reader can relate the content of the text to the information provided by the pictures. In this paper we present a study that aims to investigate this link in multimodal instructions. Instructions are omnipresent in our lives. They guide us through processes like using an appliance, constructing or repairing an object, or preparing a meal. Instructions often consist of a series of descriptions and depictions of the actions to be carried out, as is the case in recipes on the back of packages with ready-to-use ingredients for the preparation of a dish. We have developed an annotation system to analyse how the textual and pictorial modalities are combined in such process-oriented instructions.

The procedural motor actions presented in multimodal instructions (MIs) have been investigated by e.g., Dechsri et al., 1997; Michas & Berry, 2000; Van Hooijdonk & Kraemer, 2008; Iserbyt et al., 2012. Building on various models of how people process multimodal information (i.e., Mayer, 2005; Schnotz, 2005; Chandler, 2004; Chandler & Sweller, 1991), it has been shown that in performing these actions users benefit from compact multimodal information presentations in which information is conveyed through a transparent and concise distribution between modalities (Larkin & Simon, 1987; Marcus, Cooper & Sweller, 1996; McCrudden et al., 2007). Other studies suggest that the pictures are interpreted based on the information offered in the text (Hegarty & Just, 1993) and that referential cohesion between text and pictures facilitates an accurate and

efficient performance of the instructed actions (Dupont & Bestgen, 2006). This paper presents a corpus study that investigates how authors of MIs make use of compactness and cohesion in procedural instructions. We expect the multimodal presentations in our corpus to be cohesive as well as compact, because both support easy processing and use of textual and pictorial information.

Our ultimate aim is the development of evidence-based authoring guidelines for combining text and pictures in MIs. Currently, authoring seems to be based on intuitive notions (cf. Kaltenbacher, 2004), while questions about the preferred type of pictures and preferred relations between pictures and text are unanswered. To determine how authors compose MIs in terms of compactness and cohesion, we collected a corpus of multimodal cooking instructions for semi-prefabricated meals in which pictures appear in various forms: photographs, drawings, pictograms, and so forth. As in other reading-to-do texts (i.e. texts for achieving a particular goal, cf. Sticht, 1985), the pictures are often representational in that they depict parts of the accompanying text (see Levie & Lentz, 1982 for an overview of relationships between text and pictures). They may depict objects referred to in the text, but may also visualise how particular actions can be carried out. These affordances can be employed in various ways.

For instance, in the cooking domain, ‘slicing’ can be visualised by depicting a knife and (pieces of) a particular ingredient on a cutting board; but a picture may also show a pair of human hands actually slicing the ingredient with a knife on the board, as in MI I (Figure 1). We present an annotation scheme in which a categorisation of depicted actions can be matched with the actions verbalised in the MI text. This allows us to describe the compactness of an MI in terms of its distribution of instructional information between modalities as well as the MI’s cohesion between its text and pictures.

## 2. Corpus Study

### 2.1. Data

A corpus of 30 Dutch MIs for semi-prefabricated meals from various brands was collected to investigate how text and pictures are used in cooking instructions. The MIs appear on the packages and describe the preparation of the meal as a stepwise procedure with a series of text-picture combinations that are spatially aligned as in MI I (Figure 1) or linked by numbering as in MI II (Figure 2).

### 2.2. Corpus Analysis

The corpus was analysed for textual and pictorial representations of the actions involved in preparing the dishes. In the text the actions were determined by classifying the action verbs, while the notion of affordance as explained above, was used to analyse the pictures. In line with the exploratory purpose of the corpus analysis, the annotation categories were derived inductively from the MIs. This process entailed several rounds of meticulous analysis and annotation of the corpus by the second author, honed in intensive discussions of all three authors. When consensus on the annotation scheme had been reached, the second author's application of that scheme was checked by the other two authors individually. The remaining differences were satisfactorily resolved through further discussions. Four main action types each with several subtypes (see Table 1) were distinguished:

- Process*: an operation is performed on an ingredient, which changes the form of the ingredient (e.g., sliced into parts, mixed with something else);
- Heat*: an ingredient is heated in a particular way (e.g., cook, steam, stew);
- Put*: a (part of a) dish is placed in a particular space to heat or cool or to be processed further;
- Other*: other actions that involve actions (e.g., repeat an action) or processes (e.g., end a process).

Note that the categories and their subtypes are based on the MIs in the corpus. This explains the fact that the action type 'Other' does not have a subtype 'Start' to match the subtype 'Finish'.

#### 2.2.1. Text Analysis

A step can consist of multiple actions that can be recognised by the use of action verbs. Sometimes multiple instances of an action can be described with elliptic verb phrases. For instance, 'Slice the broccoli in small florets, the onion in half rings, and the pepper in stripes' (MI I), specifies three slicing actions. Only actions that constituted the core of the cooking instruction were annotated. For the sake of feasibility, conditional actions

(e.g., 'Add 400 ml milk and 250 ml water if you use fresh vegetables', MI 16) and clarifications and warnings that were not immediately consequential for the cooking process (e.g., 'Note: The closure clip is folded into the roasting bag', MI 12) were excluded from the analysis. The analysis of the actions focused on the contextualised meaning of the verbs, which may differ from their lexical meaning, for example, 'put water on' (MI 2) means to heat, not to place (no location is specified). To keep the annotation system manageable, the verb meanings were classified only with respect to the action they involve, not any effect or purpose. For instance, 'Put the Wraps in the oven for 5-10 minutes' (MI 14) is a 'Put' action, even though it implies heating; 'marinate' (MI 4) means to bring some ingredient in contact with herbs and spices and is thus classified as a 'Mix' action. 'Finish with a layer of sauce' (MI 16, Table 1) is a 'Other' action and not a 'Put' action. Similarly, 'mix the contents of the sachet 'tempura flour' with 100 ml of cold water' (MI I) is classified as 'Add' while 'Stir with a whisk to smooth batter' (MI I) is classified as 'Mix'.

#### 2.2.2. Picture Analysis

The MI corpus contains various types of pictures, such as photographs (n = 6) as in MI I (Figure 1) and drawings (n = 24) as in MI II (Figure 2) (a larger corpus would be needed to investigate any differences between those types of pictures). The pictures are categorised in terms of the affordances they visualise, which largely correspond to the categories of the verbs in the text. Of the actions presented in Table 1, four are not represented in the pictures in the corpus: 'Separate', 'Steam', 'Put somewhere for cooling' and 'Finish'. A picture can depict multiple actions. For instance in step 1 of MI I, three actions can be identified: (1) slice onion, (2) slice broccoli and (3) slice pepper. Some elements depicted in the pictures were not included in the analysis, because they did not visualise an action or were not immediately consequential for the cooking process: (1) numbering of the pictures, (2) indications of temperature, duration or quantity (e.g., '200<sup>0</sup> C' in step 1, '1 min. clock' in step 3 MI II), (3) additional measuring cups and other utensils in the background of the picture (MI 16), (4) conditional actions as depicted in step 4 in MI II, where a spicy and a less spicy variation of the wrap are presented for respectively adults and children, and (5) clarifications and warnings (e.g., exclamation marks to signal danger of burns, as in MI 12). The picture analysis focused on the contextual meaning of the picture elements. For example, a cutting board depicted above a pan with some sliced ingredients being shoved from the board by a knife was categorised as 'Add' instead of 'Slice'.




MI I: Text and pictures from package	Translated MI text
<p><b>Bereidingswijze:</b></p>  <p>1. <b>Breng een pan met ruim water aan de kook. Snij de broccoli in kleine roosjes, de ui in halve ringen en de paprika in reepjes. Snij vervolgens de kipfilet in dunne, grote plakken.</b></p> <p>2. <b>Doel de inhoud van het zakje 'sausmix' samen met 3 el ketchup en 3 el water in een kom en meng het goed. Zet de dipsaus in de koelkast tot gebruik.</b></p> <p>3. <b>Kook de 'rijst' ca. 15 min. in het kokende water en giet het daarna af. Meng ondertussen in een andere kom de inhoud van het zakje 'tempurameel' met 100 ml koud water en roer het met een garde tot een glad beslag.</b></p> <p>4. <b>Verhit een koekenpan met 4 el olie. Zorg ervoor dat de olie goed heet wordt. Haal de plakken kipfilet door het tempurabeslag, zorg dat ze rondom bedekt zijn met het beslag. Bak ze direct in de hete olie in ca. 5 min. per kant goudbruin en gaar.</b></p> <p>5. <b>Verhit ondertussen 2 el zonnebloemolie in een andere koekenpan of wok en roerbak hierin de gesneden groenten ca. 5 min. Voeg als laatste de inhoud van het zakje 'kruidenpasta' en 150 ml water hieraan toe en roer het goed door. Verwarm het geheel nog even goed.</b></p> <p>6. <b>Serveer de tempura kip met de Oosterse dipsaus, de roergebakken groenten en de rijst apart. Eet smakelijk!</b></p>	<ol style="list-style-type: none"> <li>Bring a pot of water to the boil. Slide the broccoli in small florets, the onion in half rings, and the pepper in slices. Then slice the chicken breast in large, thin slices.</li> <li>Put the contents of the sachet 'sausmix' in a bowl together with 3 tbsp of ketchup and 3 tbsp of water and mix well. Put the sauce in the fridge until use.</li> <li>Cook the 'rice' for about 15 minutes in the boiling water and then drain. Meanwhile, in another bowl, mix the contents of the sachet 'tempura flour' with 100 ml of cold water and stir with a whisk to smooth batter.</li> <li>Heat a frying pan with 4 tbsp of oil. Make sure the oil is very hot. Dip the chicken breast slices in the tempura batter, make sure they are covered all around with the batter.</li> <li>Meanwhile, heat 2 tbsp of sunflower oil in another frying pan or wok and stir-fry the vegetables for about 5 minutes. Add the contents of the sachet 'spice paste' to this and stir well. Heat everything well.</li> <li>Serve the tempura chicken with the oriental dip, the stir-fried vegetables and the rice separately. Enjoy!</li> </ol>

Figure 1: MI I 'Knorr Wereld Gerechten Krokante Specials, Tempura Kip' ('Knorr World Meals Crispy Specials, Tempura Chicken')


MI II: Text and pictures from package	Translated MI text <sup>1</sup>
 <p><b>ZELF TOEVOEGEN</b> Voor 4 personen 300 g kipfilet, in kleine blokjes 1 rode paprika, in kleine stukjes 1 ui, fijngesnipperd 1 courgette (300 g), in kleine blokjes 1/2 bekertje crème fraîche (60 ml) 150 ml water olie of boter</p> <p><b>ALGEMENE BEREIDINGSWIJZE</b></p> <ol style="list-style-type: none"> <li>Verwarm de oven voor op 200°C. Tip: verwarmen in de oven is niet nodig als je de Wraps zelf voor het vullen al even apart verwarmt.</li> <li>Verhit de olie of boter in een pan en bak hierin de kipblokjes goudbruin en gaar. Voeg de paprika, ui en courgette toe en bak deze nog even mee.</li> <li>Voeg 150 ml water, 60 ml crème fraîche en de Indiase Kruidenmix toe en verhit het geheel terwijl je omscheept nog ca. 1 minuut.</li> <li>Vul voor de milde eters alvast een aantal Wraps met het kipmengsel en rol ze op. Leg de gevulde Wraps in een ingevette ovenschaal.</li> <li>Meng de Indiase Spicemix door het resterende kipmengsel en warm het geheel goed door. Vul hiermee de overige Wraps, rol deze op en leg ze ook in de ingevette ovenschaal.</li> <li>Plaats de Wraps 5-10 minuten in de oven.</li> </ol> <p><b>DOE MEER. VARIËR</b> 1. Vervang de paprika 2. Vervang de crème</p>	<ol style="list-style-type: none"> <li>Preheat the oven to 200<sup>o</sup> C. [Tip: Preheating in the oven is not necessary if you warm the Wraps themselves separately before filling them.]</li> <li>Heat the oil or butter in a pan and fry the chicken cubes in it until golden brown. Add the pepper, onion and zucchini, and fry them a short time.</li> <li>Add 150 ml water, 60 ml crème fraîche and the Indian Spicemix and heat the whole for about 1 minute while stirring.</li> <li>For the mild eaters, fill several Wraps with the chicken mixture now and roll them up. Place the stuffed Wraps in a greased baking dish.</li> <li>Mix the Indian Spicemix with the remaining chicken mixture and heat it all well. Use it to fill the remaining Wraps, roll them up, and put them also into the greased baking dish.</li> <li>Place the Wraps in the oven for 5-10 minutes.</li> </ol> <p><sup>1</sup> Text parts excluded from our analysis are marked with text brackets.</p>

Figure 2: MI II 'Honig Familiegerecht Indiase wraps' ('Honig Family Dinner Indian Wraps')

Table 1: MI Action Types, Action Subtypes and Examples

Action Types	Action Subtypes	Examples
<b>Process</b>	1.1 Slice	1.1 Snijd de prei in ringen. ( <i>Cut the leek into rings.</i> ) (MI 10)
	1.2 Separate	1.2 Laat vervolgens goed uitlekken. ( <i>Then drain well.</i> ) (MI 10)
	1.3 Mix	1.3 Roer alles goed door elkaar. ( <i>Thoroughly stir everything together.</i> ) (MI 28)
	1.4 Other	1.4 Rol ze op. ( <i>Roll them up.</i> ) (MI 14)
<b>Heat</b>	2.1 Cook	2.1 Laat het 12 min. zachtjes doorkoken. ( <i>Let it boil gently for 12 min.</i> ) (MI 2)
	2.2 Roast	2.2 Braad hierin het vlees aan. ( <i>Roast the meat herein.</i> ) (MI 1)
	2.3 Bake	2.3 Bak de uien en champignons enkele minuten mee. ( <i>Fry the onions and mushrooms a few minutes.</i> ) (MI 1)
	2.4 Steam	2.4 Laat het geheel vervolgens ca. 5 min. met een deksel op de pan zachtjes stomen. ( <i>Then let it steam gently for about 5 minutes with a lid on the pan.</i> ) (MI 13)
	2.5 Stew	2.5 Laat het geheel op laag vuur met de deksel op de pan ongeveer 1,5 uur stoven. ( <i>Let the mixture simmer for about 1.5 hours on low heat with the lid on the pan.</i> ) (MI 1)
	2.6 Heat a space	2.6 Verwarm de oven voor op 200° C, hete lucht op 180° C. ( <i>Preheat the oven to 200° C, hot air 180° C.</i> ) (MI 13)
<b>Put</b>	3.1 Add	3.1 Voeg tenslotte nog eens 250 ml water en de inhoud van dit zakje toe. ( <i>Finally, add another 250 ml of water and the contents of this sachet.</i> ) (MI 1)
	3.2 Put somewhere for heating	3.2 Plaats de Wraps 5-10 minuten in de oven. ( <i>Put the Wraps in the oven for 5-10 minutes.</i> ) (MI 14)
	3.3 Put somewhere for cooling	3.3 Laat de taart vervolgens op een rooster afkoelen. ( <i>Then let the cake cool on the grid.</i> ) (MI 24)
	3.4 Put somewhere (no purpose given)	3.4 Serveer de risotto direct met de salade apart. ( <i>Serve the risotto immediately with the salad separately.</i> ) (MI 28)
<b>Other</b>	4.1 Repeat	4.1 Herhaal dit tot de saus op is. ( <i>Repeat until there is no more sauce.</i> ) (MI 16)
	4.2 Finish	4.2 Eindig met een laagje saus. ( <i>Finish with a layer of sauce.</i> ) (MI 16)

Table 2: Representation of actions in text and pictures (counts and percentages)

Action Types	Text & Picture		Text only		Picture only		Total	
	N	%	N	%	N	%	N	%
Process	46	27.5	83	32.7	10	32.3	139	30.8
Heat	52	31.1	65	25.6	7	22.6	124	27.4
Put	81	48.5	104	40.9	11	35.5	196	43.4
Other	0	0.0	2	0.8	3	9.7	5	1.1
Total	167	*	254	100	31	100	452	*

### 2.2.2. Text and Picture Analysis

The joint analysis of text and pictures shows which actions were presented in both text and picture, in text only, or in a picture only. This allows us to quantify the compactness of an MI in terms of the distribution of information between the modalities and to designate the cohesion between the MI's text and pictures. The representation of an action is *compact* when it occurs only in the text or only in a picture. (e.g., the text in step 1 in MI I tells the user to heat water and slice the chicken, which is not shown in the picture). The representation is *cohesive* if an action is presented in the text as well as in the picture (e.g., in step 1 in MI I the text and picture guide the user how to slice the vegetables). In some cases, text and picture present different aspects of the same action, for example, the verb 'Mix' in 'Mix the Indian spice mix with the remainder of the chicken mixture' (action type 'Mix') is combined with a picture that shows the spices falling from a sachet into the pan (action type 'Put'), and for 'Put the Wraps in the oven for 5-10 minutes' (type 'Put') the picture shows the dish in the oven (type 'Heat') (both examples are from MI II). These cases are coded as cohesive.

## 3. Corpus Results

We identified a total of 452 actions in the 30 MIs in our corpus. Table 2 shows the frequencies of textual and pictorial representations of the four main action types. Overall, text-only representations dominated with 254 actions (56%); 167 of the actions (37%) were described in the text and also presented in the pictures; and 31 actions (7%) only occurred in the pictures. In 12 of the 167 actions that were described as well as depicted, the categories for text and picture differed, because different aspects of the same action were presented. For example, the text in step 5 in MI II (see Figure 2) was annotated as 'Process' (i.e., 'Blend the Indian spice mix with the remainder of the chicken breast'), while the picture shows a 'Put' action, namely 'add the spice to the pan'. Similarly, in step 6 of MI II, the text 'place the wraps for heating in an oven' was annotated as a 'Put' action, while the picture shows a 'Heat' action ('roast the dish in a closed space'). Strictly referential cohesion (identical coding) was found in 155 of 167 actions.

The most frequent actions are 'Put' actions (43.4%). They account for 48.5% of the actions that are described as well as depicted, 40.9% of the text-only actions, but only 35.5% of the actions that are only depicted. When multiple ingredients are added (multiple 'Put' actions), they are often not all shown in the accompanying picture. 'Process' actions amount to 30.8% of all actions, but only to 27.5% of the actions that are described as well as depicted. The actions that are depicted but not described typically involve 'Process' actions that show, for instance, an ingredient sliced into pieces, while this is not mentioned in the text. Conversely, before performing the depicted 'Process' action 'Mix', ingredients need to be put in a pan or bowl, which is often only described and not

depicted. 'Heat' actions (total: 27.5%) are more often represented in text and picture or in text only than in picture only. This may be due to the fact that pictograms to indicate heat level or duration were not included in our analysis. Overall, there were 285 cases (63%) where an action was presented only in the text or only in a picture, indicating a moderate level of cohesion.

## 4. Conclusions

In this paper we have analysed the actions presented in multimodal procedural instructions to discover cohesion and compactness relations between the text and pictures of which they are composed. Our hypothesis was that the presentation of the multimodal instructions in our corpus is cohesive as well as compact, because both can facilitate understanding and use by cooks. However, our corpus analysis has shown that authors make only moderate use of cohesion and compactness between text and pictures. They present more than half of the actions only in text, while very few occur only in pictures. Authors thus seem to rely much more on the textual modality, using pictorial representations preferably in combination with text, if at all. The explanation our analysis suggests is that the textual information is more detailed and explicit, while the pictures often require more contextual inferencing. Note that this does not mean that the pictures are less helpful for the user of the instruction. In a pilot experiment (Leito et al., 2014), cooks who were given the pictures and the option to read the accompanying text, performed as well as cooks who were given text and pictures, even though they never turned the instruction sheets to read the text.

In our study we identified cohesion and compactness relations through the actions a user needs to carry out. Arguably, cohesion also results from object references (Dupont & Bestgen, 2006; Hegarty & Just, 1993), where optimal cohesion may be reached if all objects that are depicted are also referred to in the text and vice versa. Similarly, compactness may result from equality relationships and logical expansions of meaning (cf. Bateman, 2015) that may extend the user's understanding of the action instructed in the text. For instance, in our domain, the text could instruct to 'slice the chicken breast in large, thin slices', while the accompanying picture shows how the relative terms 'large' and 'thin' actually should be interpreted.

To validate our initial results and to enable further insights on relations between textual and pictorial information in the design of MIs, future work should include an evaluation of the annotation scheme to investigate inter-annotator agreement. In addition crowdsourcing experiments, using simplified annotation tasks, could be employed to investigate the structure of our annotation scheme. For instance, the action 'Add' is now modelled as a subcategory of 'Put' (as the focus is on placement, not on the way this affects the ingredients), but it might also be interpreted as a subcategory of 'Process'. Also we proposed mutual exclusiveness of the categories 'Heat' and 'Put', where 'Put somewhere for heating' is a



subcategory of ‘Put’, but not of ‘Heat’. Further consolidation of the annotation scheme should make use of lexical ontologies like WordNet (cf. Gangemi et al., 2010; Niles & Pease 2003).

The MI corpus presented in this paper (available on request) can be used for further investigations on readers’ and users’ preferences and their processing of multimodal procedural instructions defined in terms of actions. For instance, cooks may find the text too long, or find themselves too busy with the task itself to read it and may miss important information if they cook solely based on the pictures. In two questionnaire studies and an experiment (Leito et al., 2014), we investigated how people judge and use such recipes and whether using instead of just reading the recipe affects the way they judge its quality. Results show that preparing a dish based on only the pictures created fewer problems than anticipated. Moreover, when a recipe had been used for cooking, its assessment tended to focus on performance, while a recipe that was only read was judged in terms of document qualities. In general, participants in our studies preferred photographs over drawings, but suggested to supplement photographs with pictograms specifying amounts and/or durations.

For the further development of authoring guidelines for combining text and pictures in MIs, the analysis of compactness and cohesion should be extended to larger corpora with various kinds of cooking instructions, including e.g. variations in the complexity of the meal preparation and in the intended users. Importantly, the investigation should be extended to dynamic instructions as presented e.g. on YouTube or the cooking tasks in the Saarbrücken Corpus of Textually Annotated Cooking Scenes (TACOS) (cf. Regneri et al., 2013).

## 5. Bibliographical References

- Bateman, J. (2014). Multimodal coherence research and its applications. In: *The Pragmatics of Discourse Coherence. Theories and Applications*. Edited by H. Gruber and G. Redeker. [Pragmatics & Beyond New Series, Vol. 254]. 145-177. Amsterdam: Benjamins.
- Chandler, P. (2004). The crucial role of cognitive processes in the design of dynamic visualizations. *Learning and Instruction, 14*, 353-357.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition & Instruction, 8*(4), 293-332.
- Dechsri, P., Jones, L. L., & Heikkinen, H. W. (1997). Effect of a laboratory manual design incorporating visual information-processing aids on student learning and attitudes. *Journal of Research on Science and Teaching, 34*, 891-904.
- Dupont, V., & Bestgen, Y. (2006). Learning from technical documents: The role of intermodal referring expressions. *Human Factors, 48*(2), 257-264.
- Gangemi A., Guarino N., Masolo C., O. A. *Interfacing WordNet with DOLCE: towards OntoWordNet*. Cambridge: Cambridge University Press, 2010.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*(6), 717-742.
- Iserbyt, P., Mols, L., Elen, J., & Behets, D. (2012). Multimedia design principles in the psychomotor domain: The effect of multimedia and spatial contiguity on students’ learning of basic life support with task cards. *Journal of Educational Multimedia and Hypermedia, 21*(2), 111-125.
- Kaltenbacher, M. (2004). Perspectives on multimodality: From the early beginnings to the state of the art. *Information Design Journal + Document Design, 12*(3), 190-207.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*(1), 65-100.
- Leito, S., Redeker, G. & Van der Sluis, I. (2014). The use of text and pictures in cooking instructions. TABU-dag, 12-13 June 2014, University of Groningen.
- Levie, W. H. & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology, 30*(4), 195-233.
- Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*(1), 49-63.
- Mayer, R.E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Red.), *The Cambridge Handbook of Multimedia Learning* (pp. 31-48). Cambridge: Cambridge University Press.
- McCrudden, M. T., Schraw, G., Lehman, S., & Poliquin, A. (2007). The effect of causal diagrams on text learning. *Contemporary Educational Psychology, 32*(3), 367-388.
- Michas, I. C., & Berry, D. C. (2000). Learning a procedural task: Effectiveness of multimedia presentations. *Applied Cognitive Psychology, 14*, 555-575.
- Niles, I. and Pease, A. (2003). Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Procs. of the International Conference on Information and Knowledge Engineering (IKE'03)*, Las Vegas, Nevada, June 23-26, 2003.
- Regneri, M., Rohrbach, M., Wetzels, D., Thater, S., Schiele, B. & Pinkal, M. (2013). Grounding Action Descriptions in Videos. *Transactions of the Association for Computational Linguistics, 1*, 25-36.
- Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Red.), *The Cambridge Handbook of Multimedia Learning* (pp. 49-69). Cambridge: Cambridge University Press.
- Schrifer, K. (1997). *Dynamics in Document Design: Creating Texts for Readers*. New York: Wiley.
- Sticht, T. (1985). Understanding readers and their uses of texts. In T. M. Duffy & R. Waller (Red.), *Designing Usable Texts* (pp. 315-339). London: Academic Press.
- Tversky, B. (2011). Visualizing Thought. *Topics in Cognitive Science, 3*, 499-535.
- Van Hooijdonk, C., & Kraemer, E. (2008). Information modalities for procedural instructions: The influence of text, pictures, and film clips on learning and executing RSI exercises. *IEEE Transactions on Professional Communication, 51*, 50-62.

# An Abstract Syntax for ISOSpace with its <moveLink> Reformulated

Kiyong Lee

Korea University  
Seoul 137-767, Korea  
ikiyong@gmail.com

## Abstract

ISOSpace (ISO 24617-7, 2014) introduces the movement link, tagged <moveLink>, to annotate how motions are related to spatial entities in language. As pointed out in SemAF principles (ISO 24617-6, 2016), ISOSpace overlaps SemAF-SR (ISO 24617-4, 2014) that treats semantic roles in general. It also fails to conform to the *link* structure  $\langle \eta, E, \rho \rangle$  as formulated in Bunt et al. (2016). To resolve these problems, we first construct the general abstract syntax  $\mathcal{ASyn}$  of annotation structures on which the abstract syntax  $\mathcal{ASyn}_{isoSpace}$  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. (2010), 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.

**Keywords:** abstract syntax,  $\mathcal{ASyn}_{isoSpace}$ , complex basic entity, motion, path, event-path, <moveLink>

## 1. Introduction<sup>1</sup>

As part of an ISO international standard on semantic annotation, ISOSpace (ISO 24617-7, 2014) provides an abstract syntax, represented in UML diagrams, two concrete syntaxes, and a set of guidelines (Annex A) for the annotation of spatial entities and motions in language. It specifies:

- (1) a. how to annotate spatial entities such as places, paths, and spatially involving non-locational objects and motions and other non-motion events in language and also
- b. how to annotate and represent their relations in a concrete format, either XML or predicate-logic-like form.

ISOSpace treats spatial (e.g., “in”, “at”, “north of”), motion (e.g., “from” and “to”), and measure (e.g., “5 miles”) signals as its basic entities.

These signals trigger [i] topological, [ii] orientational, [iii] movement, or [iv] measure relations. The topological and orientational relations relate spatial entities, called *figures*, to other spatial entities, called *grounds*.<sup>2</sup> In contrast, the movement relation relates motions to spatial entities, which are often of the type *path*, whereas the measure relation relates spatial measures such as a distance to locations. The standard then lists attributes and their possible values for each of the six basic entity types:

- (2) **place, path, non-locational spatial entity, motion, non-motion event**, and three subtypes of the **signal**.

<sup>1</sup>To appear in H. Bunt(ed.), *Proceedings of isa-12, the 12th Joint ACL-SIGSEM and ISO Workshop on Interoperable Semantic Annotation*, pp.xx-yy. A satellite workshop (W-4) of the 10th Edition of the Language Resources and Evaluation Conference (LREC2016), Portorož, Slovenia.

<sup>2</sup>Talmy (1975) and his subsequent works adopted the gestalt terms *figure* and *ground* to relate some physical object moving or located with respect to its event-path or site: e.g., *The pen<sub>figure</sub> fell off the table<sub>ground</sub> or the pen<sub>figure</sub> lay on the table<sub>ground</sub>*.

It also lists attributes and their possible values for the four kinds of relations, called *links*, that hold between entity structures:

- (3) the **qualitative spatial link** (<qsLink>), the **orientational link** (<oLink>), the **movement link** (<moveLink>), and the **measure link** (<mLink>).

This paper is especially concerned with the movement link (<moveLink>). It is the core of ISOSpace, treating motions and their relations to spatial entities. This link is, however, found structurally different from the other links such as <qsLink> or <oLink>. The current formulation of <moveLink> in ISOSpace (2014) fails to conform to the general structure of link, as formulated in Bunt (2010), SemAF principles (ISO 24617-6, 2016), and Bunt et al. (2016):

- (4)  $\langle \eta, E, \rho \rangle$ ,  
where  $\rho$  is a relation between an entity structure  $\eta$  and a set  $E$  of entity structures.<sup>3</sup>

Furthermore, as is again pointed out in SemAF principles, the task of <moveLink> in ISOSpace overlaps SemAF-SR (ISO 24617-4, 2014) that specifies how to annotate semantic roles. The current version of the movement link (<moveLink>) thus weakens its descriptive power, failing to justify its role as an independent link.

**Proposed Modifications:** To resolve these problems, we first construct the general abstract syntax  $\mathcal{ASyn}$  of annotation structures on which the abstract syntax  $\mathcal{ASyn}_{isoSpace}$  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, which was introduced in the earlier versions of ISOSpace (Pustejovsky et al., 2010). We treat **event-path** as a genuine basic entity type in  $\mathcal{ASyn}_{isoSpace}$ , while implementing it as such into a concrete syntax.

<sup>3</sup>Treating the second argument of a relation  $\rho$  as a *set* allows the arity of  $\rho$  to be greater than or equal to 2 (e.g.: *between*).

We finally reformulate the movement link (<moveLink>) as relating the mover of a motion-event to an event-path, as triggered by that motion-event. At the same time, we formulate the **event-path** to be specified with path-related information: this information includes properties such as @trigger, @motion-signals, @begin-point (source), @endpoint (goal), @midpoints, @path, @direction, and shape<sup>4</sup>. This requires the modification of the set @ of specifications associated with data types in the abstract syntax.

All of these modifications require an **event-path** to be defined as a *complex basic entity type* like the **path** type which takes other basic entities as its values. They also require the over-all definition of property assignments @ associated with data types in *ASyn<sub>isoSpace</sub>* that are related to the movement link. With the abstract syntax modified as such, we can implement the movement link in *ASyn<sub>isoSpace</sub>* into an XML-based concrete syntax by grounding the element, tagged <moveLink>, to the complex basic entity type **event-path**.

We also illustrate how the newly formulated movement link (<moveLink>) interacts with the other links in ISOspace. The qualitative spatial link (<qsLink>) and the orientational link (<oLink>) of ISOspace relate event-paths to locations as their *ground* with various topological or directional information. Such information is triggered by spatial or motion signals interacting with motion predicates, especially those called *path verbs* of various motion predicate classes.<sup>5</sup>

The organization of the paper develops as follows: Section 2. Review of <moveLink> in ISOspace (2014), Section 3. Abstract Syntax Proposed, Section 4. Restoring the Event-path, Section 5. Reformulation of the Movement Link, and Section 6. Concluding Remarks.

## 2. Review of <moveLink> in ISOspace(2014)

As is stated in the current version of ISOspace (2014), the movement link (<moveLink>) “connects motion events with mover participants.”<sup>6</sup> Hence, it is a binary relation. It is also stated: “The other attributes of the <moveLink> tag are then used to specify any obvious information about components of the event-path as well as any motion-signals.”<sup>7</sup> Here is the list of the attributes and their possible values for the <moveLink> that should implement such a connection.<sup>8</sup>

<sup>4</sup>For the discussion of the properties @direction and @shape of event-paths, see Zwarts and Winter (2000) and Bohnemeyer (2012)’s vector spatial semantics.

<sup>5</sup>The list of predicate classes (e.g., MOVE-INTERNAL, MOVE-EXTERNAL, DEVIATE, CROSS and others) was proposed by Muller (1998) and then modified by Pustejovsky and Moszkowicz (2008) on the basis of other proposals.

<sup>6</sup>See ISO 24617-7 (2014), 8.4.3, p.18.

<sup>7</sup>See ISO 24617-7 (2014), 8.4.3 <moveLink>, p.18, and A.6.4.1 General.

<sup>8</sup>See ISO 24617-7, (2014), A.6.4.2 Movement link attributes. The list is expressed in extended BNF (ISO/IEC 14977, 1996).

### (5) List A.12 Attributes for the <moveLink> Tag

```
attributes = identifier, [trigger],
[source], [goal], [midPoint], [mover],
[ground], [goalReached], [pathID],
[motionSignalID], [comment];
```

The list given above, however, fails to represent a relation between a motion and its participants in an explicit way. None of the listed attributes is required to refer to such a relation or to a motion and its participants as the arguments of this relation. Among the attributes listed, @identifier is the only attribute required to be specified. All of the attributes other than @identifier are, on the other hand, listed as optional, referring to the semantic roles of motion participants.

Nevertheless, the movement link (<moveLink>) in ISOspace (2014) can easily be modified to conform to the general link structure,  $\langle \eta, E, \rho \rangle$ . For this modification, the list of attributes associated with the link is simply revised, as shown below:

```
(6) attributes = identifier, motion,
participant, relType, [trigger],
[motionSignalID], [goalReached],
[comment];
reltype = "source | goal | midPoint |
mover | ground | pathID";
```

The revised list treats the attributes @motion and @participant as required attributes, each representing an entity structure. The attribute @relType is also a required attribute, representing a relation between these two entities, motion and its participant. A motion and one of its participants stand for  $\eta$  and a singleton  $E$ , respectively, in  $\langle \eta, E, \rho \rangle$ , while the attribute @relType specifies what  $\rho$  is. This list thus allows the movement link to conform to the general link structure, as is required by SemAF principles (ISO 24617-6, 2016).

The revised list also allows each annotated link to carry the same information as the annotation represented by the earlier list. Here is an illustration, showing how the revised <moveLink> applies to the annotation of an example given below:

(7) a. Mia<sub>se1</sub> [grew up]<sub>e1</sub> in<sub>ss1</sub> Busan<sub>pl1</sub>, but will move<sub>m1</sub> to<sub>ms1</sub> Seoul<sub>pl2</sub> next spring.

```
b. <qsLink xml:id="qsL1" figure="#e1"
ground="#pl1" relType="IN"
trigger="#ss1"/>
<moveLink xml:id="mvL2" trigger="#m1"
goal="#pl1" mover="#se1"
goalReached="no"/>
```

c. Revised:

```
<moveLink xml:id="mvL3" motion="#m1"
participant="#se1" relType="mover"/>
<moveLink xml:id="mvL2" motion="#m1"
participant="#pl2" relType="goal"
trigger="#ms1" goalReached="no"/>
```

The structure of <qsLink> in (b) conforms to the general link structure  $\langle \eta, E, \rho \rangle$ , where figure="#e1",

ground="#pll", and relType="IN" correspond to  $\eta$ ,  $E$ , and  $\rho$ , respectively. The link is interpreted as stating that the event of (Mia's) growing up is grounded in the city of Busan, as triggered by the spatial signal  $in_{ss1}$ .

As pointed out earlier, the current version of the <moveLink> of ISOSpace (2014) fails to conform to the general link structure. Nonetheless, the link is correctly understood as carrying motion-triggered information. It should contain information about the goal of the motion and its mover as well as the state of its transition, expressed by the attribute @goalReached.

