Facilitating the Analysis of Discourse Phenomena in an Interoperable NLP Platform

Riza Theresa Batista-Navarro*, Georgios Kontonatsios*, Claudiu Mihăilă*, Paul Thompson*, Rafal Rak, Raheel Nawaz, Ioannis Korkontzelos, and Sophia Ananiadou

The National Centre for Text Mining,
The University of Manchester,
131 Princess Street, Manchester M1 7DN, UK
{rbatistanavarro,gkontonatsios,cmihaila,pthompson,
rrak,rnawaz,ikorkontzelos,sananiadou}@cs.man.ac.uk

Abstract. The analysis of discourse phenomena is essential in many natural language processing (NLP) applications. The growing diversity of available corpora and NLP tools brings a multitude of representation formats. In order to alleviate the problem of incompatible formats when constructing complex text mining pipelines, the Unstructured Information Management Architecture (UIMA) provides a standard means of communication between tools and resources. U-Compare, a text mining workflow construction platform based on UIMA, further enhances interoperability through a shared system of data types, allowing free combination of compliant components into workflows. Although U-Compare and its type system already support syntactic and semantic analyses, support for the analysis of discourse phenomena was previously lacking. In response, we have extended the U-Compare type system with new discourse-level types. We illustrate processing and visualisation of discourse information in U-Compare by providing several new deserialisation components for corpora containing discourse annotations. The new U-Compare is downloadable from `http://nactem.ac.uk/ucompare`.

Keywords: UIMA, interoperabilty, U-Compare, discourse, causality, coreference, meta-knowledge.

1 Introduction

One of the most important outcomes of recent research into biomedical text mining is the large number of newly created, manually annotated corpora. Such corpora were originally designed for training and evaluating systems that perform specific tasks. However, such resources can additionally provide support for other tasks. Data reuse is both highly demanded and occurs frequently, as it saves important amounts of human effort, time and money. For instance, the

* The authors have contributed equally to the development of this work and production of the manuscript.

A. Gelbukh (Ed.): CICLing 2013, Part I, LNCS 7816, pp. 559–571, 2013.

GENIA corpus [1], which initially contained only named entity annotations, has subsequently been extended by different researchers and groups to include event annotations and meta-knowledge information [2].

Corpora containing more complex types of semantic information, including discourse phenomena, have begun to appear recently. *Discourse* is defined as a coherent sequence of textual zones (e.g., clauses and sentences). Different zones contribute different information to the discourse, and are connected logically by *discourse relations*, which are characterised by labels such as “causal”, “temporal” and “conditional”. In turn, these allow more complex knowledge about the facts mentioned in the discourse to be inferred. Discourse relations can be either explicit or implicit, depending on whether or not they are represented in text using overt *discourse connectives* (or *triggers*). Discourse annotations cover several types of information that attempt to explain how the structure of text leads to a coherent discourse. They include:

– classification of previously mentioned textual zones and structured events (e.g., biological processes) according to their general interpretation (e.g., background knowledge, hypotheses, observations, etc.) in the discourse
– determining which textual zones are linked together in terms of discourse relations (e.g., causality) and how these links should be characterised
– resolving coreference within or across sentences.

Understanding discourse-level information is important for several tasks, such as automatic summarisation and question answering [3, 4]. Due to the complex nature of discourse phenomena, their annotation is extremely time-consuming and hence there is an urgent need to make these annotated corpora available to the community in a format that is readily usable.

As mentioned above, annotated corpora constitute an important resource for the development of NLP tools, e.g., tokenisation, part-of-speech tagging and syntactic parsing. Individually, these tools do not usually address a complete NLP task, but when combined together into workflows, they become much more powerful. For example, each of the above-mentioned tool types is a prerequisite for the next: part-of-speech tagging must be preceded by tokenisation, whilst tokens with parts-of-speech are needed as input to syntactic parsers. Discourse analysis, a more complex task, requires a range of linguistic pre-processing of input texts, and hence a pipeline of tools must be run prior to executing a discourse analysis tool.

