Building frame-based corpus on the basis of ontological domain knowledge

He Tan
Institutionen för
datavetenskap
Linköpings universitet
Sweden
`he.tan@liu.se`

Rajaram Kaliyaperumal
Institutionen för
medicinsk teknik
Linköpings universitet
Sweden
`rajka625`

Nirupama Benis
Institutionen för
medicinsk teknik
Linköpings universitet
Sweden
`nirbe455@student.liu.se`

Abstract

Semantic Role Labeling (SRL) plays a key role in many NLP applications. The development of SRL systems for the biomedical domain is frustrated by the lack of large domain-specific corpora that are labeled with semantic roles. Corpus development has been very expensive and time-consuming. In this paper we propose a method for building frame-based corpus on the basis of domain knowledge provided by ontologies. We believe that ontologies, as a structured and semantic representation of domain knowledge, can instruct and ease the tasks in building the corpora. In the paper we present a corpus built by using the method. We compared it to BioFrameNet, and examined the gaps between the semantic classification of the target words in the domain-specific corpus and in FrameNet and Prop-Bank/VerbNet.

1 Introduction

The sentence-level semantic analysis of text is concerned with the characterization of events, such as determining "who" did "what" to "whom", "where", "when" and "how". It is believed to play a key role in NLP applications such as Information Extraction, Question Answering and Summarization. Semantic Role Labeling (SRL) is a process that, for each predicate in a sentence, indicates what semantic relations hold among the predicate and other sentence constituents that express the participants in the event (such as who and where). The relations are described by using a list of pre-defined possible semantic roles for that predicate (or class of predi-

cates). Recently, large corpora have been manually annotated with semantic roles in FrameNet (Fillmore et al., 2001) and PropBank (Palmer et al., 2005). With the advent of resources, SRL has become a well-defined task with a substantial body of work and comparative evaluation. Most of the work has been trained and evaluated on newswire text (see (Màrquez et al., 2008)).

Biomedical text considerably differs from the newswire text, both in the style of the written text and the predicates involved. Predicates in newswire text are typically verbs, biomedical text often prefers nominalizations, gerunds, and relational nouns (Kilicoglu et al., 2010). Predicates like *endocytosis*, *exocytosis* and *translocate*, though common in biomedical text, are absent from both the FrameNet and PropBank data (Bethard et al., 2008). Predicates like *block*, *generate* and *transform*, have been used in biomedical documents with different semantic senses and require different number of semantic roles compared to FrameNet (Tan, 2010) and PropBank (Wattarujeekrit et al., 2004). The development of SRL systems for the biomedical domain is frustrated by the lack of large domain-specific corpora that are labeled with semantic roles.

The projects, PASBio (Wattarujeekrit et al., 2004), BioProp (Tsai et al., 2006) and BioFrameNet (Dolbey et al., 2006), have made efforts on building PropBank-like and FrameNet-like corpora for processing biomedical text. Up until recently, these corpora are relatively small. Further, no general methodology exists to support domain-specific corpus construction. The difficulties include, how to discover and define

semantic frames together with associated semantic roles within the domain? how to collect and group domain-specific predicates to each semantic frame? and how to select example sentences from publication databases, such as the PubMed/MEDLINE database containing over 20 million articles? In this paper, we propose that building frame-based lexicon for the domain can be strongly instructed by domain knowledge provided by ontologies. We believe that ontologies, as a structured and semantic representation of domain-specific knowledge, can instruct and ease all the above tasks.

The paper proceeds as follows: first we explain our method how ontological domain knowledge instructs the main tasks in building a frame-based lexicon. This is followed by the related work. In section 4, we present a "study case" of the method. We built a frame *Protein Transport* containing text annotated with semantic roles. The construction is carried out completely under the supervision of the domain knowledge from the Gene Ontology (GO) (Ashburner et al., 2000). We evaluated it to the frame *Protein_transport* in the BioFrameNet and examined the gaps between the semantic classification of the target words in the domain-specific corpus and in FrameNet and PropBank/VerbNet. Finally, we conclude our work.

2 The Method

The FrameNet project is the application of the theory of *Frames Semantics* (Fillmore et al., 1985) in computational lexicography. Frame semantics begins with the assumption that in order to understand the meanings of the words in a language, we must first have knowledge of the background and motivation for their existence in the language and for their use in discourse. The knowledge is provided by the conceptual structures, or *semantic frames*. In FrameNet, a semantic frame describes an event, a situation or a object, together with the participants (called frame elements (FE)) involved in it. A word evokes the frame, when its sense is based on the frame. The relations between frames include *is-a*, *using* and *subframe*.

