Argument Mining for Educational Applications using Discourse and Diagrams

Diane Litman

Professor, Computer Science Department
Senior Scientist, Learning Research & Development Center
Co-Director, Intelligent Systems Program

University of Pittsburgh
Pittsburgh, PA USA

Summer School on Argumentation: Computational and Linguistic Perspectives
September 8, 2014

Argument Mining

- "… techniques and methods for analyzing real data in natural arguments which will ultimately help us to automatically recognize and extract argumentative structures." *[12th ArgDiaP Conf., May 2014]*

- "… a relatively new challenge in corpus-based discourse analysis that involves automatically identifying argumentative structures within a document, e.g., the premises, conclusion, and argumentation scheme of each argument, as well as … relationships between pairs of arguments." *[1st Workshop on Argumentation Mining, ACL Conf., June 2014]*

- "… exploits the techniques and methods of natural language processing … for semi-automatic and automatic recognition and extraction of structured argument data from unstructured … texts." *[SICSA Workshop on Argument Mining, July 2014]*

Tutorial Goals

- ***Argument mining*** (from text) for teaching and assessing ***argumentative writing***
- Argument Mining Algorithms via NLP
  - Discourse Analysis
  - Data-Driven Approaches
- Empirical Evaluation Methods
- Case Studies of Educational Applications
  - Elementary, high school, and university student writing data

Why teach argumentative writing?

- In argumentative or scientific writing, authors need to:
  - communicate a position, claim, or hypothesis,
  - use it to frame reasons and evidence presented,
  - integrate others' arguments, anticipate/refute counterarguments (Durst, 1987; Mitchell, 2001; Nystrand & Graff, 2001; Yeh, 1998).
- Studies show students:
  - lack competence in argument writing (Oostdam, et al., 1994; Oostdam & Emmelot, 1991).
  - express opinions that agree/disagree with isolated statements (Ryan & Norris, 1991) but
  - do not integrate their arguments into a high-level structure or coherent position (Keith, Weiner, & Lesgold, 1991).
  - cannot distinguish between problems of
    - presenting arguments well vs.
    - generating good arguments (see Chryssafidou and Sharples, 2003).
  - Even if compose-aloud protocols show students mentally connect position statement & supporting details, *connections not evident in writing* (Durst, 1987).

 4

NLP Challenges of Educational Applications

- Noisy data (e.g. adult learners, children)
- Real-time algorithms; robust at scale
- Meaningful independent variables
- Similar challenges may arise in other emerging applications (e.g. legal text retrieval, scientific text analysis, opinion mining for user-generated content regarding products and politics)

Outline

- **Argument Mining (from Text)**
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - Tree-like structures
- Resources
  - Corpora
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- Looking Forward

Argument Mining: The Problem

(This tutorial section is from [Peldszus & Stede, 2013])

Argument Mining

- *"…the automatic discovery of an argumentative text portion, and identification of the relevant components of the argument presented there."*

Subtasks

1. **Segmentation:** Break the text down into minimal units of analysis, henceforth called 'argumentative discourse units' (ADUs).
2. **Segment Classification:** Determine the role that each ADU is playing for the argumentation.
3. **Relation Identification:** Establish relations between individual ADUs, possibly leading to a complete tree or graph structure, or to an instantiated schema of sorts.
4. **Argument Completion:** Steps 2 and 3 may involve the postulation of 'implicit' ADUs, which the analyzer constructs in order to achieve a complete structural description.

Argument Mining Subtasks: Framing and Prior Work

A. A general discourse perspective

B. Argument mining proper

1. Segmentation

- Similar to finding elementary discourse units (EDUs) in relation-based discourse theories
  - Sentences
  - Clauses
  - Constituents (e.g. prepositional phrases)
- However, ADUs may be larger than EDUs
  - Not all EDU relations are argument-relevant

2. Segment Classification

- **Minimal:** premises versus conclusions
  - may be sufficient for certain applications
- **Genre approaches:** portions of a text are analyzed in terms of their contribution to overall function
  - A text is broken down into 'content zones'
  - Genres are characterized by specifying
    - An inventory of mandatory and optional zones
    - Constraints and preferences on the linear order of zones

"Argumentative Zoning" for Scientific Papers [Teufel and Moens, 2002]

- Content zone inventory (for sentence segments)
    - **Aim:** Research goal of the paper
    - **Textual:** Statements about section structure
    - **Own:** Description of the authors' work (methodology, results, discussion)
    - **Background:** Generally accepted scientific background
    - **Contrast:** Comparison with other work
    - **Basis:** Statements of agreement with other work
    - **Other:** Description of other researchers' work

  Note: zones are more discourse genre rather than argument schema dependent
- Example features (for automatic zone classification via Naïve Bayes)
    - Sentence position and length
    - Formulaic expressions
    - Syntactic properties (voice, tense, model auxiliaries)
    - Context (zone of prior sentence)
    - Verb semantics
- Performance results
    - F-measures (details later!) from 86% (**Own**) to 26% (**Contrast**)

3. Relation Identification

- Explicit Relations
    - Overtly marked (e.g. connectives, cue phrases)
    - *The book never appeared, because the publisher had gone bankrupt.*
- Implicit Relations
    - Involve world knowledge and inference
    - *The book never appeared. The publisher had gone bankrupt.*

Discourse Relations particularly relevant to Argument Mining

- Causal Relations
  - Similar to argumentative support such as evidence
  - Research topic in *Why…* question-answering, bioinformatics
  - Often implicit
- Contrastive Relations
  - Similar to rebuttal and counter-rebuttal
  - Research topic in textual entailment, bioinformatics
  - Typically explicit

4. Argument Completion

- The task of discourse parsing (e.g. deriving an RST tree - details later) is similar to that of detecting an argument structure (e.g. a diagram)

