Enhancing the AI2 Diagrams Dataset Using Rhetorical Structure Theory

Tuomo Hiippala and Serafina Orekhova
Department of Languages, Digital Geography Lab
University of Helsinki, Finland
tuomo.hiippala, serafina.orekhova @helsinki.fi

Abstract

This paper describes ongoing work on a multimodal resource based on the Allen Institute AI2 Diagrams (AI2D) dataset, which contains nearly 5000 grade-school level science diagrams that have been annotated for their elements and the semantic relations that hold between them. This emerging resource, named AI2D-RST, aims to provide a drop-in replacement for the annotation of semantic relations between diagram elements, whose description is informed by recent theories of multimodality and text-image relations. As the name of the resource suggests, the revised annotation schema is based on Rhetorical Structure Theory (RST), which has been previously used to describe the multimodal structure of diagrams and entire documents. The paper documents the proposed annotation schema, describes challenges in applying RST to diagrams, and reports on inter-annotator agreement for this task. Finally, the paper discusses the use of AI2D-RST for research on multimodality and artificial intelligence.

Keywords: multimodality, diagrams, rhetorical structure theory, artificial intelligence

1. Introduction

The Allen Institute AI2 Diagrams dataset (hereafter AI2D)[1] contains nearly 5 000 grade-school level science diagrams, which have been annotated for their elements and the semantic relations that hold between them (Kembhavi et al., 2016). The AI2D dataset was initially developed with two emerging computer vision tasks in mind: diagram understanding and visual question answering. Both tasks are challenging, as diagrams are inherently multimodal: they frequently combine various diagrammatic elements, such as arrows and lines with natural language, two-dimensional graphic elements, illustrations and photographs, and draw a multitude of semantic relations between them.

AI2D describes the semantic relations between diagram elements using a set of relations drawn from the framework for describing diagrams proposed in Engelhardt (2002). This paper proposes an alternative scheme for describing the semantic relations that hold between diagram elements, building on Rhetorical Structure Theory, or RST for short (Mann and Thompson, 1988; Taboada and Mann, 2006). To distinguish the original AI2D dataset from the emerging language resource described in this paper, we will adopt the name AI2D-RST to refer to the RST-enhanced dataset in the following discussion.

Although RST was originally developed for describing the organization of entire texts, the framework has also been applied to the generation of diagrammatic representations (André and Rist, 1995; Bateman et al., 2001). More recently, RST has been used to describe discourse structures in various multimodal artefacts ranging from newspapers and journals (Kong, 2013; Taboada and Habel, 2013) and product packaging (Thomas, 2014) to tourist brochures (Hiippala, 2015) and health care posters (Zhang, 2018). When combined with additional layers of description to capture the logical structure of content, its visual appearance and layout, even relatively small RST-annotated corpora have been able to reveal patterns characteristic to the structure of the aforementioned multimodal artefacts.

This suggests that RST can provide descriptions that are sufficiently fine-grained to bring out distinctions in the structure of various multimodal artefacts. As AI2D already contains annotations for the diagram elements and their layout, we assume that the application of RST to the dataset will provide multimodality researchers with a valuable resource for studying the structure of diagrammatic representations. Researchers working in the domain of artificial intelligence, in turn, may evaluate whether diagram understanding and visual question answering algorithms can learn better from RST annotation than the schema originally used for AI2D.

The paper itself is structured as follows: after introducing the AI2D dataset, we evaluate its original annotation from the viewpoint of multimodality research. Next, we proceed to describe the changes introduced in AI2D-RST and the challenges in applying RST to diagrams, and report on an experiment measuring inter-annotator agreement. Finally, we consider how the emerging resource may be used in research on multimodality and artificial intelligence.

2. The Original AI2D Dataset

Explicating the motivation for developing AI2D, Kembhavi et al. (2016, 235) note that research on computer vision has mainly focused on photographic images, while "rich visual illustrations", such as diagrams and information graphics, have received relatively little attention. In diagrams, this richness emerges from the combination of multiple modes of communication: diagrammatic elements, such as arrows and lines, are typically used alongside natural language and various types of images ranging from photographs to illustrations, and combined into meaningful ensembles by taking advantage of the layout space (Bateman et al., 2017, 281). As Kembhavi et al. (2016) observe, this makes diagrams radically different from photographic images, and for this reason, their computational processing involves an entirely different set of problems.