Each of the two new <moveLink>s in (d) conforms to the general structure of link, as the <qsLink> in (b) does. These two links together carry the same information about the motion of moving as the link in (c) does. They are thus semantically equivalent.

Before trying to resolve these problems with the concrete syntax of ISOSpace (2014), we need to formulate its abstract syntax of ISOSpace. This is missing in the current version of ISOSpace (2014).

### 3. Abstract Syntax Proposed

#### 3.1. Overview

An abstract syntax<sup>9</sup> specifies information on an object language in *abstract* terms, focusing on syntactically or semantically relevant features only. The meta-language that formulates an abstract syntax may vary. It may consist of a conceptual inventory of data types in informal terms, a list of set-theoretic or algebraic definitions, a list of data or feature specifications represented in BNF or a so-called meta-model with graphic representations using a language like UML.

For the formulation of an abstract syntax  $\mathcal{ASyn}$  for ISOSpace, we adopt the principles of its construction that were presented in Bunt (2010), Bunt (2011), ISO 24617-6 (2016), Bunt et al. (2016) in general. We, however, follow Lee (2012) and Lee (2013) in formalizing it in algebraic terms that are often used in defining formal grammars. We also use ISO/IEC 14977 (1996) extended BNF as a meta-language to specify various features of data types because BNF is expressively more powerful than simple set-theoretic listing. We first formulate the general structure of an abstract syntax  $\mathcal{ASyn}$  for an annotation structure (3.2). We then discuss how annotation structures are generated (3.3). Finally, we present an instantiated version of the abstract syntax,  $\mathcal{ASyn}_{isoSpace}$  (3.4), for the annotation of spatial information, especially related to the event-paths triggered by motion-events.

#### 3.2. General Structure of the Abstract Syntax

Given a fragment  $L_i$  of a language as primary data for annotation, the general structure of an abstract syntax  $\mathcal{ASyn}$  for an annotation structure or language can be formally defined as a tuple:

<sup>9</sup>In DOL (OMG, 2016), the term *abstract syntax* is understood to be a parse tree term for a “language for representing documents in a machine-processable way”, while the term *concrete syntax* is a “serialization or specific syntactic encoding of such a language”.

(8)  $\langle M, E, L, @ \rangle$ , where

1.  $M$  is a set of (possibly null or non-contiguous) strings of character segments, called “markables”, in  $L_i$ ,
2.  $E$  is a finite set of *entity types*,
3.  $L$  is a finite set of *link types*,
4. @ specifies information associated with each of the basic entity types in  $E$  and each of the link types in  $L$ .<sup>10</sup>

For semantic annotation, its markables in  $M$  are strings of character segments which are identified as tokens, words or phrases in a fragment of a language, given as its primary data. This is so because semantic annotation normally presupposes that its input data has been preprocessed by word segmentation or morpho-syntactic analysis.

Empty strings are allowed as markables in  $M$ . They represent so-called *non-consuming tags* with their use licensed in ISOSpace (2014).<sup>11</sup> As will be shown in various illustrations in Section 5., event-paths are treated as non-consuming or empty tags, represented by  $\emptyset_i$ .

$E$  and  $L$  are very small sets, each consisting of a very small number of entity types or link types. As will be shown in the following subsections 3.3. and 3.4., the sets  $M$ ,  $E$ , and  $L$  are interrelated by associated information @. A particular function  $@_i$  in @ assigns an entity type  $e$  in  $E$  to each markable  $m$  in  $M$ , generating an event structure  $\langle e, m \rangle$  such that  $@_i(m) = e$ . @ also defines a set of (binary) relations  $\rho$  over the set  $\langle M, E \rangle$  of event structures for each link type in  $L$ .

#### 3.3. Generation of Annotation Structures

The abstract syntax  $\mathcal{ASyn}$  generates annotation structures, just like grammars generate phrasal (parse) trees. An annotation structure consists of two substructures: *entity structures* and *link structures*.

**Entity structures** are generated by  $\langle M, E, @ \rangle$ , a substructure of  $\mathcal{ASyn}$ . For each markable  $m$  in  $M$ , an assignment function  $@_i$  in @ specifies its *entity type*, listed in  $E$ :

(9)  $@_i : M \rightarrow E$ ,

which means that each markable in  $M$  is assigned an entity type in  $E$ .

Each typed markable, called *entity structure*, can be uniquely identified at the level of representation by a concrete syntax so that it can be referred to in annotation. We call the result of such assignments to a markable a *core entity structure*.

<sup>10</sup>Each of the specifications can be represented in ISO/IEC 14977 (1996) extended BNF as a function  $@_i$  over  $E$  and  $L$  that assigns a value type to each of the properties associated with each markable  $m$  in  $M$ . The particular names of properties and their value types mentioned in the specifications are not fixed, but may vary for each concrete syntax or dialect.

<sup>11</sup>See Definition 3.11 “Terms and definitions” and Annex A.3.4 “Special section: Non-consuming tags”, pp.29-30.

(10) Definition: **D1 core entity structure**

Given a set  $I$  of indices, a set  $M$  of markables, a set  $E$  of entity types, and an assignment  $@_i$ , the **core entity structure** is defined to be a tuple:  $\langle i, e, m \rangle$  generated by  $@_i$  applying to  $I$ ,  $M$ , and  $E$ , where  $i$ ,  $e$ , and  $m$  are members of  $E$  and  $M$ , respectively.

The core entity structure may be serialized in XML with an identifier (ID) as below:

- (11) a. `<entity xml:id="ID" type="Type" target="IDREF"/>`  
 b. Or simply,  
`<Type xml:id="ID" target="IDREF"/>`

For this serialized *concrete core* entity structure, additionally required or implied information is specified by  $@_i$ , depending on the needs of a particular annotation. Here is an illustration showing how such information is added to the core entity structure:

- (12) a. Mia will move to Busan<sub>w5</sub> next spring.  
 b. `<place xml:id="p11" target="#w5"/>`  
 c. `<place xml:id="p11" target="#w5" ana="#fs1"/>`  
`<fs xml:id="fs1">`  
`<f name="type" value="city"/>`  
`<f name="form" value="nam"/>`  
`<f name="country" value="KOR"/>`  
`</fs>`  
 d. Or simply,  
`<place xml:id="p11" target="#w5" type="city" form="nam" country="KOR"/>`

Instead of (c), the simplified representation (d) is adopted in ISOSpace (2014).

Some entity structures refer to other entity structures by allowing some of their attributes to take other entity structures as values. We call such a type of entity structures *complex entity type*.

(13) Definition: **D2 complex basic entity type**

A basic entity type is **complex** if at least one of the attributes for an entity structure has an entity structure of another type as value.

Unlike entities of the type **place** (e.g. *Busan*, *Korea*, *city*), entities of the type **path** have endpoints and midpoints, although these points may not be explicitly mentioned. The Massachusetts Turnpike, for instance, is known to stretch 138 miles from I 90/ Berkshire Connector (West Stockbridge), Canaan, NY, to Route 1A next to Logan International Airport, downtown Boston with several exits to major cities such as Springfield, Worcester, and Boston. The **path** here is treated as a complex basic entity type because its attributes such as endpoints and midpoints refer to all or some of these places as values.

**Link structures** are generated by  $\langle E, L, @ \rangle$ , a substructure of  $\mathcal{ASyn}$ . As formulated in Bunt et al. (2016), each link structure is of the following form:

- (14)  $@_i : L \rightarrow L_{st}$ ,  
 such that  $L_{st}$  is of the form  $\langle \eta, E, \rho \rangle$ ,  
 where  $\eta$  is an entity structure,  $E$  a set of entity structures, and  $\rho$  a relation between them.

This then turns into the following alternative form:

- (15) For each link type  $\tau$  in  $L$  and an assignment  $@_i$  in  $@$ ,  
 $@_i(\tau) = \langle \eta, E, \rho \rangle$ .

Given a specific  $@_i$ , the above specification, for instance, validates the following concrete representation in XML:

- (16) `<qsLink xml:id="qsL1" figure="bench" ground="{tree, rock}" relType="between"/>`

### 3.4. Abstract Syntax of ISOSpace

The abstract syntax  $\mathcal{ASyn}_{isoSpace}$  of ISOSpace is a particular instantiation of  $\mathcal{ASyn} = \langle M, E, L, @ \rangle$  such that:

- (17)  $\mathcal{ASyn}_{isoSpace}$  consists of:
1.  $M$  is a set of (possibly null or non-contiguous) sequences of tokens or words that refer to objects of the entity types specified in  $E$ .
  2.  $E$  consists of:
    - **spatial entity**, which is sub-categorized into a location or non-location type, such that the location type is subtyped into **place**, **path**, and **event-path**,<sup>12</sup>
    - **motion** and **non-motion event** as subtypes of **eventuality**,<sup>13</sup>
    - **spatial signal**, **motion signal**, and **measure signal** as subtypes of **signal**.
  3.  $L$  consists of
    - **qualitative spatial link**,
    - **orientation link**,
    - **movement link**, and
    - **measure link**.
  4.  $@$  is to be specified for each of the types in  $E$  and  $L$  in extended BNF (ISO/IEC 14977, 1996) separately.

To the lists of  $E$  and  $L$  given in ISOSpace (2014)<sup>14</sup>, there is one new entity type added: **event-path**. Each event-path is triggered by a motion-event.

<sup>12</sup>The non-location type is tagged `<spatialEntity>` in ISOSpace.

<sup>13</sup>The non-motion event type is tagged `<event>` in ISOSpace.

<sup>14</sup>See Clause 7.2 Abstract syntax for the ISOSpace annotation structure and Figure 1 - Schematic metamodel of ISOSpace.

### 3.5. Specification Assignments @

Other big differences between the version of ISOSpace (2014) and the new version proposed here are shown in the way of defining specification assignments @ to the new entity type **event-path** and the **movement link**. The **event-path** is designed to take over much of the information carried by the earlier version of **movement link**.

Here are the specifications for the movement link and the event-path. They assign the type of a value such as IDREF or CDATA, but not specific values to each of the properties characterizing the movement link and the event-path. Instead of a simple list in set-theoretic terms, each list is represented in extended BNF. Their justification is discussed subsequently in section 4. and section 5.

#### Specification @<sub>mvL</sub> of the Movement Link:

- (18) Specification of the Movement Link (<moveLink>)  
**properties** = figure, ground, relType;  
**figure** = IDREF; \* ID of an object that undergoes a change in its location\*  
**ground** = IDREFs; \* IDs of event-paths triggered by the motion\*  
**relType** = CDATA; \* TRAVERSE or predicate classes

Each link structure which is generated by the specification assignment @<sub>mvL</sub> to the movement link conforms to the general link structure  $\langle \eta, E, \rho \rangle$ , specified in Bunt et al. (2016). The **figure** specifies the entity structure  $\eta$  and the **ground** a set  $E$  of entity structures.<sup>15</sup> The **relType** specifies the motional relation  $\rho$  between them.

#### Specification @<sub>ep</sub> of the Event-path:

- (19) Specification of the Event-path (<epath>)  
**properties** = target, trigger, motionSignals, [begin-point], [endpoint], [midpoints], [path], [goalReached], [direction], [shape];  
**target** = NULL; \*non-consuming tag\*  
**trigger** = IDREF; \*ID of a <motion> that triggered the link\*  
**motionSignals** = IDREF; \*ID of motion signals\*  
**begin-point** = IDREF; \*ID of a place\*  
**endpoint** = IDREF; \*ID of a place\*  
**midpoints** = IDREFs; \*IDs of a place\*  
**path** = IDREF; \*ID of a path traversed by the mover of a motion traverses as part of the event-path\*  
**goalReached** = BOOLEAN;  
**direction** = CDATA; \*ID of motion signals such values as UP, DOWN, FORWARD, BACKWARD, EAST, NORTH\*  
**shape** = CDATA; \*STRAIGHT, CIRCULAR, CURVED, ZIGZAG, etc.\*

This list is exactly the same as that of the complex basic entity type **path** (<path>) except that the **event-path**

<sup>15</sup>The names of properties are not part of an abstract syntax. The property names @figure and @ground, for instance, may be replaced by Langacker (2008)'s terms @trajector and @landmark, respectively.

type has two additional properties. One is a required property @trigger and the other, four optional properties @path, @goalReached, @direction, and @shape. The properties @direction and @shape are optional properties that can contribute to doing vector spatial semantics (Zwarts and Winter, 2000).

## 4. Restoring the Event-path

### 4.1. Overview

We propose to restore the event-path, which had been introduced by Pustejovsky et al. (2010) in ISOSpace version 1.3e, into the newly proposed abstract syntax of ISOSpace. We use it as a basis for the reformulation of the movement link, as shown in subsections 3.4 and 3.5. Now, in this section, we try to motivate the use of event-paths in annotating motion-events.

The notion of **event-path** is very much related to the so-called axioms of motion and event-path, as discussed by Mani and Pustejovsky (2012) and Pustejovsky and Yocum (2013). We discuss these axioms in subsection 4.2. and then the notion of *ground* as a reference location for an event-path in subsection 4.3.

### 4.2. Two Axioms Extended

Here are two axioms of motions which are part of the abstract syntax of ISOSpace, as claimed by Pustejovsky and Yocum (2013).<sup>16</sup>

- (20) a. Axiom 1: Mover Participants  
 Every motion-event involves a mover.  
 $\forall e \exists x [motion-event(e) \rightarrow mover(x, e)]$
- b. Axiom 2: Event Paths  
 Every motion-event involves an event-path.  
 $\forall e \exists p [motion-event(e) \rightarrow [event-path(p) \wedge loc(e, p)]]$

These axioms presuppose the following definitions<sup>17</sup>:

- (21) Definitions:  
**D3 mover**: participant in a motion-event that undergoes a change in its location<sup>18</sup>  
**D4 path**: non-null sequence of locations (places)  
**D5 event-path**:  
*D5a Formal*: path which is directed, finite, and bounded with a begin-point, an endpoint, and a sequence of midpoints between them  
*D5b Functional*: path, triggered by a motion-event, that traces or represents the locational (physically necessary spatio-temporal) transition or trajectory of some object, called *mover*, of a motion-event

<sup>16</sup>The logical forms for the two axioms are copied verbatim from Pustejovsky and Yocum (2013).

<sup>17</sup>See Pustejovsky and Yocum (2013).

<sup>18</sup>Langacker (2008) (p.356) introduces **mover** as one of the six archetypal roles associated with actions and events, while defining it as "anything that moves (i.e. changes position in relation to its external surroundings)". He also treats the mover as a *trajector* in contrast to a *landmark* that provides a ground for the activity or motion of a *trajector*. These two terms, *trajector* and *landmark*, correspond to the terms *figure* and *ground* in our use related to motion-events.

Consider three examples below:

- (22) a. John<sub>mover</sub> walked from Boston to Cambridge.  
 b. An arrow<sub>mover</sub> hit the target.  
 c. John pushed a big rock<sub>mover</sub> over the hill.

As shown above, the mover is not necessarily the agent or cause of a motion, called *causa movendi*. Whatever their semantic roles are, all these movers above have the characteristics of moving from one location to another. Hence, to understand what is meant by **mover**, some locational change of an object must be implied from a motion.

By the two definitions given above, the **mover** in Axiom 1 is understood to be locationally related to the **event-path** in Axiom 2. By Axiom 1, an object  $x$  is related to a motion-event  $e$  and then by Axiom 2 the motion-event  $e$  to an event-path  $p$  with the relation *loc*. Thereby, the mover  $x$  is related to the path  $p$  *locally* provided that transitivity is assumed to hold here.

**Proposition:** To express such a relation between the mover and the path in more explicit terms, we make the following proposition:

- (23) Every motion-event has a path to which it is anchored, and the mover traverses that path.  
 $\forall e \exists \{p, x\} [motion-event(e) \rightarrow [event-path(p, e) \wedge \wedge mover(x, e) \wedge traverse(x, p)]]$ <sup>19</sup>

This proposition also needs the following set of definitions:

- (24) **D6 traverse:**  
**D6a** binary relation between an object  $x$  and a path  $p$  such that  $traverse(x, p)$  holds if and only if, for any path  $p$ , represented as  $\langle l_0, \dots, l_k \rangle$  with two endpoints  $l_0$  and  $l_k$ , and any object  $x$ , each of the locations of  $x$ , represented as  $l(x)_i$ , in its transition from one location to another, corresponds to each location  $l_i$  in  $p$ .  
**D6b** For an object  $x$  and a path  $p$  such that  $p$  is a sequence  $\langle l_0, \dots, l_k \rangle$ ,  $\sigma(traverse(x, p))$  implies:  
 $\forall t_i \in N [t_0 \preceq t_k \rightarrow [loc(x, t_0), \dots, \vee loc(x, t_k)]]$ .

### 4.3. Ground as a Reference Location

We assume the mover of a motion to be a *figure*, as suggested by Talmy (1975). For its interpretation then, each event-path or traversal of the mover of a motion-event requires a reference location, either a place or a path, called *ground*.

Consider two simple cases:

- (25) a. John<sub>figure</sub> cycles seriously in the gym<sub>ground</sub> everyday.  
 b. John<sub>figure</sub> walked through the park<sub>ground</sub>.

<sup>19</sup>*event-path* is here treated as a relation between a path and an event because, unlike a (static) path, an event-path is created by a motion-event.  $loc(e, p)$  holds if and only if  $traverse(x, p)$  holds for each  $l$  in  $p$ .

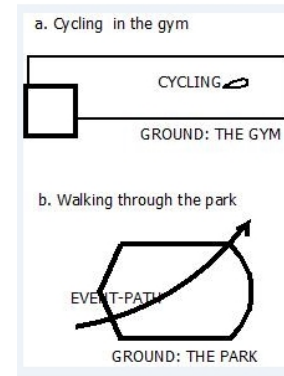


Figure 1: Ground as Reference Points

John's cycling in (a) is a genuine, but non-translocational motion-event: here John keeps moving, although he may be staying on the cycling machine in the gym without moving obviously from one place to another. The event-path of John's cycling is understood as being contained in the gym, whatever that event-path might look like. This interpretation is obtained in reference to the gym, the ground or reference-point of the event-path.

Unlike (a), John the mover as a figure in (b) moves from one location to another and then to another. All these locations are, however, contained in one and the same place, namely in the park. Characterized by the motion signal *through*, the containment relation in (b) is more complex than that of (a). The park does not contain the entire event-path of John's walking. It only contains a subpart of the event-path with its midpoints overlapping at least two points on the boundary of the park as a ground.

These interpretations are possible only if we assume that the two grounds, *gym* and *park*, are used as reference-points for the interpretation of the two event-paths for (a) and (b), respectively. See Figure 1.

## 5. Reformulation of the Movement Link

### 5.1. Overview

We propose that the movement link should conform to the general link structure  $\langle \eta, E, \rho \rangle$ , constructed in the abstract syntax  $\mathcal{ASyn}_{isoSpace}$  of ISOSpace, where

- (26) a.  $\eta$  is an entity structure of the **spatial entity** type functioning as the *mover* of a motion-event and as its *figure*,  
 b.  $E$  is a singleton containing an entity structure of the **event-path** type functioning as a *ground*, and  
 c.  $\rho$  is a relation over  $\eta$  and  $E$  triggered by a motion-event.<sup>20</sup>

### 5.2. Two Entity Types: Figure and Ground

The movement link is triggered by a motion-event. In the modified version of ISOSpace, this link is viewed as relating the mover of that motion-event to an event-path traversed

<sup>20</sup>The notions *mover* and *event-path* are defined in D3 and D5, respectively.

by the mover. The *mover* and **event-path** are then treated as the *figure* and *ground* of the movement link, respectively.<sup>21</sup>

(27) a. John walked around the park.

b.  $John_{se1:figure} walked_{m1:trigger} around\ the\ park$   
 $\emptyset_{ep1:ground}$ .

The event-path is introduced as a non-consuming tag.<sup>22</sup>

### 5.3. The Relation Type

The abstract syntax  $\mathcal{ASyn}_{isoSpace}$  of ISOSpace specifies the value of the relation type  $\rho$  to be CDATA, allowing any possible values. In a concrete syntax proposed here, we specify this value to be TRAVERSE,<sup>23</sup> a single value for each of the <moveLink> instances. If a mover  $x$  traverses a path  $p$ , then  $x$  goes through  $p$  by being located at its begin-point, midpoints, or endpoint, sequentially as time progresses.

There is no guarantee that the mover might reach the endpoint or else it might be staying at the begin-point. The mover may follow some static path such as a road or mountain trail. In such a case, the event-path overlaps the static path, but this overlap may only be partial, thus these two paths being not identical.

### 5.4. Illustrations

**Event-paths Related to Locations:** The following examples show how event-paths are related to locations (places). This relation is captured by the qualitative spatial link (<qsLink>).

(28) a. John walked around the park. [MOVE INTERNAL]

b.  $John_{se1:mover/figure} walked_{m1:trigger}$   
 $around_{ss1} the\ park_{pl1:ground} yesterday\ \emptyset_{ep1}$ .

c. 

```
<epath xml:id="ep1" target=""
trigger="#m1"/>
<moveLink xml:id="mv1" figure="#se1"
ground="#ep1" relType="TRAVERSE"/>
<qsLink xml:id="qsL1" trigger="#ss1"
figure="#ep1" ground="#pl1"
relType="IN"/>
```

d.  $[walk(e) \wedge mover(j, e) \wedge event-path(p, e)$   
 $\wedge traverse(j, p) \wedge park(l) \wedge in(p, l)]^{24}$

Here is a related example:

(29) a. John drove around the park. [MOVE EXTERNAL]

b.  $John_{se2} drove_{m2} around_{ss2} the\ park_{pl2}\ \emptyset_{ep2}$ .

<sup>21</sup>In Mani and Pustejovsky (2012), the *mover* is treated as the figure of a movement link.

<sup>22</sup>See again Definition 3.11 in Clause 3 "Terms and definitions" and its use in Annex A.3.4, pp.29-30, in ISOSpace (2014).

<sup>23</sup>See Definition D6 **traverse**.

<sup>24</sup> $in(p, l)$  is a simplification that ignores a time factor and should be understood as stating that every location (place) in the path  $p$  is contained in the location  $l$ .

c. 

```
<epath xml:id="ep2" target=""
trigger="#m2"/>
<moveLink xml:id="mvL2" figure="#se2"
ground="#ep2" relType="TRAVERSE"/>
<qsLink xml:id="qsL1" trigger="#ss2"
figure="#ep2" ground="#pl2"
relType="DC"/>
```

d.  $[walk(e) \wedge mover(j, e) \wedge event-path(p, e)$   
 $\wedge traverse(j, p) \wedge park(l) \wedge outSideOf(p, l)]$

The event-path of John's driving is outside of (DC: disconnected) the park, while the event-path of John's walking is inside (IN) the park.<sup>25</sup> They are both captured by <qsLink>.

**Path-related Information, Source and Goal:** The following examples show how event-paths carry path-related information such as information about the source and the goal of an event-motion. The begin-point and the endpoint of an event-path corresponds to the source and the goal of a motion-event that triggers that event-path. Event-paths may also contain information about the state of reaching the goal or not.

(30) a. John drove from Boston to New York. [LEAVE-REACH]

b.  $John_{se3:figure} drove_{m3:trigger} from\ Boston_{pl2}$   
 $to\ [New\ York]_{pl3}\ \emptyset_{ep3:ground}$ .

c. 

```
<epath xml:id="ep3" target=""
trigger="#m3" begin-point="#pl2"
endpoint="#pl3" goalReached="YES"/>
<moveLink xml:id="mvL3" figure="#se3"
ground="#ep3" relType="TRAVERSE"/>
```

d.  $[drive(e) \wedge mover(j, e) \wedge event-path(p, e) \wedge$   
 $traverse(j, p) \wedge begin-point(pl_2, p)$   
 $\wedge endpoint(pl_3, p) \wedge reach(j, pl_3)]$

In (b) above, an event-path is introduced and marked up as a non-consuming tag  $\emptyset_{ep1:ground}$ . It is also treated as the ground, while the motion  $drove_{m1}$  is the trigger and its mover, the figure.

By axiom 2, a motion-event creates an event-path. Furthermore, by definition each event-path as a finite directed path is characterized by its begin-point (source) and its endpoint (goal), possibly with midpoints, while representing the course of movement that the mover traverses. The movement link then relates the mover of a motion to this event-path. Such information in (d) is represented in (c) above.

Consider two more related examples:

(31) a. John arrived in New York. [REACH class]

b.  $John_{se1} arrived_{m2} in\ [New\ York]_{pl2}\ \emptyset_{ep2}$ .

c. 

```
<epath xml:id="ep2" target=""
trigger="#m2" endpoint="#pl2"
goalReached="YES"/>
<moveLink xml:id="mvL2" figure="#se1"
ground="#ep2" relType="traverse"/>
```

<sup>25</sup>See the definitions of DC and IN are given in RCC8<sup>+</sup> by Randall et al. (1992).



- d.  $[arrive(e) \wedge mover(j, e) \wedge event-path(p, e) \wedge traverse(j, p) \wedge endpoint(pl_2, p) \wedge reach(j, pl_2)]$
- (32) a. John left Boston. [LEAVE class]
- b.  $John_{se1} left_{m3} Boston_{pl1} \emptyset_{ep3}$ .
- c. `<epath xml:id="ep3" target="" trigger="#m3" begin-point="#pl1"/>  
<moveLink xml:id="mv3" figure="#se1" ground="#ep3" relType="TRAVERSE"/>`
- d.  $[leave(e) \wedge mover(j, e) \wedge event-path(p, e) \wedge traverse(j, p) \wedge begin-point(pl_1, p)]$

These two examples as well as the earlier one (case of the LEAVE-REACH motion class) have the identical <moveLink> annotation. They only differ in the annotation of their respective event-paths. The first event-path has a full specification of the path from its begin-point to the endpoint. The other two have a partial specification with one specifying the endpoint and the other specifying the begin-point.

**Directionality of Event-paths:** In its formal definition **D5a**, an event path is defined to be a *directed* finite path. Some event-paths may carry information about its directionality. This information is conveyed by the motion signal associated with the motion predicate that refers to a motion-event. Here are examples:

- (33) a.  $John_{se2} climbed_{m2} up_{ms2} the\ hill_{pl2} \emptyset_{ep2} \emptyset_{pl3}$ .
- b. `<epath xml:id="ep2" trigger="#m2" endpoint="#pl3" goalReached="yes" direction="#ms2"/>  
<oLink xml:id="qsL2" figure="#pl3" ground="#pl2" relType="ABOVE"/>  
<qsLink xml:id="qsL2" figure="#ep2" ground="#pl2" relType="EC"/>  
<moveLink xml:id="mvL2" figure="#se2" ground="#ep2" relType="TRAVERSE"/>`
- c.  $[climb(e) \wedge mover(j, e) \wedge event-path(p, e) \wedge traverse(j, p) \wedge endpoint(pl_3, p) \wedge reach(j, pl_3) \wedge above(pl_3, pl_2) \wedge dir(p) = upward \wedge ec(pl_3, pl_2)]$

This example contains two non-consuming tags: an event-path  $\emptyset_{ep2}$  and a place  $\emptyset_{pl3}$ . The non-consuming tag  $\emptyset_{pl3}$  is understood to be the goal of the event-path, namely the top of the hill. The spatial relation ABOVE between the place  $\emptyset_{pl3}$  and the hill $_{pl2}$  can be captured by an orientation link (<oLink>). The motion signal  $up_{ms2}$  marks up the directionality of the event-path  $\emptyset_{ep2}$ . The event-path, on the other hand, is marked up as being externally connected (EC) to the hill $_{pl2}$ .

- (34) a.  $The\ glacier_{p6} melted_{m8} down_{ms9} [the\ valley]_{p7} \emptyset_{pl6} \emptyset_{ep6}$ .
- b. `<motion xml:id="m8" target="#token3" motionType="manner" motionClass="follow" motionSense="intrinsicChange"/>`

`<epath xml:id="ep6" trigger="#m8" path="#p7" endpoint="#pl6" direction="#ms9" goalReached="yes"/>  
<moveLink xml:id="mvL8" figure="#p6" relType="TRAVERSE"/>  
<qsLink xml:id="qsL8" figure="#p6" ground="#p7" relType="EC"/>`

- c.  $[melt(e) \wedge mover(g, e) \wedge event-path(p, e) \wedge traverse(j, p) \wedge endpoint(pl_6, p) \wedge reach(g, pl_6) \wedge dir(p) = downward \wedge ec(p_6, p_7) \wedge p_7 \sqsubseteq ep_6]$

**Event-paths Referring to Static Paths:** As in ISO-Space-spec-1.3e (Pustejovsky et al., 2010), the event-path can refer to a path, traversed by the mover of a motion, as part of it. Here is an example:

- (35) a.  $John_{se2} drove_{m2} to\ Worcester_{pl2} on\ [the\ Massachusetts\ Turnpike]_{path:p2} \emptyset_{ep2}$ .
- b. `<epath xml:id="ep2" target="" trigger="#m2" endpoint="#pl2" path="#p2" goalReached="yes"/>  
<moveLink xml:id="mvL9" figure="#se2" ground="#ep2" relType="TRAVERSE"/>`
- c.  $[drive(e) \wedge mover(j, e) \wedge event-path(p_1, e) \wedge traverse(j, p_1) \wedge endpoint(pl_2, p) \wedge reach(j, pl_2) \wedge path(p_2) \wedge p_2 \sqsubseteq p_1]$

**Cases Movers not Mentioned:** Sometimes a mover is not explicitly mentioned. Its movement link can still be annotated as illustrated below:

- (36) a.  $Traveling_{m4:trigger} to\ Syria_{pl4} has\ become\ impossible. \emptyset_{se4} \emptyset_{ep4}$
- b. `<epath xml:id="ep4" target="" trigger="#m4" endpoint="#pl4"/>  
<moveLink xml:id="mvL4" trigger="#m4" figure="#se4" ground="#ep4" relType="TRAVERSE"/>`
- c.  $\neg \diamond [travel(e) \wedge mover(x, e) \wedge event-path(p, e) \wedge traverse(x, p) \wedge named(l, Syria) \wedge country(l) \wedge endpoint(l, p)]$

(a) mentions no mover. This is expressed as an unbound variable  $x$ . As in other cases, unbound variables are to be interpreted as being bound existentially.

## 6. Concluding Remarks

This paper has proposed to restore the event-path (<epath>), introduced in an earlier version (Pustejovsky et al., 2010) of ISOSpace, as a complex basic entity type into modified ISOSpace. The earlier version had the event-path as a basic entity type, but without the movement link. In contrast, ISOSpace (2014) introduced the movement link to take the place of the event-path at the concrete level of annotation. The *ASyn<sub>isoSpace</sub>* proposed in this paper retains both the **event path** and the **movement link** and implements both of them into a concrete syntax.

As a complex basic entity type, the event-path carries various kinds of path-related information. It is always triggered by a motion-event, but also by motion signals very often. Such information is represented by a set of specification assignments about the *begin-point (source)*, *endpoint (goal)* or *midpoints* of an event-path as well as the static *path* traversed by the mover of a motion-event, as explicitly referred to in text.

On the basis of the event-path, we have reformulated the movement link (<moveLink>). It relates the mover of a motion-event as a *figure* to the event-path as a *ground*, with some movement relation like TRAVERSE. This formulation has fully accommodated the two axioms on motions and event-paths, introduced by Pustejovsky and Yocum (2013). It also conforms to the formally defined abstract syntax of ISOSpace,  $ASyn_{isoSpace}$ , for the annotation structures of spatial information and motion-events.

Four topics have been left out for the future work. First, we did not manage to work on a compositional semantics of  $ASyn_{isoSpace}$ . Second, an earlier version of this paper introduced a new link, called the *path link*, that relates event-paths to static paths with various geometric relations such as *parallel*, *intersect*, *fork (split)*, and *merge*. We claimed that this was needed to complement the topological and orientational links involved in the movement link. Third, we have specified the two properties *direction* and *shape* for the event-path, but these need be further discussed in reference to vector spatial semantics for the annotation of spatial information. Finally, we understand that motions in the physical world involve both space and time, thus requiring each of them to form a series of unified spatio-temporal locations on four dimensions. This thus requires the integration of ISO-TimeML (ISO 24617-1, 2012) and ISOSpace (ISO 24617-7, 2014) for the annotation of spatio-temporal information involving motions. Or else these two should be made interoperable, as discussed in Lee (2012) and Lee (2013).

## 7. Acknowledgements

I owe thanks to Harry Bunt, James Pustejovsky, Roland H. Hausser, Jae-Woong Choe, Jongbok Kim, Chonwon Park, Suk-Jin Chang, Chinwoo Kim, and several anonymous reviewers for reading the pre-final version of this paper. This does not mean that they all agree with my proposal.

Bohnenmeyer, Jürgen. 2012. A vector space semantics for reference frames in Yucatec. In Elizabeth Bogal-Allbritten (ed.), *Proceedings of the sixth meeting on the Semantics of Under-Represented Languages in the Americas (SULA 6 and SULA-Bar)*, pp.15-34. Amherst: GLSA Publications.

Bunt, Harry. 2010. A methodology for designing semantic annotation languages exploiting semantic-syntactic isomorphisms. In Alex C. Fang, Nancy Ide, and Jonathan Webster (eds.), *Proceedings of the Second International Conference on Global Interoperability for Language Resources (ICGL2010)*, pp.29-46. Hong Kong.

Bunt, Harry. 2011. Abstract syntax and semantics in semantic annotation, applied to time and events. Revised

version of Introducing abstract syntax + semantics in semantic annotation, and its consequences for the annotation of time and events. In E. Lee and A. Yoon (eds.), *Recent Trends in Language and Knowledge Processing*, pp.157-204. Hankukmunhwa, Seoul.

Bunt, Harry, Volah Petukhova, Andrei Malchanau, and Kars Wijnhoven. 2016. The Tilburg DialogBank corpus. *Proceedings of 10th Edition of the Language Resources and Evaluation Conference (LREC2016)*, pp. xx-yy. Portorož, Slovenia.

Donnelly, Kevin and Hongwei Xi. 2005. Combining higher-order abstract syntax with first-order abstract syntax in ATS. *Proceedings of the 3rd ACM SIGPLAN Workshop on Mechanized Reasoning about Languages with Variable Binding (MERLIN '05)*, pp. 58-63.

ISO. ISO 24617-1:2012(E) *Language resource management - Semantic annotation framework - Part 1: Time and events (SemAF-Time, ISO-TimeML)*. The International Organization for Standardization, Geneva.

ISO. ISO 24617-4:2014(E) *Language resource management - Semantic annotation framework - Part 4: Semantic roles (SemAF-SR)*. ISO. The International Organization for Standardization, Geneva.

ISO. ISO 24617-7:2014(E) *Language resource management - Semantic annotation framework - Part 7: Spatial information (ISOSpace)*. The International Organization for Standardization, Geneva.

ISO. ISO 24617-6:2016(E) *Language resource management - Semantic annotation framework - Part 6: Principles of semantic annotation (SemAF principles)*. The International Organization for Standardization, Geneva.

ISO. ISO/IEC 14977:1996 *Information technology - Syntactic metalanguage - Extended BNF*. The International Organization for Standardization and the International Electrotechnical Commission, Geneva.

ISO/IEC. ISO/IEC 24707:2007 *Information technology - Common Logic (CL): a framework for a family of logic-based languages*. The International Organization for Standardization and the International Electrotechnical Commission, Geneva.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford University Press, Oxford.