Ontology is a formal representation of knowledge of a domain of interest. It has concepts that represent sets or classes of entities within a domain. It defines different types of relations among concepts. Intuitively, ontological concepts and their relations can be used as the frame-semantic descriptions imposed on a lexicon.

A large number of ontologies have been developed in the domain of biomedicine. Many of them contain concepts that comprehensively describe a certain domain of interest, such as GO. GO biological process ontology, containing 20,368 concepts, provides the structured knowledge of biological processes that are recognized series of events or molecular functions. For example, the concept GO:0015031 protein transport defines the scenario, *"the directed movement of proteins into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore"*. It is a subclass of GO:0006810:transport and GO:0045184:establishment of protein localization. The class has 177 descendant classes in *is-a* hierarchies. A *Protein Transport* frame can be effectively described by using these classes and relations between them.

In many cases ontological terms can be seen as phrases that exhibit underlying compositional structures (McCray et al., 2002; Ogren et al., 2005). Figure 1 presents the compositional structures of 9 direct subclasses describing various types of protein transport. They provide that `translocation`, `import`, `recycling`, `secretion` and `transport` are the possible predicates, evoking the protein transport event. The more complex expressions, e.g. translocation of peptides or proteins into other organism involved in symbiotic interaction (GO:0051808), express participants involved in the event, i.e. the entity (`peptides or proteins`), destination (`into other organism`) and condition (`involved in symbiotic interaction`) of the event.

So far, we, using these classes and relations between them, have partly defined the semantic frame *Protein Transport*, decided the participants involved in the event, and listed the domain-specific words evoking the frame. The complete frame description can be given after studying all the related classes and their relations. Lastly, collecting example sentences will be based on knowledge based search engine for biomedical text, like GoPubMed (Doms and Schroeder, 2005). As such, domain knowledge
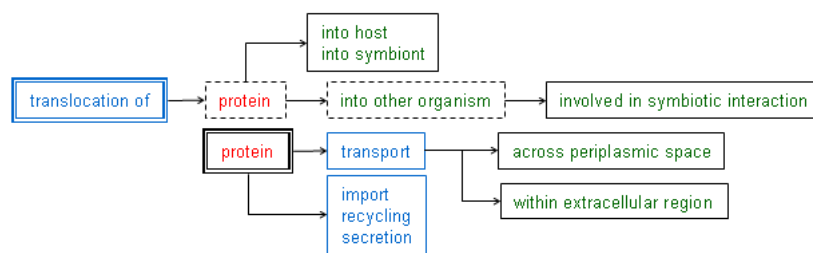
Figure 1: A concise view of 9 GO terms describing *Protein Transport*. We use the modified finite state automaton (FSA) representation given in (Ogren et al., 2005). Any path that begins at a start state, represented by double solid borders, and ends in an end state, represented by a single solid border, corresponds to a term. The nodes with a dashed border are neither start states nor end states.

provided by ontologies, such as GO biological process ontology and molecular function ontology, and pathway ontologies, can instruct us in building large frame-based corpora for the domain.

We outline the aspects of how ontologies instruct building a frame-based corpus:

1. The structure and semantics of domain knowledge in ontologies constrain the frame semantics analysis, i.e. decide the coverage of semantic frames and the relations between them;

2. Ontological terms can comprehensively describe the characteristics of events/scenarios in the domain, so domain-specific semantic roles can be determined based on terms;

3. Ontological terms provide a list of domain-specific predicates, so the semantic senses of the predicates in the domain are determined;

4. The collection and selection of example sentences can be based on knowledge-based search engine for biomedical text.

3 Related Work

The PropBank project is to add a semantic layer on the Penn Treebank (Marcus et al., 1994). For each unique verb sense, a set of semantic roles is defined at its accompanying syntactic realizations. The VerbNet project (Kipper et al., 2000) systematically creates English verb entries in a lexicon with syntactic and semantic information, referring to Levin verb classes. It made efforts to classify individual verbs in PropBank into VerbNet classes, based on patterns of usage (Kingsbury and Kipper, 2003).