Argument (as opposed to Discourse) Mining

- Map text (portion) to (graphical) arg. structure, *e.g.*
  - Hybrid analysis [Saint-Dizier, 2012; Green, 2010]
  - Legal argumentation [Palau and Moens, 2009]
    - Classify *sentences* as *argumentative* or *non-argumentative*
    - Classify *argumentative sentences* as *premise* or *conclusion*
    - Link *premises* and *conclusions* into *structures*
  - Persuasive essays
    - Argumentative discourse structure [Stab and Gurevych, to appear]
      - Identify *argument components* (claims, premises) using multiclass classificaton
      - Classify a pair of argument components as support or not *(relations)*
    - Separating *organizational elements* from content [Madnani et al, 2012]
      - Claim and premise word sequences (i.e. *shell expressions*)
      - Does not yet include argument components or relations
  - Argument scheme classification (e.g., from example, from cause to effect) [Feng and Hirst, 2011]
    - Assumes *segments* already classified as *premise/conclusion*
    - Reconstruct *unstated premises*
  - My work on educational essay data (case studies later in this tutorial)

Outline

- Argument Mining (from Text)
- **Computational Discourse**
  - **Functional structures**
  - Predicate-argument structures
  - Tree-like structures
- Resources
  - Corpora
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- Looking Forward

Functional Structure

- Texts within a given genre – e.g.,
  - news reports
  - scientific papers or abstracts
  - etc.

  generally share a similar structure, that is independent of topic and reflects the function played by each of their parts

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Example

- Well-known in academia is the multi-part structure of scientific papers (and also their abstracts)
  - **Objective** (aka Introduction, Background, Aim, Hypothesis)
  - **Methods** (aka Study Design, Methodology, etc.)
  - **Results** (aka Outcomes)
  - **Discussion**
  - Optionally, **Conclusions**
- **N.B.** Not every sentence within a section need realize the same function

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Functional Structure

- Automatic annotation of functional structure is seen as benefitting:
  - Information extraction: Certain types of information are likely to be found in certain sections [Moens]
  - Extractive summarization: More "important" sentences are more likely to be found in certain sections
  - Sentiment analysis: Words that have an objective sense in one section may have a subjective sense in another [Taboada]
  - Citation analysis: A citation may serve different functions in different sections [Teufel]

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Functional Structure

- Computational approaches to functional structure and segmentation assume that:
  - The function of a segment relates to that of the discourse as a whole.
  - While relations may hold between sisters (eg, *Methods* constrain *Results*), only sequence has been used in modelling.
  - Function predicts more than lexical choice:
    - indicative phrases such as "results show" (-> *Results*)
    - indicative stop-words such as "then" (-> *Method*).
  - Functional segments usually appear in a specific order, so either sentence position is a feature used in modelling or sequential models are used.

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Functional Structure

- The internal structure of segments has usually been ignored in high-level functional segmentation
- But given the results of work in fine-grained modelling of functional structure, not surprising that Hirohata et al [2008] found that
  - Properties of the first sentence of a segment differ from those of the rest.
  - Modelling this leads to improved performance in high-level functional segmentation.

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Labelled Biomedical Abstracts

- Much function-based modelling has been on biomedical text, where texts with explicitly labelled sections serve as free training data for segmenting unlabelled texts.
  - **BACKGROUND:** Mutation impact extraction is a hitherto unaccomplished task in state of the art mutation extraction systems. . . . **RESULTS:** We present the first rule-based approach for the extraction of mutation impacts on protein properties, categorizing their directionality as positive, negative or neutral. . . . **CONCLUSION:** . . . Our approaches show state of the art levels of precision and recall for Mutation Grounding and respectable level of precision but lower recall for the task of Mutant-Impact relation extraction. . . . **[PMID 21143808]**

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Analyzing and Scoring Student Essays

- The structure of a student's essay contributes to its quality
  - The **main point** of an essay should come before text that acts to **support** it.
  - Downgrade essay if it doesn't.

<Introductory material> In Korea, where I grew up, many parents seem to push their children into being doctors, lawyers, engineer etc. **</Introductory material> <Main point>**Parents believe that their kids should become what they believe is right for them, but most kids have their own choice and often doesn't choose the same career as their parent's**. </Main point> <Support>** I've seen a doctor who wasn't happy at all with her job because she thought that becoming doctor is what she should do. That person later had to switch her job to what she really wanted to do since she was a little girl, which was teaching. **</Support>**

[Burstein et al 2003]

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Exercise: Annotating Functional Structure

Functions: **Conclusion** (summarize entire argument)**, Introductory material** (context in which thesis, main points, or conclusion are to be interpreted)**, Irrelevant, Main Point** (author's main message in conjunction with thesis), **Support (**provide evidence supporting main points, thesis, conclusion**), Thesis** (writer's position statement)

"You can't always do what you want to do," my mother said. She scolded me for doing what I thought was best for me. It is very difficult to do something that I do not want to do. But now that I am mature enough to take responsibility for my actions, I understand that many times in our lives we have to do what we should do. However, making important decisions, like determining your goal for the future, should be something that you want to do and enjoy doing.

I've seen many successful people who are doctors, artists, teachers, designers, etc. In my opinion they were considered successful people because they were able to find what they enjoy doing and worked hard for it. It is easy to determine that he/she is successful, not because it's what others think, but because he/she have succeed in what he/she wanted to do.