In order to allow such multi-step pipelines to be built with the minimum effort, it is desirable to combine the different NLP tools in a straightforward manner. Unfortunately, for many existing tools, such combination is not trivial due to the fact that tools are implemented in different programming languages, and have differing input and output formats. Evaluation of tools against gold standard corpora may also be problematic if the annotation format is different from the output of the tool. Thus, creating and evaluating text processing pipelines can require a considerable amount of effort from developers, in order to allow different resources (i.e., tools and corpora) to “communicate”. The lack of ease with which resources can be reused can be a major obstacle to the rapid creation of complex NLP applications.

The Unstructured Information Management Architecture (UIMA) [5] is a framework that offers a solution to the problem of interoperability of NLP resources by providing a standardised mechanism by which resources communicate. However, UIMA does not impose any recommended set of types, and developers are left to define their own types.

U-Compare is a UIMA-based, text mining workflow construction platform that enables interoperability of NLP components by defining a *sharable* type system. This covers a wide range of annotation types that are common to many NLP components. The type system was designed such that there are several hierarchical levels which can be extended according to the individual requirements of different components, yet it is possible for components to be compatible at higher levels of the hierarchy. A further feature of U-Compare is its graphical user interface, via which users can rapidly create NLP pipelines using a simple drag-and-drop mechanism, with no requirement for programming skills. A key functionality of U-Compare is its evaluation and comparison mechanism that allows users to identify optimal NLP pipelines [6]. Users can also visualise agreements and disagreements between different pipelines directly in the text.

U-Compare contains the world's largest repository of type-system compatible modules [7]. Users can freely combine these components into pipelines that form syntactic (e.g., part-of-speech tagging, dependency parsing) or semantic (e.g., named entity recognition, event extraction) aggregate applications. However, despite this large library of components, U-Compare did not previously support the analysis of discourse phenomena. In this paper, we describe our work to extend U-Compare in order to facilitate the construction of NLP pipelines that annotate discourse phenomena in free text. The contributions of this paper can be summarised as follows: (a) expansion of the U-Compare type system to model discourse phenomena, (b) facilitation of the visualisation of discourse annotations, and (c) implementation of data deserialisers (readers) for various corpora containing discourse phenomena annotation, which can be used within UIMA processing pipelines and within U-Compare in particular. As a consequence, U-Compare provides a convenient environment in which to test and evaluate discourse analysis components. Given the wide range of available lower-level processing tools available in U-Compare, most tools that produce the types of information required as input to discourse components are present in the U-Compare library. Annotations produced by the new discourse components are visualised automatically, and different pipelines can be easily evaluated against various discourse phenomena corpora.

2 Discourse Phemomena and Corpora

In this section, we discuss in greater detail the three specific types of discourse phenomena that have been the focus of our extensions to U-Compare, i.e., causality, coreference and characterisation/interpretation of discourse segments. We subsequently explain the characteristics of the annotated corpora that have been made interoperable with other U-Compare components.

2.1 Causality in the Biomedical Domain

Statements regarding causal associations have long been studied in general language, mostly as part of more complex tasks. Comparatively little work has been carried out in the biomedical domain, although causal associations between biological entities, events and processes are central to most claims of interest [8]. Despite this, a unified theory of causality has not yet emerged, be it in general or specialised language. Many tasks, such as question answering and automatic summarisation, require the extraction of information that spans across several sentences, together with the recognition of relations that exist across sentence boundaries, in order to achieve high levels of performance. Take, for instance, the case in example 1, where the trigger *Therefore* signals a justification between the two sentences: because "a normal response [...] glutamic acid residues", the authors believe that the "regulation of PmrB [...] some amino acids".

(1) In the case of PmrB, a normal response to mild acid pH requires not only a periplasmic histidine but also several glutamic acid residues.
Therefore, regulation of PmrB activity may involve protonation of one or more of these amino acids.

Not all causality relations are as obvious as the previous one, where the trigger is explicit and is typically used to denote causality. In example 2, there is an implicit discourse causal association between the first half of the sentence, "This medium lacked [...] 10 mM MgCl2", and the latter half, "which represses [...] PmrA-activated genes". This is due to the fact that, generally, bacterial gene expression could be affected by specific properties of growth media, such as pH and concentration of metals.

(2) This medium lacked Fe3+ or Al3+, the only known PmrB ligands (Wosten et al., 2000), and contained 10 mM MgCl2, which represses expression of PmrA-activated genes (Soncini and Groisman, 1996; Kox et al., 2000).