Lee, Kiyong. 2012. Interoperable Spatial and Temporal Annotation Schemes. *Proceedings of The Joint isa-7, SRSL-3 and I2MRT Workshop on Interoperable Semantic Annotation*, 61-68. The Eighth Edition of Language Resources and Evaluation Conference (LREC 2012) Satellite Workshop, Istanbul.

Lee, Kiyong. 2013. A model structure for the construction of annotation schemes for their interoperability. In Harry Bunt (ed.), *Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation - isa-9*, pp.15-24. March 1920, 2013, Potsdam, Germany.

Mani,INDERJEET and James Pustejovsky. 2012. *Interpreting Motion: Grounded Representations for Spatial Language*. Oxford University Press, Oxford.

Muller, Philippe. 1998. A qualitative theory of motion based on spatiotemporal primitives. In A.G. Cohn, L.K.

- Schubert, and S.C. Shapiro (eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR98)*. Morgan Kaufmann, San Mateo, CA.
- Muller, Philippe. 2002. Topological spatio-temporal reasoning and representation. *Computational Intelligence*, 18(3):420-450. Oxford University Press, Oxford.
- OMG. 2016. *The Distributed Ontology, Modeling, and Specification Language (DOL)*, Version 1.0. Object Management Group.
- Pfenning, Frank and Conal Elliott. 1988. Higher-order abstract syntax. *Proceedings of the ACM-SIGPLAN '88 Symposium on Programming Language Design and Implementation*, pp.199-208.
- Pustejovsky, James and Jessica L. Moszkowicz. 2008. Integrating motion predicate classes with spatial and temporal annotations. *Proceedings of COLING 2008*, pp.95-98. Manchester, UK.
- Pustejovsky, James, Jessica L. Moszkowicz, and Marc Verhagen. 2010. ISO-Space Specification: Version 1.3 (October 5, 2010). includes discussion notes from the Workshop on Spatial Language Annotation, the Airlie Retreat Center, VA, September 26-29, 2010.
- Pustejovsky, James and Zachary Yocum. 2008. Capturing motion in ISO-SpaceBank. In H. Bunt (ed.), *Proceedings of isa-9*, pp. 25-34. Potsdam, Germany
- Randell, David A., Zhan Cui, and Anthony G. Cohn. 1992. A spatial logic based on regions and connection. *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, pp. 165-175. Morgan Kaufman, San Mateo, CA.
- Talmy, Leonard. 1975. Figure and ground in complex sentences. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, pp. 419-430.
- Talmy, Leonard. 1983. How language structures space. In Herbert Pick and Linda Acredolo (eds.), *Spatial Orientation: Theory, Research, and Application*. Plenum Press. Reprinted in *Toward a Cognitive Semantics* Vol.1, Ch. 3. The MIT Press.
- Zwarts, Joost and Yoad Winter. 2000. Vector space semantics: A model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information*, 9.2: 171-213.

## 8. Copyrights

The Language Resource and Evaluation Conference (LREC) proceedings are published by the European Language Resources Association (ELRA). They include different media that may be used (i.e. hardcopy, CD-ROM, Internet-based/Web, etc.).

## Discourse Markers and Disfluencies

### Integrating Functional and Formal Annotations

Ludivine Crible

Université catholique de Louvain  
1 Place Blaise Pascal, 1348 Louvain-la-Neuve, Belgium  
E-mail: ludivine.crible@uclouvain.be

#### Abstract

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.

**Keywords:** discourse markers, disfluency, annotation integration

### 1. Introduction

Spoken language in its most natural, spontaneous forms is characterized by the frequent – yet mostly unnoticed – occurrence of so-called disfluencies, which are generally considered to be cues of ongoing processes of language production and comprehension (e.g. Alter & Oppenheimer, 2009). Disfluencies generally include filled and silent pauses, repetitions, truncations, false starts and reformulations, taking up the seminal typology by Shriberg (1994). The formal and functional diversity of these elements is the direct consequence of the multifaceted nature of the abstract constructs of fluency and disfluency, which are in fact two sides of the same coin, hence the terminological choice of “(dis)fluency markers” in the remainder of this paper, to take this ambivalence into account. Concretely, the same device (e.g. a repetition) can either be perceived as strategic or disruptive depending on a wide range of linguistic (e.g. position, co-occurrence patterns) and extralinguistic (settings, speaker profile) factors. As a result, to date, corpora annotated with (dis)fluency markers are rarely comparable since they do not always include the same types of elements, and even within one type do not always follow the same definitions. Crible (in press) has shown that the same problem applies to the functional category of discourse markers (henceforth DMs) where it has been resolved by means of a corpus-based operational definition and annotation procedure (see Crible & Zufferey, 2015). DMs can be broadly defined as syntactically optional, metadiscursive cues constraining the interpretation of discourse by signaling coherence relations, topic structure and/or interpersonal strategies (Schiffrin, 1987; Brinton, 1996). DMs are here considered to be a type of (dis)fluency markers, although their inclusion is not consensual (e.g. Eklund, 2004; Beliao & Lacheret, 2013) nor always consistent with the exhaustive definition of the

category stated above (cf. the use of closed-lists of DMs in Strassel, 2003 or Pallaud, Rauzy & Blâche, 2013).

This paper reports the methodological aspects of an annotation protocol which integrates formal identification of (dis)fluency markers and a multi-layered description of DMs featuring, among others, semantic-pragmatic variables such as their domain and function in context. The challenges and merits of this integration will be illustrated by a comparison of clustering tendencies between different functions of DMs in a French-English comparable dataset, thus uncovering form-meaning patterns. In the following sections, background and key notions will be briefly outlined (Section 2); the data and annotation procedure will be presented (Section 3); results of the quantitative study will be discussed in Section 4, before concluding on some methodological perspectives (Section 5).

### 2. (Dis)fluency and Discourse Markers in Corpus Linguistics

#### 2.1 Fluency and Disfluency in Native Speech

This study follows a componential approach to (dis)fluency according to which different features or markers contribute to the relative fluency of discourse depending on their frequency, combination patterns and contextual distribution. It is argued that a fluent/disfluent interpretation is always context-bound, and that all markers in the typology are potentially fluent, thus refraining from any premature bias or interpretation during the first stages of the analysis (cf. Section 3.2). Although applicable to different populations, the present definition targets native speakers, thus effectively excluding learner and pathological (dis)fluency from the scope of this paper. In the absence of a reference standard for native speakers (as opposed to the native-like target for learners), language use

in native speech can only be assessed relative to social and contextual expectations.

Most contributions to the study of L1 (dis)fluency are either corpus-based or experimental, usually focusing on specific markers in one register (e.g. Rendle Short, 2004 on filled pauses in academic discourse; Fung, 2007 on repetitions in business discourse). However, a number of exhaustive annotation campaigns in recent years have provided substantial contributions to the field, such as Shriberg (1994), Meteer et al. (1995), Besser & Alexandersson (2007), Dister (2007) or Götz (2013) among others.

## 2.2 Defining and Annotating DMs

DMs have been extensively studied in the past thirty years in a variety of frameworks, methods and languages, which results in a lack of consensus both at the theoretical level for defining the boundaries of the category, and at the operational level for annotating various features of their behavior in corpus data (see Crible, in press for a full review). The major disagreement probably lies in the divide between i) relational DMs or “connectives” such as semantic uses of *so* or *because*, which are sometimes excluded from the category as in Lewis (2006), and ii) non-relational DMs, most of which are speech-specific such as non-propositional uses of *you know* or *sort of*, thus absent from most written-based accounts (e.g. Sanders, Spooren & Noordman, 1992; Fraser, 1999).

Current research is mostly focused on designing cognitively valid and operationally robust categories for the annotation of DM functions, in both speech and writing<sup>1</sup>. This endeavour faces numerous issues, namely particularism (language- or medium-related specificities), varying granularity, poor replicability and the intrinsic under-specification and multifunctionality of language. Major corpus-based frameworks include the Penn Discourse TreeBank (Prasad et al., 2008), Rhetorical Structure Theory (Mann & Thompson, 1988; Taboada & Mann, 2006), Segmented Discourse Representation Theory (Asher & Lascarides, 2003) and other functional taxonomies for speech such as González (2005) or Cuenca (2013).

Apart from their function(s), authors have been concerned with other features of DMs behavior especially their syntactic integration. In particular, the MDMA project (Bolly et al., 2015; in press) has designed a coding scheme based on French corpus data covering the following variables: part-of-speech category, position in the dependency unit, syntactic mobility, basic semantic value, procedural vs. conceptual meaning, presence of a contiguous pause and position in the speech turn. Another ongoing endeavour is that of the Val.Es.Co group (Briz & Pons Bordería, 2010) which combines information about the position, type of host-unit and function of the DM in a corpus of Spanish conversations.

<sup>1</sup> See the program of the ISCH COST Action IS1312 “TextLink : Structuring Discourse in Multilingual Europe” (chair L. Degand) at <http://textlink.ii.metu.edu.tr/>. See also the ISO 24617-8 standard for discourse relations (Prasad & Bunt, 2015).

## 2.3 DMs as (Dis)fluency Markers

DMs are often studied for their role in cognitive processing and overall (dis)fluency, although not always in these terms (e.g. facilitating interpretation, enhancing cooperation between participants, sounding natural and convincing, etc.). In fact, experts in the study of DMs in spoken language rarely deal with this cognitive aspect of their use and functions, let alone in a crosslinguistic perspective (see Fox Tree & Schrock, 1999 for an exception on English DM oh). However, their important role in online production and their high contextual variation advocate for their treatment as (dis)fluency markers similar to filled pauses or editing expressions.

While DMs and (dis)fluency have been extensively studied in the past as independent phenomena, combining DM-level and disfluency-level annotations however has, to the author’s knowledge, never been carried out before at such a fine-grained level of informativeness as what is proposed in the present model, especially regarding the annotation of syntactic and pragmatic features of DMs. 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 but also highly relevant to the investigation of form-meaning patterns. Concretely, this integration combines a syntagmatic or “horizontal” level (identification of (dis)fluency markers based on formal features only) with a multi-layered “vertical” level focusing on several features of DMs.

Both annotation levels were designed following the general principles of flexibility (to different languages, registers, modes, even technical formats and theoretical frameworks) and exhaustivity in the selection of observed phenomena. While exhaustivity can hardly be evaluated on open-class categories such as DMs or (dis)fluency markers, several applications of the present model to different corpora can vouch for its flexibility: the (dis)fluency typology has been tested on spoken French (Grosman, in press; Crible, in press) and English native and nonnative corpora (Dumont, 2014) as well as French Belgian Sign Language (Notarrigo, 2016); the functional taxonomy for DMs has been used by Dobrovoljc (2016) in Slovene and Gabarró-López (forthc.) in French Belgian Sign Language.

## 3. Data and Procedure

### 3.1 The *DisFrEn* Dataset

For this study, a French-English comparable dataset has been sampled from several existing corpora<sup>2</sup> in order to cover eight contextual settings in similar proportions in the two languages: conversations, private phone calls, face-to-face interviews, radio interviews, classroom lessons, sports commentaries, political speeches and news broadcasts. *DisFrEn* comprises 15 hours of speech and 163,620 words

<sup>2</sup> For reasons of limited space, only the main corpus in each language will be mentioned here: ICE-GB (Nelson, Wallis & Aarts, 2002) for English and VALIBEL (Dister et al., 2009) for French.

in total. It is segmented at word-level, sound-aligned and all transcriptions are provided with their audio file. All files have been formatted to be manually annotated under EXMARaLDA (Schmidt & Wörner, 2009), an open source software for multi-layered annotation, as can be seen in Figure 1.

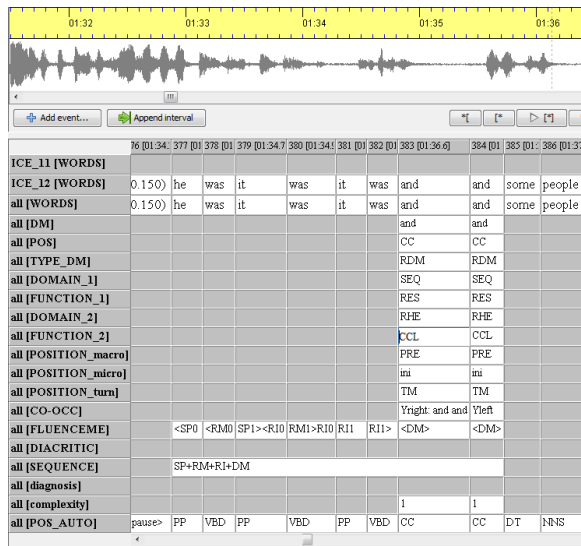


Figure 1: The EXMARaLDA interface

### 3.2 Formal Annotation of (Dis)fluency Markers

The “horizontal” level follows the typology and procedure detailed in Crible et al. (2016). It distinguishes between ten types of (dis)fluency markers for spoken language, as well as other secondary phenomena such as deletion or misarticulation. Their identification is entirely based on formal and structural criteria alone, making no *a priori* distinctions between potentially fluent or disfluent markers: all types of repetitions, pauses, DMs, etc. are selected. However, this annotation refrains from any judgment of grammaticality (unless explicitly noticed by the speakers themselves), in order to avoid any reference to a linguistic norm, as opposed to the approach taken by Besser & Alexandersson (2007) who use grammatical correctness as selective criterion for their “uncorrected” category. The complete list of markers can be found in Table 1.

Innovative features include a bracketing and numbering system that enables the annotation of complex nested structures with high precision. Example 1 shows a sequence of filled and unfilled pauses, two DMs and an identical repetition. The last row summarizes the marker types by order of appearance, and it is manually assigned during the annotation.

(1)

(0.200)	uhm	so	she	and	she
<UP>	<FP>	<DM>	<RI0	<DM>	RI1>
				<WI>	
[UP+FP+DM+RI]					

The distinction between DMs, filled pauses (FPs) and editing terms (ET) is particularly relevant to the issue of

integrating different levels of annotation. While some authors consider FPs such as *uhm* to be DMs (e.g. Tottie, 2015), in the present approach they had to be distinguished in order to be compatible with the literature on (dis)fluency. Similarly, some editing expressions such as *I don't know* or *oops* share some characteristics with DMs (i.e. grammatically optional, metadiscursive meaning), thus calling for more prescriptive criteria to keep the two types of markers apart.

Tags	Markers
UP	unfilled pause (seconds)
FP	filled pause
DM	discourse marker
ET	editing term
FS	false-start
TR	truncation
RI	identical repetition
RM	modified repetition
SP	propositional substitution
SM	morphological substitution
Diacritics	
AR	misarticulation
WI	embedded fluenceme
OR	change of order
Related elements	
IL	lexical insertion
IP	parenthetical insertion
DE	deletion

Table 1: Typology of (dis)fluency markers based on Crible et al. (2016).

A number of post-treatment operations were necessary to synthesize all possible combinations of markers in the data into quantitatively manageable categories. Several macro-labels with different levels of granularity, as well as numeric variables (e.g. length of the sequence, number of marker types etc.) were semi-automatically assigned to the whole dataset. The most coarse-grained categorization of (dis)fluent sequences amounts to six hierarchical levels based on the content of the sequence: DMs alone; pauses (UP/FP + DMs); interruptions (TR/FS + pauses and DMs); repetitions (RI/RM + pauses and DMs); substitutions (SM/SP + repetitions, pauses and DMs); and mixed sequences (interruptions/repetitions/substitutions + all others). These groupings made no use of DM-based annotations, which are processed separately.

### 3.3 Multi-layered annotation of discourse markers

The “vertical” level follows the corpus-based annotation scheme by Crible (in press). Identification and annotation of DMs are entirely manual, bottom-up and aim at exhaustivity (as opposed to closed-list selections). The annotation procedure is quite unrestricted and allows the analyst to listen to the audio as often as necessary. Each



variable is annotated on a separate layer (cf. Figure 1). The scheme contains five formal variables and three functional ones which only apply to the DMs previously identified as part of the “horizontal” (dis)fluency annotation (i.e. any item that meets the DM definition, and nothing else). The variables are:

- formal: part-of-speech; position in the dependency structure; position in the clause; position in the speech turn; co-occurrence with another DM;
- functional: (non-)relational type; domain; function.

The most informative (at a semantic-pragmatic level) and challenging annotation is the functional interpretation based on a list of thirty values grouped in four domains (or macro-functions) which can be found in Table 2. Because of space constraints, only the four domains will be briefly defined here: ideational functions are semantic relations between real-world events; rhetorical functions express the speaker’s subjectivity towards their discourse; sequential functions cover turn exchange and topic structure; interpersonal functions concern the speaker-hearer relationship. This taxonomy has been extensively tested and validated with satisfying intra-rater reliability ( $K=0.74$ ) (see Crible & Zufferey, 2015 for a discussion of inter-annotator agreement).

Ideational	Rhetorical	Sequential	Interpersonal
cause	motivation	punctuation	monitoring
consequ.	conclusion	opening	face-saving
concession	opposition	closing	disagreeing
contrast	specification	topic-resume	agreeing
alternative	reformul.	topic-shift	elliptical
condition	relevance	quoting	
temporal	emphasis	addition	
exception	comment	enumeration	
	approx.		

Table 2: Functional Taxonomy for DMs in speech based on Crible (in press).

All annotations are extracted variable by variable, using EXMARaLDA’s built-in concordancer EXAKT. In order to retrieve (dis)fluency markers that are clustered in a single sequence (i.e. contiguously co-occurring markers), horizontal annotations are extracted at sequence-level, so that each line in the concordancer corresponds to one sequence, possibly containing more than one (dis)fluency markers. Through the use of unique identification codes, each sequence is semi-automatically matched with the DM(s) it contains, and vice versa, so that both annotation levels can be cross-tabulated. For instance, the sequence “so (0.480) and once” (ID code: ClaEN4t7) contains three DMs, all of which receive a corresponding ID code (ClaEN4t7DM9, ClaEN4t7DM10 and ClaEN4t7DM11). This system, while time-consuming and not required for general corpus results, is necessary for more qualitative analyses and offers a structured representation of the two annotation layers.

#### 4. Clusters of (Dis)fluency and Discourse Markers

The annotations carried out in DisFrEn provide the material for a wide range of research questions and make it possible to draw detailed DM-specific, (dis)fluency-specific and/or function-specific profiles, by querying the dataset for any item of interest, in several registers of French and English. The integration of formal and functional annotations is particularly useful to uncover patterns of objective features that corroborate semantic-pragmatic categorization by showing contrasts and preferences. The analyses in this section will mainly make use of the six hierarchical macro-labels presented in Section 3.2 which summarize the types of (dis)fluency markers contained in a sequence, and cross-tabulate them with DM-based annotations.

Starting with the generic level of functional domains (viz. ideational, rhetorical, sequential and interpersonal), some clustering tendencies can be observed. Table 3 shows that clusters of pauses and DMs (or “P-sequences”) are the most frequent type overall, followed by DMs alone (“D-sequences”), leaving the other types of more complex (dis)fluency markers to very low frequencies.

	Ideat.	Rhet.	Sequ.	Interp.	Total
Pauses	42%	42%	59%	47%	47%
DMs	45%	39%	24%	31%	36%
Repet.	7%	9%	8%	9%	8%
Interrupt.	3%	5%	5%	8%	5%
Mixed	1%	3%	2%	3%	2%
Substit.	2%	2%	2%	2%	2%

Table 3: Distribution of functional domains by (dis)fluent sequences.

The ideational domain is the only one where D-sequences are the most frequent type (44.71%). For the interpersonal, rhetorical and sequential domains, clusters of pauses and DMs are preferred. This first result could indicate that the objectivity of ideational functions seems to be paired with a high integration in the speech flow (i.e. no pause boundaries before or after the DM). On the other hand, sequential functions show the biggest proportion of P-sequences (almost 60%), which can be related to the structuring and punctuating role of these functions (e.g. topic-shift, enumerating). This clustering tendency of pauses with sequential DMs could be expected from previous findings on the high frequency of (dis)fluency markers at discourse boundaries (e.g. Roberts & Kirsner, 2000). It should be noted that these differences are observed on a subset of the data which only includes turn-medial occurrences, since turn-initial and turn-final positions necessarily exclude the possibility to co-occur with a pause at their left and right periphery, respectively. Another potentially interesting difference is the slight reversal between D- and P-sequences across ideational and rhetorical DMs. These two domains contain several functions which only differ in their degree of subjectivity, such as cause (objective/semantic) vs. motivation (subjective/pragmatic), consequence vs conclusion,

condition vs. relevance, etc. The preference for pause clusters in rhetorical functions could be connected to well-established hypotheses on the higher cognitive complexity of subjective relations (e.g. Canestrelli, Mak & Sanders, 2013) and/or to the larger scope of these relations, usually applying to larger or more distant segments. However, any stronger conclusion would require to investigate these differences at a more fine-grained level, taking more variables into account.

In this perspective, the following results focus on specific functions, thus acknowledging the variation within a single domain. These analyses report the distribution of the top seven functions, which are the only ones with more than 200 occurrences in each language: addition (continuation in the same topic without any other value), monitoring (checking for understanding and attention), specification (elaboration with more detail or an example), opposition (pragmatic contrast or concession), temporal (chronological relation), consequence (logical effect of a previous situation) and conclusion (pragmatic result, includes summaries). Table 4 shows that only the temporal function (as well as specification, to a lesser extent) occurs more frequently as DMs alone than clustered with pauses, which could again be interpreted in terms of greater prosodic integration. On the other hand, the monitoring function shows the larger proportions of R- (repetitions) and F-sequences (interruptions), which could in turn be a symptom of disfluency, since these types of (dis)fluency markers are generally more intrusive and less directly functional than pauses.

A more fine-grained view of the content of the sequences shows that, for all seven functions, the top ten sequences are clusters of DMs, unfilled and filled pauses in various configurations (see Crible, Degand & Gilquin, in press for a detailed study of these clusters), in considerably higher proportions than all the other types of (dis)fluency markers. Only additive DMs are more frequent clustered with another marker (DM or other) than alone, while monitoring DMs stand out as rarely preceded by an unfilled pause, a configuration which is very frequent for all other functions (second most frequent type). This last result on the monitoring function can be explained by the typically final position of these markers (as in example 2), which in turn results in a higher frequency of pauses at their right periphery, compared to the other functions.

- (2) I only put the alarm back twice **you know (0.567)** it's really good (PhoEN5166)

Going a step further, this type of comparison between functions can even be refined DM by DM. For instance, looking at the occurrences of *so* signaling either a relation of consequence (objective) or of conclusion (subjective), the tendency identified above regarding the contrast between ideational and rhetorical functions seems to be confirmed: subjective *so* is much more frequently clustered with pauses (69.64%) than its objective uses (39.34%). However, the data doesn't support the general conclusion that subjective relations always trigger more clustering with (dis)fluency markers such as pauses, since it is for instance not the case for *because* which always prefers D-sequences regardless of its objective vs. subjective meaning.

## 5. Conclusion

This paper presented the methodological and technical aspects of the integration of two annotation layers: a horizontal annotation of (dis)fluency markers based on formal criteria, and a vertical annotation of syntactic and functional characteristics of DMs. By keeping these two levels of analysis independent during the annotation procedure, this model avoids circularity (e.g. only selecting DMs that are perceived by the analyst as disfluent), strives towards exhaustivity, and therefore provides a rich resource for the investigation of (dis)fluency and discourse markers in several registers of French and English.

The quantitative study which illustrates the analytical potential of *DisFrEn* has generated tentative interpretations of the relative fluency of some functions: ideational DMs appear to be more prosodically integrated than the other domains (i.e. fewer co-occurrences with pauses); monitoring DMs are potentially more disfluent than other functions (i.e. they co-occur more frequently with intrusive (dis)fluency markers); the objective-subjective distinction is not systematically associated with different clustering tendencies. Ongoing research is working towards the validation of these observations across language and context variation, as well as their extension to other DM functions and their refinement by positional variables.

Function	Pauses	DMs	Repet.	Interrupt.	Mixed	Substit.	Total
Addition	60,54%	24,06%	7,61%	3,59%	2,01%	2,19%	100,00%
Monitoring	45,42%	31,54%	9,64%	7,84%	3,43%	2,12%	100,00%
Specification	42,11%	43,42%	7,07%	4,44%	1,48%	1,48%	100,00%
Opposition	49,80%	36,00%	7,00%	3,40%	2,20%	1,60%	100,00%
Temporal	39,63%	47,43%	7,39%	3,90%	0,62%	1,03%	100,00%
Consequence	47,51%	41,21%	6,07%	1,52%	1,30%	2,39%	100,00%
Conclusion	55,20%	28,64%	8,31%	2,31%	4,39%	1,15%	100,00%
<b>Total</b>	<b>50,09%</b>	<b>34,33%</b>	<b>7,63%</b>	<b>3,98%</b>	<b>2,17%</b>	<b>1,79%</b>	<b>100,00%</b>

Table 4: Distribution of the Seven Most Frequent Functions by (Dis)fluent Sequences.



In addition to these methodological and empirical results, the present work offers a first step towards a more general purpose, which is to show how theoretical and methodological decisions (such as functional categorization and annotation of DMs) can be motivated by empirically-sound clusters of structural characteristics. This endeavour is in line with cognitive corpus linguistics (Arppe et al., 2010) and usage-based cognitive semantics (Glynn, 2010) whereby theory and data feed each other by validating certain abstract groupings through their actual use in native language.

## 6. Acknowledgements

This research benefits from the support of the ARC-project “A Multi-Modal Approach to Fluency and Disfluency Markers” granted by the Fédération Wallonie-Bruxelles (grant nr. 12/17-044).

## 7. Main References

- Alter, A. & Oppenheimer, D. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13:219–235.
- Arppe, A.; Gilquin, G.; Glynn, D.; Hilpert, M. & Zeschel, A. (2010). Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora*, 5(1), pp. 1–27.
- Asher, N. & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.
- Beliao, J. & Lacheret, A. (2013). Disfluency and discursive markers: When prosody and syntax plan discourse. In R. Eklund (Ed.), *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS) 2013*, 54(1), pp. 5–8.
- Bolly, C.; Crible, L.; Degand, L. & Uygur-Distexhe, D. (2015). MDMA. Identification et annotation des marqueurs discursifs “potentiels” en contexte. *Discours* 16, pp. 3–32.
- Bolly, C.; Crible, L.; Degand, L. & Uygur-Distexhe, D. (in press). Towards a Model for Discourse Marker Annotation. From potential to feature-based discourse markers. In C. Fedriani & A. Sansó (Eds.), *Discourse Markers, Pragmatic Markers and Modal Particles: New Perspectives*. Amsterdam: John Benjamins.
- Brinton, L. (1996). *Pragmatic markers in English. Grammaticalization and discourse functions*. New York: Mouton de Gruyter.
- Briz, A. & Pons Borderia, S. (2010). Unidades, marcadores discursivos y posición. In O. Loureda & E. Acín (Eds.), *Los Estudios sobre Marcadores del Discurso*. Madrid: Acro/Libros, pp. 523–557.
- Canestrelli, A.; Mak, W. & Sanders, T. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes* 28(9), pp. 1394–1413.
- Crible, L. (in press). Towards an operational category of discourse markers: A definition and its model. In C. Fedriani & A. Sansó (Eds.), *Discourse Markers, Pragmatic Markers and Modal Particles: New Perspectives*. Amsterdam: John Benjamins.
- Crible, L. & Zufferey, S. (2015). Using a unified taxonomy to annotate discourse markers in speech and writing. In H. Bunt (Ed.), *Proceedings of the 11<sup>th</sup> Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-11)*, April 14<sup>th</sup>, London, UK, pp. 14–22.
- Crible, L.; Dumont, A.; Grosman, I. & Notarrigo, I. (2016). Annotation manual of fluency and disfluency markers in multilingual, multimodal, native and learner corpora. Version 2.0. *Technical report*. Université catholique de Louvain and Université de Namur.
- Crible, L.; Degand, L. & Gilquin, G. (in press). The clustering of discourse markers and filled pauses: a corpus-based French-English study of (dis)fluency. *Languages in Contrast*.
- Cuenca, M.J. (2013). The fuzzy boundaries between discourse marking and modal marking. In L. Degand, B. Cornillie & P. Pietrandrea (Eds.), *Discourse markers and modal particles. Categorization and description*. Amsterdam: John Benjamins, pp. 191–216.
- Dister, A. (2007). De la transcription à l'étiquetage morphosyntaxique – Le cas de la banque de données textuelles orales VALIBEL. PhD thesis. Université catholique de Louvain.
- Dobrovoljc, K. (2016). Annotation of multi-word discourse markers in spoken Slovene. Poster presented at *Discourse Relational Devices (LPTS 2016)*, January 24–26, Valencia, Spain.
- Dumont, A. (2014). Annotation of fluency and disfluency markers in nonnative spoken corpora. Paper presented at the *Interlanguage Annotation Workshop (Societas Linguistica Europaea - 47th Annual Meeting)*, September 11–14, Poznań, Poland.
- Eklund, R. (2004). Disfluency in Swedish human-human and human-machine travel booking dialogues. PhD thesis. Linköping Studies in Science and Technology.
- Fox Tree, J. & Schrock, J. (1999). Discourse markers in spontaneous speech: Oh what a difference an Oh makes. *Journal of Memory and Language* 40, pp. 280–295.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics* 31, pp. 931–952.
- Fung, L. (2007). The communicative role of self-repetition in a specialised corpus of business discourse. *Language Awareness* 16(3), pp. 224–238.
- Gabarró-López, S. (forthcoming). The hotchpotch of buoys: A corpus study on their use across genres and discourse functions in French Belgian Sign Language (LSFB).
- Glynn, D. (2010). Testing the hypothesis. Objectivity and verification in usage-based Cognitive Semantics. In D. Glynn & K. Fischer (Eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*. Berlin: De Gruyter Mouton, pp. 239–269.
- González, M. (2005). Pragmatic markers and discourse coherence relations in English and Catalan oral narrative. *Discourse Studies* 77(1), pp. 53–86.
- Götz, S. (2013). *Fluency in native and nonnative English speech*. Amsterdam: John Benjamins.
- Grosman, I. (in press). How do French humorists manage their persona across situations? A corpus study on their prosodic variation. In L. Ruiz-Gurillo (Ed.),

- Metapragmatics of Humor: Current Research Trends*. Amsterdam: John Benjamins.
- Lewis, D. (2006). Discourse markers in English: A discourse-pragmatic view. In K. Fischer (Ed.), *Approaches to Discourse Particles*. Amsterdam: Elsevier, pp. 43–59.
- Mann, W. & Thompson, S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8(3), pp. 243–281.
- Meteer, M.; Taylor, A.; MacIntyre, R. & Iver, R. (1995). Disfluency annotation stylebook for the Switchboard corpus. *Technical report*. Linguistic Data Consortium.
- Notarrigo, I. (2016). Les marqueurs de (dis)fluence en Langue des Signes de Belgique Francophone (LSFB). PhD thesis. Université de Namur.
- Pallaud, B.; Rauzy, S. & Blâche, P. (2013). Identification et annotation des auto-interruptions et des disfluences dans les corpus du CID. *Unpublished technical report*. Laboratoire Parole et Langage.
- Prasad, R. & Bunt, H. (2015). Semantic relations in discourse: The current state of ISO 24617-8. In H. Bunt (Ed.), *Proceedings of the 11<sup>th</sup> Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-11)*, April 14<sup>th</sup>, London, UK, pp. 80–92.
- Prasad, R.; Dinesh, N.; Lee, A.; Miltsakaki, E.; Robaldo, L.; Joshi, A. & Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of LREC, June 2008, Marrakech, Morocco*, pp. 2961–2968.
- Rendle-Short, J. (2004). Showing structure : using um in the academic seminar. *Pragmatics* 14(4), pp. 479–498.
- Roberts, B. & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes* 15(2), pp. 129–157.
- Sanders, T.; Spooren, W. & Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Processes* 15, pp. 1–35.
- Schiffirin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schmidt, T. & Wörner, K. (2009). EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics* 19(4), pp. 565–582.
- Shriberg, E. 1994. Preliminaries to a theory of speech disfluencies. PhD thesis. University of California at Berkeley.
- Strassel, S. (2003). Simple metadata annotation specification v.5. *Technical report*. Linguistic Data Consortium.
- Taboada, M. & Mann, W. (2006). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies* 8(3), pp. 423–459.
- Tottie, G. (2015). *Uh* and *um* in British and American English: Are they words? Evidence from co-occurrence with pauses. In N. Dion, A. Lapierre & R. Torres Cacoulios (Eds.), *Linguistic variation: Confronting Fact and Theory*. New York: Routledge, pp. 38–54.
- (2009). Du corpus à la banque de données. Du son, des textes et des métadonnées. L'évolution de la banque de données textuelles orales VALIBEL (1989-2009). *Cahiers de Linguistique* 33(2), pp. 113–129.
- Nelson, G.; Wallis, S. & Aarts, B. (2002). *Exploring natural language: Working with the British component of the International Corpus of English*. Amsterdam: John Benjamins.

## 8. Language Resource References

- Dister, A.; Francard, M.; Hambye, P. & Simon, A.-C.

# ISO DR-Core (ISO 24617-8): Core Concepts for the Annotation of Discourse Relations

Harry Bunt\* and Rashmi Prasad\*\*

\*Tilburg Center for Cognition and Communication (TiCC), Tilburg University, Tilburg, Netherlands

\*\*Department of Health Informatics and Administration,

University of Wisconsin-Milwaukee, Milwaukee, USA

harry.bunt@uvt.nl; prasadr@uwm.edu

## Abstract

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.

**Keywords:** discourse relation annotation, ISO standard, interoperability

## 1. Introduction

In a discourse, which comes into play when communication involves a sequence of clauses or sentences in a text, or utterances in a dialogue, a major aspect of the understanding comes from how the events, states, facts, and propositions mentioned in the discourse are related to each other. Understanding such relations, such as *Cause*, *Contrast*, and *Condition*, contribute to what is called the 'coherence' of the discourse. They can be realized explicitly, by means of certain words and phrases (often called 'discourse connectives'), or they can be implicit and have to be inferred on the basis of the discourse context and world knowledge.

Existing annotation frameworks exhibit two major differences in their underlying assumptions: the representation of discourse structure, and the semantic classification of discourse relations. Notwithstanding these differences, there are also strong compatibilities. Based on an analysis of differences and commonalities, ISO DR-Core (ISO 24617-8: 2016) forms the first part of an effort to develop an international standard for the annotation of discourse relations.<sup>1</sup> This first part aims to: (1) establish desiderata for the interoperable annotation of discourse relations; (2) specify a way of annotating discourse relations that is compatible with existing and emerging ISO standard annotation schemes of semantic information; and (3) provide clear and mutually consistent definitions of a set of 'core' discourse relations which are commonly found in some form in existing approaches to discourse relations and their annotation. Together, (2) and (3) form a 'core annotation scheme' for discourse relations. ISO DR-Core does not aim at providing a fixed and

exhaustive set of discourse relations, but rather at providing an open, extensible set of relations. It also discusses certain issues that it leaves open, as they require further study in collaboration with other efforts, in particular with the European COST action TextLink.

Drawing on the commonalities found across existing frameworks, ISO DR-Core defines 20 core discourse relations and provides mappings of these relations to other annotation schemes. With respect to discourse structure, ISO DR-Core provides specifications for a low-level annotation of discourse relations, with the idea that (a) the description at this low level is what is well understood and can be unequivocally defined; (b) extensions to represent higher-level discourse structure will be possible where desired; and (c) it will allow for annotations to be compatible across frameworks, even when they are based on different theories of discourse structure.

The ISO 24617-8 core annotation scheme can be used in three different situations:

- for annotating discourse relations in natural language corpora;
- for defining mappings between annotations made using different frameworks or annotation schemes;
- as a target representation of automatic methods for shallow discourse parsing, for summarization, and for other NLP applications.