<Introductory material>In Korea, where I grew up, many parents seem to push their childre into being doctors, lawyers, engineer etc. **</Introductory material> <Main point>** Parents believe that their kids should become what they believe is right for them, but most kids have their own choice and often doesn't choose the same career as their parent's **</Main point> <Support>** I've seen a doctor who wasn't happy at all with her job because she thought that becoming doctor is what she should do. That person later had to switch her job to what she really wanted to do since she was a little girl, which was teaching. **</Support>**

Parents might know what's best for their own children on a daily basis, but deciding a long term goal for them should be one's own decision of what he/she likes to do and wants to do**.**

Answer: Annotating Functional Structure

Functions: **Conclusion, Introductory material, Irrelevant, Main Point, Support, Thesis**

<**Introductory material>** “You can’t always do what you want to do,” my mother said. She scolded me for doing what I thought was best for me. It is very difficult to do something that I do not want to do. **</ Introductory material> <Thesis>** But now that I am mature enough to take responsibility for my actions, I understand that many times in our lives we have to do what we should do. However, making important decisions, like determining your goal for the future, should be something that you want to do and enjoy doing **</Thesis>**

<Introductory material> I’ve seen many successful people who are doctors, artists, teachers, designers, etc. **</Introductory material> <Main point>** In my opinion they were considered successful people because they were able to find what they enjoy doing and worked hard for it **</Main point> <Irrelevant>** It is easy to determine that he/she is successful, not because it’s what others think, but because he/she have succeed in what he/she wanted to do.**<Irrelevant>**

<Introductory material>In Korea, where I grew up, many parents seem to push their childre into being doctors, lawyers, engineer etc. **</Introductory material> <Main point>** Parents believe that their kids should become what they believe is right for them, but most kids have their own choice and often doesn’t choose the same career as their parent’s **</Main point> <Support>** I’ve seen a doctor who wasn’t happy at all with her job because she thought that becoming doctor is what she should do. That person later had to switch her job to what she really wanted to do since she was a little girl, which was teaching. **</Support>**

<Conclusion>Parents might know what’s best for their own children on a daily basis, but deciding a long term goal for them should be one’s own decision of what he/she likes to do and wants to do**. </Conclusion>**

Topic Structure

- Another type of (typically) linear structure
- Will not be covered by this tutorial

Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - **Predicate-argument structures**
  - Tree-like structures
- Resources
  - Corpora
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- Looking Forward

Predicate-Argument Structures

- Discourse has structure arising from semantic and pragmatic relations that hold between the referents of its clauses.
- These "higher-order" pred-arg structures (aka *discourse relations* or *coherence relations*) are often explicitly signalled by a discourse connective
  – a conjunction like *because* or *but*,
  – a discourse adverbial like *nevertheless* or *instead*.

  though they may be signalled by other means, like *that means, what if*, etc. [Prasad et al, 2008]).

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Predicate-Argument Structures

- Coherence relations can also be conveyed through adjacency between clauses or sentences (aka implicit connectives).
  - *Viewers may not be cheering, either. (implicit= REASON) Soaring rights fees will lead to an even greater clutter of commercials*. [wsj 1057]

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Predicate-Argument Structures

- The Penn Discourse TreeBank (PDTB) (more later) is currently the largest resource manually annotated for discourse connectives, their arguments, and the senses they convey
- Snapshot
  - Explicit: 18459 tokens
  - Implicit: 16224 tokens

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Predicate-Argument Structures

- Computational models assume that:
  - Each predicate/relation has two arguments.
  - The arguments can be distinguished
    - syntactically, where the arg syntactically attached to an explicit connective is called **arg2**, and the other**, arg1** [Prasad et al, 2008].
    - semantically, where one arg of any *Causal* relation is the **cause**, and the other, the **result**) [Oza, 2009]
    - positionally, where **arg1** of an implicit connective always precedes **arg2** [Prasad et al, 2008].

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Predicate-Argument Structures

- The structure is not necessarily a tree:
  - A single span may serve as an argument to multiple relations (ie, have incoming edges from different nodes).
- The structure may only be a partial cover of the text.

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Serving as an arg to multiple relations

- In times past, life-insurance salesmen targeted heads of household, meaning men, but ours is a two-income family and accustomed to it. So if anything happened to me, I'd want to leave behind enough so that my 33-year-old husband would be able to pay off the mortgage . . . [Lee et al., 2006]

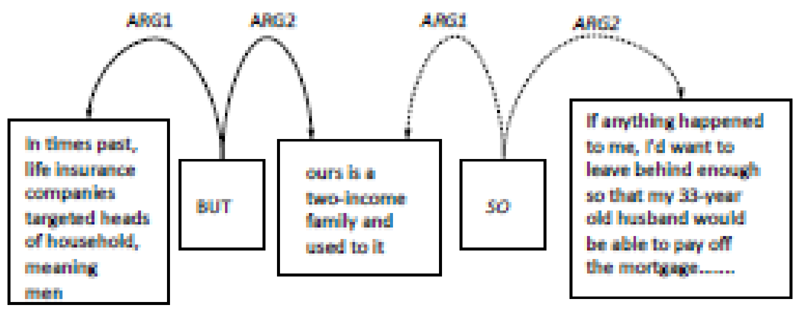
Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Serving as an arg to multiple relations

- In times past, life-insurance salesmen targeted heads of household, meaning men, but ours is a two-income family and accustomed to it. So if anything happened to me, I'd want to leave behind enough so that my 33-year-old husband would be able to pay off the mortgage . . . [Lee et al., 2006]

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Serving as an arg to multiple relations



Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Partial connectivity – Disconnected structures

- The early omens, we admit, scarcely suggest so wholesome an outcome. ___The Fleet Street reaction was captured in the Guardian headline, "Departure Reveals Thatcher Poison." ___ British politicians divide into two groups of chickens, those with their necks cut and those screaming the sky is falling. ___ So far as we can see only two persons are behaving with a dignity recognizing the seriousness of the issues: Mr. Lawson and Sir Alan Walters . . . . [wsj 0553]

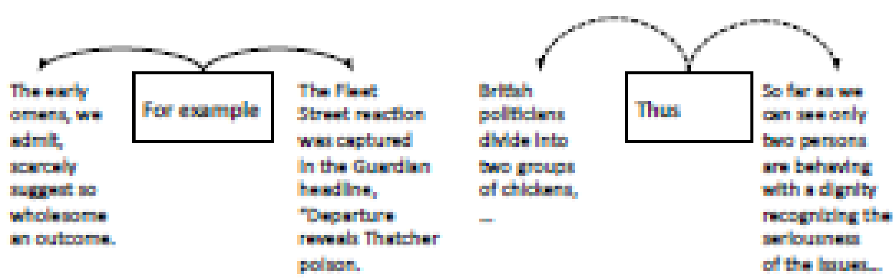
Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Partial connectivity – Disconnected structures