Amongst the large number of corpora that have been developed for biomedical text mining, several include the annotation of statements regarding causal associations, such as GENIA [1] and GREC [9]. However, these corpora do not include an exhaustive coverage of causal statements and the granularity of the annotation of such statements is rather limited. Other corpora, e.g., BioCause [10] and BioDRB [11], contain annotations of causality and other discourse relations in biomedical journal articles, and here we focus on these.

2.2 Coreference

Coreference is a phenomenon involving linguistic expressions referring to a unique referent [12]. It is often associated with the phenomenon of *anaphora*, which is characterised by an expression (*anaphor*) whose interpretation depends on an entity previously mentioned in the same discourse (*antecedent*). Coreference and anaphora are extensively used in both spoken and written discourse, and are central to discourse theories such as the Centering Theory [13], a theory of local discourse coherence.

The task of grouping all coreferring mentions in text into respective coreference chains is known as *coreference resolution. Anaphora resolution*, on the other hand, is the process of determining the antecedent of an anaphor. Whilst the output of anaphora resolution is a set of anaphor-antecedent pairs, that of coreference resolution is a set of coreference chains which can be treated as equivalence classes. Despite this difference, an overlap between them may be observed in several cases. Often, an anaphor and its antecedent are coreferential (i.e., they have the same referent) and may be placed in the same coreference chain, as in the following example:

(3) John gave his wife a necklace for her birthday. She thanked him for it.

Example 3 contains the following coreference chains: {*John, his, him*}, {*his wife, her, She*} and {*necklace, it*}. The anaphor and antecedent in each of the anaphoric pairs {*his, John*}, {*her, his wife*}, {*She, his wife*}, {*him, John*} and {*it, necklace*} fall within the same chain. In some scenarios, however, they do not fall within the same chain, such as in the following example:

(4) Peter received his paycheque yesterday but John didn't get one.

Example 4 includes the anaphoric pairs {*his, Peter*} and {*one, his paycheque*}. Whilst *his* and *Peter* refer to the same person and thus belong to the same coreference chain, *one* and *his paycheque* do not corefer since the paycheque that John is expecting is a different entity from Peter's paycheque. The coreference chains in this example, therefore, are: {*Peter, his*}, {*his paycheque* (Peter's)}, {*yesterday*}, {*John*} and {*one* (John's paycheque)}.

The phenomenon of coreference has diverse representation formats: in some corpora, coreferring mentions are represented as coreference chains; in most of them, however, they are linked in the same way as an anaphor and its antecedent. In the 2005 Automatic Content Extraction (ACE'05) corpus [14], coreferring mentions pertaining to the same entity/relation/event are grouped together, forming a coreference chain. In contrast, in GENIA MedCo [15], HANAPIN [16] and the BioNLP Shared Task [17] corpora, a pair of mentions referring to the same entity are linked like an anaphor and its antecedent are.

Both the coreference and anaphora resolution tasks are non-trivial and are still considered unsolved, especially in scientific text. Providing NLP researchers with access to tools and resources for these tasks is hence desirable for the development and improvement of coreference and anaphora resolution systems.

2.3 Meta-knowledge

There have been several efforts to annotate and automatically recognise aspects of the characterisation or interpretation of textual discourse segments.

Collectively, these different aspects have been defined as *meta-knowledge* [2]. Recognition of meta-knowledge is important for tasks such as isolating new knowledge claims in research papers [18], maintaining models of biomedical processes [19] or curating biomedical databases [20]. The following sentences expressing similar information illustrate different aspects of meta-knowledge:

(5) It is *known* that leukotriene B4 (LTB4) affects the expression of the proto-oncogenes c-jun and c-fos.
(6) We *examined* whether leukotriene B4 (LTB4) affects the expression of the proto-oncogenes c-jun and c-fos.
(7) *Previous studies* have shown that leukotriene B4 (LTB4) does *not* affect the expression of the proto-oncogenes c-jun and c-fos.
(8) These results *suggest* that leukotriene B4 (LTB4) *partially* affects the expression of the proto-oncogenes c-jun and c-fos.