[1] Available at http://allenai.org/plato/diagram-understanding/ (accessed 13 February 2018).

Relationship	Definition	Count
ARROWHEAD ASSIGNMENT	An arrow head associated to an arrow tail	18541
INTRA-OBJECT LABEL	A text box naming the entire object	16281
INTER-OBJECT LINKAGE	Two objects related to one another via an arrow	15802
INTRA-OBJECT LINKAGE	A text box referring to a region within an object via an arrow	15664
INTRA-OBJECT REGION LABEL	A text box referring to a region within an object	2002
IMAGE TITLE	The title of the entire image	1458
IMAGE MISC	Decorative elements in the diagram	1148
IMAGE SECTION TITLE	Text box that serves as a title for a section of the image	807
IMAGE CAPTION	A text box that adds information about the entire image, but does not serve as the image title	690
ARROW DESCRIPTOR	A text box describing a process that an arrow refers to	681

Table 1: Semantic relations defined between diagram elements in the AI2D annotation by Kembhavi et al. (2016, 239). The column on the right-hand side gives their number of occurrences in the AI2D dataset.

To drive forward research on computational processing of diagrams, Kembhavi et al. (2016, 242–243) present a dataset containing 4 907 grade-school level science diagrams, which are described using 150 000 annotations that capture their elements, semantic interrelations and position in the layout. In addition, the dataset contains 15 000 multiple choice questions about the content of the diagrams for experiments involving diagram understanding and visual question answering.

The diagrams, which AI2D provides as PNG images accompanied by their annotations in JSON, were scraped from Google Images using the chapter headings of primary school science textbooks for grades one to six as seed terms. The result is a diverse dataset containing diagrams from various sources, ranging from professionally-produced diagrams in school textbooks to diagrams produced for learning materials by the teachers or by the students themselves. Figure 1 illustrates one diagram in the dataset, which clearly belongs to the first category of professionally-produced diagrams.

The diagrams were annotated by workers on Amazon Mechanical Turk[2], a crowdsourcing platform frequently used to create datasets for AI research (Kovashka et al., 2016). Due to their complex structure, the diagram annotation process was broken down into separate stages to ensure agreement between annotators. These stages involved, for instance, identifying diagram elements, categorising them, labelling their interrelations and answering multiple choice questions about the content of the diagram (Kembhavi et al., 2016, 243).

Building on the annotated corpus, Kembhavi et al. (2016, 239) propose to represent diagrams using graphs, which they refer to as *Diagram Parse Graphs* (hereafter DPG). DPG uses nodes to represent diagram elements, such as blobs (illustrations), text boxes, arrows and arrowheads, while the edges between nodes represent relationships between the elements. These relationships, which are listed in Table 1, are drawn from the framework developed by Engelhardt (2002) for analysing the syntax and semantics of maps, charts and diagrams.

That being said, Kembhavi et al. (2016) use the AI2D

[2] `https://www.mturk.com/`

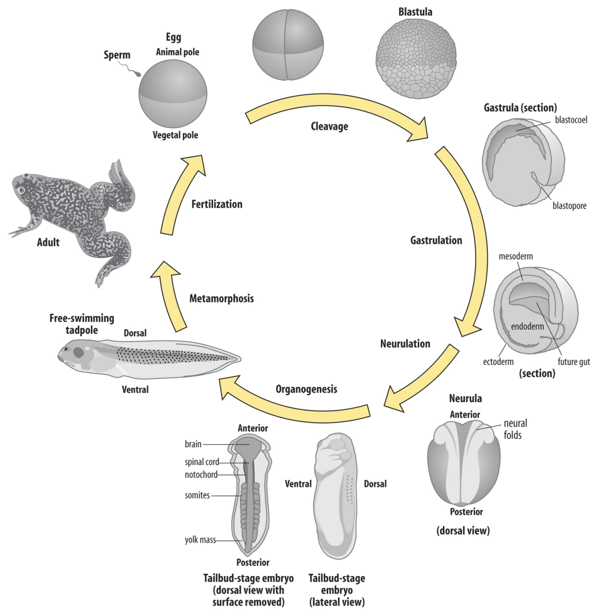

Figure 1: Diagram number 2240 in the AI2D dataset

dataset to train deep neural networks for two distinct tasks, which also reflect the original distinction defined by Engelhardt (2002): *syntactic parsing* and *semantic interpretation*. Whereas syntactic parsing refers to the task of learning to infer a DPG that captures the diagram structure, semantic interpretation is concerned with interpreting a DPG to answer questions about diagram content. The goal of AI2D is to enable and support the development and evaluation of algorithms for both tasks.