- The early omens, we admit, scarcely suggest so wholesome an outcome. Implicit=for example The Fleet Street reaction was captured in the Guardian headline, "Departure Reveals Thatcher Poison." NoRel British politicians divide into two groups of chickens, those with their necks cut and those screaming the sky is falling. Implicit=thus So far as we can see only two persons are behaving with a dignity recognizing the seriousness of the issues: Mr. Lawson and Sir Alan Walters . . . . [wsj 0553]

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Partial connectivity – Disconnected structures



Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Exercise: Annotating Pred-Arg Structures

- **Explicit connectives:** Specific connectives that indicate a discourse relation.
- **Implicit connectives:** There is no connective, but a relation can be inferred. Remember to propose an explicit connective that could capture this relation.
- **NoRel:** No discourse relation can be perceived.
- **Arg1/Arg2:** For explicit connectives, Arg2 is bound to the connective and Arg1 is the other part of the relation. In all the other cases, the order of the arguments is linear.
- **Sense of connectives:** The type of the relation (e.g. cause, comparison, condition)

Examples to annotate [Prasad et al., 2008]: connective, *Arg1*, **Arg2**

1. Some have raised their cash positions to record levels. High cash positions help buffer a fund when the market falls.
2. Jacobs is an international engineering and construction concern. Total capital investment at the site could be as much as $400 million, according to Intel.
3. It was a far safer deal for lenders since NWA had a healthier cash flow and more collateral on hand.
4. Domestic car sales have plunged 19% since the Big Three ended many of their programs Sept. 30.

Answer: Annotating Pred-Arg Structures

- **Explicit connectives:** Specific connectives that indicate a discourse relation.
- **Implicit connectives:** There is no connective, but a relation can be inferred. Remember to propose an explicit connective that could capture this relation.
- **NoRel:** No discourse relation can be perceived.
- **Arg1/Arg2:** For explicit connectives, Arg2 is bound to the connective and Arg1 is the other part of the relation. In all the other cases, the order of the arguments is linear.
- **Sense of connectives:** The type of the relation (e.g. cause, comparison, temporal)

Examples to annotate [Prasad et al., 2008]: connective, *Arg1*, **Arg2**

1. *Some have raised their cash positions to record levels.* Implicit=BECAUSE **High cash positions help buffer a fund when the market falls.** (causal)
2. *Jacobs is an international engineering and construction concern.* NoRel **Total capital investment at the site could be as much as $400 million, according to Intel.**
3. *It was a far safer deal for lenders* since **NWA had a healthier cash flow and more collateral on hand.** (causal)
4. *Domestic car sales have plunged 19%* since **the Big Three ended many of their programs Sept. 30.** (causal and temporal)

Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - **Tree-like structures**
- Resources
  - Corpora
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- Looking Forward

Tree-like Structures

- Discourse structures mainly as trees, although some more complex graph structures
- This tutorial focuses only on RST
  - associated corpus and parsers

Introduction to RST (Rhetorical Structure Theory)

Slides modified from the RST Web site
http://www.sfu.ca/rst/

Maite Taboada and Manfred Stede
May 2009

Rhetorical Structure Theory

- Created as part of a project on Natural Language Generation at the Information Sciences Institute (www.isi.edu)

- Central publication
  - Mann, William C. and Sandra A. Thompson. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text, 8* (3), 243-281.

- More recent overview
  - Taboada, Maite and William C. Mann. (2006). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies, 8* (3), 423-459.

- For many more publications and applications, visit the bibliography on the RST web site

Principles

- Coherent texts consist of minimal units, which are linked to each other, recursively, through rhetorical relations
  - Rhetorical relations also known, in other theories, as coherence or discourse relations
- Coherent texts do not show gaps or non-sequiturs
  - Therefore, there must be some relation holding among the different parts of the text

45

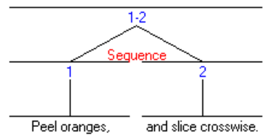
Components

- Units of discourse
  - Texts can be segmented into minimal units, or spans
- Nuclearity
  - Some spans are more central to the text's purpose (nuclei), whereas others are secondary (satellites)
  - Based on hypotactic and paratactic relations in language
- Relations among spans
  - Spans are joined into discourse relations
- Hierarchy/recursion
  - Spans that are in a relation may enter into new relations

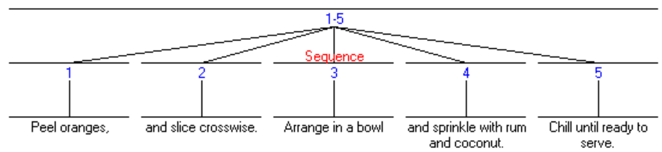
46

Paratactic (coordinate)

- At the sub-sentential level (traditional coordinated clauses)
  - Peel oranges, and slice crosswise.



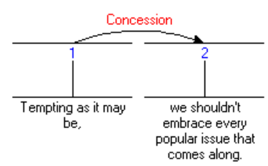
- But also across sentences
  - 1. Peel oranges, 2. and slice crosswise. 3. Arrange in a bowl 4. and sprinkle with rum and coconut. 5. Chill until ready to serve.