The italicised words and expressions highlight how the interpretations of these sentences differ in both subtle and more significant ways. In sentence (5), the word *known* indicates that generally accepted background knowledge is being expressed, whilst in sentence (6), the word *examined* denotes the very different interpretation that the truth value of the information described is under investigation. In sentence (7), a finding coming from a previous study is reported (*Previous studies*), and the presence of the word *not* indicates that this is a negated finding. In sentence (8), an analysis of results is presented, as denoted by *suggests*, whilst the word *partially* indicates that the reaction expressed by the sentence occurs at lower intensity than would be expected by default.

Most previous work on meta-knowledge annotation and recognition has been concerned with annotation of scientific texts at the sentence level with a single aspect of meta-knowledge, usually negation/speculation detection [21] or classification according to rhetorical function or general information content, using categories such as background knowledge, methods, hypotheses, experimental observations, etc. Different schemes and classifiers (both rule and machine learning-based) have been created for both abstracts [22] and full papers [23].

A small number of annotation schemes have considered more than one aspect or dimension of meta-knowledge. For example, the CoreSC annotation scheme [24] augments general information content categories with additional attributes, such as *New* and *Old*, to denote current or previous work. In order to account for the fact that longer sentences may contain several types of information, another corpus [25] annotated sentence segments with five different aspects of meta-knowledge. A related effort [2] also annotates five meta-knowledge dimensions, i.e., knowledge type, certainty level, manner, polarity and knowledge source, but at the level of *events* rather than segments. Events are structured representations of pieces of biomedical knowledge, formed of multiple discontinuous spans of text. They usually consist of an *event trigger*, i.e., a word or phrase that characterises the event, together with its arguments, e.g., theme, agent, cause, location, etc. Like segments, there may be multiple events in a single sentence. Annotation of meta-knowledge at the event level can be used to train event extraction systems that allow constraints to be specified not only in terms of event types/participants, but also in terms of meta-knowledge interpretations of individual events [26].

3 Related Work

Text mining platforms alleviate the requirement for programming and technical skills and thus minimise the effort required to conduct *in silico* experiments. Text mining frameworks that support the construction of workflows define a common, standardised and well-defined architecture for developing pipelined applications and are becoming highly popular and widely applicable. Of the numerous text mining workflow construction platforms currently available [27–30], we focus on those that enable interoperability of resources and allow the representation of discourse phenomena.

GATE [28] is a workflow construction framework that has been used for the development of various text mining applications (e.g., automatic summarisation, cross-lingual mining and ontology building). It supports the representation and processing of the discourse phenomenon of coreference. However, GATE mainly focusses on assisting in specific programming tasks (i.e., large collection of software libraries) rather than interoperability of text mining resources. It also lacks a common, hierarchical data type system. In contrast, U-Compare comes with a library of heterogenous components (i.e., native components written in various languages or web services) which can communicate with each other under one common type system.

Heart of Gold [29] is an XML-based architecture primarily aimed at the development of text analysis pipelines. It focusses on the interoperability of components, allowing tools developed using different programming languages to be combined. However, it does not support the processing nor representation of discourse phenomena.

Specifically focussed on the analysis of medical records, cTAKES [27] was developed as another UIMA-based workflow construction platform. As in the case of U-Compare, cTAKES defines UIMA types, but it supports the representation and analysis of only two discourse phenomena, namely coreference and negation.

Argo [30] is a web-based, UIMA-compliant workflow construction platform that supports the U-Compare type system. The two platforms are therefore fully compatible (i.e., components and workflows are interchangable between U-Compare and Argo). Currently, Argo does not support the analysis of discourse phenomena. However, due to the compatibility of the two platforms, the type system extensions described here are also applicable to Argo.

4 Discourse Phenomena in U-Compare

In this section, we discuss how the U-Compare type system has been extended to allow the representation of the discourse phenomena that we are considering, namely, causality, coreference and meta-knowledge.

4.1 Type System Extensions

The U-Compare type system can be divided into three hierarchies, namely syntactic (e.g., part-of-speech tags), document (e.g., title, author) and semantic (e.g., named entities) levels. We have introduced a new sub-hierarchy of semantic annotations that allows us to model discourse phenomena. The extended U-Compare type system is depicted in Figure 1. As can be observed from the figure, the newly created *Discourse phenomena hierarchy* has been added within the *Semantic hierarchy*. In this new hierarchy, `DiscoursePhenomenon` is the top level discourse data type, from which various annotations descend. These include three classes of annotations: one for events (`EventAnnotation`), one for coreference (`CoreferenceAnnotation`) and one for discourse relations (`DiscourseAnnotation`).