3. Evaluating AI2D from the Perspective of Multimodality Research

The work of Kembhavi et al. (2016) assumes that resolving the relations that hold between the elements of a diagram is crucial for their computational understanding. In research on multimodality, these relations are often discussed under the heading of text-image relations, or more broadly, relations that hold between contributions from different modes of expression. Bateman (2014a) provides an extensive re-

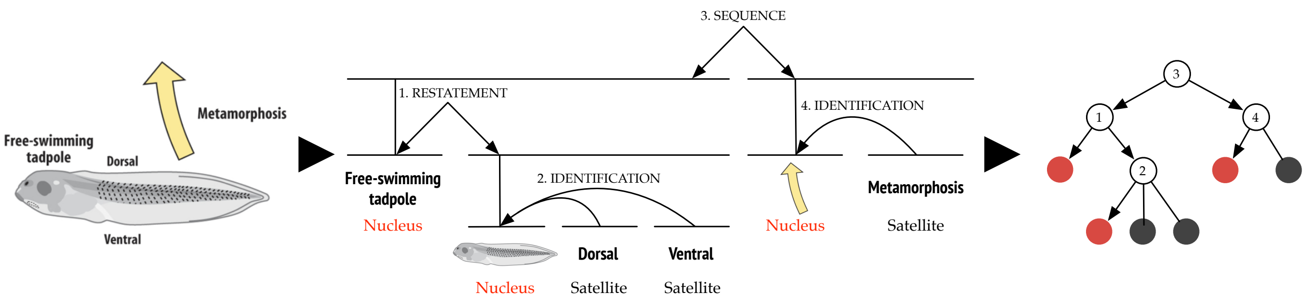

Figure 2: A partial RST analysis attempting to capture the "phase-and-process" structure of the diagram in Figure 1, represented using a schema for visualising RST relations. The straight lines indicate *multinuclear* relations between nuclei (such as RESTATEMENT and SEQUENCE), whereas the curved lines indicate *mononuclear* relations between nuclei and satellites (such as IDENTIFICATION). These relations are grouped together under *spans*, which are indicated by the horizontal lines. The information carried by spans, which reflects how diagram elements are chunked together, is carried over to the graph in the form of nodes that join related nuclei and satellites together. These nodes are coloured white, whereas nuclei are red and satellites are black.

view of research in this area, which may also be used to evaluate the semantic relations defined within AI2D from a multimodal perspective.

To begin with, several of the relations defined in AI2D appear to share roughly the same semantic function: Kembhavi et al. (2016, 245) acknowledge this explicitly by stating that relations 1–6 in Table 1 serve the purpose of relating one or more elements to one another. There seems to be, however, significant overlap among these relations, particularly between INTRA-OBJECT LABEL, INTRA-OBJECT REGION LABEL, INTRA-OBJECT LINKAGE and ARROW DESCRIPTOR, as they all appear to serve the general purpose of *identifying* some object.

It should be noted that such a semantic relation – identification – may be realised in different ways, either using an explicit diagrammatic element such as an arrow or a line, or spatially using layout, by placing the identifier and the identified close to each other. Both alternatives may be found on both inner and outer circle of the diagram shown in Figure 1 (cf. e.g. *Blastula* and *Gastrula (section)*). From a multimodal perspective, the question is whether these relations could be reduced into fewer, more generic categories, which would also unify the description.

Furthermore, the relation of ARROWHEAD ASSIGNMENT does not indicate a semantic, but rather a logical relation between diagram elements, indicating which elements belong together. Although defining such relations explicitly may be necessary for the current algorithms developed for diagram understanding, corpus-driven frameworks developed for describing complex multimodal artefacts such as the one presented in Bateman (2008) and its extension in Hiippala (2015) argue against conflating different descriptions of multimodal structure. This demarcation is intended to enable pulling apart different types of structure, in order to provide a clearer view of their distinct contribution to the meaning of the diagram.