The FrameNet project collects and analyzes the corpus (the British National Corpus) attestations of target words with semantic overlapping. The attestations are divided into semantic groups, noting especially the semantic roles of each target words, and then these small groups are combined into frames.

Ontologies have been put under the spotlight for providing the framework for semantic representation of textual information, and thus a basis for text mining systems (Spasic et al., 2005; Ashburner et al., 2008). Up to recently, TM systems mainly use ontologies as terminologies to recognize biomedical terms, by mapping terms occurring in text to concepts in ontologies, or use ontologies to guide and constrain analysis of NLP results, by populating ontologies. In the latter case, ontologies are more actively used as a structured and semantic representation of domain knowledge.

The FrameNet project links Semantic Types (ST) of FrameNet to the Suggested Upper Merged Ontology (SUMO) classes (Scheffczyk et al., 2006). The main function of ST is to indicate the basic typing of fillers of semantic roles, e.g. "Sentient" defined for the semantic role "Cognizer" in the frame "Cogitation". The goal of the work is to combine frame semantics in FrameNet and the formal world knowledge from SUMO, for improving FrameNet capability for deductive reasoning.

BioFrameNet is a domain-specific FrameNet extension. Its *intracellular protein transport* frames are mapped to the Hunter Lab [1] knowledge base (HLKB) protein transport classes. The frame elements are taken from HLKB slots. BioFrameNet

---

[1] Website for Hunters Bioinformatics research lab: http://compbio.uchsc.edu/.

76

considered a collection of Gene References in Function (GRIF) texts that are annotated by the HLKB protein transport classes in the knowledge base. Predicates are extracted from this collection of GRIF texts.

PASBio and BioProp are the projects that aim to produce definitions of Predicate Argument Structure (PAS) frames. They do not offer a direct linking of the predicates or their arguments to domain or general ontologies. PASBio used a model for a hypothetical signal transduction pathway of an idealized cell, to motivate verb choices. BioProp annotated the arguments of 30 frequent biomedical verbs found in the GENIA corpus (Kim et al., 2003).

4   Case Study: *Protein Transport* **Frame**

In this section we present the frame *Protein Transport*. The frame is built completely based on the domain knowledge provided by the piece of GO describing the event. The core structure of the frame is the same as that of FrameNet. The description of the scenario evoked by the frame is provided, along with a list of the frame elements and their definitions. A list of lexical units (LUs) that evoke the frame is provided. In addition, example sentences that contain at least one of the LUs, are given annotations using definitions of the frame. The annotations follow FrameNet's guidelines for lexicographic annotation, described in (Ruppenhofer et al., 2005).

4.1   The Frame

Resources. The description of the frame uses the scenario defined in GO:0015031 protein transport from the GO biological process ontology. It is a subclass of GO:0006810 transport and GO:0045184 establishment of protein localization. The class has 177 descendant classes. A total of 581 class names and synonyms are collected for the study. In addition to that from GO concepts, synonyms are also gathered by querying the UMLS Metathesaurus (Schuyler et al., 1992).

Frame. The definition (see Table 1) follows the definition of GO:0015031 protein transport.

Frame Elements. By studying all the names and synonyms (we call them "term" in the paper), we defined all possible FEs for the frame (see Table 2). The first 4 FEs are considered as core FEs. Ta-

"This frame deals with the cellular process in which a protein or protein-complex, the *Transport_Entity*, moves from the *Transport_Origin* to a different location, the *Transport_Destination*. Sometimes the *Transport_Origin* and *Transport_Destination* are not specified or are the same location. The *Transport_Entity* undergoes directed movement into, out of or within a cell or between cells or within a multicellular organism. This activity could be aided or impeded by other substances, organelles or processes and could influence other cellular processes."

Table 1: The frame definition.

ble 3 gives the number of the GO terms that indicate the FEs. For instance, in the term GO:003295 B cell receptor transport within lipid bilayer, `lipid bilayer` is the location `within` which protein transport happens. The term GO:0072322 protein transport across periplasmic space describes the path along which protein transport occurs. The term GO:0043953 protein transport by the Tat complex specifies a molecule that carries protein during the movement. GO:0030970 retrograde protein transport, ER to cytosol indicates the direction (`retrograde`) of the movement. An attribute (`SRP-independent`) of the event is described in the term GO:0006620 SRP-independent protein-membrane targeting ER.