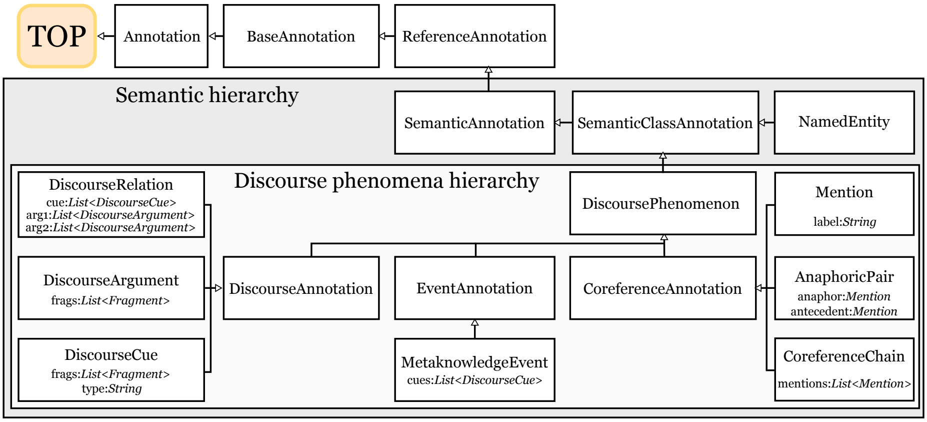

Fig. 1. The extended U-Compare type system

A discourse annotation is modelled using two text-bound data types, whilst a third type links the previous two. `DiscourseCue` marks discourse connectives that can either cover spans of text, in the case of explicit triggers, or have a length of zero, in the case of implicit triggers. In the latter case, the trigger that best fits the situation (if provided by annotators) is displayed in the annotation panel of U-Compare. `DiscourseArgument` represents a span of text that plays a role in the discourse relation. Both types are modelled using `Fragments` in order to allow the representation of discontinous annotations. `DiscourseRelation` combines the previous types to create a discourse relation. It contains three lists, one for discourse connectives and two for each of the arguments of the relation.

Using this abstract hierarchy of data types, it becomes possible to directly model a variety of discourse relations, such as causality, temporality or conditionality. Moreover, relations from different resources become interoperable and can be seamlessly combined and processed by existing text analysis components in U-Compare. Figure 2 illustrates a causal relation from the BioCause corpus displayed in the U-Compare GUI. As can be noticed, the two arguments and the trigger are underlined. Moreover, there are two arrows connecting the trigger with each of the arguments. The arrows also display the role of each argument in the relation.

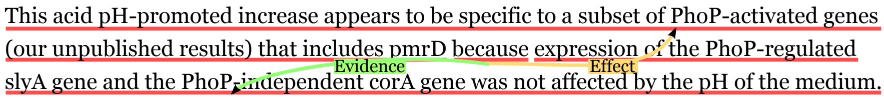

Fig. 2. BioCause Causality

In the case of coreference, we have introduced three data types under `CoreferenceAnnotation`. `Mention` represents any expression annotated in text, whilst `AnaphoricPair` consists of two `Mention` objects corresponding to an anaphor and its antecedent. In Figure 3, for example, sample coreference annotations in the HANAPIN corpus [16] are represented in the GUI as three `AnaphoricPair` objects. The last data type is `CoreferenceChain`, which consists of a list of `Mention`s referring to the same entity.

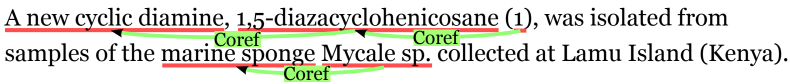

Fig. 3. HANAPIN Coreference

`CoreferenceChain` was used, for instance, in representing the grouping of the underlined mentions in Figure 4, which all belong to the same coreference chain in the ACE'05 corpus. The newly added data types allow for the representation of different types of coreference annotations found in various corpora, enabling U-Compare to interpret them regardless of whether they come in the form of coreference chains or anaphoric pairs.