In the case of AI2D, having multiple layers of annotation at hand – one for the logical structure, and one for the semantic relations – would also remove the need for relations such as IMAGE MISC, which is used to indicate decorative elements in the diagrams (see the last item in Table 1). If these descriptions were represented using multiple graphs, this would simply mean that some nodes present in the graph representing logical structure would be absent from the graph representing semantic relations. Because AI2D already contains rich annotations covering aspects of both logical and semantic structure, the existing annotation may be used as a basis for generating separate descriptions as a part of AI2D-RST.

4. Towards AI2D-RST

To clearly separate and describe different kinds of multimodal structures typically found in diagrams, AI2D-RST leverages the rich annotation provided by the original AI2D dataset to produce two distinct types of Diagram Parse Graphs (DPG). These DPGs, whose descriptions are also connected using cross-referenced identifiers, account for two kinds of multimodal structures in diagrams: DPG-L for the logical structure, and DPG-R for the rhetorical structure. The motivation for representing these structures using two different graphs is to make them more focused, that is, avoiding the problems associated with conflated descriptions of multimodal structure discussed in Section 3.

To begin with DPG-L, the purpose of this graph is to indicate which elements are generally presented as belonging together in the logical structure, without making assumptions about any semantic relations that may hold between them. As pointed out above, DPG-L covers all elements identified in the original AI2D annotation, such as those that do not necessarily contribute to the rhetorical structure, such as decorative elements or authorship attributions identified using the IMAGE MISC relation. This representation bears close resemblance to the original DPG defined in AI2D, as the analytical decomposition of diagrams remains the same in AI2D-RST.

DPG-R, in turn, contains only the nodes that participate in the rhetorical structure, whose relations are described using Rhetorical Structure Theory (Taboada and Mann, 2006), the multimodal extension of RST proposed in Bateman (2008). Essentially, this extension contains the original relations defined within 'classical' RST and a set of 'sub-nuclear' relations necessary for decomposing fragments of discourse, which classical RST would treat as a single analytical unit (Bateman, 2008, 162). Finally, certain diagram-specific additions to DPG-R are presented shortly below in Section 5.

Another feature that DPG-R introduces to the description is what RST terms *nuclearity*. This assumption states that some parts of a text or a multimodal ensemble act as *nuclei*, which carry the most relevant meanings for the communicative task at hand, whereas optional parts – *satellites* – enhance them by providing additional information. Figure 2 illustrates this by defining RST relations between the elements and representing these relations using a graph. Here the assumption is that algorithms developed for semantic interpretation of diagrams could learn to attend to the elements that carry information relevant to the task at hand, for instance, by searching for sequences among nuclei and additional information among satellites.

The application of RST also requires information on the chunking of elements, that is, how they are grouped together in the diagram. As Taboada and Mann (2006, 427) point out, RST relations are applied recursively to the artefact under analysis in order to capture the intended communicative effect, which may be achieved using a combination of rhetorical relations. This phenomenon is well-known from the analysis of entire multimodal documents, in which the chunks often constrain the process of interpretation by limiting the possible RST relations to elements belonging to the same chunk (Hiippala, 2015, 168). However, diagrams present certain requirements for RST analysis, which are discussed below.

5. Applying RST to AI2D

5.1. Feasibility

To assess the feasibility of using RST to describe semantic relations in AI2D, we sampled the data without replacement for 545 semantic relations and annotated them for both rhetorical relations and their nuclearity using RST. This number amounted to roughly 1% of semantic relations reported in Table 1, excluding the relationship of ARROWHEAD ASSIGNMENT, which we considered to belong to the logical structure, as set out in Section 3.

In order to measure the level of agreement between our annotations, we used the common metrics surveyed in Artstein and Poesio (2008), such as Scott's π and Krippendorff's α, as implemented in Natural Language Toolkit (NLTK) (Bird et al., 2009). Table 2 shows the scores for average observed agreement and the aforementioned metrics for RST relations and nuclearity. In the case of nuclearity, we assigned the role of *nucleus*, *satellite* or *none* (for elements with no RST relation) to the elements acting as the origin or destination of a semantic relation, as defined in the original AI2D annotation.