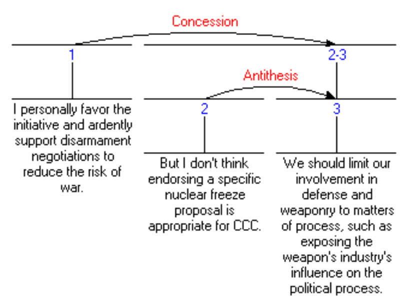
47

Hypotactic (subordinate)

- Sub-sentential Concession relation



- Concession across sentences
  - Nucleus (spans 2-3) made up of two spans in an Antithesis relation



48

Relations

- Hold between 2 non-overlapping text spans
- Most hold between a nucleus and a satellite, although also multi-nuclear
- Consisist of
  1. Constraints on the Nucleus
  2. Constraints on the Satellite
  3. Constraints on the combination of Nucleus & Satellite
  4. The Effect

49

Example: Evidence

- Constraints on the Nucleus
  - The reader may not believe N to a degree satisfactory to the writer
- Constraints on the Satellite
  - The reader believes S or will find it credible
- Constraints on the combination of N+S
  - The reader's comprehending S increases their belief of N
- Effect (the intention of the writer)
  - The reader's belief of N is increased

- Assumes a written text and readers/writers; extensions for spoken language
- Definitions of common relations available from the RST web site

50

Relation types

- **Subject matter**: relate content of text spans
  - Cause, Purpose, Condition, Summary
- **Presentational**: more rhetorical in nature; meant to achieve some effect on the reader
  - Motivation, Antithesis, Background, Evidence
  - Particularly relevant for argumentation!

51

Relation names (in M&T 1988)

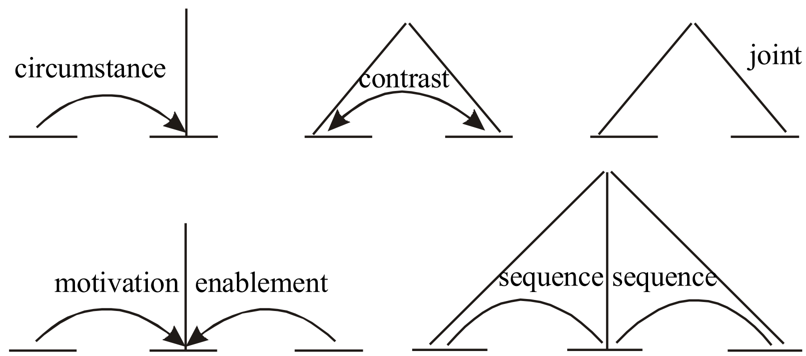
Circumstance	**Antithesis and Concession**
Solutionhood	**Antithesis**
Elaboration	**Concession**
Background	**Condition and Otherwise**
Enablement and Motivation	**Condition**
Enablement	**Otherwise**
Motivation	**Interpretation and Evaluation**
Evidence and Justify	**Interpretation**
Evidence	**Evaluation**
Justify	**Restatement and Summary**
Relations of Cause	**Restatement**
Volitional Cause	**Summary**
Non-Volitional Cause	**Other Relations**
Volitional Result	**Sequence**
Non-Volitional Result	**Contrast**
Purpose	

Other classifications are possible, and longer / shorter lists have been proposed

52

Schemas

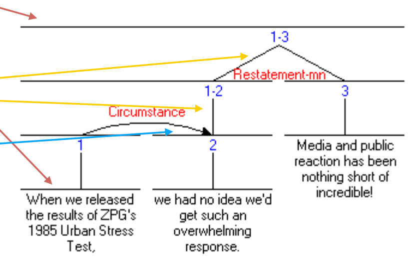
- They specify how spans of text can co-occur, determining possible RST text structures



53

Graphical representation

- A horizontal line covers a span of text (possibly made up of further spans
- A vertical line signals the nucleus or nuclei
- A curve represents a relation, and the direction of the arrow, the direction of satellite towards nucleus



- RST tool (for drawing diagrams)
  - http://www.wagsoft.com/RSTTool/

54

How to do an RST analysis

1. Divide the text into units
   - Unit size may vary, depending on the goals of the analysis
   - Typically, units are clauses
2. Examine each unit, and its neighbours. Is there a clear relation holding between them?
3. If yes, then mark that relation (e.g., Condition)
4. If not, the unit might be at the boundary of a higher-level relation. Look at relations holding between larger units (spans)
5. Continue until all the units in the text are accounted for
6. Remember, marking a relation involves satisfying all 4 fields (especially the Effect). The Effect is the plausible intention that the text creator had.

55

Some issues

- Problems in identifying relations
  - Judgments are plausibility judgments. Two analysts might differ in their analyses
- Definitions of units
  - Vary from researcher to researcher, depending on the level of granularity needed
- Relations inventory
  - Many available
  - Each researcher tends to create their own, but large ones tend to be unmanageable

56

Applications

- Writing research (more later!)
  - How are coherent texts created
  - RST as a training tool to write effective texts
- Natural Language Generation
  - Input: communicative goals and semantic representation
  - Output: text
- Rhetorical/discourse parsing (more later!)
  - Rendering of a text in terms of rhetorical relations
  - Using signals, mostly discourse markers
- Corpus analysis
  - Annotation of text with discourse relations (Carlson et al. 2002)
  - Application to spoken language (Taboada 2004; references in Taboada and Mann 2006)
- Relationship to other discourse phenomena
  - Between nuclei and co-reference
- For more applications (up to 2005 or so):
  - Taboada, Maite and William C. Mann. (2006). Applications of Rhetorical Structure Theory. *Discourse Studies, 8* (4), 567-588.

57

RST for Argument Representation? [Peldszus and Stede, 2013]

- Nucleus-satellite distinction is crucial [Azar, 1999]
  - 5 RST relations have the 2 roles needed in an argumentative relationship (segments for conclusion & its argument)
    - Motivation, Antithesis, Concession, Evidence, Justify
- Hybrid approach [Green, 2010]
  - RST relations plus other annotations
    - E.g., Toulmin roles, argumentation schemes
- While there are parallels between presentational relations and argumentative moves, there are also limitations of "pure" RST

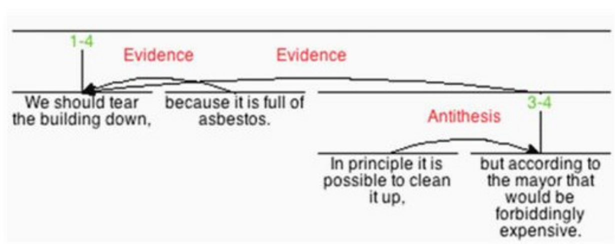
Exercise: Annotating Tree-Like Structures

- *We should tear the building down, because it is full of asbestos. In principle it is possible to clean it up, but according to the mayor that would be forbiddingly expensive*.