## 2. Basic concepts

This section provides a very brief comparison of the most important frameworks, focusing on those that have been used as the basis for annotating discourse relations in corpora, in particular, the theories of discourse coherence developed by Hobbs (Hobbs, 1990) and Kehler (1995); Rhetorical Structure Theory (Mann

<sup>1</sup>This paper may be regarded as an update and complement of Prasad and Bunt (2015), henceforth PB'15, which describes the state of developing ISO 24617-8 in early 2015.

and Thompson, 1988); the cognitive account of coherence relations by Sanders et al (Sanders et al., 1992); Segmented Discourse Representation Theory (Asher and Lascarides, 2003); and the annotation framework of the Penn Discourse Treebank (Prasad et al., 2008, 2014). The section ends with a summary of the main assumptions that underlie ISO DR-Core.

### 2.1. Representation of discourse structure

One difference between existing frameworks for representing discourse relations concerns the representation of structure. For example, the RST Bank (Carlson et al., 2003) assumes a tree representation to subsume the complete text of the discourse; the Discourse Graphbank (Wolf and Gibson, 2005), based on Hobbs' theory of discourse allows for general graphs that allow multiple parents and crossing, while the DISCOR corpus (Reese et al., 2007) and the ANNODIS corpus (Afantenos et al., 2012), based on SDRT (Asher and Lascarides, 2003), use directed acyclic graphs that allow for multiple parents, but not for crossing. Some frameworks are theory-neutral with respect to discourse structure, including the PDTB (Prasad et al., 2008) and DiscAn (Sanders and Scholman, 2012), both of which annotate individual relations and their arguments without combining these to form a structure that encompasses the entire text. *ISO DR-Core takes a theory-neutral stance, annotating only low-level discourse relations that can then be annotated further to project a higher-level tree or graph structure, depending on one's theoretical preferences.* (Note, however, that no constraints are assumed that would prevent selecting larger spans of text as realizations of arguments, including single- or multi-paragraph long text, or subdialogues – see Section 2.7.) From the point of view of interoperability, low-level annotation can serve as a pivot representation when comparing annotations based on different theories.

### 2.2. Semantic description of discourse relations

Some frameworks, such as SDRT, Hobbs' theory, PDTB, and Sanders et al's theory, describe the meaning of discourse relations in 'informational' terms, i.e., in terms of the content of the arguments; RST, on the other hand, provides definitions in terms of the intended effects on the hearer/reader. *In ISO DR-Core, discourse relation meaning is described in informational terms, with the idea that a mapping can be created from the ISO core relations to those present in various existing classifications, including those that define relations in intentional terms. These mappings are provided in Section 3.*

### 2.3. Pragmatic variants of discourse relations

With the exception of Hobbs (1990), all frameworks distinguish relations when one or both of the arguments involve an implicit belief or a dialogue act that takes

scope over the semantic content of the argument. This is motivated by examples like (1), where John's sending of the message did not cause him to be absent from work, but rather that it caused the speaker/writer to believe that John is not at work.

- (1) John is not at work today, because he sent me a message to say he was sick.

This distinction is known in the literature as the 'semantic-pragmatic' distinction (Sanders et al. (1992)); as the 'ideational-pragmatic' distinction in Redeker (1990); and as the 'content-metataalk' distinction in SDRT. Some frameworks, such as that of Sanders et al., allow this distinction for all relation types; others, like the PDTB and RST only admit it for some. Since we believe that the choice should in the end be determined by what is observed in corpus data *ISO DR-Core allows this distinction for all relation types. However, ISO DR-Core does not encode this distinction on the relation, but on the arguments of the relation*, because in all cases what is different is not the relation itself, but rather that the arguments require an inference of a belief or dialogue act that is implicit in the text. In the semantic representation of (1), for example, the two arguments related by a *Cause* relation are a belief (namely that John is not at work today) and an event, where the inferred belief concerns the first argument, not the relation between the arguments. The annotation scheme thus conforms to the *semantic* representation of the relation and its arguments.

### 2.4. Hierarchical classification of discourse relations

All existing frameworks group discourse relations together to a greater or lesser degree, but they differ in how the groupings are made. Reconciliation of groupings across frameworks is difficult, since they arise from differences in what is taken to count as semantic closeness. *The solution adopted in ISO DR-Core is to initially provide a 'flat' set of core relations.* In some cases, an ISO relation can turn out to be a more general case of more fine-grained relations in some other framework. As noted in Prasad and Bunt (2015), an advantage of a flat set is that it can serve as a pivot representation between frameworks, especially between those that group relations differently. A disadvantage, especially for the ISO 24617-8 set of core relations, is that in some cases a relation may turn out to be a more general case of more fine-grained relations in some framework. However, note that the ISO core relation set is part of an ongoing effort and we envisage further extensions to the relation set. Furthermore, an extensions to the core annotation scheme with a well-motivated taxonomical structure is planned to be elaborated in concertation with the TextLink project.

## 2.5. Representation of (a)symmetry of relations

Virtually all existing frameworks embody a representation of whether a discourse relation is symmetric or asymmetric; for example, the Contrast relation is symmetric whereas the Cause relation is asymmetric. Most annotation schemes encode asymmetry in terms of the textual ordering and/or syntax of the argument realizations. Thus, in Sanders et al's classification, where the argument span ordering is one of the 'cognitive' primitives underlying the scheme, the relation Cause-Consequence captures the 'basic' order for the semantic causal relation, with the cause appearing before the effect, whereas the relation Consequence-Cause is used for the reversed order. In the PDTB, argument spans are named Arg1 and Arg2 according to syntactic criteria, including linear order.

*In ISO DR-Core, annotations abstract over the linear ordering for argument realizations, since this is not a semantic distinction. Asymmetry is represented by specifying the argument roles in the definition of each relation, arguments bearing relation-specific semantic roles.* For example, in the Cause relation, defined as 'Arg1 provides a reason for Arg2' (see Table 1), the text span named Arg1 is the one that provides the reason in the Cause relation, irrespective of linear order or any other syntactic consideration, and Arg2 corresponds to what constitutes the result in the relation. This representation can be effectively mapped to other schemes for representing asymmetry, and in no way obfuscates the differences in linear ordering of the arguments, which is easily determined by pairing the argument role annotations with the text span annotations, as in the examples (2) - (5) in Section 4.3. Linear ordering has a bearing for claims that different versions of an asymmetric relation may not have the same linguistic constraints, for example, in terms of linguistic predictions for the discourse that follows (Asher et al., 2007).

## 2.6. Relative importance of arguments for text meaning/structure

Some frameworks, namely RST, Hobbs' theory, and SDRT distinguish relations or arguments in terms of their 'relative importance' for the meaning or structure of the text as a whole. In RST, one argument of an asymmetric relation is labeled the 'nucleus', whereas the other is labeled 'satellite'. *In ISO DR-Core, the relative role of arguments for the text (meaning or structure) as a whole is not represented directly, but because of the explicit identification of the roles of the arguments in each relation definition, such a layer of representation can be derived using the relation-specific argument roles.*

## 2.7. Syntactic form, extent and (non-)adjacency of arguments

Concerning the kinds of syntactic forms the realization of an argument can have, all frameworks agree that the typical realization of an argument is as a clause, but some allow for certain non-clausal phrases as well. *In ISO DR-Core, constraints are placed on the semantic nature of arguments rather than on their syntactic form. That is, an argument of a discourse relation must denote a certain type of abstract object.* Two related issues have to do with how complex the realizations of arguments can be syntactically, and whether the realizations should be adjacent in the discourse. *ISO DR-Core remains neutral on both these issues and does not specify any constraints on the extent or adjacency of argument realizations.*<sup>2</sup>

## 2.8. Summary: Assumptions of ISO DR-Core

In summary, the following basic concepts and assumptions underlie ISO DR-Core.

- A discourse relation is a relation expressed in discourse (written, spoken, or multimodal) between abstract objects, such as events, states, conditions, and dialogue acts.
- Discourse relations can be expressed explicitly in text/speech or can be implicit. The annotation of implicit relations may optionally include the specification of a connective that could express the inferred relation.
- A discourse relation takes two and only two arguments. Arguments can be shared by different relations.
- The meaning of discourse relations is described in informational terms.
- Pragmatic aspects of meaning involving beliefs and dialogue acts as arguments are represented as a property of arguments, rather than of discourse relations.
- Discourse relations are categorized as a flat set of relations.
- Annotations are at a low level; ISO DR-Core is agnostic towards the nature of the global structure of a text or dialogue.
- Asymmetrical relations are represented with relation-specific argument role labels.
- The relative importance of a relation's arguments with respect to the text as a whole is not represented as such.
- No a priori assumptions are made concerning constraints on syntactic form, syntactic complexity, or textual adjacency of expressions that may realize the arguments of a discourse relation.

<sup>2</sup>Despite the flexibility for these argument features in the current ISO model, we note that for a fully interoperable annotation scheme it is important for a consensus to be established for well-defined constraints on arguments.

	ISO DRel	Symmetry	Relation and Argument-Role Definitions
1.	Cause	Asymmetric	Arg1 provides a reason for Arg2 to come about or occur.
2.	Condition	Asymmetric	Arg1 is an unrealized situation which, when realized, would lead to Arg2.
3.	Negative Condition	Asymmetric	Arg1 is an unrealized situation which, when not realized, would lead to Arg2.
4.	Purpose	Asymmetric	Arg1 enables Arg2.
5.	Manner	Asymmetric	Arg1 is a way in which Arg2 comes about or occurs.
6.	Concession	Asymmetric	An expected causal relation between Arg1 and Arg2, where Arg1 is expected to cause Arg2, is cancelled or denied by Arg2.
7.	Contrast	Symmetric	One or more differences between Arg1 and Arg2 are highlighted with respect to what each predicates as a whole or to some entities they mention.
8.	Exception	Asymmetric	Arg1 evokes a set of circumstances in which the described situation holds, while Arg2 indicates one or more instances where it doesn't.
9.	Similarity	Symmetric	One or more similarities between Arg1 and Arg2 are highlighted with respect to what each predicates as a whole or to some entities they mention.
10.	Substitution	Asymmetric	Arg1 and Arg2 are alternatives, with Arg2 being the favored or chosen alternative.
11.	Conjunction	Symmetric	Arg1 and Arg2 bear the same relation to some other situation evoked in the discourse. Their conjunction indicates that they are doing the same thing with respect to that situation, or are doing it together.
12.	Disjunction	Symmetric	Arg1 and Arg2 are alternatives, with either one or both holding.
13.	Exemplification	Asymmetric	Arg1 describes a set of situations; Arg2 is an element of that set.
14.	Elaboration	Asymmetric	Arg1 and Arg2 are the same situation, but Arg2 contains more detail.
15.	Restatement	Symmetric	Arg1 and Arg2 are the same situation, but described from different perspectives.
16.	Synchrony	Symmetric	Some degree of temporal overlap exists between Arg1 and Arg2. All forms of overlap are included.
17.	Asynchrony	Asymmetric	Arg1 temporally precedes Arg2.
18.	Expansion	Asymmetric	Arg2 provides further description about some entity or entities in Arg1, expanding the narrative forward of which Arg1 is a part, or expanding on the setting relevant for interpreting Arg1. The Arg1 and Arg2 situations are distinct.
19.	Functional dependence	Asymmetric	Arg2 is a dialogue act with a responsive communicative function; Arg1 is the dialogue act(s) that Arg2 responds to.
20.	Feedback dependence	Asymmetric	Arg2 is a feedback act that provides or elicits information about the understanding or evaluation by one of the dialogue participants of Arg1, a communicative event that occurred earlier in the discourse.

Table 1: ISO set of core discourse relations

### 3. ISO Core Discourse Relations

Table 1 presents the set of core ISO discourse relations. The level of granularity is motivated by the consideration that these relations cover what has been more or less successfully implemented in various annotation efforts to date. However, this set is by no means fixed and can be augmented if necessary. As discussed in Section 2.5., the semantic roles of the arguments are built into the definition of each relation; labels for the semantic roles are listed in Table 2.

The set of ISO core discourse relations takes into account the work of annotating relations in (spoken) dialogue that has resulted in ISO standard 24617-2 for dialogue act annotation. The coherence relations that are found in written text are also found in spoken dialogue, both within speaker turns (where they contribute to the coherence of what is said in a turn) and between speaker turns (Petukhova et al., 2011; Riccardi et al., 2016; Tonnelli et al., 2010).<sup>3</sup> But more important for

the coherence of spoken dialogue is that the participants respond to each other. Many dialogue acts are inherently ‘responsive’ (or ‘backward-looking’), such as *Answer*, *Confirmation*, *Disconfirmation*, *Agreement*, *Disagreement*, *Correction*, *Accept Offer*, *Reject Suggestion*, *Address Request*, *Accept Apology*, and many others. Dialogue acts with such a function can only be understood in relation to what it is that they respond to. ISO 24617-2 therefore annotates such dialogue acts not only as having a certain communicative function, but also as having a ‘functional dependence’ relation with one or more previous dialogue acts. Similarly for feedback acts, which do not only have a feedback function but also a ‘feedback dependence’ relation to what it is that they provide or elicit information about (see Petukhova et al., 2011).

The annotation of coherence relations in (spoken) dialogue that have their basis in the use of ‘responsive’ dialogue acts is defined in the annotation scheme of ISO

<sup>3</sup>ISO DR-Core has been applied in the annotation of dis-

course relations in dialogues in the DialogBank; see Bunt et al., 2016).

24617-2. For example, the question-answer relation, which is sometimes considered as a discourse relation, is annotated as shown in (2), where speaker P2 answers a question by speaker P1. In (2b), which is represented in the XML-based format of the ISO 24617-2 Dialogue Act Markup Language (DiAML), the markables #s1 and #s3 identify the stretches of speech corresponding to P1's question and P2's answer, respectively. (See example (6) in Section 4.4 for the annotation of P1's second utterance.) The characterization of P2's contribution as being an answer in combination with the specification of the functional dependence relation with P1's question captures this coherence relation.

- (2) a. P1: Is it safe to put my camera through here?  
       It's a very expensive camera you know.  
       P2: Yes, that's perfectly safe.
- b. <dialogAct id="a1" target="#s1" sender="#p1" addressee="#p2" dimension="task" communicativeFunction="question" />  
    <dialogAct id="a3" target="#s3" sender="#p1" addressee="#p2" dimension="task" />  
    communicativeFunction="answer"  
    functionalDependence="#a1"/>

Tables (3) - (5) show equivalences between the ISO DR-Core relations and seven well-known taxonomies for discourse relations: RST and the RST Treebank; SDRT and the DISCOR and ANNODIS schemes; the PDTB taxonomy and the classification of Sanders et al.. It also draws on the experiences with discourse relation annotation in multiple languages and genres (Carlson et al., 2003; Wolf and Gibson, 2005; Prasad et al., 2008; Oza et al., 2009; Prasad et al., 2011; Zuferey et al., 2012; Zhou and Xue, 2012; Mladová et al., 2008; Afantenos et al., 2012; Sanders and Scholman, 2012), among others, and on other attempts to construct mappings between annotation schemes (Benamara and Taboada, 2015; Lapshinova et al., 2015). The correspondences shown are based on a comparison of the relation definitions provided in the various frameworks. From RST, we have also included a few of the presentational relations since they cover the same kinds of examples, although we note that the presentational type of meaning is described in RST to capture speaker intentions (i.e., speaker's belief of the intended effect on the hearer), so the correspondence with these presentational relations in RST is not strict. On the other hand, it may be possible to view this subset of the presentational relations as also subject-matter relations.

## 4. Annotation of Discourse Relations in XML

### 4.1. Overview

The annotations of discourse relations in ISO DR-Core are designed in accordance with ISO 24617-6,

Principles of semantic annotation<sup>4</sup>, which implements the distinction between annotations and representations that is made in the Linguistic Annotation Framework (ISO 24612). Accordingly, the definition of an annotation language consists of three parts:

1. an abstract syntax, which specifies a class of 'annotation structures' as set-theoretical constructs, independent of any particular representation format, in accordance with a conceptual view as expressed in a metamodel;
2. a formal semantics, describing the meaning of the annotation structures defined by the abstract syntax;
3. a concrete syntax, specifying a reference format for representing the annotation structures defined by the abstract syntax.

Abstract and concrete syntax are related through the requirements that the concrete syntax is *complete* and *unambiguous* relative to the abstract syntax. Completeness means that the concrete syntax defines a representation for every structure defined by the abstract syntax; unambiguity means that every expression defined by the concrete syntax represents one and only one structure defined by the abstract syntax. A representation format defined by a concrete syntax which has these two properties is called *ideal*. An important point of this approach is that *any ideal representation format is convertible through a meaning-preserving mapping to any other ideal representation format*.

Figure 1 presents the metamodel that expresses the conceptual view underlying ISO DR-Core and outline its abstract and concrete syntax. The semantics of ISO DR-Core annotations, which is defined through a translation into discourse representation structures (DRSs), is outlined in the appendix of PB'15.

Note that annotators only have to deal with the *concrete* DReML syntax; the abstract syntax mainly has a theoretical significance for proving the convertibility of the DReML scheme to other annotation schemes and representations and vice versa; see ISO 24617-2 and Bunt (2015). The semantics of the annotations, which is defined for the abstract syntax and is inherited by its concrete representations, is relevant for the extraction of content from DReML annotated resources and for inferring with DReML annotations.

### 4.2. Metamodel

Of central importance in the annotation of discourse relations are evidently the relations and their arguments, and they take central stage in the metamodel shown in Figure 1. Discourse relations are linked to relation arguments through argument roles. The arguments themselves can be of various types, as indicated by the link from relation arguments to argument types. ISO

<sup>4</sup>See Bunt (2015) for a summary description of ISO 24617-6.

	Discourse relation	Argument role labels
1	Cause	Reason, Result
2	Concession	Expectation-raiser, Expectation-denier
3	Elaboration	Broad, Specific
4	Restatement	n.a.
5	Condition	Antecedent, Consequent
6	Negative Condition	Negated-Antecedent, Consequent
7	Contrast	n.a.
8	Similarity	n.a.
9	Expansion	Foreground, Entity-description
10	Purpose	Goal, Enablement
11	Manner	Means, Achievement
12	Exception	Regular, Exclusion
13	Substitution	Disfavoured-alternative, Favoured-alternative
14	Conjunction	n.a.
15	Disjunction	n.a.
16	Exemplification	Set, Instance
17	Synchrony	n.a.
18	Asynchrony	Before, After
19	Functional dependence	Antecedent-act, Dependent-act
20	Feedback dependence	Feedback-scope, Feedback-act

Table 2: Role labels for arguments of ISO DR-Core discourse relations

ISO DR-Core	RST	RST Treebank
Cause	Vol. cause, Non-vol. cause, Vol. result, Non-vol. result, Evidence, Justify	Cause, Consequence, Result Evidence, Explanation-argumentation, Reason
Condition	Condition	Condition, Contingency, Hypothetical
Negative Condition	Otherwise	Otherwise
Purpose	Purpose	Purpose
Manner	–	Manner, Means
Concession	Concession	Concession, Antithesis, Preference
Contrast	Contrast	Comparison
Exception	–	–
Similarity	–	Analogy, Proportion
Substitution	Antithesis	–
Conjunction	Joint	List
Disjunction	Joint	Disjunction
Exemplification	Elaboration (set-member)	Elaboration set-member, Example
Elaboration	Elaboration (general-specific, whole-part, Elaboration (abstract-instance, process-step)	Conclusion, Elaboration-general-specific, Conclusion, Elaboration-general-specific, Elaboration-part-whole, Elaboration-process-step, summary
Restatement	Restatement	–
Synchrony	–	Temporal-same-time
Asynchrony	Sequence	Temporal-before, Temporal-after, Sequence, Inverted-sequence
Expansion	Elaboration (object-attribute)	Elaboration object-attribute, Elaboration additional

Table 3: Mapping between discourse relations in ISO DR-Core, RST, and RST Treebank

DR-Core assumes that two types of arguments have to be distinguished (possibly with subtypes): ‘situations’, which include eventualities (events, states, processes,...), facts, conditions, as well as negated eventualities (as in “*Mary smiled at John, but she didn’t smile back*”), and dialogue acts involved in ‘pragmatic’ interpretations of discourse relations (as in “*Carl is a fool; he beats his wife*”) (cf. Ginzburg, 2011).

The fact that a discourse relation can be explicit or im-

plicit is reflected in the indication ‘0..1’ at the tip of the arrow from discourse relations to markables. The dotted arrows at the bottom indicate possible links to another layer of annotation, concerned with the identification of the source to which a discourse relation or (one or both of) its arguments may be attributed.

#### 4.3. Abstract and concrete syntax

The abstract syntax of ISO DR-Core annotations consists of (a) a specification of the concepts from which



ISO DR-Core	SDRT	DISCOR	ANNODIS
Cause	Explanation, Result	Explanation, Result	Explanation, Result
Condition	Consequence	Consequence	Conditional
Negative Condition	Consequence	Consequence	Conditional
Purpose	Explanation	Explanation	Goal
Manner	Elaboration	Elaboration	Elaboration
Concession	Contrast	Contrast	Contrast
Contrast	Contrast	Contrast	Contrast
Exception	–	–	–
Similarity	Parallel	Parallel	Parallel
Substitution	–	–	–
Conjunction	Continuation	Continuation	Continuation
Disjunction	Alternation	Alternation	Alternation
Exemplification	Elaboration	Elaboration	Elaboration
Elaboration	Elaboration	Elaboration	Elaboration
Restatement	Elaboration	Elaboration	Elaboration
Synchrony	–	–	–
Asynchrony	Narration	Narration, Precondition	Narration, Flashback
Expansion	Background, Elaboration	Background, Elaboration Commentary	Background, Entity-Elaboration Comment
–		Attribution, Source	Attribution, Frame, Temporal-location

Table 4: Correspondences between ISO DR-Core, SDRT, DISCOR and ANNODIS

ISO DR-Core	PDTB	Sanders et al/DiscAn
Cause	Reason, Result, Justification	Causal-Semantic-Basic-Positive Causal-Semantic-NonBasic-Positive Causal-Pragmatic-Basic-Positive Causal-Pragmatic-NonBasic-Positive
Condition	Hypothetical, General, UnrealPast, UnrealPresent, FactualPast, FactualPresent	Causal-Semantic-Basic-Positive Causal-Semantic-NonBasic-Positive Causal-Pragmatic-Basic-Positive Causal-Pragmatic-NonBasic-Positive
Negative Condition	Condition	–
Purpose	Result	Causal-Pragmatic-Basic-Positive Causal-Pragmatic-NonBasic-Positive
Manner	– –	AdditiveSemantic-Basic-Positive AdditiveSemantic-NonBasic-Positive
Concession	Expectation, Contra-Expectation	Causal-Semantic-Basic-Positive , Additive-Semantic-Negative
Contrast	Juxtaposition, Opposition	Additive-Semantic-Negative
Exception	Exception	Additive-Semantic-Negative
Similarity	Conjunction	Additive-Semantic-Positive
Substitution	Chosen Alternative	Additive-Semantic-Negative
Conjunction	Conjunction, List	Additive-Semantic-Positive
Disjunction	Disjunctive, Conjunctive	Additive-Semantic-Negative
Exemplification	Instantiation	Additive-Semantic-Positive
Elaboration	Generalization, Specification	Additive-Semantic-Positive
Restatement	Equivalence	–
Synchrony	Synchronous	–
Asynchrony	Precedence, Succession	–
Expansion	EntRel	Additive-Semantic-Positive

Table 5: Correspondences between ISO DR-Core, the PDTB, and Sanders et al./DiscAn

annotations are built up, and (b) a specification of the possible ways of combining these elements into annotation structures.

An annotation structure is a set of *entity structures*, which contain semantic information about a region of primary data, and *link structures*, which describe a se-

mantic relation between two such regions. An entity structure is either (1) a relation entity structure, which is a pair  $\langle m_i, r_j \rangle$  consisting of a markable  $m_i$ , and a discourse relation  $r_j$ , or (2) an argument entity structure, which is a pair  $\langle m_k, t \rangle$  consisting of a markable and an argument type. A link structure

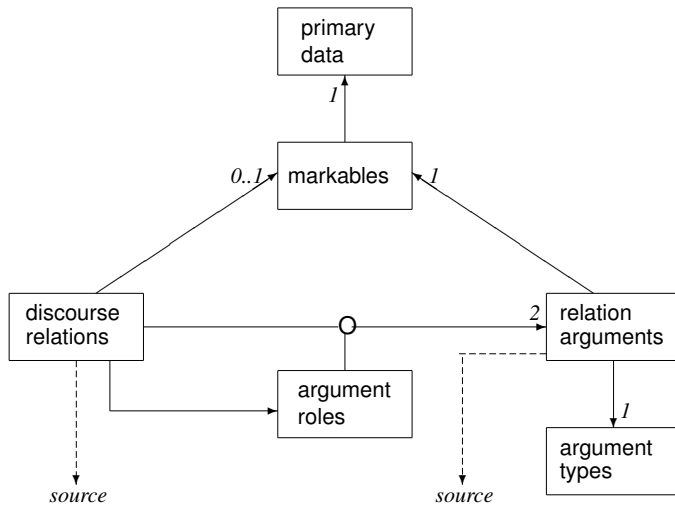


Figure 1: Metamodel for the annotation of discourse relations.

captures the information that two arguments participates in a discourse relation in a certain roles, such as a triple  $\langle \rho_{cause}, \varepsilon_1, \varepsilon_2 \rangle$  consisting of a relation entity structure  $\rho_{cause} = \langle m, cause \rangle$  and two argument entity structures, participating in the argument roles  $\alpha(cause) = \langle reason, result \rangle$  as defined by the argument role assignment function  $\alpha$  which is specified in the abstract syntax (see Table 2 for this specification).

#### DReML representation

The DReML concrete syntax uses four special XML elements in order to allow compact representations:

1. the elements `dRel` and `drArg` are defined for representing discourse connectives and their arguments, respectively;
2. the elements `explDRLink` and `implDRLink` are defined for representing explicit and implicit discourse relations, respectively.

The examples (3) and (4) illustrate the use of DReML to represent the annotation of an explicit and an implicit *Cause* relation, respectively. The markables `m1` and `m3` correspond to the clauses “*John fell*” and “*Bill pushed him*”; the markable `m2` corresponds to the discourse connective “*because*” in (3). Sequences like “`arg1=#e1 arg1Role=result`” support an annotation like (3) to be semantically interpreted as the DRS  $\langle r, x, y, cause(r), reason(r, x), result(r, y) \rangle$ . For clarity the argument type ‘event’ is specified in these examples (a subtype of ‘situation’), but this value may be left unspecified, which is interpreted as not requiring an inferred belief (in contrast with example (1)).

- (3) John fell because Bill pushed him.  
`<drArg xml:id="e1" target="#m1" type="event"/>`  
`<dRel xml:id="r1" target="#m2" rel="cause"/>`  
`<drArg xml:id="e2" target="#m3" type="event"/>`  
`<explDRLink rel="#r1" result="#e1" reason="#e2"/>`

- (4) John fell. Bill pushed him.  
`<drArg xml:id="e1" target="#m1" type="event">`  
`<drArg xml:id="e2" target="#m3" type="event"/>`  
`<implDRLink rel="cause" result="#e1"`  
`reason="#e2"/>`

Note that the representations using this DReML form are just an abbreviation of a standard XML expression, such as the following representation of (3):

- (5) John fell because Bill pushed him.  
`<fs xml:id="e1">`  
`<f name="target"><value="#m1"/></f>`  
`<f name="type"><value="event"/></f>`  
`</fs>`  
`<fs xml:id="r1">`  
`<f name="target"><value="#m2"/></f>`  
`<f name="rel"><value="cause"/></f4;4`  
`</fs>`  
`<fs xml:id="e2">`  
`<f name="target"><value="#m3"/></f>`  
`<f name="type"><value="event"/></f>`  
`</fs>`  
`<fs xml:id="dr1">`  
`<f name="result"><value="#e1"/></f>`  
`<f name="reason"><value="#e2"/></f>`  
`<f name="rel"><value="#r1"/></f>`  
`</fs>`

#### 4.4. Annotation of discourse relations in dialogue

The discourse relations defined in ISO DR-Core are not only relevant within the ISO DR-Core annotation scheme, but can also be used to annotate rhetorical relations in spoken or multimodal dialogue according to the ISO 24617-2 annotation scheme for dialogue act annotation (using the DiAML markup language). The following example illustrates this.<sup>5</sup> The speaker

<sup>5</sup>For more examples see the DialogBank resource at <https://dialogbank.uvt.nl>

in (6a) first asks whether it is safe to put his camera through the X-ray machine at an airport security check and subsequently motivates his question by telling that his camera is a very expensive one. Following ISO 24617-2 this can be annotated as in (6b), where #p1 and #p2 indicate the two participants, and the markables #s1 and #s2 identify the functional segments "Is it safe to put my camera through here" and "It's a very expensive camera you know", respectively. A <rethoricalLink> element relates the *Inform* act to the *Question* act as its 'antecedent' through a *Cause* relation, indicating moreover that the *Inform* act is the *reason* argument of that relation. If the order of the two dialogue acts would be the other way round, as in "I have a very expensive camera. Is it safe to put that through here?", then the rhetorical relation would be annotated as "cause\_result".

- (6) a. Is it safe to put my camera through here? It's a very expensive camera you know.  
 b. <dialogAct id="a1" target="#s1" sender="#p1" addressee="#p2" dimension="task" communicativeFunction="question" />  
 <dialogAct id="a2" target="#s2" sender="#p1" addressee="#p2" dimension="task" />  
 <communicativeFunction="inform" />  
 <rethoricalLink dact="#a2" rhetoAntecedent="#a1" rhetoRel="cause.reason" />

Note that DiAML representations like the one shown here are a compact form of XML, abbreviating more lengthy standard XML expressions, just like a DRelML representation such as (3) abbreviates (5).

The representation in (6) illustrates the DiAML-representation of a 'pragmatic' *Cause* relation among dialogue acts. Dialogue acts are related though a 'semantic' *Cause* relation if there is a causal relation between their respective semantic contents. The DiAML representation with 'rhetorical links' cannot distinguish between 'pragmatic' and 'semantic' variants of a discourse relations. By combining the annotation schemes of ISO 24617-2 and ISO DR-Core, in particular of their unabbreviated representations (like (5)), we can both distinguish 'pragmatic' and 'semantic' variants as well as 'mixed' variants of a discourse relations without needing specifications like 'argType="dialogueAct"'. To avoid a length XML expression, we illustrate this by combining in (7b) the abbreviations of DRelML and DiAML to compactly represent a 'mixed' cause relation in the sense of the fact they "they don't have a fixed place" is why P1 says that he can never find them (where markable s3 identifies the word "because" and s4 the stretch "they don't have a fixed place").

- (7) a. P1: I can never find my remote control.  
 P2: That's because they don't have a fixed place.  
 b. <dialogAct id="a1" target="#s1" sender="#p1" addressee="#p2" dimension="task"

```
communicativeFunction="inform" />
<dialogAct id="a2" target="#s2" sender="#p2"
  addressee="#p1" dimension="task" />
  communicativeFunction="inform" />
<dRel xml:id="r1" target="#s2" rel="cause"/>
<drArg xml:id="e2" target="#s4" />
<explDRLink rel="#r1" result="#da1"
  reason="#e2"/>
```

## 5. Concluding Remarks

In this paper we have summarized ISO 24617-8 ('ISO DR-Core'), the first part of an effort to establish an interoperable annotation scheme for semantic relations in discourse. On the basis of an analysis of a range of theoretical approaches and annotation efforts a clear delineation of the scope of the ISO effort was made, restricting the effort to local, low-level relations with a solid theoretical and empirical basis. The ISO principles for linguistic annotation in general and semantic annotation in particular were applied to design annotations and representations for discourse relations. Future work will aim to remove some of the limitations, in particular aiming to develop a well-motivated taxonomy of discourse relations in collaboration with the TextLink project.

## Acknowledgement

This work was partially supported by NSF grant IIS-1421067.

## References

- Afantenos, S., N. Asher, F. Benamara, M. Bras, C. Fabre, L.-M. Ho-Dac, A. L. Draoulec, P. Muller, M.-P. Pry-Woodley, L. Prvot, J. Rebeyrolle, L. Tanguy, M. Vergez-Couret, and L. Vieu (2012). An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*.
- Asher, N. and A. Lascarides (2003). *Logics of conversation*. Cambridge University Press.
- Asher, N., L. Prévot, and L. Vieu (2007). Setting the background in discourse. *Discours. Revue de linguistique, psycholinguistique et informatique* (1).
- Benamara, F. and M. Taboada (2015). Mapping different rhetorical relation annotations: A proposal. In *Proceedings of the 4th Joint Conference on Lexical and Computational Semantics (\*SEM)*, Denver.
- Bunt, H. (2015). On the principles of interoperable semantic annotation. In *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, pp. 1–13.
- Bunt, H., V. Petukhova, K. Wijnhoven, A. Fang, and A. Malchanau (2016). The DialogBank. In *Proceedings LREC 2016, Portoroz*, Paris. ELDA.

- Carlson, L., D. Marcu, and M. E. Okurowski (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In J. van Kuppevelt and R. Smith (Eds.), *Current Directions in Discourse and Dialogue*, pp. 85–112. Kluwer Academic Publishers.
- Crible, L., L. Degand, and A.-C. Simon (2016). Interdependence of annotation levels in a functional taxonomy for discourse markers in spoken corpora. Paper presented at the 2nd TextLink Action Conference. Budapest.
- Ginzburg, J. (2011). *Situation semantics and the ontology of natural language*. Berlin: De Gruyter.
- Hobbs, J. R. (1990). *Literature and Cognition*. Menlo Park, Cal.: CSLI/SRI.
- ISO (2016a). *ISO 24617-6, Semantic annotation framework (SemAF) Part 6, Principles of semantic annotation*. Geneva: ISO.
- ISO (2016b). *ISO 24617-8, Semantic annotation framework (SemAF) Part 8, Semantic relations in discourse*. Geneva: ISO.
- Lapshinova, E., A. Nedoluzhko, and K. Kunz (2015). Cross languages and genres: Creating a universal annotation scheme for textual relations. In *Proceedings of the Workshop on Linguistic Annotations, NAACL-2015*, Denver, CO.
- Mann, W. C. and S. A. Thompson (1988). Rhetorical structure theory. Toward a functional theory of text organization. *Text* 8(3), 243–281.
- Mladová, L., S. Zikanova, and E. Hajicová (2008). From sentence to discourse: Building an annotation scheme for discourse based on Prague Dependency Treebank. In *Proc. 6th Int. Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech.
- Muller, P. and L. Prévot (2008). The rhetorical attachment of questions and answers. In *Meaning, Intentions, and Argumentation. CSLI Lecture Notes 186*. University of Chicago Press.
- Oza, U., R. Prasad, S. Kolachina, S. Meena, D. M. Sharma and A. Joshi (2009). Experiments with annotating discourse relations in the Hindi Discourse Relation Bank. In *Proceedings of the 7th International Conference on Natural Language Processing (ICON-2 009)*, Hyderabad, India.
- Petukhova, V., L. Prévot, and H. Bunt (2011). Discourse relations in dialogue. In *Proceedings 6th Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation (ISA-6)*, Oxford, pp. 18–27.
- Prasad, R. and H. Bunt (2015). Semantic relations in discourse: The current state of ISO 24617-8. In *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, London, U.K., pp. 80–92.
- Prasad, R., N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of 6th International Conference on Language Resources and Evaluation (LREC2008)*.
- Prasad, R., S. McRoy, N. Frid, A. Joshi, and H. Yu (2011). The biomedical discourse relation bank. *BMC bioinformatics* 12(1), 188.
- Prasad, R., B. Webber, and A. Joshi (2014). Reflections on the Penn Discourse Treebank, comparable corpora, and complementary annotation. *Computational Linguistics* 40(4), 921–950.
- Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14(3), 367–381.
- Reese, B., P. Denis, N. Asher, J. Baldridge, and J. Hunter (2007). Reference manual for the analysis and annotation of rhetorical structure. Unpublished Ms. <http://comp.ling.utexas.edu/discor/>.
- Riccardi, G., E. Stepanov, and S. Chowdhury (2016). Discourse connective detection in spoken conversation. In *Proceedings 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)*.
- Sanders, T. J. and M. Scholman (2012). Categories of coherence relations in discourse annotation. Presented at the International Workshop on Discourse Annotation. Utrecht Institute of Linguistics, Universiteit Utrecht.
- Sanders, T. J. M., W. P. M. Spooren, and L. G. M. Noordman (1992). Toward a taxonomy of coherence relations. *Discourse Processes* 15, 1–35.
- Tonelli, S., G. Riccardi, R. Prasad, and A. Joshi (2010). Annotation of discourse relations for conversational spoken dialogs. In *Proc. 7th Int. Conference on Language Resources and evaluation (LREC 2010)*.
- Wolf, F. and E. Gibson (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics* 31(2).
- Zhou, Y. and N. Xue (2012). PDTB-style discourse annotation of Chinese text. In *Proceedings 50th Annual Meeting of the Association for Computational Linguistics*, pp. 69–77.
- Zufferey, S., L. Degand, A. Popescu-Belis, T. Sanders, et al. (2012). Empirical validations of multilingual annotation schemes for discourse relations. In *Proceedings 8th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-8)*, pp. 77–84.