For annotating the sample, we used a total of 10 out of 35 available RST relations to describe how the elements relate to each other: CONTRAST, EFFECT, ELABORATION, IDENTIFICATION, INTERPRETATION, PREPARATION, PROPERTY-ASCRIPTION, RESTATEMENT, SEQUENCE and TITLE. We also used an additional relation, NONE, to mark cases in which no relation was deemed to hold between the diagram elements, providing a final number of 11 categories. The chance-corrected scores for measuring agreement between our annotation for RST relations, namely Krippendorff's α (0.7245) and Scott's π (0.7242) suggest that we may draw only tentative conclusions about semantic relations in diagrams on the basis of our initial annotation.

Metric	Relation	Origin	Destination
Average agreement	0.7835	0.8128	0.9083
Krippendorff's α	0.7245	0.5490	0.4520
Scott's π	0.7242	0.5486	0.4515

Table 2: Inter-annotator agreement between annotators

The results are nevertheless promising, as it should be noted that there was a considerable difference in our level of experience in applying RST, as the second author received only minimal training before the experiment, which may have affected the α and π scores for relations. The low scores for nuclearity (origin/destination), in turn, reflect the presence of just three categories, which obviously increases the possibility of chance agreement.

In order to provide a measure of agreement that would take into account the differences in our expertise in applying RST, we follow Das et al. (2017) and report precision, recall and F1-scores for expert (first author) vs. novice (second author) annotation in Table 3, as implemented in scikit-learn (Pedregosa et al., 2011).

	Precision	Recall	F1-score
RST relation	0.78	0.78	0.78
Origin role	0.81	0.81	0.81
Destination role	0.93	0.91	0.92

Table 3: Prevalence-weighted macro-average scores for precision, recall and F1 for expert vs. novice annotator

The precision, recall and F1-scores appear promising despite the limited training, and are likely to be improved by revising the annotation manual for AI2D-RST as the work proceeds. This involves extending RST to diagrammatic representations by redefining some of the relations and their definitions, as we will describe below.

5.2. Extending RST

Evaluating inter-annotator agreement, in connection with close analyses of selected examples, helped to identify a core set of RST relations applicable to the description of semantic relations between diagram elements, which are given in Table 4. At this point, however, it is also useful to highlight some challenges in applying RST to diagrams. To begin with, classical RST was designed to capture relations that hold between sequential units of discourse in written language, mainly at the level of clause and beyond. The assumption of sequentiality, however, which is at the

Relationship	Description	Source
EFFECT	A generic mononuclear relation for describing processes that take place between entities, which are often reinforced using lines or arrows (see Figure 3). The affected entity acts as the nucleus, while the origin of the effect acts as the satellite.	AI2D RST
ELABORATION	A more extensive verbal description, such as a phrase or a clause, which provides more specific information about some entity or its part(s).	Classical RST
IDENTIFICATION	A short text segment, such as a single noun or a noun group, which identifies an entity or its part(s). A typical example would be a label for a part of an entity (see Figure 2).	GeM RST
PROPERTY-ASCRIPTION	A mononuclear relation between an entity (nucleus) and something predicated of that entity (satellite).	GeM RST
RESTATEMENT	A multinuclear relation holding between two entities that could act as a substitute for each other, such as the name of an entity and its visualisation (see Figure 2).	Classical RST
SEQUENCE	A multinuclear relation indicating a temporal or spatial sequence holding between two or more entities.	Classical RST
TITLE	A text segment acting as the title for the entire diagram or its parts.	GeM RST

Table 4: Common RST relations encountered in preliminary studies of applying RST to the AI2D dataset. The column on the right gives the work in which the relation was originally defined. Classic RST refers to the foundational work in Mann and Thompson (1988) and Taboada and Mann (2006), whereas GeM RST refers to the multimodal extension presented in Bateman (2008). AI2D RST, in turn, refers to the work presented here.

heart of classical RST, rarely holds for entire multimodal artefacts or their parts (Hiippala, 2015, 50–51).