Predicates. All lexical units in the frame are listed in Table 4. The first row gives the head of the GO terms (noun phrases). The number in the bracket indicates the number of GO terms with the head. If the verb derived from a head, can be used to describe the event that is expressed by the head, it is also included as a LU. GO terms, such as *related* and *broader* synonyms, may be not considered for collecting predicates. For example, fat body metabolism, a *broad* synonym of GO:0015032 storage protein import into fat body, is not considered.

Example Sentences. The example sentences are retrieved from the PubMed/MEDLINE database by using the GoPubMed (Doms and Schroeder, 2005), a knowledge-based search engine for biomedical text. The sentences to be annotated, are always the most relevant and from the latest publications. For

FEs	definition
Transport_Entity (TE)	Protein or protein complex which is undergoing the motion event into, out of or within a cell, or between cells, or within a multicellular organism.
Transport_Origin (TO)	The organelle, cell, tissue or gland from which the Transport_Entity moves to a different location.
Transport_Destination (TDS)	The organelle, cell, tissue or gland to which the Transport_Entity moves from a different location.
Transport_Condition (TC)	The event, substance, organelle or chemical environment which positively or negatively directly influences or is influenced by, the motion event. The substance organelle does not necessarily move with the Transport_Entity
Transport_Location (TL)	The organelle, cell, tissue or gland where the motion event takes place when the origin and the destination are the same or when origin or destination is not specified.
Transport_Path (TP)	The substance or organelle which helps the entity to move from the Transport_Origin to the Transport_Destination, sometimes by connecting the two locations, without itself undergoing translocation
Transport_Transporter (TT)	The substance, organelle or cell crucial to the motion event, that moves along with the Transport_Entity, taking it from the Transport_Origin to the Transport_Destination.
Transport_Direction (TDR)	The direction in which the motion event is taking place with respect to the Transport_Place, Transport_Origin, Transport_Destination or Transport_Location.
Transport_Attribute (TA)	This describes the motion event in more detail by giving information on how (particular movement, speed etc.) the motion event occurs. It could also give information on characteristic or typical features of the motion event.

Table 2: The frame elements

#T	578	50	159	95	41	27	6	2	1
FEs	TE	TO	TDS	TC	TL	TP	TT	TDR	TA

Table 3: The number of the GO terms that describe the frame elements

the head of GO terms	delivery (1), egress (2), establishment of ... localization (19), exit (2), export (20), import (88), recycling (2), release (1), secretion (226), sorting (4), targeting (68), trafficking (1), translocation (76), transport (100), uptake (5)
LUs	delivery.n, deliver.v, egress.n, establishment of ... localization.n, exit.n, exit.v, export.n, export.v, import.n, import.v, recycling.n recycle.v, release.n, release.v, secretion.n, secrete.v, sort.v, sorting.n, target.v, targeting.n, translocation.n, translocate.v, transport.v, transport.n, trafficking.n, uptake.n

Table 4: The lexical units

[L.pneumophila$_{Transport\_Origin}$|NP.Ext] [**translocate**$_{predicate}$] [more than 100 effector proteins$_{Transport\_Entity}$|NP.Obj] [into host cytoplasm$_{Transport\_Destination}$|PP[into].Dep] [using Dot/Icm T4BSS$_{Transport\_Path}$|VPing.Dep], [modulating host cellular functions$_{Transport\_Condition}$|VPing.Dep] to establish a replicative niche within host cells.
(PMID: 20949065)

Table 5: An example sentence: the three layers of annotations are given as FE|PT.GF.

LUs derived from one head, we acquired sentences by using the GO terms with the head. The query starts from using the most general GO terms. In the case that the number of query results is huge, more specific terms are used instead. Minimally, 10 sentences are gathered for each LU, if applicable. In cases when only specific GO terms are available and the number of query results is too small, we generalize the query term. For example, the lexical units, `release.n` and `release.v`, are derived and only derived from GO:0002001 renin secretion into blood stream's synonym renin release into blood stream. No query result returns for the GO term. The general term "protein release" is used as the query term instead.

Table 5 shows an example sentence for the frame. For each sentence annotated, we mark the target LU, and collect and record syntactic and semantic information about the relevant frame's FEs. For each FE, three types of annotation are gathered. The first layer is the identity of the specific FE. In cases when the FE is explicitly realized, phrase type (PT, for example NP) and grammatical function (GF) of the realization are annotated. The GFs describe the ways in which the constituents satisfy abstract grammatical requirements of the target word. In cases when the FE is omitted, the type of its null instantiation is recorded. These three layers for all of the annotated sentences, along with complete frame and FE descriptions are used in summarizing valence patterns for each annotated LU.