# Feedback Matters: Applying Dialog Act Annotation to Study Social Attractiveness in Three-Party Conversations

Benjamin Weiss, Stefan Hillmann

Technische Universität Berlin  
Ernst-Reuter-Platz 7, 10587 Berlin, Germany  
benjamin.weiss@tu-berlin.de, stefan.hillmann@tu-berlin.de

## Abstract

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.

**Keywords:** back-channel, conference calls, likability

## 1. Introduction

Social attractiveness between interlocutors – whether we like or dislike somebody – represents one fundamental interpersonal attitude and is typically formed quickly as part of a first global impression. Such a first impression is important for the development of a relationship in the future (Levinger and Snoek, 1972) and quite persistent and reliable (Ambady and Skowronski, 2008; Curhan and Pentland, 2007; Harris and Garris, 2008). For the case of vocal signals, very brief speech samples of a few seconds of duration already result in consistent ratings of likability on the listeners side (Gravano et al., 2011; Weiss, 2015). However, social attractiveness is mostly studied for established relationships. For well acquainted people, some key influencing factors like physical attractiveness, reciprocal liking, similarity, or proximity have already been identified (Aronson et al., 2009). Also, personality seems to play a role, for both, acquainted (van der Linden et al., 2010) and unacquainted people (Back et al., 2011).

For non-acquainted people, not only the face, voice and conversational dynamics seem to affect social attraction (Ambady and Skowronski, 2008; Curhan and Pentland, 2007), but also the body movements, friendliness of facial expressions, strength of voice, or originality of verbal content, e.g., cues of an extrovert personality (Back et al., 2011).

Such results indicate the relevance of surface cues for the formation of social attractiveness in conversation. Accordingly, also verbal behavior in conversation has been studied on the surface level: Numbers of filled pauses and contractions, for example, correlate positively with 3rd party ratings of recorded conversations, whereas numbers of interjections, interruptions, and – surprisingly – back-channels correlate negatively (Gravano et al., 2011). For interlocutors with just the role of a follower in these conversations, taking turns by causing overlap instead of a pause are rated more positively.

In contrast to these results, social attractiveness correlates positively with the number of back-channels for the role of the caller in dyadic telephone calls (Vinciarelli et al., 2011). The concept of cohesiveness is related to social attractiveness, and was studied for the four-person scenarios of the AMI corpus (Lai et al., 2013). Within a number of extracted parameters, averaged post-meeting ratings of cohesiveness are negatively correlated with the number of interruptions, proportion of silence in turn-taking (both in line with Gravano et al. (2011)). In this study, also dialog acts were annotated. Cohesiveness is negatively correlated with the number of the dialog act “eliciting information”, and positively with “providing assessments” and “comments about understanding”.

Apart from Lai et al. (2013), dialog acts are usually not operationalized for studying the effect of conversational strategies/behavior on subjective ratings. Either the aforementioned surface parameters are extracted (semi-) automatically, or a qualitative approach is chosen, applying external ratings of communicative style (Brandt, 1979; Norton and Pettegrew, 1977) or instructing a confederate interlocutor (Goldbrand, 1981; Baker and Ayres, 1994; Kohn and Dipboye, 1998). Therefore, the aim of this paper is to continue the approach of Lai et al. (2013) by studying the relationship between selected dialog acts and ratings of social attractiveness by applying the ISO 24617-2:2012 (2012) to scenario-based three-party-telephone conferences.

## 2. Conversation Experiment

A laboratory situation was considered to be most appropriate in order to collect likability ratings and parameters of interaction behavior in a controlled manner with regard to situation and topics, but also to ensure high recording quality for manual annotation. Three persons were recorded, verbally interacting according to prepared scenarios. The expected benefit of triads over dyads is the possibility to

compare pair-wise effects to the likability of a speaker averaged over both interlocutors.

Altogether, 39 persons took part in this experiment (9 women, 30 men, aged 36.2 years, SD=12.2). The triads did not know each other in advance. The participants were all experienced in telephone conferences to ensure familiarity with the situation and technology. As a requirement, all had conducted at least three conferences within the last 12 months. On average, the participants stated to have conducted 34.7 telephone conferences during their lifetime. All participants of a group were instructed to the procedure together and then sent to their individual sound proofed room (ITU-T Rec. P.800, 1996). From this personal meeting, participants gained a first impression of each other including visual information.

The three rooms were connected by a conferencing system implemented in PureData (Puckette, 2007). It provided broadband connection with intensity attenuation of 23.1 dB SPL (test signal of 61.3 dB SPL). Closed headsets of the type Beyerdynamics DT 290 were used by the participants. Prior to the actual session a first training scenario was conducted. Following this, issues concerning the procedure could be clarified, but the subjects did not leave their individual room. Nine different scenarios were taken from the collection described in ITU-T P Suppl. 26 (2012): These semi-structured tasks provide business conversations with topics like choosing a conference location or songs of a music album. Each conversation was initiated by a melody, as the experimental set-up did not allow for a real call initiation.

For each scenario, every participant receives different information to contribute or ask for in order to stimulate the conversation. All aspects are “solved” in this manner by the three participants. Although the scenarios provide various job descriptions, no specific (conversational) roles are defined. During annotation, we gained the impression that roles of leading or following the conversations was taken individually, if at all, and its conversational consequences are thus reflected in the resulting conversational parameters. The training scenario was the same for all groups, whereas the actual scenario accounting for the analysis varied. The conversations analyzed here are only the first block of an experiment with optimal network conditions. The later blocks investigated the effects of transmission delay (Schoenberg, 2015), which is why the scenario was randomized.

After the conversation, each participant was asked to state each partner’s likability, as well as personal attention to the call and overall quality of the transmission. The likability ratings after the training are used as basis to control for the first impression established so far. The scale used was continuous with the two antonyms “very likable” and “very unlikeable” afterwards transformed to numeric values between zero and ten.

Unfortunately, some participants failed to rate the likability in some cases; maybe due to the cover story of speech quality assessment: For some data, participants did not fill-in the likability scale: From the 39 participants, 3 interlocutors did not rate at all, and 4 participants missed one rating. Thus, the final data set comprises 68 data points to be re-

lated to dialog act frequencies.

### 3. Dialog Act Annotation

Based on the provided ELAN<sup>1</sup> literal transcriptions (Weiss and Schoenberg, 2014), manual annotations of dialog act classes were conducted. As we applied the ISO recommendation (ISO 24617-2:2012, 2012) for the first time, we decided against separate annotations and  $\kappa$  values, but for consequent discussions and counter-insurance to ensure a single strategy. Two students of linguistics conducted the annotation under supervision of Weiss and Schoenberg (2014).

Based on the referenced literature, feedback and meta-communication was in focus of this study. Therefore, the annotation included the following dimensions as separate ELAN tiers with the specified labels of communicative functions. All labels include identifiers of the speakers to separate the three interlocutors, and in special cases also the ID of the addressee:

1. Allo-Feedback
  - Allo Positive
  - Allo Negative
2. Auto-Feedback 1
  - Auto Positive
  - Auto Negative
3. Auto-Feedback 2
  - Auto-Summary<sup>2</sup>
  - Auto-Repetition<sup>3</sup>
4. Turn Management 1
  - *NEW*: Failed-Turn-Grab<sup>4</sup>
  - Turn-Assign (ID1 to ID2)
5. Turn Management 2
  - *NEW*: “Self-Induced Stop”<sup>5</sup>
  - Turn-Grab (ID1 to ID2)
6. Own- and Partner-Communication-Management
  - Own-Communication-Management<sup>6</sup>
  - Partner-Communication-Management<sup>7</sup> (ID1 to ID2)
7. Contact-Management
  - Contact Check
  - Contact Indication

Dimension 7 was solely annotated for the purpose to test for delay-induced differences (Schoenberg (2015), Chap. 7), and is not analyzed here. Some dimension were annotated on two separate tiers to facilitate the work flow.

<sup>1</sup>EUDICO Linguistic Annotator:  
<http://tla.mpi.nl/tools/tla-tools/elan/>.

<sup>2</sup>“Rephrasing” ISO 24617-2:2012 (2012), p. 39.

<sup>3</sup>“Echo” ISO 24617-2:2012 (2012), p. 39.

<sup>4</sup>A /turnKeep/ with attempted Turn-Grab.

<sup>5</sup>Ending a turn, not /retraction/.

<sup>6</sup>Self-Correction or Retraction.

<sup>7</sup>Correct-Misspeaking or Completion.

However, during the planning phase of the annotation, not only the dimensions and functions to annotate were defined. But also, it was decided to define two new labels in order to cover aspects of interest. This may be caused by the limited practical experience in applying the ISO standard. The dialog act of *failed Turn-Grab* was introduced to cover both successful and not-successful interruptions with the goal of taking the floor. The */turnKeep/* function was not annotated, and seems to cover in particular pre-planned keeps of the turns, instead of reactions to */turnGrabs/*. The *Self-Induced Stop* was defined to cover all instances of unfinished turns, for which the repairing function */retraction/* was not considered appropriate, as this label indicates withdrawing from the own contributions in the same turn.

Deciding on the extent of annotations and applying the ISO norm was a challenging task indeed. During this process and during the first annotation session, several adjustments were made and examples collected as references. In particular, the semantic annotations turned out to require much more resources in comparison to annotations of surface structures (such as speaker changes w/o overlapped speech, identifying back-channeling from turn, transcriptions).

#### 4. Results

Annotations were exported from ELAN to R for analysis, along with the likability ratings. Interestingly, there is no relevant agreement between two raters on the third interlocutor (Intra-Class-Correlation,  $ICC=.23$ ,  $p=.11$ ). This holds also for reciprocal ratings ( $ICC=-.26$ ,  $p=.98$ ), which is unexpected, as it is not in line with results for reciprocal social attraction (Aronson et al., 2009; Kenny, 1996). As a result, interdependence does not have to be considered (Kenny, 1996), which is why we apply simple linear models.

The lacking consistency between the two raters may be caused by individual interpretations of the scale, as with only two ratings per participant, no normalization can be conducted. Still, no averaging of the ratings was conducted, but instead each rating was taken into account individually as dependent variable, although most extracted parameters are thus doubled.

According to Kenny (1996), perceiver effects (variances in ratings between raters) should to be separated from target effects (variances of ratings between targets). Although, his model requires at least groups of four people, this basic idea is implemented here, as observable behavior like dialog acts can be considered either as externalizations of a rater, or of a target (Back et al., 2011).

In order to inspect the data, initial linear models were fitted including all 11 dialog act classes. The model with likability ratings associated to the target (Is there a relation between someone *rated* as likable and his/her dialog acts?) is not significant ( $F(11,56)=0.93$ ;  $p=.52$ ). However, associating likability ratings to the rater (Is there a relation between someone *rating* positively and his/her own dialog acts?) gives significant results (Table 1). As this model is just not significant ( $F(11,56)=1.96$ ;  $p=.05$ ), stepwise inclusion was conducted applying the AI-Criterion, which results in a new model with four predictors (Table 2). The second model is significant ( $F(4,63)=4.74$ ;  $p=.002$ ) and has

the same three significant variables in common with the first model. It explains 23% of the variance.

Predictor	t-value	p-value
(Intercept)	1.04	0.085.
Allo-feedback positive	-0.47	0.007**
Allo-feedback negative	-1.76	0.020*
Auto-feedback positive	0.08	0.026*
Auto-feedback negative	0.25	0.322
Auto-Repetition	-0.14	0.178
Auto-Summary	0.20	0.520
Self-Stop	-0.31	0.240
Turn-Grab	-0.13	0.581
Failed-Turn-Grab	0.01	0.957
Own Communic. Man.	0.02	0.634
Partner Communic. Man.	0.26	0.517

Table 1: Linear model of perceiver ratings with all dialog act classes.

Predictor	t-value	p-value
(Intercept)	3.688	0.000474***
Allo-feedback positive	-3.24	0.002**
Allo-feedback negative	-2.92	0.005**
Auto-feedback positive	3.36	0.001**
Auto-Repetition	-1.75	0.085.

Table 2: Linear model with step-wise inclusion (AI-Criterion).

#### 5. Discussion and Conclusion

This exploratory study aims at finding candidates from hand-annotated dialog acts correlating with likability ratings. As a result, a model was found for the ratings given by the interlocutors. The number of conversations analyzed is rather low for such a statistical approach, which might cause the low amount of variance explained. That the raters' data shows a relationship with observable behavior is in line with results from Back et al. (2011), who found that extra-version and self-centering of the raters correlates positively with the ratings giving to others. Therefore, investigating the relationship between dialog act occurrences and personality as well as likability will be conducted in the future.

From the dialog acts taken into account, numbers of allo- and auto-feedback were found as possible candidates to likers' ratings. This kind of feedback annotated here was not considered in an earlier analysis, where only surface information were available, not separating different kind of feedback (Weiss and Schoenberg, 2014). There, only the target ratings were analyzed, for which turn-changes with overlap correlated positively and amount of turns and speaking time negatively, suggesting a positive impact of smooth turn-changes and non-dominant or efficient conversational participation.

Positive Auto-feedback indicates successful communication and was thus expected to show a positive relation

with likability. However, the negative impact of Allo-feedback might either reflect an attitude or stance towards the interlocutors that is negatively interpreted in these very easy tasks, or it might hint at earlier communication problems. Likability ratings as overall measure of social attractiveness cannot be analyzed further on this matter, but as social attractiveness is proposed to be also affected by physical attraction and task-attraction (for collaborative settings), a more elaborate questionnaire should be applied in the follow-up experiment (McCroskey and McCain, 1974). Another issue is to properly entangle cause and effects for conversational behavior affecting liking or already reflecting it during the course of a conversation, which is not possible by simply aggregating over time.

The broader aim of this approach is to explain the interrelationship between social attractiveness and conversational behavior. Counts of dialog act occurrences could also be used to identify conversational strategies/styles in dialog and be related to personality of the interlocutors. For this, 3rd-party evaluations of the recorded and annotated conversations would be required along with self-assessments in the future, including annotation of more dimensions than presented here. From a practical point of view, audio calls represent an ideal domain for this subject as recruiting participants from different locations is easy and would allow to control for acquaintance and visual first impressions.

## 6. Acknowledgments

We want to thank Charlotte Buchsbaum and Anne Katzur for their tremendous support in annotating the database. This work was financially supported by the Deutsche Forschungsgemeinschaft DFG (German Research Community), grant WE 5050/1-1.

- Nalini Ambady et al., editors. (2008). *First Impressions*. Guilford Press, New York.
- Aronson, E., Wilson, T., and Akert, R. (2009). *Social Psychology*. Prentice Hall, 7 edition.
- Back, M., Schmuckle, S., and Egloff, B. (2011). A closer look at first sight: Social relations lens model analysis of personality and interpersonal attraction at zero acquaintance. *European Journal of Social Psychology*, 25:225–238.
- Baker, A. and Ayres, J. (1994). The effect of apprehensive behavior on communication apprehension and interpersonal attraction. *Communication Research Reports*, 11:45–51.
- Brandt, D. (1979). On liking social performance with social competence: Some relations between communicative and attributions of interpersonal attractiveness and effectiveness. *Human Communication Research*, 5:223–226.
- Curhan, J. and Pentland, A. (2007). Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92:802–811.
- Goldbrand, S. (1981). Imposed latencies, interruptions and dyadic interaction: Physiological response and interpersonal attraction. *Journal of Research in Personality*, 15:221–232.
- Gravano, A., Levitan, R., Willson, L., Beňuš, Š., Hirschberg, J., and Nenková, A. (2011). Acoustic and prosodic correlates of social behavior. In *Proc. Inter-speech*, pages 97–100.
- Harris, M. and Garris, C. (2008). You never get a second chance to make a first impression: behavioral consequences of first impressions. In N. Ambady et al., editors, *First Impressions*, pages 147–170. Guilford Press, New York.
- ISO 24617-2:2012. (2012). Language resource management – Semantic annotation framework (SemAF), Part 2: Dialogue acts.
- ITU-T P Suppl. 26. (2012). Scenarios for the subjective evaluation of three-party audio telemeetings quality. International Telecommunication Union, Geneva.
- ITU-T Rec. P.800. (1996). Methods for subjective determination of transmission quality. International Telecommunication Union, Geneva.
- Kenny, D. (1996). Models of non-independence in dyadic research. *Journal of Social and Personal Relationships*, 13:279–294.
- Kohn, L. and Dipboye, R. (1998). The effect on interview structure on recruiting outcomes. *Journal of Applied Social Psychology*, 28:821–843.
- Lai, C., Carletta, J., and Renals, S. (2013). Modelling participant affect in meetings with turn-taking features. In *Proc. Workshop of Affective Social Speech Signals*.
- Levinger, G. and Snoek, J. (1972). *Attraction in relationship: A new look at interpersonal attraction*. General Learning Press, Morristown, N.J.
- McCroskey, J. and McCain, T. (1974). The measurement of interpersonal attraction. *Speech Monographs*, 41:261–266.
- Norton, R. and Pettegrew, L. (1977). Communicator style as an effect determinant of attraction. *Communication Research*, 4:257–282.
- Puckette, M. (2007). The theory and technique of electronic music. <http://puredata.info/>.
- Schoenenberg, K. (2015). *The Quality of Mediated-Conversations under Transmission Delay*. Ph.D. thesis, Technische Universität Berlin.
- van der Linden, D., Scholte, R., Cillessen, A., te Nijenhuis, J., and Segers, E. (2010). Classroom ratings of likeability and popularity are related to the Big Five and the general factor of personality. *Journal of research in personality*, 44:669–672.
- Vinciarelli, A., Salamin, H., Polychroniou, A., Mohammadi, G., and Origlia, A. (2011). From nonverbal cues to perception: personality and social attractiveness. In *COST'11 Proceedings of the 2011 international conference on Cognitive Behavioural Systems*, page 60–72.
- Weiss, B. and Schoenenberg, K. (2014). Conversational structures affecting auditory likeability. In *Proc. Inter-speech*, pages 1791–1795.
- Weiss, B. (2015). Akustische Korrelate von Sympathieurteilen bei Hörern gleichen Geschlechts. In *Proc. ESSV*, pages 165–171.



## Time Frames: Rethinking the Way to Look at Texts

Andreea Macovei, Dan Cristea

Faculty of Computer Science, "A.I. Cuza" University of Iași

General Berthelot 16 700483 Iași

E-mail: {andreea.gagea, dcristea}@info.uaic.ro

### Abstract

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.

**Keywords:** time frames, temporal ruptures, event ordering

### 1. Introduction

Attention towards detection of intrinsically connected sequences of events in texts is relatively new in computational linguistics. Anchored in the need to represent, extract and abstract narrative structures from streams of news, especially when gathered from different sources of information, a line of research seems to evolve towards detection of timelines and storylines. Timelines are representations of chronologically ordered events in time for a specific entity (Chambers, 2011).

Timeline extraction requires event detection and classification, extraction of temporal relations, coreference resolution of entities and events, event factuality, name entity recognition and temporal expression recognition and normalization (Caselli et al., 2015). Systems as VUA-Timeline (Minard et al., 2014) are developed for extracting cross-document timelines.

**Storylines**, on the other hand, are more complex representations, intended to take into account temporal, causal and subjective dimensions. A storyline comprises the entire sequence of significant events exposed by the narrator or by his/her characters. But, despite the chronological order of events or the ordinary flow of the story, flashbacks, temporal ruptures, flash forwards, etc. can appear and in this way, interrupt the storyline.

A storyline can be represented as the merger between individual timelines where two or more entities (characters) are involved in at least one relevant event (Laparra et al., 2015).

The necessity to determine storylines is strengthened by the fact that starting with information specialists and finishing with readers, all of them need to select large amounts of information in order to find stories, to monitor events that involve one or more participants, to reconstruct cases, etc. If the storylines are represented by schemas, these representations can reveal specific information for better selections and innovative methods used in processing texts.

In literary texts, such as novels and memories, the authors

often change the current direction of time, include flashbacks, commute the story on a completely new axis, or modify the perspective through which a story unfolds.

To deal with such linguistic phenomena, we introduce the notion of **time frame**, as a sequence of anchored or unanchored events, belonging to a delimited period of time, although the limits might be vaguely mentioned, if at all. The events of a time frame are not necessarily presented in a chronological order in the text and, as the text unfolds, switches between different time frames might also occur. Most often, the reader is aware of crossing a temporal border, even if, sometimes, with some delay during reading. Our aim is to determine if it is possible to find out clues that may indicate the transition from one time frame to another one, in order to formulate some directions for a process aiming to identify them automatically.

Literary theoreticians consider that different time frames are intentionally used by authors to introduce ambiguities and to raise the suggestive power of their stories. Often, switches between time frames can be considered a particularity of a text that would force the reader to zigzag back and forth between different story levels. The automatic identification of time frames can be an important step in disambiguating a text, reordering events, deciphering temporal relations and the general organisation of the discourse.

This paper describes a tentative approach to define and recognise in free texts time borders which accommodate groups of related events called time frames. Determining the time frames can be a premise in the automatic ordering of events in texts where no temporal indications appear or their appearance is scarce.

We intend thus to broaden the sphere of temporal annotations and extend the data structures and attributes of TIMEML, the standard mark-up language for annotating events in a text. The automatic identification of time frames represents a further step in temporal information extraction and in natural language processing, that could precede or go intertwined with the operation of

determining the order of events. Moreover, the final outcome has a much ampler benefit than building a lattice of partially ordered events in time: it is intended to draw and relate the time frames one with respect to the others, thus building a general overview of the text organisation.

## 2. Time Frames in Texts

We focus on sequences of events which are not necessarily paired with temporal information, as given by the TIMEML standard. As the dimensions of time can be expressed on 3 axis – the real time (of the reader), the discourse time (following strictly the text) and the time of the story (in which the time inversions that appear in the text are reordered) – we state our interest as focussing the discourse time, as our intention is to decipher temporal and semantic relations established among different time frames, intentionally placed by the writer.

Time frames often denote flashbacks, producing temporal ruptures which bring into attention things that happened before the current flow of the story. A reader usually deciphers without difficulty these flashbacks, which she/he will connect later to the developing story, thus reconstructing the complete intended mis-en-scene.

The following examples highlight the existence of multiple time frames in two short texts.

*Example 1* (source: a Facebook post<sup>1</sup>):

[<sub>1</sub> *Between the two rounds of preparing tomato sauce and a quick chat on Facebook, I remembered that* <sub>1</sub>] [<sub>2</sub> *I put Teodor Baconschi's book "Facebook. Factory narcissism" among the books labelled "to necessarily read"* <sub>2</sub>], [<sub>3</sub> *published this year by Humanitas* <sub>3</sub>]. [<sub>2</sub> *I could not stop myself from reading it until the end.* <sub>2</sub>]

A closer look to this short text reveals three different time frames emphasized by several sequences of events: the first frame includes the events indicated by the preparation of the tomato sauce, the chat and the remembrance; the second frame exposes the moment when the book was placed on a shelf; and the third frame is represented by the publication of Teodor Baconschi's book. The final sentence brings the reader back onto the second frame (because there is little chance that the reading of the mentioned book would be made in sequence with the events on the first frame). The square brackets and their small attached figures make visible the 3 frames.

*Example 2* (source: a novel<sup>2</sup>):

[<sub>1</sub> *Someone told me once that* <sub>1</sub>] [<sub>2</sub> *he had taken a bus full of odours and noises to return to another city. At some point, the noises reverberated, reminding him the leitmotif of the movie* <sub>2</sub>] [<sub>3</sub> *that foreshadowed a warning for a possible nuclear war; on the east coast of Australia, the wind of a dead ocean struck the window on which was hung a bottle of Coca-Cola, which in turn struck a Morse*

*signal at each burst.* <sub>3</sub>]

This example shows another disposal of time frames: in the first frame the author tells, sometimes in the past, a story about a character, mentioned here as "someone"; in the second frame, this "someone" remembers a movie; and the third frame develops the story in the movie.

A time frame encompasses one or more sequences of events, temporal scenes or episodes belonging to a specific period of time with clear or vague limits that the reader discovers or establishes once he/she continues to follow the storyline.

The example below shows an intersection of two time frames within a paragraph. The adverb phrase (*it was a long time ago*) clearly separates the first time frame from the next one.

*Example 3* (source: a novel<sup>3</sup>)

[<sub>1</sub> *Wait! Adam wanted to call him, to yell at Karl to return. Do not go, this is what he wanted to scream. But he remained silent and motionless, hidden in dense the foliage full of thorns. Now he was able to master and keep silence, held his breath and counted slowly to ten.* <sub>1</sub>] [<sub>2</sub> *It was a long time since he had learned to control such a fear.* <sub>2</sub>]

## 3. Types of Time Frames and Lexical Features Announcing Transitions

In this section we suggest a categorisation of time frames. The intention is to study the possible transitions between types and their signalling clues in the language. Our investigation till now has put in evidence the following types: NAR – the narration frame (where the time flows constantly ahead); REM – the frame of remembers belonging to a character (also a narration frame, but whose time limits are back in time with respect to a preceding narration type frame); SUP – the supposition frame (where the time is vaguely attached to a plausible, wanted or unwanted, world); GEN – the general knowledge frame (where there is no time anchor, only statements about generally accepted things); FIC – a fiction, an invented reality, like in a movie, a play or a novel (also a narration frame, but whose time limits have no connection with respect to the current story time).

Aiming to detect the borders between time frames belonging to these types, we started to investigate if there are textual clues which signal the different types of transitions.

Table 1 shows the types of time frames and the cue expressions announcing transitions for the three examples presented above.

<sup>1</sup> The post in Romanian can be found on: <https://www.facebook.com/teodor.baconschi/posts/949492091778957>.

<sup>2</sup> The example is a fragment from Octavian Paler's book "Life on a Platform".

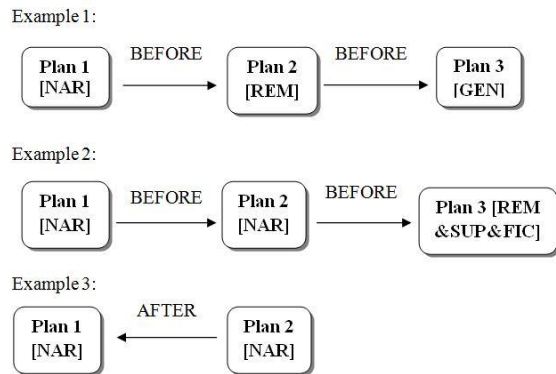
<sup>3</sup> Example 3 represents a fragment of Tash Aw's book, *Map of the Invisible World*.

	Plan no.	Type of plan	Transitions	Cue phrases announcing transitions
Example 1	[1]	NAR	[1] → [2]	<i>remembered that</i>
	[2]	REM	[2] → [3]	
	[3]	NAR		
Example 2	[1]	NAR	[1] → [2]	<i>told... that reminding... that, possible, movie</i>
	[2]	REM	[2] → [3]	
	[3]	REM&SUP&FIC		
Example 3	[1]	NAR		<i>It was a long time</i>
	[2]	NAR	[2] → [1]	

**Table 1: Types of time frames and cue phrases announcing transitions**

As seen in the Example 2 of the table, it is possible for a frame to belong to more types (here REM, SUP and FIC, because the story of a film could develop on a fictive world which is remembered or narrated by a character). Our study is at an initial stage, but we can remark already that verbs such as *remember, tell, say etc.* are good candidates to announce a REM frame. Also, other possible cue features that signal transitions, noticed on the investigated examples, could be: the change of verbs' time, some nouns, adverbs and adverbial locutions.

Figure 1 shows the relations existent among time frames, in the three examples. Nodes of the graphs represent frames and edges represent interconnections as given by temporal relations.



**Figure 1: Graph representations showing interconnections of time frames**

#### 4. Evaluation

We have conducted an evaluation of our work on a very small scale, for the time being. Our interest was to find out if the concept of various time frames in a text can really be

perceived by readers, i.e. if more readers have the same representation of time frames. So far, our aim was to identify and collect examples of time frames, to clarify to a number of students this notion and to train them to recognize different pieces of text as belonging to the same time frame.

In order to measure the agreement between annotators, several groups of three students received a fragment of Tash Aw's book, *Map of the Invisible World*<sup>4</sup>, and were asked to determine if the received text exposed one or more time frames and to annotate them according to the already-mentioned classes. Also, they had to mark borders for each time frame and to indicate whether cue phrases announcing the transition between time frames exist. The results show that in most cases, two out of three students had a similar representation of time frames including identical borders and cues, while the third one had difficulty in detecting and delimiting time frames. The text was specially chosen for its richness of time frames and frequent switches between them. As this is an incipient research, we considered the preliminary results obtained with our students encouraging, to the point to go further towards annotating a larger corpus within the lines shown in this paper.

#### 5. Relation with other Work

The field of extracting temporal information is well-studied: from a delimitation of time-denoting expressions (Schildder and Habel, 2001) as explicit references (precise dates: *06.02.2010*), indexical references (temporal expressions which depend on a given index time: *today, next Monday*) and vague references (*three days ago, in the summer, in several weeks*) and the ontology of complex events (Mele and Sorgente, 2011), there is a growing interest on ordering the events on a time axis and on determining the storyline. In a story, the common chronological order of the events is not observed (Vossen *et al.*, 2015) and each story may contain more than one *fabula* (by *fabula*, we mean logically and chronologically related events that are caused or experienced by participants in events).

On the other hand, the notion of narrative container (Pustejovsky and Stubbs, 2011) is introduced in order to delimitate the events appeared in a text without any explicit temporal anchor. The challenging work of these authors emphasizes the importance of the text style and genre in the attempt to fix in time not explicitly anchored events.

We believe that, more often than not, the events belonging to the same time frame are ordered temporally, even if different segments of the same time frame are not contiguously displayed in the text. An annotation of the events (EVENT tags) belonging to the same time frame (following the TimeML or ISO-TimeML specifications) is thus necessary. Temporal links (TLINK tags) establish relationships between two or more events and order the

<sup>4</sup> We thank to the Humanitas Publishing House for offering us the Romanian version of the book for research purposes.

events in time. The temporal expressions as dates, durations, times, etc. (TIMEX3 tags) will bring anchoring information that will help to position different time frames or segments of time frames in time.

## 6. Conclusion

In this paper we propose a new way of looking at a text that combines elements of time analysis and text structure. The approach resides on the identification of segments of texts, called time frames, that are individualised by coherent placements of sequences of events, observations about certain phenomena, places, general knowledge, etc., which are temporally situated on different time intervals or characterising opinions of different characters. In the unfolding of a text, time frames could be interrupted and interleaved, but their relationship can be represented as a graph that structures the text with respect to places, moments or intervals of time, characters and situations, being them real, supposed or fictive.

Identification of time frames is important in deciphering the structure of the discourse. Apart from being relatively well delimited in time, time frames could be related or not among them with respect to the time axis.

There is much work to be done in the future. We intend to do several things: continue this analysis by completing the classification of types, inventorying possible signals for frame transitions, proposing annotation conventions in view of a corpus analysis and performing comparative annotation in order to see to what degree different annotators have similar opinions about time frames.

The technique of frame story or story within a story could be a premise in detecting the style of each author, the creativity of her/his writing and the level of involvement in order to attract the interest of a reader. This has much to do with our time frames.

Working on literary texts, our challenge is to investigate complex text structures, rich in flashbacks and constructions about fictive worlds, frequent ruptures of time, and other types of time frames that could raise new ideas and formalisations, with the final goal to develop a process that would do an automatic identification.

This research highlights a preparatory analysis that can end up in the development of a tool capable to identify and graphically represent time frames in a text. This will be a tremendous step forward towards the goal of mirroring in the machine the human capacity of deep understanding of a text.

## 7. Bibliographical References

- Caselli T., A. Fokkens, R. Morante, P. Vossen: "What happened to ...?" Entity-based Timeline Extraction, in proceedings of 25th Meeting of Computational Linguistics in the Netherlands (CLIN2015), February 5-6, 2015, University of Antwerp, Belgium.
- Chambers, N., & Jurafsky, D. (2008, June). Unsupervised Learning of Narrative Event Chains. In *ACL* (Vol. 94305, pp. 789-797).
- Laparra, E., Aldabe, I., & Rigau, G. (2015). From TimeLines to StoryLines: A preliminary proposal for

- evaluating narratives. *ACL-IJCNLP 2015*, 50.
- Mele, F., & Sorgente, A. (2011). A Formalism for Temporal Annotation and Reasoning of Complex Events in Natural Language. *DART@ AI\* IA*, 771.
- Minard, A. L., Speranza, M., Agirre, E., Aldabe, I., van Erp, M., Magnini, B., ... & Kessler, F. B. (2015, June). Semeval-2015 task 4: Timeline: Cross-document event ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)* (pp. 778-786).
- Pustejovsky, J., & Stubbs, A. (2011, June). Increasing informativeness in temporal annotation. In *Proceedings of the 5th Linguistic Annotation Workshop* (pp. 152-160). Association for Computational Linguistics.
- Schilder, F., & Habel, C. (2001, July). From temporal expressions to temporal information: Semantic tagging of news messages. In *Proceedings of the workshop on Temporal and spatial information processing-Volume 13* (p. 9). Association for Computational Linguistics.

# ECAT: Event Capture Annotation Tool

Tuan Do, Nikhil Krishnaswamy, James Pustejovsky

Computer Science Department, Brandeis University  
Waltham, Massachusetts USA  
{tuandn,nkrishna,jamesp}@brandeis.edu

## Abstract

This paper introduces the Event Capture Annotation Tool (ECAT), a user-friendly, open-source interface tool for annotating events and their participants in video, capable of extracting the 3D positions and orientations of objects in video captured by Microsoft's Kinect® hardware. The modeling language VoxML (Pustejovsky and Krishnaswamy, 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 tool'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.