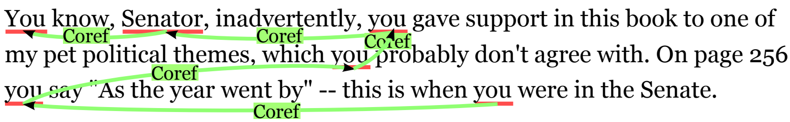

Fig. 4. ACE'05 Coreference

The extensions to the U-Compare type system for meta-knowledge are geared towards its representation at the event level. This information is encoded by extending the existing `Event` annotation type with a new type `MetaKnowledgeEvent`, which includes attributes to store the values of all five meta-knowledge dimensions. In addition, the `DiscourseCue` type (previously introduced in the context of causality) is used to annotate different types of meta-knowledge (e.g., negation or certainty level cues). There is a *type* attribute, which allows the dimension that the cue represents (e.g., certainty level) to be encoded.

The representation of event-based meta-knowledge annotation in U-Compare is illustrated in Figure 5. Two event triggers are underlined, i.e., "activation" and "increases". Each of the two events has different meta-knowledge information associated with it. The word "partial" shows the intensity (manner cue) of the "activation" event. For the "increases" event, the word "suggests" gives information about both the knowledge type (KT) and certainty level (CL) of the event: it shows that the event is the result of an analysis based on the previous "activation" event and that there is a lack of complete certainty about the event on the part of the author.

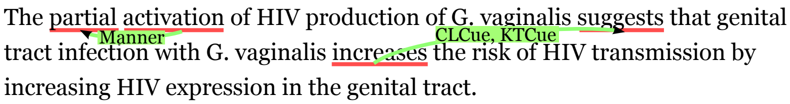

Fig. 5. Meta-knowledge

4.2 From Resource Readers to Semantic Labellers

We have made a number of tools available through U-Compare that automatically read various discourse annotated corpora, create UIMA annotations that include the newly introduced types described above, and visualise these annotations in U-Compare. The new corpus readers are capable of handling several corpora annotated with the discourse phenomena previously mentioned. These include the GENIA Event Corpus [2], BioCause [10], BioDRB [11], ACE'05 (Coreference) [14], BioNLP Shared Tasks [17], GENIA MedCo [15] and HANAPIN [16]. The above resources are interoperable with U-Compare's existing processing components. They can be combined into pipelines using a drag-and-drop graphical user interface, without any requirement for programming skills. The only prerequisite is that the input types of a component are compliant with the output types of preceding components.

We present a use case in which heterogeneous components have been combined with the aim of determining the semantic labels of otherwise semantically empty expressions (e.g., pronouns) in the GENIA MedCo corpus.

The created pipeline consists of the following components:

1. Collection Reader: GENIA MedCo Coreference Reader
2. Syntactic components: OSCAR Tokeniser [6]
3. Semantic components: ABNER NLPBA [31] and OSCAR Maximum Entropy Markov Model [6] Named Entity Recognisers (NERs)

Figure 6 illustrates an example document from the corpus annotated by this pipeline. With the incorporation of NERs into the pipeline, "NAC" and "DC", for example, have been identified as chemical molecules (CM), allowing users to infer that the pronouns *its* and *their* also refer to chemical molecules. If it is desired that coreferential expressions inherit the semantic types predicted by the NERs, our new Coreference Annotation Merger component can be optionally included in the pipeline to propagate labels to the coreferential expressions. Such propagation of semantic labels will allow users to subsequently extract relations or events involving coreferential expressions [32].