4.2 Evaluation

4.2.1 Compared to BioFrameNet

We compared this frame to the frame *Protein_transport* in BioFrameNet [2]. The frame involves the phenomenon of intracellular protein transport. BioFrameNet considered a collection of GRIF texts that describe various types of intracellular protein transport phenomena. The GRIFs texts are annotated by HLKB protein transport classes. All the 5 HLKB protein transport classes are arranged in *is-a* hierarchy. The description of the top level class *protein transport* is taken from the definition of GO:0015031 protein transport which is a su-

perclass of GO:0006886 intracellular protein transport in GO. For the frame, BioFrameNet provides definitions for 4 FEs, including *Transported_entity*, *Transport_origin*, *Transport_destination* and *Transport_locations*. The proposed FEs are taken from the slot definitions in the HLKB classes.

Table 6 illustrates the difference between the LUs in the 2 frames. The LUs that are not included in our corpus, can be classified into two groups. The first group include the LUs `enter.v`, `redistribution.n`, `return.v`, and `traffic.n`. They or their nominals are absent from GO biological process ontology terms. The second group includes those appear in GO, but in the terms that are not included in descendants of GO:0015031 protein transport.

The LUs, `endocytosis.n`, `internalization.n`, `recruitment.n`, do not appear in the descendants of GO:0015031 protein transport, but appear in GO terms that indeed describe protein transport event. `endocytosis` is the head of 9 GO terms, among which 2 concepts indeed describe an endocytotic process of protein (e.g.GO:0070086 ubiquitin-dependent endocytosis). 3 GO terms have `internalization` as the head. They all describe protein transport event (e.g. GO:0031623 receptor internalization). `recruitment.n` occurs in GO:0046799 recruitment of helicase-primase complex to DNA lesions and GO:0046799 recruitment of 3'-end processing factors to RNA polymerase II holoenzyme complex, which describe the movement of protein complex to another macro molecule.

The LUs, `efflux.n`, `entry.n`, `exocytosis.n`, `migrate.n`, `mobilization.n`, `move.v`, `movement.n`, `shuttle.n` and `shuttling.v`, appear in GO terms that are descendants of GO:0006810 transport. They are used to describe various kinds of transport events that protein is not involved in.

`shift.n` only occurs in GO:0003049 regulation of systemic arterial blood pressure by capillary fluid shift. `capillary fluid shift` describes a kind of transport event. `relocation.n` and `relocate.v` only appear in GO:0009902 chloroplast relocation which is considered as a kind of organelle organization.

Example Sentences. The number of example sentences for each lexical unit in BioFrameNet re-

LUs only in Bio-Frame-Net	efflux.n, endocytosis.n, enter.v, entry.n, exocytosis.n, internalization.n, migrate.v, mobilization.n, move.v, movement.n, recruitment.n, redistribution.n, relocate.v, relocation.n, return.v, shift.n, shuttle.v, shuttling.n, traffic.n
LUs in both corpus	delivery.n, exit.v, export.n, import.n, recycle.v, recycling.n, release.n, targeting.n, trafficking.n, translocate.v, translocation.n, transport.n, transport.v
LUs only in our corpus	deliver.v, egress.n, establishment of ... localization.n, exit.n, export.v, import.v, release.v, secretion.n, secrete.v, sort.v, sorting.n, target.v, uptake.n

Table 6: The comparison of LUs in the 2 frames

lies on the existing collection of GRIFs in HLKB. The number of annotated sentences for each LU ranges from 1 to over 200. 207 GRIFs use the LU `translocation.n`, and 10 GRIFs use `transport.v`.

In our corpus, minimally for each LU 10 annotated sentences are gathered, if applicable. Tables 7 and 8 show the realizations of the FEs for the LUs `translocation.n` and `translocate.v`. The second columns give the number of times that the FE is realized in the 10 sentences. The PT and GF layers and the number of times they occur are given in the last columns, in the format of PT GF (number of occurrences). There are differences between the valence patterns of two corpus. We notice that example sentences in BioFrameNet mainly describe about protein. Although protein transport is described, different topics may be covered in the sentences in our corpus.