    – Constructed text from [Peldszus and Stede, 2013]

Answer: RST Analysis [Peldszus & Stede 2013]

Figure 9. RST analysis for a short text



- Curved lines connect satellite to nucleus
    - arrowhead points to nucleus, also indicated by vertical line
- Horizontal lines demarcate larger segments
    - fused from smaller segments

Other Approaches: Joint Modeling

- The assumption that discourse has multiple structures that should be modelled jointly goes back to Grosz & Sidner [1986], who proposed three inter-connected structures for discourse
- Example: Joint Functional Segmentation & Pred-Arg modelling
  - The ART/CoreSC corpus contains fine-grained (sentence-level) functional annotation of core components of scientific investigations.
  - www.aber.ac.uk/en/cs/research/cb/projects/art/art-corpus/

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Computational Discourse: Summing Up

- Current computational models of discourse structure are tied, more or less, to empirical data
  - tree-like discourse structures *(RST in this tutorial)*
  - conventionalized functions *(genre examples in this tutorial)*
  - predicate-argument structures *(PDTB in this tutorial)*
  - abstract topics
  - entity mentions
- Since natural coherent discourse involves them all, joint modelling may lead to better understanding/models
- As for modelling blog posts, tweets, . . . ???
- As for modelling argumentation directly . . . ???

Slide modified from "Discourse Structures and Language Technology" by Bonnie Webber, 2011

Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - Tree-like structures
- **Resources**
  - **Corpora**
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- Looking Forward

The Need for Annotated Corpora

- Required for supervised machine learning
  – Reliability / inter-rater agreement also desirable

Corpora with Discourse Annotations

- Penn Discourse Treebank (PDTB) [Prasad et al. 2008]
  - http://www.seas.upenn.edu/~pdtb/
    - 1 million words of Wall Street Journal (aligned with Penn Treebank)
    - Annotated for discourse relations and their two arguments
    - Distribution of relations in PDTB-2.0

PDTB Relations	No. of tokens
Explicit	18459
Implicit	16224
AltLex	624
EntRel	5210
NoRel	254
Total	40600

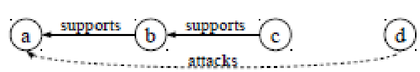
    - Distribution of sense tags (highest level of hierarchy)

"CLASS"	Explicit (18459)	Implicit (16224)	AltLex (624)	Total
"TEMPORAL"	3612	950	88	4650
"CONTINGENCY"	3581	4185	276	8042
"COMPARISON"	5516	2832	46	8394
"EXPANSION"	6424	8861	221	15506
Total	19133	16828	634	36592

- Similar resources are being created for languages other than English
- Only a subset of generic discourse relations are relevant for argumentation, and those that are relevant are very general

Relevance for Argumentation?
[Stab et al. 2014]

"Everybody should study abroad$_a$. It's an irreplaceable experience if you learn standing on your own feet$_b$ ***since*** *you learn living without depending on anyone else$_c$.* ***But*** *one who is living overseas will of course struggle with loneliness, living away from family and friends$_d$."*



One claim *(a)* and three premises:
- (Implicit justify) relates argument components *a* and *b*
- **since** (explicit CAUSE) relates argument components *b* and *c*
- **but** (explicit CONTRAST) relate argument components *c* and *d*

Corpora with Discourse Annotations

- RST Discourse Treeback [Carlson et al., 2002, 2003]
  - https://catalog.ldc.upenn.edu/LDC2002T07
    - 385 Wall Street Journal articles from the Penn Treebank
  - While RST analysis can be a useful first step, prior argumentation work has typically needed to supplement the RST annotations
  - Other mismatches (e.g. RST only considers adjacent relationships)

Argumentation Text/Dialog Corpora

- AIFdb/AraucariaDB Corpus [Reed, 2005; Budzynska et al. 2014]
  - http://www.arg-tech.org/index.php/projects/araucariadb/
  - http://www.arg.dundee.ac.uk/aif-corpora/

+ Several text genres, used by other research projects
- Reduced rather than authentic text

Argumentative Writing Corpus

- Argument Annotated Essays [Stab & Gurevych, 2014]
  - 90 persuasive essays with annotations of argument *components* and argumentative *relations*

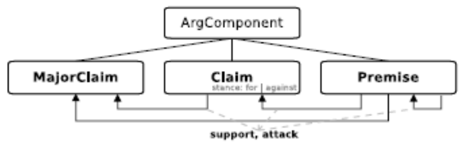


Figure 1: Argument annotation scheme including argument components and argumentative relations indicated by arrows below the components.

- Data and Annotation Guidelines
  - https://www.ukp.tu-darmstadt.de/data/argumentation-mining/argument-annotated-essays/

Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - Tree-like structures
- Resources
  - Corpora
  - **Software**
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- Looking Forward