For the analysis of written texts, this assumption enabled classical RST to control the number of potential relations drawn between different units of discourse, as this prevented drawing relations between non-adjacent units of discourse in written text. In a complete and well-formed description of rhetorical structure, each discourse unit would participate in a single relation, and textual progression would naturally structure the discourse units into a recursive organisation (Mann et al., 1992).

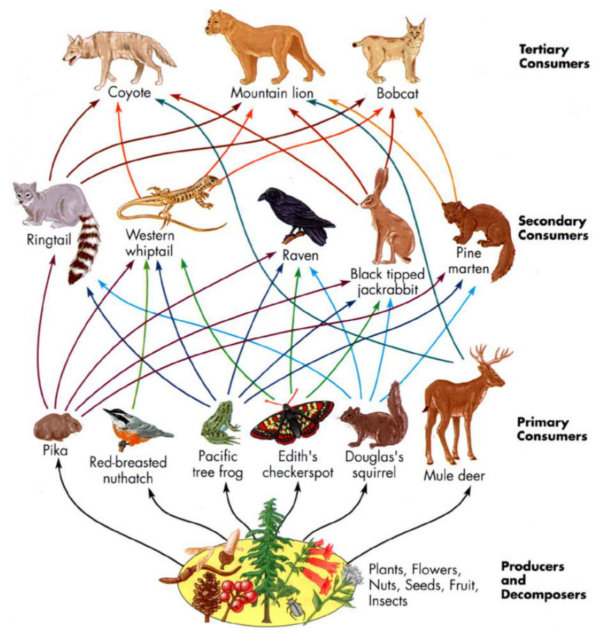

Figure 3: Diagram number 1807 in the AI2D dataset

In multimodal extensions of RST, the sequentiality assumption has been discarded in favour of alternative criteria such as spatial proximity (Bateman, 2008, 158). Applying RST to diagrams is likely to require additional criteria to complement spatiality. *Connectedness* emerges as one criterion for addressing graph-like diagrams, in which diagram elements participate in multiple relations, which do not necessarily respect the criterion of spatiality.

This may be exemplified using Figure 2, which shows how several diagram elements, namely the combinations of an illustration and its caption, relate to multiple elements of the same kind, which are not positioned close to each other in the diagram layout (see e.g. the arrow drawn from *Mule deer* to *Coyote*).

To capture the relations that build on the property of connectedness, we introduce a generic relation termed EFFECT, which describes any relation between two interconnected elements that affect each other in some way (see Table 4). The target of this effect is marked as the nucleus, whereas the origin acts as the satellite. Introducing this relation allows taking on graph-like diagrams, whose elements participate in multiple rhetorical relations. This, however, requires an alternative, graph-like means of visualising RST structures, as these interconnections are difficult to visualise using hierarchical trees. Finally, it should be noted that we do not incorporate the diagram elements responsible for signalling EFFECT, such as arrows and lines, into the description of rhetorical relations, but consider them as a part of the logical structure (see Section 4).

6. AI2D-RST as a Multimodal Resource

AI2D-RST is intended as a resource for researchers working on multimodality and artificial intelligence. The research community focusing on multimodality has long called for larger datasets, which would enable empirical research in the manner of corpus linguistics (Kaltenbacher, 2004; Bateman, 2014b; O'Halloran et al., 2018). AI2D-RST takes the first step towards this long-term goal, enabling the empirical study of diagrammatic representations and testing hypotheses about their multimodal characteristics against a sufficiently large dataset. If such a resource is found useful, this may also lead to the adoption of crowd-sourcing techniques for generating low-level annotations for multimodal corpora in the future.

For the AI community, AI2D-RST offers a dataset containing mixed annotations, sourced from non-experts in the form of original, crowd-sourced annotations describing the diagram elements, and from experts in the form of the RST annotation capturing the relations between these elements. Moreover, the RST-based description in AI2D-RST provides the AI research community with the first annotation schema informed by theories of multimodality. This novel resource may be used to evaluate whether expert annotations improve the performance of algorithms for the tasks defined by Kembhavi et al. (2016).