4.2.2 Predicates in FrameNet and PropBank/VerbNet

We examined the gaps between the semantic classification of the LUs (or only verbs) in the frame, and in FrameNet and PropBank/VerbNet. Around half of the LUs from the frame are absent from FrameNet data. 5 LUs are used in describing protein transport event, with the same semantic sense as in FrameNet. We identified the FEs for *Protein Transport* frame based on the domain knowledge. The

FEs	#	Realizations
TE	10	PP[of] Dep (6); NP Dep (3); Poss Gen (1);
TO	1	PP[from] Dep (1);
TDS	7	A Dep (2); PP[into] Dep (2); PP[to] Dep (3);
TC	6	NP Ext (5); NP dep (1);
TL	2	PP[in] Dep (1); A Dep (1);
TP	1	PP[across] Dep (1);
TT	0	-
TD	0	-
TA	1	AJP Dep (1);

Table 7: FE realizations for annotations with `translocation.n`

FEs	#	Realizations
TE	10	PP[than].Dep (1); NP Ext (6); NP Obj (3);
TO	4	PP[from] Dep (2); PP[of] Dep (1); NP Ext (1);
TDS	9	PP[to] Dep (6); PP[into] Dep (3);
TC	6	NP Ext (1); PP[upon] Dep (2); PP[prior to] Dep (1); PP[during] Dep (1); VPing Dep (1); VPbrst Dep (1); VPfin Dep (1);
TL	0	-
TP	4	NP Ext(3); VPing Dep (1)
TT	0	-
TD	0	-
TA	2	PP[with] Dep (1); AVP Dep (1)

Table 8: FE realizations for annotations with `translocate.v`

LUs	FrameNet	SS
egress.n, establishment of ... localization, export.n, localization.n, localize.v, recycling.n, recycle.v, targeting.n, translocation.n, translocate.v, trafficking.n, uptake.n	-	-
delivery.n, deliver.v	Delivery	√
exit.v	Departing	√
export.v	Sending	√
	Exporting	
	Import_export	
import.n	Importance	
import.v	Importing	
	Import_export	
release.n, release.v	Releasing	
secrete.v	Emitting	√
sort.n	Type	
sort.v	Differentiation	
target.v	Aiming	
transport.n, transport.v	Bringing	√

Table 9: Predicates in FrameNet: If the predicate is used with the same semantic sense as in the FrameNet's frame, "semantic sense (SS)" is checked.

verbs	VerbNet	PropBank
translocate	-	-
deliver, transport	send-11.1	with the same semantic sense
secrete	-	
exit	escape-51.1	with different semantic sense
release	free-80.1	
sort	classify-29.10	
target	confront-98	
export, import, localize, recycle	-	

Table 10: Verbs in PropBank/VerbNet

logical process ontology [3]. We compared the frame *Protein Transport* to the frame *Protein_transport* in BioFrameNet, and examined the gaps between the semantic classification of the target words in the domain-specific corpus and in FrameNet and Prop-Bank/VerbNet.

In the future, we aim to extend the corpus to cover other biological events. GO ontologies will be the main resource to provide domain knowledge, but also other ontologies, such as pathway ontologies will be considered as important domain knowledge resources. The identification of frames and the relations between frames are needed to be investigated. In addition, we will study the definition of STs in the domain corpus and their mappings to classes in top domain ontologies, such as BioTop (Beißwanger et al., 2008).

References

Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan

number of FEs and their definitions are very different from FrameNet data. Other LUs are used with different semantic senses.

Except `translocate`, all verbs are included in PropBank data. Half of the verb senses have been classified into VerbNet classes. Only 3 verbs are used with the same sense as in describing protein transport event.

5 Conclusion

In this paper we propose a method for building frame-based corpus for the domain of biomedicine. The corpus construction relies on domain knowledge provided by ontologies. We believe that ontological domain knowledge can instruct us and ease the tasks in building the corpora. We built a corpus for transport event completely on basis of the piece of domain knowledge provided by GO bio-

---

[3]The corpus is publicly available on `http://www.ida.liu.se/~hetan/bio-onto-frame-corpus`

T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin and Gavin Sherlock. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25-29.

Michael Ashburner, Ulf Leser and Dietrich Rebholz-Schuhmann (Eds.). 2008. *Ontologies and Text Mining for Life Sciences: Current Status and Future Perspectives*. 24.03-28.03, Dagstuhl Seminar Proceedings.