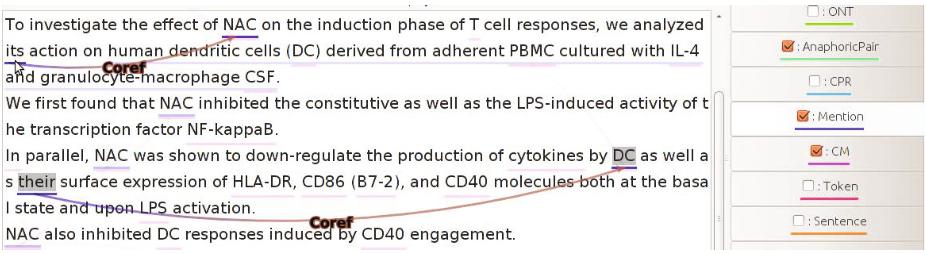

Fig. 6. Sample document with results of our semantic labelling pipeline

5 Conclusions and Future Work

In this paper, we have presented extensions that were implemented on top of the existing workflow construction platform U-Compare to allow the processing and visualisation of discourse phenomena. We have introduced a new hierarchy of annotation types, now integrated into the type system of U-Compare, that models various discourse phenomena, including causality, coreference and meta-knowledge. Furthermore, we have implemented readers for several annotated corpora that can read and map the annotations to U-Compare's type system and display complex relations (e.g., coreference chains and causal relations) for easy interpretation. Our goal is to enhance the interoperability of such resources so that they can be seamlessly incorporated into text mining pipelines that process and analyse discourse phenomena. Moreover, building on the interoperability of U-Compare, we have shown in practice how discourse annotated corpora can become compatible with U-Compare's processing tools. As a case study, we have designed a pipeline that automatically assigns semantic labels to otherwise semantically empty expressions (e.g., pronouns) in the GENIA MedCo corpus.

It should be noted that automatically produced annotations cannot be included directly in gold standard annotated corpora. As post-processing work, curators need to manually remove erroneous tags. As future work, we plan to integrate tools that aim to reduce the human effort in producing or enriching gold-standard annotated corpora into U-Compare. Such tools [33] are incrementally trained on annotations produced by curators in order to correctly predict unseen annotations in text.

Acknowledgement. The work described in this paper has been funded by the MetaNet4U project (ICT PSP Programme, Grant Agreement: No 270893) and the JISC-funded "Integrated Social History Environment for Research (ISHER)-Digging into Social Unrest" project, which is part of the "Digging into Data Challenge."

References

1. Kim, J.D., Ohta, T., Tsujii, J.: Corpus annotation for mining biomedical events from literature. BMC Bioinformatics 9, 10 (2008)
2. Thompson, P., Nawaz, R., McNaught, J., Ananiadou, S.: Enriching a biomedical event corpus with meta-knowledge annotation. BMC Bioinformatics 12, 393 (2011)
3. Marcu, D.: The Theory and Practice of Discourse Parsing and Summarization. MIT Press, Cambridge (2000)
4. Sun, M., Chai, J.Y.: Discourse processing for context question answering based on linguistic knowledge. Knowledge-Based Systems 20, 511–526 (2007)
5. Ferrucci, D., Lally, A.: UIMA: an architectural approach to unstructured information processing in the corporate research environment. Natural Language Engineering 10, 327–348 (2004)
6. Kolluru, B., Hawizy, L., Murray-Rust, P., Tsujii, J., Ananiadou, S.: Using Workflows to Explore and Optimise Named Entity Recognition for Chemistry. PLoS ONE 6, e20181 (2011)
7. Kano, Y., Baumgartner Jr., W.A., McCrochon, L., Ananiadou, S., Cohen, K.B., Hunter, L., Tsujii, J.: U-Compare: share and compare text mining tools with UIMA. Bioinfomatics 25, 1997–1998 (2009)
8. Kleinberg, S., Hripcsak, G.: A review of causal inference for biomedical informatics. Journal of Biomedical Informatics 44, 1102–1112 (2011)
9. Thompson, P., Iqbal, S., McNaught, J., Ananiadou, S.: Construction of an annotated corpus to support biomedical information extraction. BMC Bioinformatics 10, 349 (2009)
10. Mihăilă, C., Ohta, T., Pyysalo, S., Ananiadou, S.: BioCause: Annotating and analysing causality in the biomedical domain. BMC Bioinformatics 14, 2 (2013)
11. Prasad, R., McRoy, S., Frid, N., Joshi, A., Yu, H.: The Biomedical Discourse Relation Bank. BMC Bioinformatics 12, 188 (2011)
12. Jurafsky, D., Martin, J.H.: Speech and Language Processing, 2nd edn. Prentice Hall Series in Artificial Intelligence. Prentice Hall (2008)
13. Grosz, B.J., Weinstein, S., Joshi, A.K.: Centering: A Framework for Modeling the Local Coherence of Discourse. Comp. Ling. 21, 203–225 (1995)
14. Walker, C.: ACE 2005 Multilingual Training Corpus (2006)
15. Su, J., Yang, X., Hong, H., Tateisi, Y., Tsujii, J.: Coreference Resolution in Biomedical Texts: a Machine Learning Approach. In: Ashburner, M., Leser, U., Rebholz-Schuhmann, D. (eds.) Ontologies and Text Mining for Life Sciences: Current Status and Future Perspectives. Dagstuhl Seminar Proceedings, vol. 08131 (2008)
16. Batista-Navarro, R.T.B., Ananiadou, S.: Building a coreference-annotated corpus from the domain of biochemistry. In: Proceedings of BioNLP 2011, pp. 83–91 (2011)
17. Stenetorp, P., Topić, G., Pyysalo, S., Ohta, T., Kim, J.D., Tsujii, J.: BioNLP Shared Task 2011: Supporting Resources. In: Proceedings of the BioNLP Shared Task 2011 Workshop, pp. 112–120. ACL (2011)
18. Sandor, A., de Waard, A.: Identifying Claimed Knowledge Updates in Biomedical Research Articles. In: Proceedings of the Workshop on Detecting Structure in Scholarly Discourse (DSSD), pp. 7–10 (2012)
19. Oda, K., Kim, J.D., Ohta, T., Okanohara, D., Matsuzaki, T., Tateisi, Y., Tsujii, J.: New challenges for text mining: mapping between text and manually curated pathways. BMC Bioinformatics 9, S5 (2008)
20. Yeh, A., Hirschman, L., Morgan, A.: Evaluation of text data mining for database curation: lessons learned from the KDD Challenge Cup. Bioinformatics 19, 331–339 (2003)