To support further research on diagrams and potential applications of the dataset, we also provide tools for visualising the annotation in both AI2D and AI2D-RST, and their respective Diagram Parse Graphs, in addition to the annotation tool used for creating the AI2D-RST corpus. These tools, written in Python 3.6, and the AI2D-RST corpus will be available online.[3]

7. Conclusion

This paper has presented an emerging multimodal language resource based on the AI2 Diagrams dataset (Kembhavi et al., 2016), which contains nearly 5000 grade-school level diagrams for developing algorithms that can process the structure and contents of diagrams. The emerging resource described in this paper, named AI2D-RST, enriches the annotation contained in the original AI2D dataset by using Rhetorical Structure Theory (RST) to describe the relations holding between elements participating in the diagram. The motivation for developing AI2D-RST is to provide an alterative annotation informed by recent research on multimodality, which has shown RST to be a powerful analytical tool for describing semantic relations, particularly when combined with descriptions of multimodal structure, which are already in place in AI2D. In the long run, AI2D-RST is expected to contribute to research on multimodality and artificial intelligence, improving the understanding of diagrams for both humans and computers.

8. Bibliographical References

André, E. and Rist, T. (1995). Generating coherent presentations employing textual and visual material. *Artificial Intelligence Review*, 9:147–165.

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Bateman, J. A., Kamps, T., Reichenberger, K., and Kleinz, J. (2001). Towards constructive text, diagram, and layout generation for information presentation. *Computational Linguistics*, 27(3):409–449.

Bateman, J. A., Wildfeuer, J., and Hiippala, T. (2017). *Multimodality: Foundations, Research and Analysis – A Problem-Oriented Introduction*. De Gruyter Mouton, Berlin.

Bateman, J. A. (2008). *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Palgrave Macmillan, London.

Bateman, J. A. (2014a). *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. Routledge, London and New York.

Bateman, J. A. (2014b). Using multimodal corpora for empirical research. In Carey Jewitt, editor, *The Routledge Handbook of Multimodal Analysis*, pages 238–252. Routledge, London and New York, second edition.

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly, Sebastopol, CA.

Das, D., Stede, M., and Taboada, M. (2017). The Good, the Bad, and the Disagreement: Complex ground truth in rhetorical structure analysis. In *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, pages 11–19. Association for Computational Linguistics.

Engelhardt, Y. (2002). *The Language of Graphics: A Framework for the Analysis of Syntax and Meaning in Maps, Charts and Diagrams*. Ph.D. thesis, Institute for Logic, Language and Computation, University of Amsterdam.

Hiippala, T. (2015). *The Structure of Multimodal Documents: An Empirical Approach*. Routledge, New York and London.

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.

Kembhavi, A., Salvato, M., Kolve, E., Seo, M., Hajishirzi, H., and Farhadi, A. (2016). A diagram is worth a dozen images. In *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV*, pages 235–251, Cham. Springer.

Kong, K. C. C. (2013). A corpus-based study in comparing the multimodality of Chinese- and English-language newspapers. *Visual Communication*, 12(2):173–196.

Kovashka, A., Russakovsky, O., Fei-Fei, L., and Grauman, K. (2016). Crowdsourcing in computer vision. *Foundations and Trends in Computer Graphics and Vision*, 10(3):177–243.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Mann, W. C., Matthiessen, C. M., and Thompson, S. A. (1992). Rhetorical structure theory and text analysis. In William C. Mann et al., editors, *Discourse Description: Diverse Linguistic Analyses of a Fund-raising Text*, pages 39–78. Benjamins, Amsterdam.

O'Halloran, K. L., Tan, S., Pham, D.-S., Bateman, J. A., and Vande Moere, A. (2018). A digital mixed methods research design: Integrating multimodal analysis with data mining and information visualization for big

[3] The resources will be made available at: `https://www.github.com/DigitalGeographyLab/AI2D-RST`

data analytics. *Journal of Mixed Methods Research*, 12(1):11–30.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Taboada, M. and Habel, C. (2013). Rhetorical relations in multimodal documents. *Discourse Studies*, 15(1):65–89.

Taboada, M. and Mann, W. C. (2006). Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8(3):423–459.

Thomas, M. (2014). Evidence and circularity in multimodal discourse analysis. *Visual Communication*, 13(2):163–189.

Zhang, K. (2018). *Public Health Education Through Posters in Two World Cities: A Multimodal Corpus-Based Analysis*. Ph.D. thesis, The Hong Kong Polytechnic University.