Steven Bethard, Zhiyong Lu, James H Martin and Lawrence Hunter. 2008. Semantic Role Labeling for Protein Transport Predicates. *BMC Bioinformatics*, 9:277.

Elena Beißwanger, Stefan Schulz, Holger Stenzhorn and Udo Hahn. 2008. BioTop: An upper domain ontology for the life sciences: A description of its current structure, contents and interfaces to OBO ontologies. *Applied Ontology*, 3(4):205-212.

Adress Doms and Michael Schroeder. 2005. GoPubMed: exploring PubMed with the Gene Ontology. *Nucleic Acids Research*, 33:W783-786.

Andrew Dolbey, Michael Ellsworth and Jan Scheffczyk. 2006. BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies. *The proceedings of KR-MED*, 87-94.

Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2).

Charles J. Fillmore, Charles Wooters and Collin F. Baker. 2001. Building a Large Lexical Databank Which Provides Deep Semantics. *The Pacific Asian Conference on Language, Information and Computation*.

Halil Kilicoglu, Marcelo Fiszman, Graciela Rosemblat, Sean Marimpietri and Thomas C. Rindflesch. 2010. Arguments of nominals in semantic interpretation of biomedical text. *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing (BioNLP'10)*.

Jin-Dong Kim, Tomoko Ohta, Yuka Teteisi and Jun'ichi Tsujii. 2003. GENIA corpus - a semantically annotated corpus for bio-textmining. *Bioinformatics*. 19(suppl. 1):180-182.

Paul Kingsbury and Karin Kipper. Deriving Verb-Meaning Clusters from Syntactic Structure. *Workshop on Text Meaning, held in conjunction with HLT/NAACL 2003*.

Karin Kipper, Hoa Trang Dang and Martha Palmer. 2000. Class-Based Construction of a Verb Lexicon. *AAAI-2000 Seventeenth National Conference on Artificial Intelligence*.

Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz and Britta Schasberger. 1994. The Penn Treebank: annotating predicate argument structure. *In Proceedings of the workshop on Human Language Technology (HLT '94)*.

Alexa T. McCray, Allen C. Browne and Olivier Bodenreider. 2002. The Lexical Properties of the Gene Ontology (GO). *Proceedings of AMIA Symposium* , 504-508.

Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski and Suzanne Stevenson. 2008. Semantic role labeling: an introduction to the special issue. *Computational Linguistics*, 34(2).

Philip V. Ogren, K. Bretonnel Cohen and Lawrence Hunter. 2005. Implications of compositionality in the gene ontology for its curation and usage. *Pacific Symposium on Biocomputing*, 10:174-185.

Martha Palmer, Daniel Gildea and Paul Kingsbury. 2005. The Proposition Bank: an annotated corpus of semantic roles. *Computational Linguistics*, 31:71-105.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson and Jan Scheffczyk. 2005. ICSI. *FrameNet II*: Extended Theory and Practice.

Irena Spasic, Sophia Ananiadou, John McNaught and Anand Kumar. 2005. Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in Bioinformatics*, 6(3):239-251.

Jan Scheffczyk, Adam Pease and Michael Ellsworth. 2006. Linking FrameNet to the SUMO Ontology. *International Conference on Formal Ontology in Information Systems (FOIS 2006)*.

Peri L. Schuyler, William T. Hole, Mark S. Tuttle and David D. Sherertz. 1992. The UMLS Metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association*, 81(2):217-22.

He Tan. 2010. A study on the relation between linguistics-oriented and domain-specific semantics. *Proceedings of the 3rd International Workshop on Semantic Web Applications and Tools for the Life Sciences*.

Richard Tzong-Han Tsai, Wen-Chi Chou, Ying-Shan Su, Yu-Chun Lin, Cheng-Lung Sung, Hong-Jie Dai, Irene Tzu-Hsuan Yeh, Wei Ku, Ting-Yi Sung and Wen-Lian Hsu. 2006. BIOSMILE: adapting semantic role labeling for biomedical verbs: an exponential model coupled with automatically generated template features. *Proceedings of the 2005 Workshop on Biomedical Natural Language Processing (BioNLP'06)*.

Tuangthong Wattarujeekrit, Parantu K Shah and Nigel Collier. 2004. PASBio: predicate-argument structures for event extraction in molecular biology. *BMC Bioinformatics*, 5:155.