**Keywords:** event capture, event annotation, motion capture

## 1. Introduction

Much existing work in video annotation has focused on capturing objects from video in a purely two-dimensional format (i.e. tracking pixels) as in (Goldman et al., 2008), among others, or in capturing human body positioning in 3D for pose and gesture recognition (Kipp et al., 2014). We seek to wed these two types of capabilities by extracting the positions and orientations of objects and human body-rigs in video captured by the Microsoft Kinect®. These objects can be annotated as participants in a recorded motion event and this labeled data can then be used to build a corpus of *multimodal semantic simulations* of these events that can model object-object, object-agent, and agent-agent interactions through the durations of said events. This library of simulated motion events can serve as a novel resource of direct linkages from natural language to event visualization. We rely on the Kinect's capacity for body recognition and object tracking to produce output in the form of annotated object movement over time, allowing us to create an abstract representation of the denoted event.

The Kinect's depth field stream facilitates improved tracking of human movement, as reflected in the Kinect SDK's skeleton and face tracking performance (Livingston et al., 2012). The depth field provides a way to apply two-dimensional object tracking methods to a three-dimensional environment, which allows us to annotate captured video with a labeled event and its participants *with* their 3D positions throughout the event's duration. We can directly map from ECAT's output into VoxML, which was created specifically for modeling visualizations of objects and events. This mapping allows us to recreate the captured event instance in a simulated environment, and to begin compiling a library of labeled events and their participant objects simulated in 3D space, allowing in turn for the possibility of learning automatic discrimination of events from the motions of their participants.

ECAT is released as open source and it is available at <https://github.com/tuandnvn/ecat>.

## 2. Functionality

We use Kinect Sensor v2 for Windows which supports resolutions of up to HD 1920px × 1080px (RGB video) and 512px × 424px × 8 meters (depth). The latest SDK also supports 25 joint points of body tracking, and face tracking.

### 2.1. Capture and Input

For ECAT, we created our own capture and compression functionality rather than use the Kinect SDK's default functionality due to the large size of the resultant raw data files. Kinect capture automatically recognizes human bodies. Other objects may be manually marked by annotators. Once a video is captured and loaded, annotators may play it back and edit it. This may include removing an incorrectly recognized human body rig from the scene or cropping the video clip. The video clip may include frames beyond the interval of the captured event.

The default RGB color image and depth stream data are saved as separate video files. Body-tracking data is saved along with a scheme file specifying the name and index of every recognized joint in the body rig, how they are connected<sup>1</sup> and how they can be projected onto the RGB video. Additionally, users can import a property scheme file specifying what properties each object type can support, allowing them to modify the set of annotatable fields.

### 2.2. User Interface

Figure 1 shows the ECAT GUI. The various components are enumerated below.

1. Project management panel. Each project can hold multiple captured sessions.
2. Video display. For displaying either the color video or grayscale depth field video, and locating objects of interest in the scene—e.g., the table outlined in green in Figure 1.
3. Object annotation controller. Yellow time scrub bars show when each tracked object appears in the video.

<sup>1</sup>A human body rig is always a directed rooted tree whose nodes and edges form roughly the shape of a human stick figure.

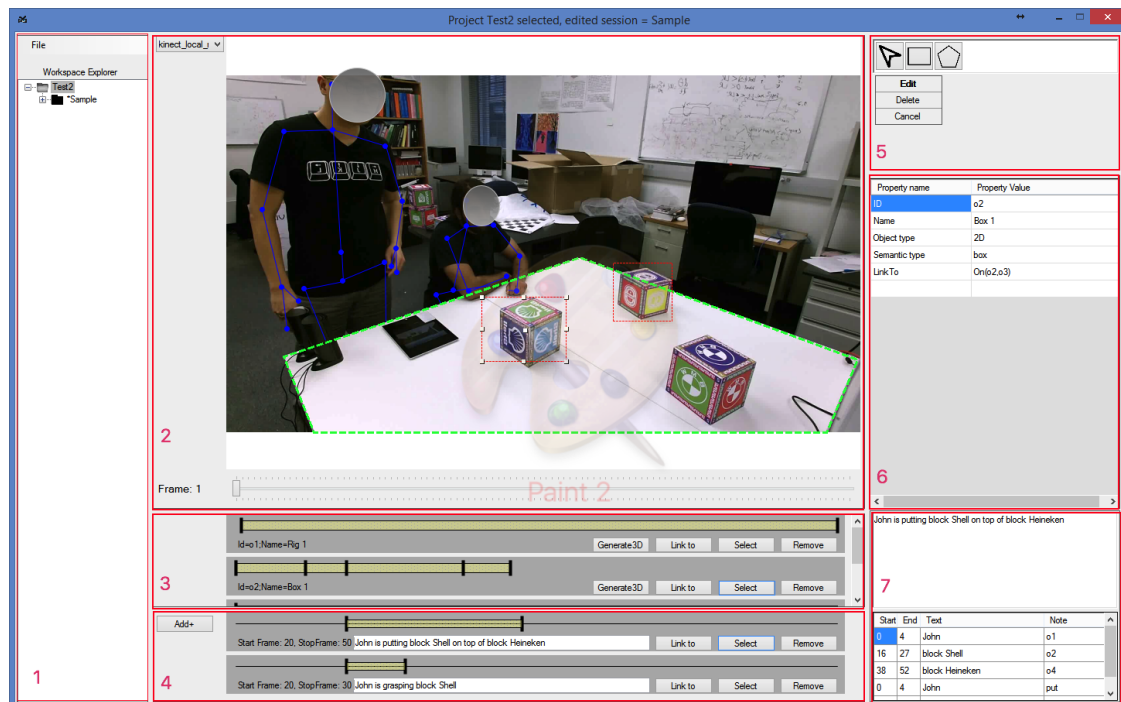


Figure 1: ECAT GUI. The left panel allows annotators to manage their captured and annotated sessions. Recognized human rigs display as blue skeletons. Marked object bounds display in color.

Black ticks mark frames where an annotator has drawn a bounding polygon around the object using the object toolbox (item 5). *Link to* links the selected object to another using a specified spatial configuration. *Generate3D* generates the selected object's tracking data using the depth field.

4. Event annotation controller. Time scrub bars here show the duration of a marked event. Users provide a text description for the event, or use *Link to* to link the selected event to another captured event as a subevent. ECAT supports marking events that comprise multiple non-contiguous segments. Due to space constraints not all annotated subevents are visible in this screenshot.
5. Object toolbox. Annotators can manually mark an in-video object with a bounding rectangle or arbitrary polygon. Marked bounds can be moved across frames as the object moves.
6. Object property panel. Data about a selected object shows here, such as ID and name.
7. Event property panel. The selected event's properties, including type and participants, show here, and the event can be linked to a VoxML event type.

Users can easily specify objects of interest in the scene, generate 3D tracking data, add or change object properties, and link them to VoxML objects. Events can be annotated with both natural language and a parametrized semantic markup, and linked to VoxML semantic programs.

### 2.3. Object Annotation

ECAT supports two ways of marking objects in a video. One is to import objects that have been automatically tracked using other libraries, such as human body rigs recognized by Kinect SDK. The other is to annotate locations of objects on the RGB video stream. Annotators mark the locations of objects at the beginning and end of an interval, and ECAT provides semi-automatic tracking using the depth field data and the iterative closest point method (Besl and McKay, 1992) to track the object's three-dimensional location. The output of the tracking algorithm can be either a point cloud or a parametric format if the object's shape can be approximated as a simple geometry (e.g., an orange or apple could be modeled as a spheroid, the tracking output being just the position of the object's center, and a radius).

An object's `objectType` field can be set to either `2D` or `3D`. Objects must be given an ID, Name and `semanticType`. We address usage of `semanticType` in Section 3.

Annotators may also mark relations between objects. For example, in Fig. 1, two blocks are on top of the table. Users can link a block object and the table object and specify the relation between the objects as "on," resulting in a predicate  $On(Block_1, Table)$  that is interpretable as a VoxML RELATION entity. Annotators could modify the available set or specify a different set of available relation predicates by importing a predicate scheme file. By default, ECAT supports the following binary predicates: *On*, *In*, *Attach.to*, *Part.of*.



## 2.4. Event Annotation

In principle, there are at least two ways to annotate an event associated with a video or video subinterval: (a) IDing an event type from an existing ontology or semantic resource, such as FrameNet (Baker et al., 1998); or (b) describing the event in natural language. We currently use the latter approach for filling an event’s `text` field, but we are working toward incorporating ontologies with the `event` tag information, addressed in section 4.

As mentioned in section 2.2., ECAT allows annotation of event-subevent relations. Thus an overarching event may be annotated as *put*, but it contains the subevents *grasp*, *hold*, *move*, and *ungrasp*, which may overlap with some subsection of the main event and each other.

## 3. Links to VoxML

Entities modeled in VoxML can be objects, programs, attributes, relations, or functions. The VoxML OBJECT is used for modeling nouns, while PROGRAM is used for modeling events. The `semanticType` field of an object captured in ECAT, filled in with free NL input, can be linked to objects annotated in VoxML if objects with the specified label exist in the VoxML-based lexicon (the *voxicon*). An object of `semanticType=block` can be linked in a 3D scene to a VoxML object denoted by the lexeme *block*, linking the captured object to all the ontological and semantic data provided by the VoxML markup (e.g. an object marked with `semanticType=stack` will be assigned, in the ECAT-to-VoxML mapping, all the VoxML knowledge of what a “stack” is). Objects whose `objectType=3D` can then be placed or moved within such a scene according to the `Location` and `Rotation` tags from the video annotation. Thus ECAT annotation can be used to recreate an equivalent scene in a VoxML-based 3D environment.

The `semanticType` field of an annotated event can be attached to the motion of the objects in the scene that correspond to the event’s participants. Thus, using the scene above as an example, the interaction of the *body\_rig* object and the *block* objects can explicitly be marked as a *put* event, and the same object/agent motions can be recreated in a 3D scene, allowing for the creation of a linked dataset of annotated videos and procedurally generated scenes. This dataset could then be used to train machine-learning algorithms to discriminate motion events based on the motions of an event’s respective participants in 3D space.

## 4. Output

Body rigs are saved as objects with `semanticType=body_rig`. They are ID’ed (`id=o1` as seen in Fig. 2) and can be given an alias for the user’s ease (here *John*).

Annotated objects are treated similarly, assigned an ID, a name, and a semantic type. Here *o2* is the Shell logo block from Fig. 1. Object locations and relative spatial relations can be annotated by frame. At frame 1, *o2* is on the table (*o3*) while by frame 50, it has been put on the other block (*o4*), so the corresponding `LinkTo` tags are `On(o2, o3)` and `On(o2, o4)`, respectively. By default, ECAT supports the relations *On*, *In*, *Part\_of*, and *Attach\_To*, where an

object is in a parent-child relationship with another object, such as when a body rig’s hand is carrying a block. annotations denote events, with participants as referents (refs). In Fig. 3, *o1*, *o2*, and *o4* are refs, while *a1*’s event’s `semanticType=put`, marking the three above objects as the “put” event’s participants. An annotation’s `superEvent` indicates super/subevent relationships, so that *a2*, a “grasp” event, is noted as a subevent of “put” *a1*.

Both objects and annotations can be mapped to VoxML representations, for instance as in Fig. 4 below, which shows a VoxML representation of *put*, an event annotated in Fig. 3.

```
<object id="o1" name="Rig 1" alias="John"
objectType="3D" generate="provided" source="body.xml"
sourceScheme="bodyScheme.xml" semanticType="body_rig">
</object>

<object id="o2" name="Block Shell" objectType="2D"
generate="manual" semanticType="box">
<marker type="Location" frame="1" shape="Rectangle"
LinkTo="On(o2,o3)"> 50, 50, 10, 10 </marker>
<marker type="Location" frame="50" shape="Rectangle"
LinkTo="On(o2,o4)"> 70, 50, 10, 10 </marker>
<tracking depth="depth.avi" track="tracked.out"/>
</object>
```

Figure 2: Object output format.

```
<annotation id="a1">
<duration startFrame="20" endFrame="50" />
<text>John is putting block Shell
on top of block Heineken</text>
<refs>
<ref start="0" end="4" refTo="o1" />
<ref start="16" end="27" refTo="o2" />
<ref start="38" end="52" refTo="o4" />
</refs>
<events>
<event start="8" end="15" semanticType="put"/>
...

<annotation id="a2" superEvent="a1">
<duration startFrame="20" endFrame="30" />
<text>John is grasping block Shell</text>
...
```

Figure 3: Event annotation output format. Some subevent specifics are elided here for space.

$$\text{put} \quad \text{LEX} = \left[ \begin{array}{l} \text{PRED} = \text{put} \\ \text{TYPE} = \text{transition\_event} \end{array} \right]$$

$$\text{TYPE} = \left[ \begin{array}{l} \text{HEAD} = \text{transition} \\ \text{ARGS} = \left[ \begin{array}{l} A_1 = \text{o1} \\ A_2 = \text{o2} \\ A_3 = \text{On(o4)} \end{array} \right] \\ \text{BODY} = \left[ \begin{array}{l} E_1 = \text{grasp}(A_1, A_2) \\ E_2 = [\text{while}(\text{hold}(A_1, A_2), \\ \text{move}(A_2))] \\ E_3 = [\text{at}(A_2, A_3) \rightarrow \\ \text{ungrasp}(A_1, A_2)] \end{array} \right] \end{array} \right]$$

Figure 4: VoxML for Fig. 3’s *put* instance. *o1*, *o2*, and *o4* each point to that object’s VoxML representation.  $E_1$ ,  $E_2$ , and  $E_3$  are mapped from annotated subevents, such as *grasp* in Fig. 3.

## 5. Conclusions

Event and action detection and recognition in video is receiving increasing attention in the scientific community, due to its relevance to a wide variety of applications (Ballan et al., 2011) and there have been calls for annotation infrastructure that includes video (Ide, 2013). We have presented here a tool that provides a user-friendly interface for video annotation that is able to capture a level of detail not provided by most existing video annotation tools, provides links to existing linguistic infrastructures, and is well suited for building a corpus of event-annotated multimodal simulations for use in the study of spatial and motion semantics (Pustejovsky and Moszkowicz, 2011; Pustejovsky, 2013). For future annotation capabilities, we are planning on introducing links to existing semantic lexical resources, such as FrameNet, as well as event ontologies. More significantly, we are extending the ECAT environment to allow for annotation of much longer videos, encompassing multiple event sequences comprising narratives, including simultaneous or overlapping events that do not hold super/subevent relations between them but together make up a larger story (e.g. a man cooking dinner while a woman sets the table). This will entail enriching our specification to enable the markup of discourse connectives, linking the events in the narrative.

## Acknowledgements

This work is supported by DARPA Contract W911NF-15-C-0238 with the US Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO). Approved for Public Release, Distribution Unlimited. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. All errors and mistakes are, of course, the responsibilities of the authors.

## 6. Bibliographical References

- Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.
- Ballan, L., Bertini, M., Del Bimbo, A., Seidenari, L., and Serra, G. (2011). Event detection and recognition for semantic annotation of video. *Multimedia Tools and Applications*, 51(1):279–302.
- Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics.
- Goldman, D. B., Gonterman, C., Curless, B., Salesin, D., and Seitz, S. M. (2008). Video object annotation, navigation, and composition. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 3–12. ACM.
- Ide, N. (2013). An open linguistic infrastructure for annotated corpora. In *The People’s ½ Web Meets NLP*, pages 265–285. Springer.
- Kipp, M., von Hollen, L. F., Hrstka, M. C., and Zamponi, F. (2014). Single-person and multi-party 3d visualizations for nonverbal communication analysis. In *Proceedings*

*of the Ninth International Conference on Language Resources and Evaluation (LREC), ELDA, Paris.*

- Livingston, M. A., Sebastian, J., Ai, Z., and Decker, J. W. (2012). Performance measurements for the microsoft kinect skeleton. In *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*, pages 119–120. IEEE.
- Pustejovsky, J. and Krishnaswamy, N. (2016). Voxml: A visual object modeling language. *Proceedings of LREC*.
- Pustejovsky, J. and Moszkowicz, J. (2011). The qualitative spatial dynamics of motion. *The Journal of Spatial Cognition and Computation*.
- Pustejovsky, J. (2013). Dynamic event structure and habitat theory. In *Proceedings of the 6th International Conference on Generative Approaches to the Lexicon (GL2013)*, pages 1–10. ACL.



## A Construction-centered Approach to the Annotation of Modality

Elisa Ghia<sup>1</sup>, Lennart Kloppenburg<sup>2</sup>, Malvina Nissim<sup>2</sup>, Paola Pietrandrea<sup>3</sup>, Valerio Cervoni<sup>3</sup>

<sup>1</sup>University for Foreigners of Siena, <sup>2</sup>University of Groningen, <sup>3</sup>University of Tours and CNRS LLL  
elisaghia@gmail.com, {l.kloppenburg|m.nissim}@rug.nl, pietrandrea-guerrini@univ-tours.fr, cervoni@etu.univ-tours.fr

### Abstract

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.

**Keywords:** modality, annotation, agreement

### 1. Introduction

Modality is a pervasive phenomenon crucial to language understanding, analysis, and automatic processing, and at the same time difficult to encapsulate in one exhaustive but workable definition (Morante and Sporleder, 2012). This is reflected in the continuous efforts towards two intertwined aims, namely (i) the definition of the core and the borders of modality, and (ii) the creation of annotated data, also towards the development of automatic systems.

Indeed, modality-annotated data would benefit Natural Language Processing in at least two major aspects: (i) factuality detection, consisting in the automatic distinction between propositions that represent factual events and propositions that represent non factual ones; and (ii) opinion mining and sentiment analysis, which involve the processing of extra-propositional aspects of meaning and the detection of polarised judgements. Efforts in this sense are exemplified by recurring sentiment analysis tasks within the context of Semeval (see for example Task 9 to Task 12 in the 2015 campaign)<sup>1</sup>, as well as specific factuality tasks such as the CoNLL-2010 Shared Task on identifying hedges (Farkas et al., 2010), and data annotation towards further campaigns, not just limited to English (Minard et al., 2014; Schoen et al., 2014).

In addition to NLP applications, the annotation of modality may have important repercussions in the Corpus Linguistics field, as the techniques developed in the automatic treatment of modality can be used to improve our linguistic knowledge of modality itself. Nevertheless, shared standards for modality annotation do not exist as yet (Morante and Sporleder, 2012).

In the current contribution, we apply the model described in (Nissim et al., 2013) to epistemic modality, and we describe the development and implementation of a flexible and comprehensive scheme for the annotation of modalised constructions in transcribed dialogues. With a view to developing a flexible model for the automatic annotation of modality, we suggest that the annotation procedure follow a corpus-driven approach, as operational categories can be drawn and refined from data. Because such a model has

<sup>1</sup><http://alt.qcri.org/semeval2015/index.php?id=tasks>

to be not only theoretically sound but operational (both in terms of annotation as well as in terms of automatic processing), we propose a comprehensive annotation framework for modality which we motivate theoretically, and test its validity empirically by means of a pilot study.

### 2. Phenomena

Annotating modality may involve the identification of factuality and/or subjectivity. These two dimensions are key not only to interpreting modalised constructions but also in terms of their annotation. Indeed, to summarise what we explain in detail below, approaches that focus more on the factuality aspect of modality are target-centered, while approaches that focus more on subjectivity (including here opinion mining) are trigger-centered (see Figure 1 for an example of trigger and target).

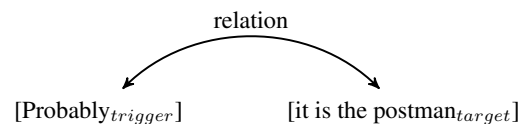


Figure 1: Trigger, target, and relation between them in a modalised context.

#### 2.1. Factuality and Target-centered Schemes

Factuality refers to the extent to which the event described in a proposition is grounded in reality. Factuality annotation is hence aimed at distinguishing linguistic material presented as a fact from other language material. This has also to do with speculation (Medlock and Briscoe, 2007) and uncertainty (Rubin, 2010; Szarvas et al., 2012; Saurí and Pustejovsky, 2012; Sanchez and Vogel, 2015; Thompson et al., 2008). When focusing on factuality, the annotation is usually target-driven. This means that the annotation procedure consists in identifying the element whose factuality has to be evaluated, i.e., the target of the factuality relation, and in providing information about that element. In the annotation of FactBank (Saurí and Pustejovsky, 2012), for example, the text is segmented in ‘events’. For each event the

schema specifies the following attributes: source, source introducing predicate, factuality value, time (see Figure 2). Building on the idea that the factuality of a semantic state can be annotated via the annotation of opinions, Wiebe et al. (2005) identify in the text the semantic entities that correspond to private states. Each private state is then annotated for intensity, attitude type, source, anchor text, target (Figure 3). In Thompson et al. (2008)’s annotation scheme the text is segmented in sentences. For each sentence the scheme specifies the certainty trigger that determines the factuality value of the sentence. For each trigger three attributes are specified: the point of view, the knowledge type and the certainty level (Figure 4).

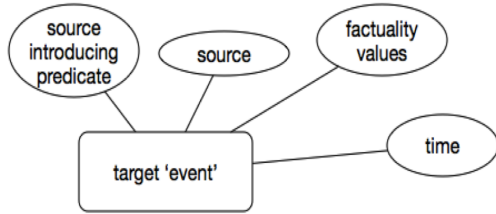


Figure 2: Overview of Saurí and Pustejovsky (2012)’s target-centered annotation scheme.

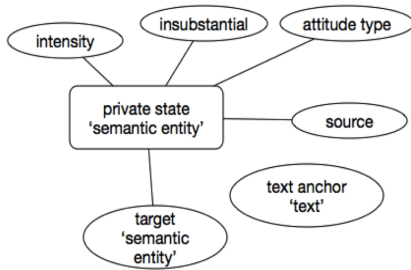


Figure 3: Overview of Wiebe et al. (2005)’s target-centered annotation scheme (in the context of opinion mining).

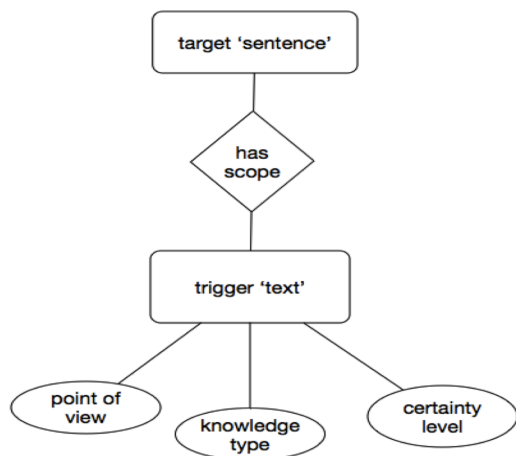


Figure 4: Overview of Thompson et al. (2008)’s target-centered annotation scheme.

## 2.2. Subjectivity and Trigger-centered Schemes

Beside factuality, modality interpretation involves the identification of subjectivity, or *extrapositional aspects of meaning*. When specifically annotating subjectivity normally a wide notion is adopted, including such components as appreciation, fear, effort, epistemic opinion. Annotation schemes that focus on this aspect, including work on sentiment analysis, adopt a trigger-centered approach to annotation, as it is the subjectivity/sentiment of triggers that is mostly informative (Wiebe et al., 2005; Rubin, 2010; Nirenburg and McShane, 2008; Hendrickx et al., 2012; Ávila et al., 2015; Baker et al., 2010).

In brief, trigger-driven annotation approaches consist in identifying in a text the linguistic elements that encode the subjective meaning. Vincze et al. (2010)’s procedure, for example, consists in annotating the ‘lexical cue’ that encodes uncertainty and in specifying for each lexical cue the following attributes: the genre and the domain of the text in which it occurs, the type of uncertainty that it encodes, its PoS and the chunk it belongs to (Figure 5). Sanchez and Vogel (2015) take the ‘hedges’ encoding the degree of commitment of the speaker as the central element of their annotation. For each hedge they specify: the syntactic type, the dependency tree over which the hedge scopes, the type of source to which the hedge has to be attributed (Figure 6). The schemes represented in Figures 2 to 6 specify different abstract syntaxes. However, from a conceptual standpoint all these schemes regard a modal relation in the same way, namely as a dyadic relation between a trigger and a target. According to the objectives of the annotation, either the trigger or the target of the relation is taken as the annotable unit.

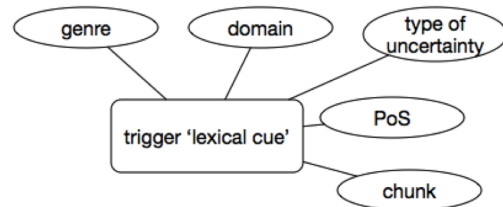


Figure 5: Overview of Vincze et al. (2010)’s trigger-centered annotation scheme.

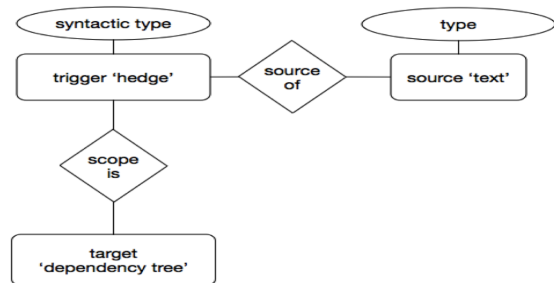


Figure 6: Overview of Sanchez and Vogel (2015)’s trigger-centered annotation scheme.

### 3. A Construction-centered Approach

Rather than as binary relations between a trigger and a target, in this contribution we view *constructions* as triadic relations between a trigger, a target and the relation between them. Accordingly, we revise Figure 1 as Figure 7.

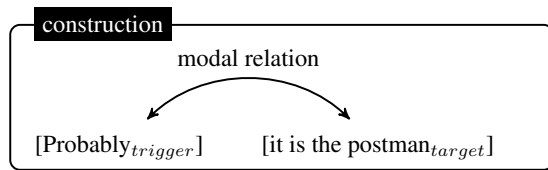


Figure 7: A construction is conceived as a trigger, a target, and a modal relation between them.

As a consequence, from a conceptual point of view, the relation between the trigger and the target has its own properties and functions, and from a practical perspective, such properties and functions have to be specified as attributes of the relation itself in an annotation scheme. In what follows, we justify this view theoretically (Section 3.1.) as well as practically (Section 4. and Section 5.).

#### 3.1. Theory

Our formalisation choice is linked to multiple factors. First, it happens quite frequently in spoken language (and it may theoretically happen in written language alike) that one and the same target can receive more than one evaluation. In Example (1):

- (1) C: *dovevano venire a leggerla quanto meno* [they were supposed to come and read it, at least]  
 A: *no anche se non veniva<no> sì dovevano veni' a leggerla* [no even if they didn't come- yes they were supposed to come and read it]  
 C: *così almeno si sapeva* [at least this is what we knew]

the same target, i.e. the proposition [they come and read it] is linked to four different truth values through four different triggers:

1. the modal “dovevano” [‘were supposed to’]
2. the pragmatic marker “no” [‘no’]
3. the pragmatic marker “sì”, [‘yes’]
4. the utterance “così almeno si sapeva” [‘at least this is what we knew’].

In other words, the target enters four different epistemic constructions, and it would not make sense to try to establish its factuality status value independently of such constructions. As a consequence, factuality evaluation cannot be conceived as a property of the target (which indeed receives several modality evaluations).

It would be equally awkward to regard modal evaluation as a property of the trigger. As the utterances in Examples (2) through (6)<sup>2</sup> show, one and the same trigger – in

<sup>2</sup>These examples are all from the EnTenTen corpus (Pomikálek et al., 2009).

this case the complement-taking predicate “I think” – triggers different types of modality to its target, according to the target’s semantic nature, whether a statement, a judgement, etc.

- (2) I think he went through a separation with his wife and I think that depressed him.
- (3) I love your wife, and I think she is beautiful!
- (4) Quite frankly, I think he has the right to make that decision.
- (5) I think you are better off fixing the “issues” one by one than going into bankruptcy.
- (6) I did have a waxing service from one other person here, but I think I will choose to stick with Simona for future waxing services from here.

Therefore, modal evaluation is better regarded as a property of the construction as a whole, i.e. as a functional property encoding the relation between the trigger and the target. Our annotation scheme is grounded in such assumption.

#### 3.2. Overview of the Annotation Scheme

We describe in this section the procedure and the scheme we adopted for our annotation task.

As a first step, we identified triggers and targets in a modalised construction in the text. Once selected, triggers were defined through the attributes form (i.e. text token), lemma, illocution (i.e. the trigger’s illocution, involving the values assertion, expression, injunction, question) and morphosyntactic category (including morphological triggers, e.g. tense/aspect marking, lexical triggers, e.g. adverbs or pragmatic markers, syntactic triggers, such as inversions for interrogatives, prosodic triggers). The target was subsequently identified and defined by its illocution (assertion, exclamation, injunction, question). The third stage in the annotation procedure involved the linking of trigger and target into a modal relation, which was further defined through the attributes direction (trigger > target or target > trigger, embedding, co-extension, extension over more turns and speakers), function (i.e. discourse function: qualifying, accepting, non accepting, checking, confirming, non confirming, informing), polarity (positive, negative, neutral), and type (type of evidence upon which the epistemic construction is grounded). See (Pietrandrea, submitted) for a theoretical justification of the annotation schema (see Figure 8 for a screenshot of the annotation schemes with all of the categories and labels).

As we have seen, different approaches are associated with different schemes which respond to different objectives. However, even distinct modal phenomena are related to each other, as they all deal with the validation of representations or, in other words, with extrapositional aspects of meaning. A flexible and broader annotation scheme could thus allow to encompass all specific schemes and support the needs of target-, trigger-, and relation-centred approaches altogether.

The comprehensive scheme was tested on the annotation of modalised constructions in spoken text. Annotation was carried out by multiple annotators on the basis of guidelines

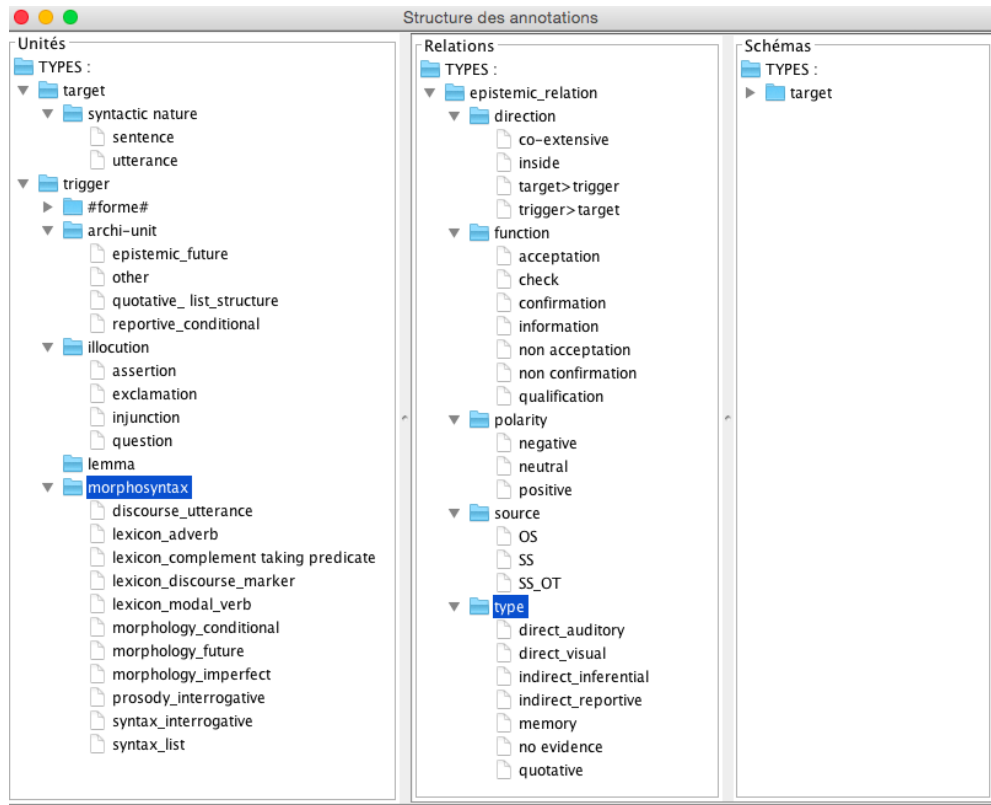


Figure 8: Screenshot of the Analec annotation tool customised for a construction-centered modality annotation. The categories and labels of the annotation scheme are all visible.

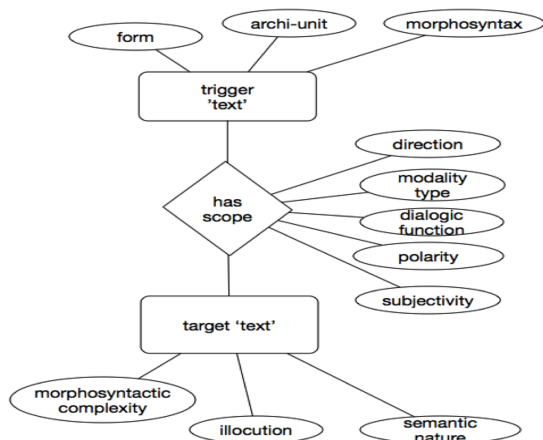


Figure 9: Overview of our construction-centred annotation scheme.

established via decision trees. A common and highly customizable annotation tool was used for manual annotation, along with shared evaluation metrics. Annotators discussed the annotation process and the operational categories at regular meetings.

After initial identification of the relevant categories by multiple annotators, the cognitive salience of such categories is recursively tested through inter-annotator agree-

ment. The model is hence refined incrementally (Glynn and Krawczak, 2014), leading to ultimate operationalisation of the categories that allow for the semantic modelling of modality.

#### 4. Annotation Experiment

With a view to testing our annotation framework for modality and its implementation on language data, annotation has proceeded along a set of successive stages: (i) (pilot) annotation by multiple expert annotators and identification of relevant categories, (ii) calculation of inter-annotator agreement to test the feasibility of a construction-based annotation (tested via alignment, see below) and the cognitive salience of the categories, (iii) refinement of the annotation scheme, (iv) second annotation phase, (v) operationalisation of the necessary categories for the semantic modelling of modality. We are describing here stages (i) and partially (ii)-(iii).

The pilot experiment is divided into two phases, and involves the annotation of spoken Italian data from the LIP corpus (De Mauro, 1993) and spoken French from the ESLO corpus<sup>3</sup>. The annotation was performed using the Analec annotation tool (Landragin et al., 2012), which produces TEI-compliant XML output, and was originally designed for the annotation of anaphoric phenomena and thus lends itself well to the task of annotating a three-way

<sup>3</sup><http://eslo.huma-num.fr/>

construction, with features for trigger, target, and relation. Appropriate categories and features were implemented via in-tool customisation of the annotation schemes (see Figure 8).

Each annotator worked individually following these steps:

1. identification of the trigger of the modal construction
2. identification of the target of the modal construction
3. identification of the relation holding between trigger and target.

For the first phase, on Italian, a total of approximately 650 constructions was annotated on a corpus section of 19,665 words consisting of six dialogic situations: a university exam, a dialogue excerpt from a television programme, a transactional exchange, two conversations among friends, one family conversation over dinner. At this stage, the alignment of constructions across annotations, needed to assess whether the judges had identified the same modalised constructions, was performed in a rather simple and shallow way, with substantial manual intervention. Note that alignment is crucial towards assessing the validity of the scheme, as freedom must not imply randomness or the impossibility to perform evaluation. Although the annotation yielded promising results on agreement, with an f-score of 0.779 over the constructions, and 53% agreement on the exact extension of the targets, the alignment procedure wasn't properly formalised in any way. For a second pilot study we therefore refined the annotation guidelines not only conceptually but also operationally in order to provide more precise instructions regarding the spans to be annotated, and we devised a more structured, more robust and at the same time more flexible procedure for aligning annotations. This procedure is described in the next section, and has been deployed on portions annotated in the second phase of the pilot study.