21. Medlock, B., Briscoe, T.: Weakly supervised learning for hedge classification in scientific literature. In: Proceedings of ACL, pp. 992–999 (2007)
22. McKnight, L., Srinivasan, P.: Categorization of sentence types in medical abstracts. In: Proceedings of the AMIA Annual Symposium, pp. 440–444 (2003)
23. Mizuta, Y., Korhonen, A., Mullen, T., Collier, N.: Zone analysis in biology articles as a basis for information extraction. Int. J. Med. Inf. 75, 468–487 (2006)
24. Liakata, M., Teufel, S., Siddharthan, A., Batchelor, C.: Corpora for the conceptualisation and zoning of scientific papers. In: Proceedings of LREC, pp. 2054–2061 (2010)
25. Wilbur, W.J., Rzhetsky, A., Shatkay, H.: New directions in biomedical text annotation: definitions, guidelines and corpus construction. BMC Bioinformatics 7, 356 (2006)
26. Miwa, M., Thompson, P., McNaught, J., Kell, D., Ananiadou, S.: Extracting semantically enriched events from biomedical literature. BMC Bioinformatics 13, 108 (2012)
27. Savova, G., Masanz, J., Ogren, P., Zheng, J., Sohn, S., Kipper-Schuler, K., Chute, C.: Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association 17, 507–513 (2010)
28. Cunningham, H., Hanbury, A., Rüger, S.: Scaling Up High-Value Retrieval to Medium-Volume Data. In: Cunningham, H., Hanbury, A., Rüger, S. (eds.) IRFC 2010. LNCS, vol. 6107, pp. 1–5. Springer, Heidelberg (2010)
29. Schäfer, U.: Middleware for creating and combining multi-dimensional NLP markup. In: Proceedings of the 5th Workshop on NLP and XML: Multi-Dimensional Markup in Natural Language Processing, pp. 81–84. ACL (2006)
30. Rak, R., Rowley, A., Black, W., Ananiadou, S.: Argo: an integrative, interactive, text mining-based workbench supporting curation. Database: The Journal of Biological Databases and Curation 2012 (2012)
31. Settles, B.: Biomedical named entity recognition using conditional random fields and rich feature sets. In: Proceedings of the International Joint Workshop on NLP in Biomedicine and its Applications, pp. 104–107. ACL (2004)
32. Gabbard, R., Freedman, M., Weischedel, R.: Coreference for Learning to Extract Relations: Yes Virginia, Coreference Matters. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 288–293. Association for Computational Linguistics, Portland (2011)
33. Tsuruoka, Y., Tsujii, J., Ananiadou, S.: Accelerating the annotation of sparse named entities by dynamic sentence selection. BMC Bioinformatics 9, S8 (2008)