This second phase focuses on French (also with a view to keep the scheme cross-linguistically valid), for which we are annotating 20,000 words in dialogues from the ESLO corpus. For this pilot experiment, 7 annotators working on a 1000 word text, annotated about 40 constructions.<sup>4</sup> For this showcase evaluation we take the annotations performed by two judges, *a* and *b* in what follows.

## 5. Evaluation

Before comparing and evaluating the values attributed to the relations and the triggers, we need to *align* the *constructions* identified by two annotators. Thus, in order to compute agreement among two annotated documents, the output XML files are subjected to preprocessing, alignment and agreement phases. The alignment phase is particularly meaningful, as it tells us whether two judges have identified the same moralised construction (independently of the specific feature values assigned to trigger/target/relation).

<sup>4</sup>Some coding details have changed, and we haven't yet transferred the whole Italian annotation to the current format, so that the more structured evaluation of alignment hasn't been performed on this data again, yet.

### 5.1. Preprocessing

The file is pre-divided into different paragraphs. Every paragraph contains either no or one/multiple trigger/target annotations, which we term *anchors* in this context. As a first step, we collect all of the anchors and extract the transcript contents between the beginning and end of an anchor, in other words: the marked up text. For example, in Figure 10, for '**u-trigger-3**', the content (text) is '*il m'a dit*', and for '**u-target-portion-3**' it is '*il travaillait pas*'. These anchor-content pairs are subsequently stored for both annotation files.

### 5.2. Alignment

The IDs of the anchors (displayed as 'id' in the XML-sample in Figure 10) of either annotation file do not correspond to each other as they obviously only obey internal consistency, and the texts are annotated separately. Therefore, we need an alignment step which matches anchors from both annotation files. Anchors can be aligned iff:

- they are of the same type (trigger or target)
- they overlap in content by at least a given proportion of lexical material, which we base on character offset. For example, for a required overlap of 50% and a token length of an anchor *A* of ten tokens, the content of the candidate anchor from the other file needs to have at least five subsequent words in common with *A* (see Section 5.4. for an example of partial overlap and a further discussion of varying overlap requirements)

This process results in a collection of pairs of aligned anchors. For example, considering annotator *a* and annotator *b*, we would have an aligned pair of trigger  $t_a$  and trigger  $t_b$ .

The final step is to iterate through the relations that judge *a* introduced and align them with relations that judge *b* introduced. In order to explain the procedure of further alignment to relations, we take judge *a* as reference, but in terms of scores it doesn't make any difference which direction we go, since  $precision_{ab} = recall_{ba}$  so that eventually  $fscore_{ab} = fscore_{ba}$ . Relations consist of a trigger and one or multiple target portions. Aligning relations is done by pairing up triggers and targets into relations introduced by judge *a* and check if the aligned counterparts of these triggers and targets by judge *b* are part of a relation as well. In case this is the case, we deem the two constructions as "the same". Note that at this stage we have not checked yet agreement on the features assigned to relations and triggers – we are just evaluating that the two judges identify in text the same modalised construction, which is a crucial step.

The alignment process between judge *a* and judge *b* results then in three sets that can be evaluated: a set of *trigger* pairs, a set of *target* pairs and a set of *relation* pairs. The agreement between judge *a* and judge *b* for a given set is expressed as the precision of annotations by judge *b* compared to those of judge *a*. Recall for this same process is computed by swapping judge *b* and judge *a*, since as hinted above, a false negative, or a relation/trigger/target which was annotated by judge *a* but not by judge *b*, turns into a false positive if this is reversed.

```

<anchor id="u-trigger-3-start" type="AnalecDelimiter" subtype="UnitStart"/>
il m'a dit
<anchor xml:id="u-trigger-3-end" type="AnalecDelimiter" subtype="UnitEnd"/>
<anchor xml:id="u-target-portion-3-start" type="AnalecDelimiter" subtype="UnitStart"/>
il travaillait pas
<anchor xml:id="u-target-portion-3-end" type="AnalecDelimiter" subtype="UnitEnd"/>

```

Figure 10: Example annotation

### 5.3. Layers of Inconsistency

In the alignment process, there are a number of layers where mismatches and actual disagreements can occur. The *paragraph layer* refers to the possibility that any paragraph annotated by judge *a* is not annotated by judge *b*, which means that all annotations that judge *a* made in this particular paragraph cannot be aligned. This was necessary since in the pilot study not all annotators completed the whole text markup, thus leaving some final portions simply unannotated. Since this does not tell us anything about conceptual agreement, only the paragraphs which were annotated by both judge *a* and *b* were considered. This stage would not be relevant if complete texts are annotated by both annotators. At the *alignment layer* we align anchors. The process can fail if there is not enough overlap or if judge *a* annotated fewer or more anchors than judge *b*, which automatically results in failed alignments. The final layer is the *relation layer*. Consider that the alignment of two relations between judge *a* and judge *b* must obey the following constraints:

1. both the target and trigger of the relation by judge *a* need to be *aligned* with counterparts from judge *b*. If one of these was not aligned, the relation alignment fails as well
2. both of the counterparts need to belong to the same relation by judge *b*.

### 5.4. Results

According to the specific procedure just described, we report agreement results for construction alignment on a sample of two files from the French data, annotated by judge *a* and judge *b*, with approximately 40 constructions found.

As mentioned, we have to evaluate two main aspects. First, whether the annotators have identified the same modalised constructions, thus whether we can align their annotations. Second, whether the features assigned to triggers and to relations according to the annotation scheme correspond between the two judges. Agreement over alignment is measured using precision/recall/f-score as we have to deal with potentially different spans. For the relations' and triggers' features we can then use Cohen's Kappa (Cohen, 1960) (or Fleiss' Kappa in case of more than two annotators) over the agreed upon constructions only, as it becomes a plain classification task. In this paper, we are only reporting alignment agreement.

Because of freedom in the annotation of the extension of anchors, as mentioned above we evaluated alignment at different percentages of overlap. This is particularly relevant

for the target portion. As for triggers, we can be very lenient, especially if the properties assigned to them by both annotators correspond.<sup>5</sup> For the pilot study that we present here, we have tested targets' overlaps in the range of 10% to 100% in terms of tokens.

To illustrate this, consider Example (7). The token overlap between the strings selected by the annotator *a* and annotator *b* is just under 50%. So setting the overlap constraint at 40% would yield an alignment and thus agreement, agreement, while setting at 50% wouldn't.

- (7) *a* = 'il ne travaillait pas'  
*b* = 'il m'a dit qu'il ne travaillait pas'

For the triggers, we have fixed the overlap at a minimum of 10%. At this level of overlap, f-score is measured at 0.87 (see Table 1), with a total of 34 aligned triggers out of 40 detected by *a* and 38 detected by *b*. By fixing this alignment for triggers, in Table 2 we report precision, recall, and the specific amount of true positives (TPs) and false negatives (FNs) at varying degrees of overlap. Results for alignment over constructions as wholes is given in Table 3.

Overlap	Prec	Rec	F1
10%	0.89	0.85	0.87
50%	0.84	0.85	0.84
100%	0.71	0.71	0.70

Table 1: Alignment agreement for *triggers* with varying amounts of overlap. Judge *a* is taken as reference in indicating precision and recall.

## 6. Conclusion

We have presented a construction-based annotation scheme for modality that is theoretically sound and empirically applicable. There are existing schemes that cover some aspects of modality annotation, but no specific shared standards, as yet, and no comprehensive framework that can encompass and account for all aspects related to (the annotation of) modality. Indeed, we believe that a comprehensive scheme for the annotation of modality needs to fulfill a set of requirements, and our proposed approach manages to obey them: (i) general flexibility (validity for all approaches); (ii) exhaustiveness (ability to encompass all

<sup>5</sup>Annotation of features following specific guidelines is underway for this part of the pilot study. Preliminary agreement over triggers' features show a Kappa of 0.72 at 10% overlap and 0.83 at 100% overlap, so that it indeed seems wise to allow for more aligned constructions while still preserving reasonable agreement.

Overlap	Prec	Rec	F1	TP/FN/TOT
10%	0.82	1.00	0.90	18/4/22
20%	0.82	1.00	0.90	18/4/22
30%	0.82	1.00	0.90	18/4/22
40%	0.82	1.00	0.90	18/4/22
50%	0.77	0.95	0.85	17/5/22
60%	0.77	0.95	0.85	17/5/22
70%	0.77	0.90	0.83	17/5/22
80%	0.77	0.81	0.79	17/5/22
90%	0.64	0.67	0.65	14/8/22
100%	0.41	0.43	0.42	9/13/22

Table 2: Alignment agreement for *targets* with varying amounts of overlap, with trigger alignment fixed at 10%. Judge *a* is taken as reference in indicating precision, recall, TPs and FNs.

Overlap	Precision	Recall	F1
10%	0.76	0.62	0.68
20%	0.76	0.62	0.68
30%	0.76	0.62	0.68
40%	0.82	0.68	0.74
50%	0.82	0.68	0.74
60%	0.82	0.68	0.74
70%	0.74	0.68	0.71
80%	0.66	0.68	0.67
90%	0.47	0.47	0.47
100%	0.26	0.25	0.25

Table 3: Alignment agreement for *constructions*, with trigger alignment fixed at 10%, and varying overlap constraints for targets. Judge *a* is taken as reference in indicating precision and recall.

specific schemes, which have to be interpreted within the larger scheme); (iii) constrained freedom (the scheme has to offer a wide set of possibilities among which only some are realised in a given scheme; the way things are realised is fixed, but the choice of what to realise is free); (iv) a shared abstract syntax for the annotation scheme; (v) a shared semantics for values; (vi) shared practices for the annotation procedure. The rather successful agreement over the identification of constructions – which we have evaluated through a rigorous alignment protocol – shows that in spite of freedom and flexibility, the scheme has a strong potential for implementation. Further evaluation of properties and features is underway, as well as further tests on yet other languages. The annotation tool that we have used is freely available and so are the annotation schemes, with a view to provide as much shared material as possible.

## 7. Bibliographical References

Luciana Beatriz Ávila, Amália Mendes, and Iris Hendrickx. 2015. Towards a unified approach to modality annotation in portuguese. In Malvina Nissim and Paola Pietrandrea, editors, *Proceedings of the IWCS Workshop on Models for Modality Annotation*, London.

Kathryn Baker, Michael Bloodgood, Mona Diab, Bonnie J. Dorr, Ed Hovy, Lori Levin, Marjorie McShane, Teruko Mitamura, Sergei Nirenburg, Christine Piatko, Owen

Rambow, and Gramm Richardson. 2010. SIMT SCALE 2009 - Modality Annotation Guidelines, Technical Report 004. Johns Hopkins University, Baltimore, MD.

J. Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37.

Tullio De Mauro. 1993. *Lessico di frequenza dell'italiano parlato*. Etas.

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of CoNLL '10: Shared Task*, pages 1–12, Stroudsburg, PA, USA.

Dylan Glynn and Karolina Krawczak. 2014. Operationalisation and robust manual annotation of non-observable usage-features. an exploratory study in english and polish. In *Workshop on Modal Categories at EMEL'14*, Universidad Complutense de Madrid.

Iris Hendrickx, Amália Mendes, and Silvia Mencarelli. 2012. Modality in text: a proposal for corpus annotation. In Nicoletta Calzolari et al., editor, *Proc. of LREC'12*, Istanbul, Turkey. ELRA.

Frederic Landragin, Thierry Poibeau, and Bernard Victorri. 2012. Analec: a new tool for the dynamic annotation of textual data. In *Proc of LREC'12*, pages 357–362.

Ben Medlock and Ted Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *ACL*, volume 2007, pages 992–999. Citeseer.

Anne-Lyse Minard, Alessandro Marchetti, and Manuela Speranza. 2014. Event Factuality in Italian: Annotation of News Stories from the Ita-TimeBank. In *Proc. of CLIC-it*, Pisa.

Roser Morante and Caroline Sporleder. 2012. Modality and negation: An introduction to the special issue. *Computational Linguistics*, 38(2).

Sergei Nirenburg and Marge McShane. 2008. Annotating modality. technical report. Technical report, University of Maryland, Baltimore County.

Malvina Nissim, Paola Pietrandrea, Andrea Sanso, and Caterina Mauri. 2013. Cross-linguistic annotation of modality: a data-driven hierarchical model. In *Proc. of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation*, pages 7–14, Potsdam, Germany, March. ACL.

Paola Pietrandrea. submitted. Epistemic constructions at work: A corpus-driven study on spoken italian dialogues. *Journal of Pragmatics*. [https://www.researchgate.net/publication/296484341\\_Epistemic\\_constructions\\_at\\_work\\_A\\_corpus\\_study\\_on\\_Italian\\_spoken\\_dialogues](https://www.researchgate.net/publication/296484341_Epistemic_constructions_at_work_A_corpus_study_on_Italian_spoken_dialogues)

Jan Pomikálek, Pavel Rychlý, Adam Kilgarrieff, et al. 2009. Scaling to billion-plus word corpora. *Advances in Computational Linguistics*, 41:3–13.

Victoria L Rubin. 2010. Epistemic modality: From uncertainty to certainty in the context of information seeking as interactions with texts. *Information Processing & Management*, 46(5):533–540.

Liliana Mamani Sanchez and Carl Vogel. 2015. A hedging

- annotation scheme focused on epistemic phrases for informal language. In Malvina Nissim and Paola Pietrandrea, editors, *Proc. of the IWCS Workshop on Models for Modality Annotation*, London.
- Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.
- Anneleen Schoen, Chantal van Son, Marieke van Erp, and Hennie van der Vliet. 2014. Newsreader document-level annotation guidelines-dutch. Technical report, Vrije Universiteit Amsterdam, TechReport 2014-8.
- György Szarvas, Veronika Vincze, Richárd Farkas, György Móra, and Iryna Gurevych. 2012. Cross-genre and cross-domain detection of semantic uncertainty. *Computational Linguistics*, 38(2):335–367.
- Paul Thompson, Giulia Venturi, John McNaught, Simonetta Montemagni, and Sophia Ananiadou. 2008. Categorising modality in biomedical texts. In *Proc. of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining*, pages 27–34.
- Veronika Vincze, György Szarvas, György Móra, Tomoko Ohta, and Richárd Farkas. 2010. Linguistic scope-based and biological event-based speculation and negation annotations in the Genia Event and BioScope corpora. In Nigel Collier et al., editor, *Proc of the Fourth International Symposium for Semantic Mining in Biomedicine, Cambridge, UK*, volume 714 of *CEUR Workshop Proceedings*.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210.



# MAE2: Portable Annotation Tool for General Natural Language Use

**Kyeongmin Rim**

Brandeis University  
Department of Computer Science  
Waltham, Massachusetts, 01002 USA  
krim@brandeis.edu

## Abstract

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.

**Keywords:** manual annotation, annotation tool, inter-annotator agreement

## 1. Introduction

Since Stubbs (2011) introduced a natural language manual annotation tool for natural language, Multi-purpose Annotation Environment, paired with its companion adjudication tool, Multi-document Adjudication Interface, the pair has been used for semantic annotations over a wide variety of text genres, including clinical text (Sun et al., 2013), scientific journal articles (Meyers et al., 2014), and chronological travel logs (Pustejovsky et al., 2015) in the process of tackling widely-known but challenging semantic inference tasks, such as sentiment analysis (Herzig et al., 2011; Di Bari et al., 2013), temporal (Uzuner et al., 2013) or spatial reasoning (Kolomiyets et al., 2013; Pustejovsky et al., 2015), and discourse relation detection (Yung, 2014). Despite of its strength of being reasonably adaptable, flexible and portable that resulted in the wide adoption in the NLP community, the original MAE and MAI left room for improvement, for example, their error-prone user interface. In this paper, we report on the improvements we have made on the newer version, Multi-document Annotation Environment 2, MAE2<sup>1</sup> which is recently released, and discuss future design plans.

## 2. Design Principles of MAE

When it comes to a semantically-driven natural language annotation task, theoretical modeling of the linguistic phenomenon behind the task quickly gets very complicated, compared to shallow structural annotations. As a result, the task designers often come up with a highly complicated specifications for semantic annotation tasks. The major cognitive loads such a complicated annotation scheme puts on human annotators often make the already intricate tasks more time-consuming and costly, and sometimes to end up with unreliable results. That being said, having a properly decomposed annotation scheme to lessen the burden

<sup>1</sup>Release packages and users' guide are available from MAE project website: <https://www.keighrim.github.io/mae-annotation>

to the human annotators is crucial in designing annotation tasks that can capture various types of semantic phenomena. However such a design always involves incremental development of the model and the scheme through repetitive model-annotate prototyping especially in the very early stage, introduced as MAMA cycle in Pustejovsky and Stubbs (2012). MAE/MAI was originally developed to provide a simple yet flexible interface to formalize task specifications separately from guidelines or mark-ups, using a slightly modified Document Type Definition (DTD) syntax, to facilitate the iterative design procedure. Furthermore, all mark-ups from MAE are exported as a stand-off XML format as defined in the DTD specification, to make separation of annotation from the primary data that being annotated. Both separation of annotation instances from the data and from the annotation structure are suggested as principles for linguistic annotation by the ISO Linguistic Annotation Framework (Ide and Romary, 2004; ISO 24612:2012, 2012). Additionally, MAE/MAI is designed to be lightweight and portable so that it provides accessible cross-platform interface for annotators who usually have little, if any, experience of the computational procedure of the annotations.

## 3. Improvements in MAE2

In the process of upgrading MAE, while maintaining its simplicity, adaptability, and portability from the previous version, we have added many new features that we expect to help both task managers and annotators in various aspects.

### 3.1. Multi-Document Annotation

The new MAE2 is providing a tab-based annotation environment (Figure 1) that can handle multiple documents simultaneously. This will help annotators to remain consistent over different documents in the corpus, by allowing them to refer back to their work in the recent past.

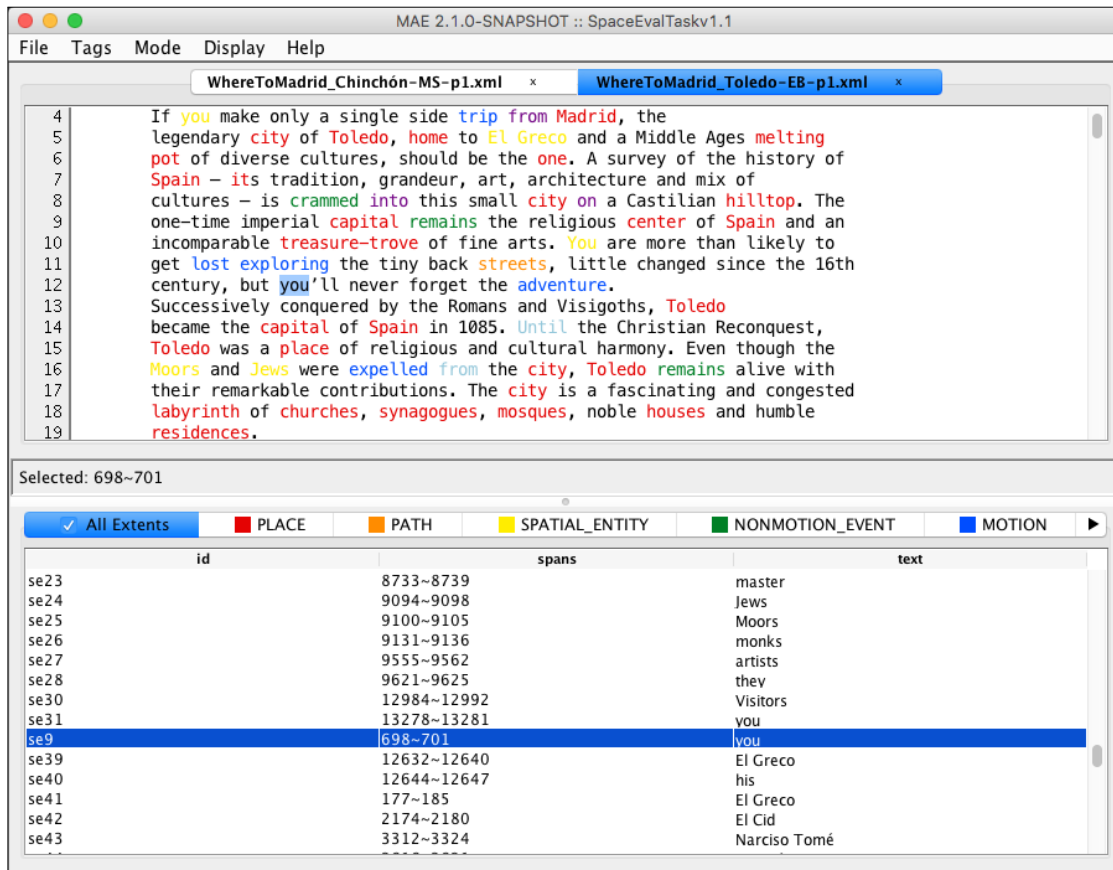


Figure 1: Multi-document Annotation Environment

### 3.2. Integrated Adjudication Interface

The original MAE and MAI were released as separate tools for annotation and adjudication respectively. This dyad implementation was not only unnecessarily increasing the learning curve by presenting two separate interfaces, but also imposing additional maintenance-wise expense with a number of redundant lines of code. Figure 2 shows the new interface for adjudication which is now integrated within a single application. Adjudication in MAE is carried out element-by-element, with the visualization for the agreement between annotators is given - at text span level as well as at attributes level. With its integration with the adjudication interface, along with a built-in IAA calculator that we will discuss in the following section, the MAE2 is expected to provide an accelerated experience for prototyping of a complex annotation task. In addition, the new MAE2 will reduce the cost of maintenance and modification by removing the major redundancy in the code base.

### 3.3. Built-in IAA Calculator

The MAE2 is equipped with an inter-annotator agreement (IAA) calculator that task managers can use to measure the reliability of the resultant annotations. The calculator is implemented based on

dkpro-statistics-agreement<sup>2</sup> library from the DKPro group (Meyer et al., 2014) that provides calculation of various agreement coefficients of  $\pi$ ,  $\kappa$ , and  $\alpha$  families that are widely adopted among the computational linguistics community as pointed out by Artstein and Poesio (2008). More importantly, the dkpro-statistics-agreement library provides a complete implementation of  $\alpha_U$  computation suggested by Krippendorff (1995) and Krippendorff (2004). The  $\alpha_U$  coefficient is designed to measure agreements of mark-ups anchored on arbitrary text spans (named as *unitization task* by Krippendorff) from multiple annotators. Since an annotation work-flow of the MAE is heavily depending on the entity-relation architecture as many semantic annotation tasks do, we believe the spans of entities indeed should be the atomic units of annotation. Thus, it is important for the MAE2 as a general-purpose tool to provide a generally applicable and commonly accepted metric to measure the agreement of text spans of entities such as the  $\alpha_U$ . However, the vanilla  $\alpha_U$  cannot be universal applied to all annotation tasks, as different types of tasks have different types of constraints in their underlying models. We are addressing this problem by offering several options on different levels of annotation elements, allowing taking such

<sup>2</sup>Available at <https://dkpro.github.io/dkpro-statistics> as of this writing.

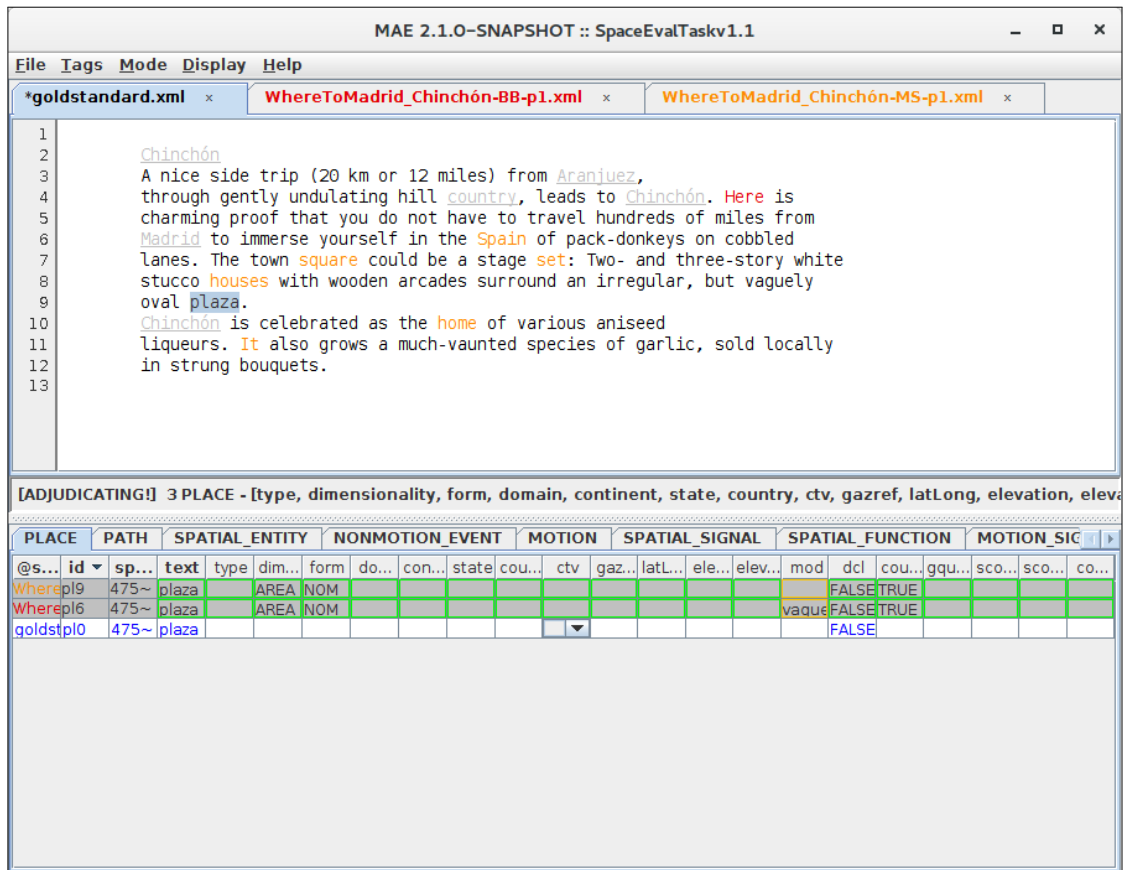


Figure 2: Integrated Adjudication Interface

task-specific constraints into account. For each tag, users can choose attributes to include in the calculation (choosing none of attributes would result in computing IAA solely upon the agreement of text spans for extent tags or spans of arguments for link tags). Additionally for extent tags, it's possible to choose to count only complete matches of text spans of mark-ups as agreements, which would make sense in a Chinese word segmentation task, for example. Also, users can opt to ignore the order of arguments when counting agreements for link tags, treating a link as a bag-of-entities as in a task of chaining co-References. This will provide ease of measuring tag-level and attribute-level agreements, and are expected to help task designers to easily spot common sources of error/underspecification from the annotation scheme to fix in the next revision of model or guidelines.

### 3.4. Marking Discontiguous Text Span

Using the new MAE2, annotators can mark up a segment or an entity that anchors over discontiguous text span. This can be particularly useful, for example, to annotate over discourse segments in conversational discourse, which often contain dysfluency fragments or interruptions inserted in the middle of segmentation units.

### 3.5. Linking N-ary Entities with Multi-slot Relations

Another improvement of the MAE2 in terms of annotation expressiveness is the multi-slot relations. Since it was only able to make directional links between binary arguments in the original MAE/MAI, to work around this limitation, a task designer should have annotators fill in extra attributes of link tags as pointers to additional arguments when she/he wants more than binary link relations. By adding capability of handling customizable multi-slot relations which can have an arbitrary number of can-be-optional arguments with custom titles as a designer wants, we expect it to be easier for designers to model complex semantic structures with many participants, as well as annotators as human agents feel more natural and intuitive to mentally model such structures.

For instance, the annotation task specification for ISO-Space (Pustejovsky et al., 2011) introduced a simple ternary spatial relation, qualitative spatial link (*QSLINK*), to capture topological relations between an event/entity and a place/location that triggered by an explicit/implicit linguistic cue phrase. However, in the dataset used at *SemEval-2015 Task 8: SpaceEval* (Pustejovsky et al., 2015) which was annotated using the old MAE/MAI under the ISO-Space specification, the implementation of *QSLINK* not only had to treat the third participant (*trigger*) not as a real

argument but as an extra attribute as a pointer to the phrase, but also had to use redundant attributes for entities (*trajectory*) and place (*landmark*) to entitle each argument. We believe that the new feature can provide a useful solution to such problems while the MAE2 still provides the plain old binary linking, which is fairly commonly used, as its default configuration to prevent unnecessary design overheads.

### 3.6. Customizable Visualization

Although the original MAE provided visualization for annotation elements based on text colors and styles, there was no practical way for annotators to control the visualization. We have added customization that allows annotators to customize the color for each element or to completely turn on or off the visualization for specific elements. We expect this feature not only to provide a more focusable annotation environment by helping annotators to easily isolate the immediate task objectivities, but also to make MAE2 more adoptable for multi-dimensional or multi-phased annotation tasks, which is a frequent case in semantic annotation.

### 3.7. Annotation Integrity and Quality Control

We have added additional means to assure the quality of annotations, which the previous MAE lacked. Any unsaved change in the working document will be notified to the annotators when they try to close the document. Any tag instance which is not completely fulfilled in respect to its required attributes will also be notified to the annotators. We added several internal validation procedures to evaluate attribute values given by the annotators, to make sure they are valid values. This is particularly important for the attributes-as-pointers (*IDREF* type attributes), as they are supposed to point to valid target tags. All these changes are expected to contribute to reducing human errors in annotations by and large, therefore, producing more robust annotations that halve efforts in post-processing including reviewing and cleaning ill-formed annotations.

### 3.8. Common Software Design Patterns

As MAE is an open source software<sup>3</sup>, we believe technical changes under the hood are worth mentioning as well, aside from all aforementioned changes and additions of features. While developing the new version, we adopted two major software design patterns: ORM and MVC. The Object-Relational Mapping pattern we used to implement the current SQLite-powered backend database operations will make it easier to expand the MAE2 into larger systems with different database architectures, in any necessary case. We also have followed the Model-View-Controller pattern to implement the frontend interface and frontend-backend interconnection. We expect the adoption of these common design patterns will enable a group who work on a specific annotation task to surgically amend the software when they need to add or modify particular features for their task.

<sup>3</sup>Both MAE and MAE2 are released under GNU General Public License v3 <http://www.gnu.org/licenses/gpl-3.0.en.html>

### 3.9. Limitations

In spite of all the shiny features we have discussed so far, MAE2 still has limitations. MAE is written as a Java standalone application and distributed as a single executable jar file along with the source code. As a result, of course, annotators should have access to a local computer system with Java Runtime Environment (JRE) installed. Nevertheless, we believe that this can be an advantage over web-based tools in a way, since desktop applications not only can offer a more native user experience, but also they are free from a network connection, thus could be faster and scale well being free from network latencies. Furthermore, cross-platform Java applications would reduce developing overhead, needless to worry about the fragmentation from various browsers and their versions currently available. However, MAE2 currently does not provide a method or protocol to side-load data from a remote server or to store annotation results remotely, and consequently annotators need to store the dataset locally to work on it. Not only this can be problematic particularly when annotating sensitive data under complicated data-use agreements as Chen and Styler (2013) pointed out, but this can also be a burden on project managers as they have to manually assemble the results from individual annotators. Another limitation of MAE2 is that, even though MAE2 now supports multi-document annotation environment, it does not support cross-document annotation. Likewise, MAE2 does not support an annotation interface for hierarchical structures such as syntactic trees, as MAE is claimed to be a lightweight, portable general-purpose tool.

## 4. Existing Annotation Tools

Before jumping into the conclusion, it is of course important to acknowledge the current existing natural language annotation tools. We present table 1 for quick comparison of features of available general-purpose annotation tools that are being widely used recently.

BRAT (Stenetorp et al., 2012) is a web-based annotation tool providing recognizable visualizations and intuitive user interface. It is probably the most successfully adopted tool in the community, partly because of its visualization feature and intuitive interface based on the visualization. However Yimam et al. (2013) pointed out that it often fails to scale for a fairly large annotation task, suffering from the overheads caused by the visualization.

Still BRAT's superior visualization has been adopted, usually as a front-end component, by other tools, and WebAnno (Yimam et al., 2013) is one of those. It managed to provide better scalability under the BRAT visualization as well as many useful features such as in-place adjudication, IAA calculator, and interface to crowd-sourcing platforms. More recently WebAnno has been extended toward a machine-aided annotation software (Yimam et al., 2014), becoming a heavy toolbox.

Anafora (Chen and Styler, 2013) is probably the annotation tool that is most closely comparable to MAE. Anafora is a multi-purpose but lightweight annotation tool. However, as it is built as a web application, though it aims to being a lightweight tool, using Anafora for an annotation

	MAE2	BRAT	WebA	Anaf
Platform	JRE	Web	Web	Web
Req. server instance	×	○	×	○
Discontiguous mark-ups	○	○	×	○
Multi-slot linking	○	×	×	○
IAA calculator	○	×	○	×
Machine-aided annotation	×	○	○	×
Collaborative annotation	×	○	×	×
Crowdsourcing integration	×	×	○	×

Table 1: Quick comparison of available tools

project will involve configuring, securing, and running a web server to keep the tool alive, unlike standalone desktop software that can readily be distributed among annotators. This holds true for many other web-based tools popular these days.

## 5. Conclusion and Future Work

In this paper, we presented the recent improvements we made in the new MAE2. The MAE2 is a general-purpose natural language annotation software with many features to assist rapid task design for a wide range of semantic annotation tasks and provide a robust annotation interface that can lessen human errors. MAE is an open source project, and the release package, the users' guide and the source code are currently available on the project website. MAE project is still actively developed as we have plans for further upgrade of MAE2. For example, the visualization for link tags still has room for improvement. We are also planning to add a graphical interface for creating and editing task definitions (DTD format). Also in compliance with LAF specification, we will add support for GrAF "dump" format (Ide and Suderman, 2007) to generate sustainable and interoperable output. Lastly, but not least, since the MAE2 has a clear limitation in terms of the way it loads/stores data as we mentioned, we are exploring an appropriate way to implement network-oriented data exchange method by either moving toward a web-based interface or implanting a network protocol within the desktop application.

## 6. Acknowledgments

This work has been supported by an NSF grant to Prof. James Pustejovsky, NSF 1147912. I would also like to show my gratitude to Prof. James Pustejovsky and Zachary Yocum for great ideas to improve MAE, also to Amber Stubbs for the great project she started, as well as to three anonymous reviewers for helpful comments.

## 7. Bibliographical References

- Artstein, R. and Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4):555–596.
- Chen, W.-T. and Styler, W. (2013). Anafora: A Web-based General Purpose Annotation Tool. In *Proceedings of the 2013 NAACL HLT Demonstration Session*, number June, pages 14–19.
- Di Bari, M., Sharoff, S., and Thomas, M. (2013). SentimentML: functional annotation for multilingual sentiment analysis. In *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities*, pages 15–22.
- Herzig, L., Nunes, A., and Snir, B. (2011). An Annotation Scheme for Automated Bias Detection in Wikipedia. In *Proceedings of the Fifth Linguistic Annotation Workshop*, number June, pages 47–55.
- Ide, N. and Romary, L. (2004). International standard for a linguistic annotation framework.
- Ide, N. and Suderman, K. (2007). GrAF: A Graph-based Format for Linguistic Annotations. In *Proceedings of the Linguistic Annotation Workshop*, pages 1–8.
- ISO 24612:2012. (2012). Language resource management – Linguistic annotation framework (LAF). Standard, International Organization for Standardization, Geneva, CH.
- Kolomiyets, O., Kordjamshidi, P., Moens, M.-F., and Bethard, S. (2013). SemEval-2013 Task 3: Spatial Role Labeling. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation*, volume 2, pages 255–262.
- Krippendorff, K. (1995). On the reliability of unitizing continuous data. *Sociological Methodology*, pages 47–76.
- Krippendorff, K. (2004). Measuring the reliability of qualitative text analysis data. *Quality & quantity*, 38:787–800.
- Meyer, C. M., Mieskes, M., Stab, C., and Gurevych, I. (2014). DKPro Agreement: An Open-Source Java Library for Measuring Inter-Rater Agreement. In *Coling 2014 (Demos)*, pages 2–6.
- Meyers, A., Lee, G., Grieve-smith, A., He, Y., and Taber, H. (2014). Annotating Relations in Scientific Articles. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 4601–4608.
- Pustejovsky, J. and Stubbs, A. (2012). *Natural language annotation for machine learning*. O'Reilly Media, Inc.
- Pustejovsky, J., Moszkowicz, J. L., and Verhagen, M. (2011). ISO-Space : The Annotation of Spatial Information in Language. In *Proceedings of the Sixth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic*

- Annotation*, pages 1–9.
- Pustejovsky, J., Kordjamshidi, P., Moens, M.-F., Levine, A., Dworman, S., and Yocum, Z. (2015). SemEval-2015 Task 8: SpaceEval. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, number 1, pages 884–894.
- Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). BRAT: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, number Figure 1, pages 102–107.
- Stubbs, A. (2011). MAE and MAI: Lightweight Annotation and Adjudication Tools. In *Proceedings of the Fifth Linguistic Annotation Workshop*, pages 129–133.
- Sun, W., Rumshisky, A., and Uzuner, O. (2013). Annotating temporal information in clinical narratives. *Journal of Biomedical Informatics*, 46(SUPPL.):S5–S12.
- Uzuner, Ö., Stubbs, A., and Sun, W. (2013). Chronology of your health events: Approaches to extracting temporal relations from medical narratives. *Journal of biomedical informatics*, 46:S1.
- Yimam, S. M., Gurevych, I., Eckart de Castilho, R., and Biemann, C. (2013). WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, volume 1, pages 1–6.
- Yimam, S. M., Biemann, C., de Castilho, R., Gurevych, I., Muhie, S., Richard, Y., and Castilho, E. D. (2014). Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 1(1):91–96.
- Yung, F. (2014). Towards a discourse relation-aware approach for Chinese-English machine translation. In *Proceedings of the ACL 2014 Student Research Workshop*, number June, pages 18–25.

## Contrastive Annotation of Epistemicity in the MULTINOT Project: Preliminary Steps

Julia Lavid, Marta Carretero, Juan Rafael Zamorano

Department of English Philology  
Faculty of Philology  
Universidad Complutense de Madrid  
Ciudad Universitaria s/n  
28040 Madrid

E-mail: julavid@filol.ucm.es, mcarrete@filol.ucm.es, juanrafaelzm@filol.ucm.es

### Abstract

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.

**Keywords:** epistemicity, annotation, English-Spanish contrasts

### 1. Introduction and Background

This paper describes the preliminary steps undertaken for the annotation of epistemic meanings in English and Spanish in the context of MULTINOT project, aimed at the creation of a high-quality, register-diversified parallel and medium-sized corpus for the English-Spanish pair. The bilingual corpus consists of originals and translated texts in both directions and is enriched with linguistic annotations which can be exploited in a number of linguistic, applied and computational contexts (see Lavid et al. 2015).<sup>1</sup> Among the linguistic annotations foreseen for the MULTINOT corpus are those pertaining to the semantic domain of modality, in general, and epistemicity, in particular, given their relevance and importance for contrastive investigations between these two languages and in applied NLP contexts such as machine translation.

The notion of epistemicity has been proposed in the literature as a conceptual domain comprising the subcategories of epistemic modality and evidentiality. The former is defined in terms of the notion of degree of certainty, degree of commitment or epistemic support, while the latter is defined in terms of the notion of source of information, evidence, justification, or epistemic justification (Boye 2012: 2). While there exist proposals for the annotation of epistemicity and other modal values for the English language, and recently there have also been proposals for other languages, such as Portuguese (see Hendrickx et al., 2012), to date there are no concrete proposals for the annotation of epistemic meanings in Spanish in a contrastive manner with English. In this

paper, therefore, we intend to fill a gap in this area, by presenting an annotation scheme which attempts to capture the functional similarities and the linguistic differences between these two languages when expressing epistemicity in different contexts.

The paper is organised as follows: we first describe the conceptual domain of epistemicity and the two subcategories which have been studied as comprising this domain, namely, evidentiality (section 2.2) and epistemic modality (section 2.3). We then present in section 3 the steps undertaken for the annotation of these categories, consisting of: a) identifying and defining the categories to be used as tags; b) designing an annotation scheme which captures both the functional-semantic dimensions of epistemicity, on the one hand, and the language-specific realisations of epistemic meanings, on the other; and c) performing a pilot agreement study to validate the reliability of the proposed annotation scheme. Finally, in section 4 we conclude with a summary and some pointers for the future.

### 2. Epistemic Meanings in English and Spanish

The notions of epistemic modality, epistemic stance and evidentiality have attracted an increasing amount of interest in the last two decades among researchers from different theoretical backgrounds. Some authors treat them as separate categories (de Haan, 2000; Marín-Arrese, 2004, 2015; Cornillie, 2009), while others include one inside the other, treating evidentiality as a subtype of epistemic modality (Palmer, 2001). Other authors advocate for treating these domains as highly overlapping in their linguistic expression (Willet, 1988; van der Auwera and Plungian; Carretero, 2004). One of the most illuminating accounts is the one provided by

<sup>1</sup> The MULTINOT project is financed by the Spanish Ministry of Economy and Competitiveness under project grant FFI2012-32201. The authors of this paper gratefully acknowledge the support provided by the Spanish authorities.

Boye, who proposes a general category, ‘epistemicity’, as a general conceptual domain which comprises two related subcategories: a) *epistemic modality*, defined in terms of degree of certainty, degree of commitment or epistemic support, and b) *evidentiality*, defined in terms of the notion of source of information, evidence, or epistemic justification. A more detailed description of these meanings is provided in section 2.1. below.

### 2.1. Evidentiality

Evidentiality is a way of qualifying the truth of a proposition by expressing the *source of the evidence* that the speaker has or claims to have at his / her disposal, for or against this truth. The sources can be divided into three types: *perceptual*, *cognitive* and *communicative* (Marín-Arrese, 2013). In many cases, the same expressions can be used with more than one mode. The modes are defined as follows:

**-Perceptual:** this category comprises non-linguistic evidence obtained through the senses. Examples for English (1) and Spanish (2) are provided below (the linguistic item expressing evidentiality is in bold in these examples and the following):

- (1) and the developing world **looks** to be 50 years behind 006
- (2) y la belleza de su entorno natural **parece** tener una espiritualidad ancestral  
the beauty of its surrounding natural seem(p-3sg)  
have(inf) a spirituality ancestral(sg)  
‘and the beauty of its surrounding environment seems to have an ancestral spirituality’

**-Cognitive:** the evidence comes from knowledge by someone different from the speaker/ writer. Cognitive evidence includes thoughts, beliefs and apprehension attributed to other people. Examples for English (3) and for Spanish (4) are provided below:

- (3) Anyone who has studied economic performance since the onset of the financial crisis in 2008, **understands** that damage to balance sheets – such as excess debt and unfunded non-debt liabilities – can cause growth slowdowns, sudden stops, or even reversals.
- (4) **Se cree** que los primeros pobladores del Imp(3sg) believe (pres-3sg) that the(pl) first(pl) inhabitants of-the (m-sg) Neolítico entraron desde el norte de África hasta Europa  
Neolithic enter (past-3pl) from the(m-sg) north of Africa until Europe  
por las tierras andaluzas through the(f-sg) land(pl) Andalusian (f-pl)

‘It is believed that the first Neolithic inhabitants went in from the North of Africa to Europe through

the Andalusian land’

A controversial issue are the speaker/writer’s cognitive processes, as in mental state predicates in the first person singular, such as *I believe*, *I guess*, *I suppose* or *I think*. Some references (Perkins, 1983; Kärkkäinen, 2003) consider them as epistemic, on the grounds that they express attitude towards the truth of the proposition (belief, but not total knowledge). Other authors consider them as evidential, in the sense that the speaker/writer’s mental state can be considered as a source of evidence (Marín Arrese, 2013). We consider that the first dimension is more salient, in the sense that these verbs in the first person express, above all, a medium degree of certainty about the truth of the proposition (higher than 50%, but not total). Therefore, these expressions are considered as epistemic, in the ‘probability’ subtype (see below). Other similar expressions are *I know* and *I remember*, which express knowledge that the proposition is true; *I remember* has a stronger component of retrieval of information in the speaker/writer’s mind. In order to give a coherent treatment to the speaker/writer’s cognitive processes, we have classified both within epistemic modality, in the subcategory ‘certainty’ (see below). In their turn, a number of expressions with these verbs in the first person, such as *as far as I know*, *as far as I remember*, and the synonym *to my knowledge*, express medium certainty due to the limitations of the speaker/writer’s knowledge, and have therefore been classified under ‘probability’.

**-Communicative:** evidence that comes from linguistic messages. In line with most of the literature, hearing of linguistic messages is considered in the ‘communicative’ subtype, not in the ‘perceptual’ subtype. The source can be specified or not. An example of the first kind is (5), in which the source is ‘she’. In (6), the source is not specified:

- (5) **Según ella**, los americanos son los  
According to she(obj-3sg) the(m-pl) American(m-pl) be (pres-3pl) the(m-pl) continuadores del Pueblo Elegido  
continuers-of-the(m-sg) People Chosen(m-sg)  
‘According to her, the Americans are the continuers of the Chosen People’
- (6) Indeed, the fragility of the global economic recovery is often **cited** as a justification to delay such action.

We exclude direct and indirect quotations as evidential expressions. A similar position may be found in Boye (2012: 32) and Chojnicka (2012: 173), who state that verbatim quotations introduce what someone else has said, while reportive evidentials qualify information as not personally witnessed and coming from other sources. It could be argued that reported speech communicates



evidentiality as a conversational implicature in the sense that, by using these constructions, the speaker/writer is indirectly qualifying the reported contents as not personally witnessed but communicated by the original speaker. That is to say, reported speech could be considered as evidential but only in a peripheral way. For this reason, we have not included it in our annotation scheme.

## 2.2. Epistemic Modality

Epistemic modality refers to the estimation of the chances of a proposition to be or become true, but not by means of qualifying evidence for or against it. The speaker/writer may express knowledge that the proposition is true or false, or else s/he may not be sure about its truth or falsity and therefore proceed in different ways: assign a degree of probability to it, express belief, or express doubt. These meanings can be grouped according to the following parameters:

- strength of the degree of probability, which determines ‘possibility’, ‘probability’ and ‘certainty’.
- reference to mental states, expressed directly by means of speech act verbs (Perkins 1983). Three categories are distinguished according to the strength of the mental state: ‘doubt’, ‘belief’ and ‘knowledge’.
- existence of an emotive meaning, expressing a favourable or unfavourable attitude towards the truth of the proposition together with a lack of certainty. This is the case of ‘apprehension’.

In the following paragraphs we will describe and illustrate each of these meanings in detail.

### 2.2.1. Possibility

This category covers low certainty, that is, around 50% probability for the proposition to be true, as in (7) and (8) below. Clauses with expressions of this category can be coordinated with clauses where the same expression qualifies the same proposition with the opposite polarity or another incompatible proposition, as illustrated by (9) and (10) below:

- (7) It will take another generation, **perhaps**, before robots have completely taken over manufacturing, kitchen work, and construction.
- (8) Es **posible** que Renfe-Operadora le facilite Be(pres-3sg) possible(sg) that Renfe-Operator you(V-Obj-sg) facilitate (pres-subj-3sg) el acceso a otras páginas web que consideramos pueden the(m-sg) access to other(f-pl)pages web that consider (pres-1pl) can(inf) ser de su interés. be(inf) of your(V) interest ‘Renfe-Operator may facilitate to you the access to other web pages that we consider might be of your interest.’

- (9) The lack of a licence in Barlow Clowes' early years **may or may not** have made a difference to the way investors' funds were handled during that time.

- (10) **Puede no** darles que pensar, pero May(3sgPres) no give-you(inf-2pV) that think(inf) but **puede que sí.** Yo, por si acaso, May(3sgPres) that yes I in case se lo indico you(2pV) it(3s-obj) indicate(1s-pres-ind) ‘Perhaps it will not make you think, but perhaps it will. Just in case, I will show you.’

### 2.2.2. Probability

This category concerns a medium degree of certainty. It includes expressions that communicate stronger probability than those of ‘possibility’, as illustrated by (11) and (12) below. They cannot be used in coordinated constructions with the opposite polarity, but can be followed by ‘but I’m not (absolutely) sure’ / ‘but I don’t know’, as in the dialogue in (13):

- (11) It can be hard to imagine what they really are, for what they really are is far beyond our ordinary experience. If you are a regular stargazer, you have **probably** seen an elusive light hovering near the horizon at twilight.
- (12) De estos últimos se cree ahora que Of these(m-pl) last(m-pl) IMP(3sg) believe(pres-3sg) now thateran un pueblo neolítico que **probablemente** be(past-impf-3pl) a(m-sg) people(sg)Neolithic(m) that probablyllegó desde el norte de África en el primer arrive(past-3sg)from the(m-sg) northof Africa in the(m-sg) first milenio antes de Cristo, millennium before of Christ ‘About these last people, now it is believed that they were a Neolithic people who probably arrived from the North of Africa in the first millennium before Christ’
- (13) Do you think that our century will be the age of surrealism? Yes, **probably, but I don't know for sure.** Deep down, I believe that our century will not be very interesting compared to other centuries.

Probability also includes expressions that estimate the chances of the propositions to be true by referring specifically to mental states –usually expressed as mental

processes (*I think, I believe, I suppose...*) as in (14) and (15) below:

(14) So now we have all the crew of the boat gathered here, except for the Norwegian, and **I believe** there is an Australian expected.

(15) [...] hay una **probabilidad** altísima de que su obra no salga de la más respetable medianía.  
There is a probability very high that his work not stand (3rpsg) from the most respectable mediocrity.  
'There is a very high probability that his work does not stand out from the most respectable mediocrity'.

We also include occurrences of mental processes (e.g. *I believe, I think* and their Spanish counterparts) when they express opinion rather than degree of probability. Stubbs (1986) considers that '*I think*' has a modal meaning when the proposition is verifiable, and a psychological meaning when it is not verifiable. However, the distinction is not easy to draw: some propositions are in theory verifiable, but not in practice unless professional statistical studies are carried out, and the speaker does not modalise the proposition with that spirit, as in (16) below:

(16) **I think** that people born in the 80s are more careful with their diet than people born in the 60s.

The category of probability also includes expressions with 'know' or 'remember', or derived constructions ('to my knowledge'), that indicate non-total certainty due to the limitations of knowledge, as in (17) and (18) below:

(17) They also acquired a very enthusiastic Deputy Headmaster who was very, very keen **as far as I remember**.

(18) **Después, que yo recuerde**,  
Later, that I remember (1-pres-subj-sg)  
he estado dos veces en el Festival de S. Sebastián.  
have been (1sg)two times in the Festival of SS.  
'Afterwards, as far as I remember, I have been twice in San Sebastián Festival'.

### 2.2.3. Certainty

This meaning involves a higher degree of commitment to the truth of the proposition than that of probability: the expressions do not admit the occurrence with 'but I'm not absolutely sure' / 'but I don't know'. The speaker / writer may know that the proposition is true, as in (19) or not, as in (20), where the event will take place in the future:

(19) Did you know Jos in those days, Grandma?  
'**Certainly** I did', she said.

(20) Es **seguro** que, al menos en los Estados Unidos, Be(pres-3sg) sure(m-sg) that at least in the(m-pl) United States el año Darwin será ocasión de que se the(m-sg) year Darwin be(fut-3sg) occasion of that pass(3sg) acentúe la crudeza de la polémica accentuate(pres-subj-3sg) the(3sg) rawness de los defensores de las tesis evolucionistas of the defenders of controversy between the(m-pl) defensores de las tesis evolucionistas y los defensores de defenders(m) of the(f-sg) theses evolucionista(pl) and the(m-pl) defenders(m) of posiciones creacionistas, positions creationist(pl)

'It is sure that, at least in the United States, the Darwin year will be an occasion for the increase of the controversy between defenders of evolutionist theses and defenders of creationist positions.'

Some expressions of certainty are often used mainly for pragmatic reasons or for reasons of information structure (see, for example, Byloo et al. 2007). For example, in (21) '*certainly*' expresses acceptance to comply with a request, and in (22) it lays emphasis on the concessive relationship between the clause where it occurs and the following coordinate clause introduced with an expression of concession. However, we believe that '*certainly*' still expresses certainty in these cases: its semantic value of certainty is compatible with the pragmatic or discourse functions mentioned above.

(21) Could I have a copy of the letter, please, can I take it up? You **certainly** can, Anne, thank you.

(22) It **certainly** was a challenge to have to teach people stuff by Steve Vai or Yngwie Malmsteen, but the most difficult was Allan Holdsworth.

This category also comprises the expressions *I know* and *I remember*, which indicate the speaker/writer's knowledge (and therefore total certainty) of the truth of the proposition, as exemplified in (23) below:

(23) Así **recuerdo** yo el testimonio de mis padres y So remember(pres-1sg) I the(m-sg) testimony of my(pl) parents and abuelos cuando hablaban sobre Julio de 1936 grandparents when talk(past-3pl) about July of 1936  
'I remember my parents' and grandparents' testimony to be like that when they talked about July 1936.'

### 2.2.4. Doubt

The category of ‘doubt’ concerns the expression of a mental state of doubt, uncertainty or lack of knowledge of the truth of the proposition, without assigning any degree of probability to it. These expressions are related to Brandt’s (2004: 6-7) stance of ‘aphony’, by which “the speaker emphatically withdraws or refrains from investing in the utterance”. Examples for English (24,25) and for Spanish (26) are provided below:

- (24) But there is **no guarantee** that gains in service-sector employment will continue to offset the resulting job losses in industry.
- (25) But, unless the proper policies to nurture job growth are put in place, it remains **uncertain** whether demand for labor will continue to grow as technology marches forward.
- (26) El caso es que no habiendo vivido guerra The(m-sg)case be(pres-3sg) that not having live(pple) war alguna, los de mi quinta **no sabemos** cómo empiezan. any(f-sg) the(m-pl) of my(sg) age not know(pres-1pl) how start(pres-3pl) ‘The point is that people my age, who have not lived any war, we do not know how they start.’

### 2.2.5. Apprehension

Apprehension is uncertainty combined with a positive or negative wish for or against the truth of the proposition (Lichtenberk 1995: 293-294). Givón (1984) calls this ‘epistemic anxiety’. Apprehension has an epistemic element and a volitional element, but has been classified here as an epistemic and not as a volitional category because the speaker / writer does not have (at least total) control over the truth of the proposition being or becoming true. Besides, apprehension has propositions under its scope, like the other types of epistemic modality. Examples of apprehension are provided in (27) and (28) below:

- (27) The upheavals resulting from momentous technological change are rarely **expected**.
- (28) Y no sé, **espero** que se And not know(1sg-pres-ind) hope(1sg-pres-ind) that it(obj-3sg) haga realidad esta buena idea. (1sg-pres-subj) reality this(f) good idea ‘And I don’t know, I hope that this good idea comes true.’

## 3. Annotation Scheme

Once the conceptual domain of epistemicity and its subcategories has been described, our next tasks are focused on: a) identifying the categories to be used as tags for annotation; and b) performing a pilot agreement study to validate the proposed tags. These two tasks are described in detail in subsections 3.1. and 3.2. below.

### 3.1. Annotation Tagsets

Given the fact that we want to compare the domain of epistemicity in two different languages, we have found it necessary to work with two tagsets to be able to identify the functional similarities and the linguistic differences between these two languages. This distinction has also been adopted by other researchers working on the cross-linguistic annotation of modality (see Nissim et al. 2013), but our proposal provides a much more detailed characterisation of epistemicity, both in terms of the functional categories and the linguistic candidates that encode these categories in English and Spanish.

On the one hand, we propose one functional-semantic tagset, which captures the epistemic meanings which occur both in English and in Spanish, as graphically displayed in table 1 below:

Epistemicity	Evidential [EV]	Perception [PE]
		Cognition [CO]
		Communication [COM]
	Epistemic modal [EM]	Possibility [POS]
		Probability [PRO]
		Certainty [CER]
		Doubt [DO]
		Apprehension [AP]

Table 1: Functional Tagset for Epistemicity in English and Spanish

This tagset is hierarchical, allowing annotators to choose more general or coarser tags when in doubt about the more delicate ones. For example, if the annotator is uncertain about whether a markable is ‘possibility’ or ‘probability’, s/he can simply tag it as ‘epistemic modal’. The abbreviated form of each tag is given in capital letters in brackets next to the full form.

On the other hand, we propose a linguistic tagset, which captures the language-specific realisations of the epistemic meanings presented in table 1 above. The tags here capture a wide variety of linguistic realisations of epistemic meanings in English and Spanish both in terms of lexicogrammatical options (LG) and in terms of the syntactic functions and constructions (SF) where the lexicogrammatical options can occur, as shown in table 2 below:

LG	SF	ENGLISH		SPANISH	
		EPIST.	EVID.	EPIST.	EVID.
Adv. [A]	Modal Adjunct [AD]	Perhaps, possibly, presumably [ADEP]	Apparently, visibly, obviously [ADEV]	Quizás, posible mente,	Por lo visto, al parecer
		Pred. Adj in impersonal matrix clause [AJIP]	It is possible, likely + that [AJIPEP]	It is evident + that indicative [AJIPEV]	Es posible que + Subj.
Pred. Adj in speaker	Pred. Adj in speaker	I am sure, certain that, doubtful	----	Estoy seguro( a) de	----

	-hearer matrix clause [AJIT]	<i>whether</i>		<i>que + Subjunctive</i>	
Adj. [AJ]	Pred Adj in to+ inf cl [AJIF]	<i>He is sure, certain, likely to ...</i>	----	----	----
	Attrib.A in NG [AJN]	<i>He is a sure winner</i> [AJNEP]	<i>He is an obvious winner</i> [AJNEV]	<i>Es una apuesta segura</i>	<i>Una corrupción evidente</i>
Noun [N]	Noun C. in imp.cl [NI]	<i>There is a possibility/ that + Ind.</i> [NIEP]	<i>There is a rumour that / + Ind</i> [NIEP]	<i>Hay (una) posibilidad ad(es) de que +Subj.</i>	<i>Hay rumores de que / una creencia en que...</i>
Verb [V]	Verb Ope.in matrix cl [VO]	<i>It might/must be true</i> [VOEP]	<i>He seems /appears to be right</i> [VOEV]	<i>Podría/ Debe ser verdad</i>	<i>Parece tener razón Pinta raro</i>
	Verbal inf. [VI]	----	----	<i>Future tense: Será verdad</i> [VIEP]	<i>Cond. verbal mood: Habría</i> [VIEV]
	Mental process (1 <sup>st</sup> p sg/pl) + that cl [MP]	<i>I/we think/believe/ hope/ guess/know it is true</i> [MPEP]	<i>I conclude/ notice/ hear + (that) indicative</i> [MEEV]	<i>Creo/ pienso/ /calculo/ sé/ que es verdad</i>	---
	Mental /verb process in passive [M/V]	----	<i>It is expected/ said / that +</i> [M/VEP]	----	<i>Se ve/dice que + indicative</i> [M/VEV]
	Verbal process with generic 3rd p.pl subject [VP]	----	<i>They say that</i> [VPEP]	----	<i>Dicen que</i>
	Rel.Proc in impersonal cl. [RP]	----	<i>It seems it was true</i> [RPEP]	----	<i>Parece que fue así</i> [RPEV]

Table 2: Linguistic tagset for Epistemicity in English and Spanish

The lexicogrammatical options are specified as a core tagset capturing the paradigmatic and more general linguistic encodings of epistemic meanings in English and Spanish (i.e., as adverb, adjective, noun or verb). The syntactic functions and constructions are specified as an extended tagset capturing the syntagmatic encodings where the lexicogrammatical options can occur in both languages. Some tags only hold for one of the languages (i.e., verbal inflection only holds for Spanish). In such cases, we provide an example of the available language and cross out the one that is not available in the other language. As in the functional-semantic tagset, these linguistic tags are also hierarchically organized and become more specific as we move to the right in the table. This is to allow the annotator to opt for a more general tag when s/he is

uncertain which tag to choose at the most specific level.

### 3.2. Pilot Agreement Study

In order to test the reliability and consistency of the functional tagset proposed for annotating epistemic meanings in English and Spanish, we carried out a pilot agreement study on a randomly selected set of one hundred and twenty sentences from the MULTINOT corpus (seventy sentences in English and fifty in Spanish). The sentences contained lexicogrammatical candidates which can typically express epistemic modality, as in (29), evidentiality, as in (30), other type of modality, as in (31), or not express modality at all, as in (32).

- (29) **perhaps** it wants to reinstitute debtor prisons for over indebted countries.
- (30) And yet, despite our obvious ability to produce much more than we need, we do not **seem** to be blessed with an embarrassment of riches.
- (31) My personal ancestors **must** have been living, or I wouldn't be here.
- (32) you **must clearly** and conspicuously disclose the nature of your connection to Sears.

The lexicogrammatical candidates included equal proportions of adjectives, nouns, adjectives, lexical verbs and modal verbs. The annotations were carried out by two expert annotators who tagged both the English and the Spanish sentences independently. Inter-annotator agreement results for the Spanish sentences are presented in table 3 below:

		Annotator A			
Annotator B		N-MODAL	EPISTEMIC	EVI-DENTIAL	MODAL (OTHER)
		N-MODAL	8	1	0
	EPIS.	2	11	0	5
	EVID.	4	1	10	0
	MOD	0	3	0	12

The number of observed agreements for Spanish examples is 41 (70.69% of the observations), and the number of agreements expected by chance is 14.6 (25.21% of the observations). The kappa value is 0.608. Therefore, the strength of the agreement is considered to be 'good'. In the case of the English sentences, the interannotator agreement results are presented in table 4 below:

		Annotator A				
Annotator B		N-MODAL	EPISTEMIC	EVIDENTIAL	MODAL (Other)	
		N-MODAL	4	1	0	0
		EPIS.	8	25	1	0
		EVID.	0	3	19	0
		N-MOD	0	0	0	12

B	DAL				
---	-----	--	--	--	--

Table 4: Interannotator agreement for English examples

Here the number of observed agreements is 60 (82.19% of the observations) and the number of agreements expected by chance is 22.3 (30.59% of the observations). The Kappa value is 0.743, which is also considered to be 'good'.

Although agreement rate is slightly higher in English, the overall result is similar in both languages, both in percentage and kappa value. The main conclusion from the results is that, even though there is still room for improvement, the distinction between the four basic categories (non-modal, epistemic, evidential and modal other than epistemic or evidential) is reasonably robust and can be replicated by different annotators to a large extent.

The tables also reveal the main areas of disagreement in English and Spanish. In Spanish the main problem is found in the distinction between epistemic and other modal meanings (more specifically, dynamic modality and the subtag of 'possibility' for epistemic modality). The reason for this is that in some contexts it is hard to tell if the sentence is describing a potential tendency or a possible development in the future, as illustrated in (33) below:

- (33) El objetivo es mantener la calidad y actualización de esta información y evitar y minimizar **posibles** errores causados por fallos técnicos.  
 'The Museum's aim is to maintain the quality of this information, update it and avoid and minimise possible errors arising from technical faults'

The second cause of disagreement is the distinction between epistemic or evidential meanings and the complete absence of modality. This was often the case with lexical items whose meaning includes a component that lends itself to an evidential interpretation, as in (34), or to an epistemic one, as in (35):

- (34) Es **natural**: si hay en el mundo un bien escaso, ése es el raro don de los genios.  
 'It's natural: if there is any scarce commodity in the world, this is the rare gift of genius'
- (35) Esto **vale** fundamentalmente para los negocios que tienen lugar en el ámbito financiero.  
 'This is true basically for deals that take place in the financial arena'

The distinction between epistemic modality and the absence of modal meaning seems to be a problematic area in English too, as it is sometimes hard to decide if a potential trigger of modality is used in a specific example to convey that modal meaning or not, as illustrated in (36) below:

- (36) We **believe** it is important for you to know how we treat the information you share with us.

To a lesser extent, discriminating between epistemic

modality and evidentiality also posed some problems in the English examples. These are often examples that involve the epistemic subtag of 'knowledge' and the evidential subtag of "cognitive", since it is sometimes difficult to say if having knowledge of something is used as irrefutable evidence or just as a way of showing certainty about the truth of the proposition, as in (37) below:

- (37) Google may warn you if it considers the app to be unsafe, or block its installation on your device if it is **known** to Google to be harmful to devices, data or users.

As for the agreement rate for the subtags defined within the broader categories of evidential and epistemic, they seem to be higher than those for the general categories. It is true that the number of examples on which our observation is made is significantly lower. This is due to the fact that here we are limited to only those cases on which there was agreement between annotators when classifying the examples as evidential or epistemic in the first place. But the cases we have suggest that it is in fact easier to distinguish between the various subtypes of epistemic or evidential modality than between epistemic, evidential, non-modal and other types of modality. The only exception is the subcategory of 'cognitive' within evidential, which was often confused with the evidential categories of 'communicative' and 'perception' in both languages. This is perhaps because perception – in particular visual perception – is often equated with comprehension in English and Spanish. The result is that it is sometimes hard to tell if a fact is visually perceived or simply mentally realized, as shown in (38) below:

- (38) Even stranger, productivity growth does not **seem** to be soaring, as one would expect

The reliability of these conclusions must be strengthened through further experiments to enlarge the body of examples on which they are based. However, we believe that this pilot study already hints at the main problematic areas in the definition of epistemicity and its subcategories. Experiments like this on a large scale will reveal when automatic annotation is feasible as well as which modal categories must be redefined.

#### 4. Summary and future work

In this paper we have presented the preliminary steps undertaken within the MULTINOT project for annotating epistemicity in English and Spanish in a contrastive manner: first, we have identified and defined the categories to be used as tags for annotation; second, we have presented an annotation scheme which captures both the functional-semantic dimensions of epistemicity and the language-specific realisations of epistemic meanings these two languages; third, we have described a pilot agreement study to empirically validate the tags proposed for annotation. The results of the agreement study indicate that the distinction between the four basic categories (non-modal, epistemic, evidential and modal other than epistemic or evidential) is reasonably robust and can be replicated by different annotators to a large extent.

Further work will focus on the annotation of a large sample of bilingual texts from the different registers of the MULTINOT corpus, including not only comparable but also parallel (translated) texts. This will, hopefully, shed light on the genre-specific preferences in the use of certain modal meanings, and on the translation tendencies which characterise this conceptual domain. We are also planning to annotate other modal meanings (deontic, dynamic, volitional) in a contrastive manner as an extension of our current work in the domain of epistemicity.

## 5. References

- Boye, Kasper (2012). *Epistemic Meaning: A Crosslinguistic and Functional-Cognitive Study*. (Empirical Approaches to Language Typology 43). Berlin & Boston: De Gruyter Mouton.
- Brandt, Per Aage (2004). Evidentiality and enunciation. A cognitive and semiotic approach. In Juana I. Marín-Arrese (ed.) *Perspectives on evidentiality and modality*. Madrid: Editorial Complutense: 3-10.
- Byloo, Pieter, Richard Kastein and Jan Nuyts (2007). On *certainly* and *zeker*. In Mike Hannay & Gerard J. Steen (eds.), *Structural-functional Studies in English Grammar*. Amsterdam & Philadelphia: John Benjamins, 35-57.
- Carretero, Marta (2004). The role of evidentiality and epistemic modality in three English spoken texts from legal proceedings. In: Juana I. Marín-Arrese (ed.) *Perspectives on Evidentiality and Modality*. Madrid: Editorial Complutense. 25-62.
- Chojnicka, Joanna (2012). Reportive evidentiality and reported speech: is there a boundary? Evidence of the Latvian Oblique. In Usonienė, A., N. Nau, I. Dabašinskienė (eds.) *Multiple Perspectives in Linguistic Research on Baltic Languages*. Cambridge Scholars Publishing. 170-192.
- Cornillie, Bert (2007). Evidentiality and Epistemic Modality in Spanish (Semi) Auxiliaries. A Cognitive Functional Approach. Berlin: Mouton de Gruyter.
- Givón, Talmy (1984). *Syntax: A Functional-typological Introduction. Volume 1*. Amsterdam / Philadelphia: John Benjamins.
- Haan, Ferdinand de (2000). Evidentiality in Dutch. In Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Berkeley Linguistics Society*, February 12-15, 1999: General Session and Parasession on Loan Word Phenomena, 74-85. Berkeley Linguistics Society.
- Kärkkäinen, Elise (2003). *Epistemic Stance in English Conversation: A Description of its Interactional Functions, with a Focus on 'I Think'*. Amsterdam: John Benjamins.
- Lavid, Julia, Arús, Jorge, DeClerck, Bernard and Hoste, Veronique (2015). Creation of a high- quality, register-diversified parallel corpus for linguistic and computational investigations. In *Current Work in Corpus Linguistics: Working with Traditionally-conceived Corpora and Beyond. Selected Papers from the 7th International Conference on Corpus Linguistics (CILC2015)*. Procedia - Social and Behavioral Sciences, Volume 198, 24 July 2015, Pages 249–256.
- Lichtenberk, Frantisek (1995). Apprehensional epistemics. In Joan Bybee and Suzanne Fleischman (eds.). *Modality in Grammar and Discourse*. Amsterdam: Benjamins. 293-327.
- Marín-Arrese, Juana I. (2004). Evidential and epistemic qualifications in the discourse of fact and opinion. In: Juana I. Marín Arrese (ed.) *Perspectives on Evidentiality and Modality*. Madrid: Editorial Complutense. 153-184.
- Marín-Arrese, Juana I. (2013). Stancetaking and inter-subjectivity in the Iraq Inquiry. Blair vs. Brown. In Juana I. Marín-Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.), *English Modality Core, Periphery and Evidentiality*. Berlin & New York: De Gruyter Mouton. 411-445.
- Nissim, Malvina, Paola Pietrandrea, Andrea Sansò & Caterina Mauri (2013). Cross-linguistic annotation of modality: A data-driven hierarchical model. In Harry Bunt (ed.) *Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation (isa-9)*, March 19-20, 2013, Postdam, Germany, 7-14.
- Perkins, Michael R. (1983). *Modal Expressions in English*. London: Frances Pinter.
- Stubbs, Michael (1986). A Matter of prolonged field work: notes towards a modal grammar of English'. *Applied Linguistics* 7(1): 1-25.
- Van der Auwera, Johan & Vladimir Plungian (1998). Modality's semantic map. *Linguistic Typology* 2: 79-124.
- Wärnsby, Anna (2006). *(De)coding Modality. The Case of Must, May, Måste and Kan*. Lund Studies in English 113. Lund: Lund University.
- Wiemer, Björn & Katerina Stathi (2010). The database of evidential markers in European languages. A bird's eye view of the conception of the database (the template and the problems hidden beneath it). *STUF* 63(4). 275-289.
- Willett, Thomas L. (1988). A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language* 12.1, 51-97.