Software Resources

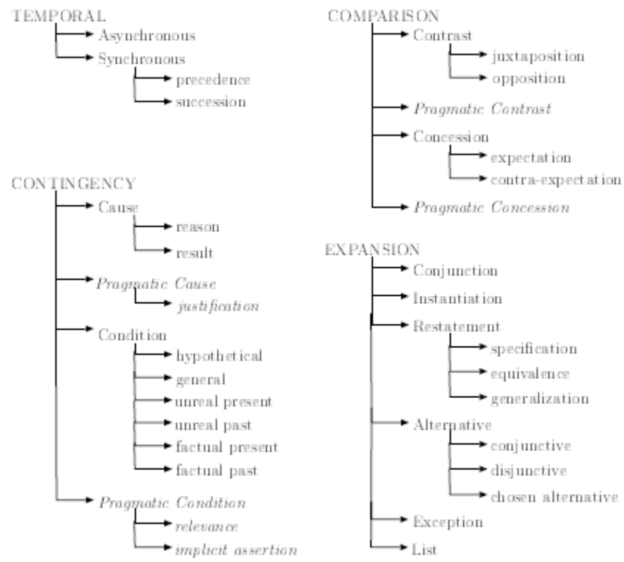
- Many general NLP resources (NLTK, Stanford, etc.)
  - Potential features for detecting implicit relations
    - e.g. can try to approximate "world knowledge" using lexical and structural features
- Discourse-specific resources
  - Typically trained on Wall Street Journal

Discourse Connectives Tagger

- http://www.cis.upenn.edu/~epitler/discourse.html
- Identifies explicit discourse connectives and senses (Expansion, Contingency, Comparison, Temporal) [Pitler & Nenkova, 2009]
  - Input: the output of automatic parsers or gold-standard parses
  - Output: syntactic trees augmented with tags indicating discourse connectives
  - Disambiguates discourse and non-discourse usages:
    - John likes to run marathons, *and* ran 10 last year alone. (Expansion)
    - My favorite colors are blue *and* green. (Non-discourse)
  - Disambiguates (PDTB) discourse senses:
    - They have not spoken to each other *since* they saw each other last fall. (Temporal)
    - I assumed you were not coming *since* you never replied to the invitation. (Contingency/Causal)

+ Genre-neutral applicability
- Not all discourse (argument) relations are explicitly marked

Hierarchy of Sense Tags



Human (h) vs. Discourse Connectives Tagger (d)

1h: Some have raised their cash positions to record levels. Implicit=BECAUSE **(causal/contingency)** High cash positions help buffer a fund when the market falls.

1d: Some have raised their cash positions to record levels. High cash positions help buffer a fund when **(temporal)** the market falls. WRONG

2h*:* Jacobs is an international engineering and construction concern. **NoRel** Total capital investment at the site could be as much as $400 million, according to Intel.

2d: Jacobs is an international engineering and construction concern. Total capital investment at the site could be as much as $400 million, according to Intel. CORRECT

3h: It was a far safer deal for lenders since **(causal/contingency)** NWA had a healthier cash flow and more collateral on hand.

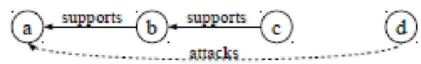
3d: It was a far safer deal for lenders since **(contingency)** NWA had a healthier cash flow and more collateral on hand. CORRECT

4h: Domestic car sales have plunged 19% since **(causal/contingency and temporal)** the Big Three ended many of their programs Sept. 30.

4d: Domestic car sales have plunged 19% since **(contingency)** the Big Three ended many of their programs Sept. 30. PARTIALLY CORRECT

Relevance for Argumentation? [Stab et al. 2014]

*"Everybody should study abroad$_a$. It's an irreplaceable experience if you learn standing on your own feet$_b$ **since** you learn living without depending on anyone else$_c$. **But** one who is living overseas will of course struggle with loneliness, living away from family and friends$_d$."*



One claim *(a)* and three premises:

- (Implicit justify) relates argument components *a* and *b* MISSING
- **since** (explicit CAUSE/CONTIGENCY) relates argument components *b* and *c*
- **but** (explicit CONTRAST/COMPARISON) relate argument components *c* and *d*

Tagger output:

- **if** (Contingency) EXTRA
- **since** (Contingency)
- **but** (Comparison)

A PDTB-Styled Discourse Parser

- http://wing.comp.nus.edu.sg/~linzihen/parser/
- Sequential pipeline includes a connective classifier, argument labeler, explicit classifier, non-explicit classifier, and attribution span labeler [Lin et al., 2014]

Human vs. PDTB Parser

<u>connective,</u> *Arg1*, **Arg2**

3h. *It was a far safer deal for lenders* <u>since</u> **NWA had a healthier cash flow and more collateral on hand.** (causal)

3d. It was *a far safer deal for lenders* <u>since</u> **NWA had a healthier cash flow and more collateral on hand.** (asynchronous/temporal) arg 1 span and connective sense are wrong

4h. *Domestic car sales have plunged 19%* <u>since</u> **the Big Three ended many of their programs Sept. 30.** (causal and temporal)

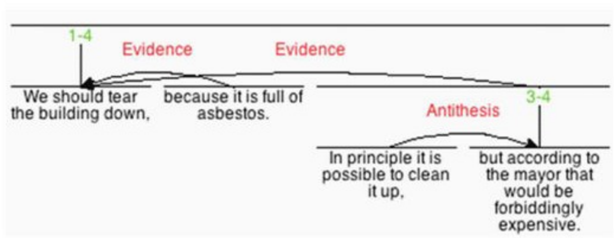
4d. *Domestic car sales have plunged 19%* <u>since</u> **the Big Three ended many of their programs Sept. 30.** (asynchronous/temporal) missing connective sense

A RST-Style Discourse Parser

- http://www.cs.toronto.edu/~weifeng/software.html
- [Feng and Hirst, 2012]

Human vs. Computer RST Analysis

Figure 9. RST analysis for a short text



Computer postulates 5 rather than 4 elementary discourse units, due to:

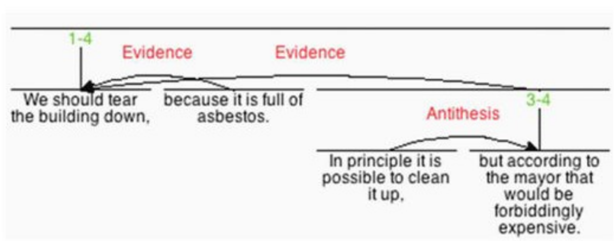
- *In principle*
- *it is possible to clean it up,*

Differences in relation identification too:

- Explanation (includes Evidence), rather than first Evidence
- Elaboration, rather than second Evidence (and nucleus span is bigger)
- Contrast rather than Antithesis

Human RST vs. Discourse Connectives Tagger

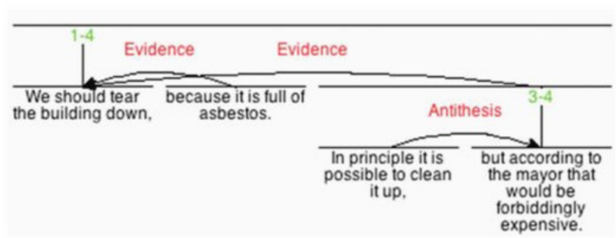
Figure 9. RST analysis for a short text



We should tear the building down, because (contingency) it is full of asbestos. In principle it is possible to clean it up, but (comparison) according to the mayor that would be forbiddingly expensive.

Human RST vs. PDTB Parser

Figure 9. RST analysis for a short text



- *We should tear the building down*, because (cause/contingency) **it is full of asbestos**.
- *We should tear the building down, because it is full of asbestos*. Implicit (cause/contingency) **In principle it is possible to clean it up, but according to the mayor that would be forbiddingly expensive.**

Discourse Connectives vs. PDTB Parser

Discourse Connectives Tagger

- We should tear the building down, because (contingency) it is full of asbestos. In principle it is possible to clean it up, but (comparison) according to the mayor that would be forbiddingly expensive.

***PDTB Discourse Parser:* connective, *Arg1*, Arg2**

- *We should tear the building down*, because (cause/contingency) **it is full of asbestos**.
- *We should tear the building down, because it is full of asbestos*. Implicit (cause/contingency) **In principle it is possible to clean it up, but according to the mayor that would be forbiddingly expensive.**

Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - Tree-like structures
- Resources
  - Corpora
  - Software
- **Evaluation**
  - **Intrinsic versus Extrinsic**
  - **Quantitative Metrics**
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- Looking Forward

Evaluation

- Standard data-driven method
  - Train model on a **training set**
  - Look at the model's performance on some new data
    - This is exactly what happens in the real world; we want to know how our model performs on data we haven't seen
  - So use a **test set**. A dataset which is different than our training set, but is drawn from the same source
    - **Cross-validation** when training + testing corpora are small
  - Then we need an **evaluation metric** to tell us how well our model is doing on the test set.

Quantifying Classification Performance

Confusion Matrix

	Claim (actual)	Other (actual)
Claim (predicted)	True positive (tp)	False positive (fp)
Other (predicted)	False negative (fn)	True negative (tn)

Accuracy:
- $(tp+tn)/(tp+tn+fp+fn)$

Precision: % of predicted labels that are correct
- $tp/(tp+fp)$

Recall: % of actual labels that are predicted
- $tp/(tp+fn)$

F-Measure (F_1 score is balanced): Harmonic mean
- 2 * $(precision * recall)/(precision+recall)$

Intrinsic (in vivo) Evaluation

- The predicted values are compared with a manually coded "Gold Standard" (the actual values)
  – They may also be compared with the predictions from a baseline model (e.g. majority class prediction)
- 100% is impossible even for human annotators

Inter-Annotator Reliability

- Kappa
  - Observed accuracy (actual agreement) corrected for chance (expected agreement)
  - (P(A) – P(E)) / (1 – P(E))
  - Online calculators
    - E.g., http://vassarstats.net/kappa.html

Extrinsic (in vitro) Evaluation

- Incorporate the prediction model in another system, and evaluate that system's performance

Exercise: Intrinsic Evaluation

	Claim (actual)	Other (actual)
Claim (predicted)	29	4
Other (predicted)	41	32

- Model: Accuracy
- Baseline : Accuracy

- Claim: Precision, Recall
- Other: Precision, Recall
- Unweighted Avg. Precision, Recall

Answer: Instrinsic Evaluation

	Claim (actual)	Other (actual)
Claim (predicted)	29	4
Other (predicted)	41	32

- Model: Accuracy = .58, Kappa = .24
- Baseline : Accuracy = .66, Kappa = 0

- Claim: Precision = .88, Recall =. 41, F = .56
- Other: Precision = .44, Recall = .89, F = .59
- Unweighted Avg. Precision = .66, Recall = .65, F = .58

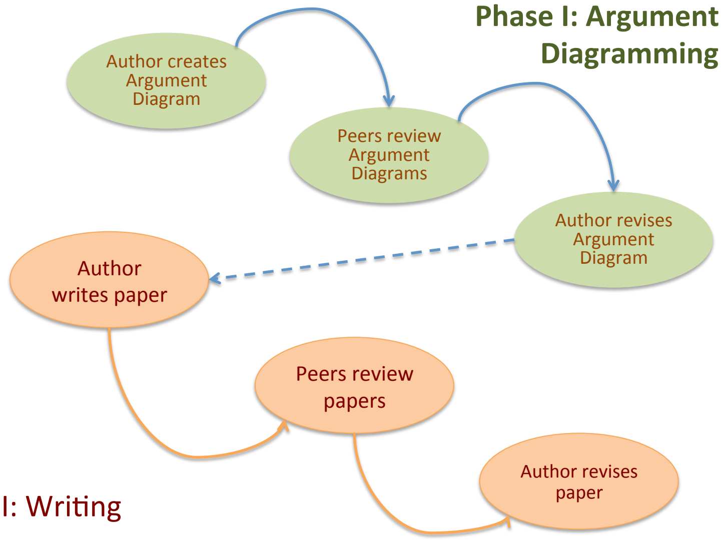
Error Analysis

- Look at the confusion matrix
- See what errors are causing problems

Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - Tree-like structures
- Resources
  - Corpora
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- **Educational Case Studies**
  - **Teaching Writing with Diagramming and Peer Review**
  - Automated Writing Assessment
- Looking Forward

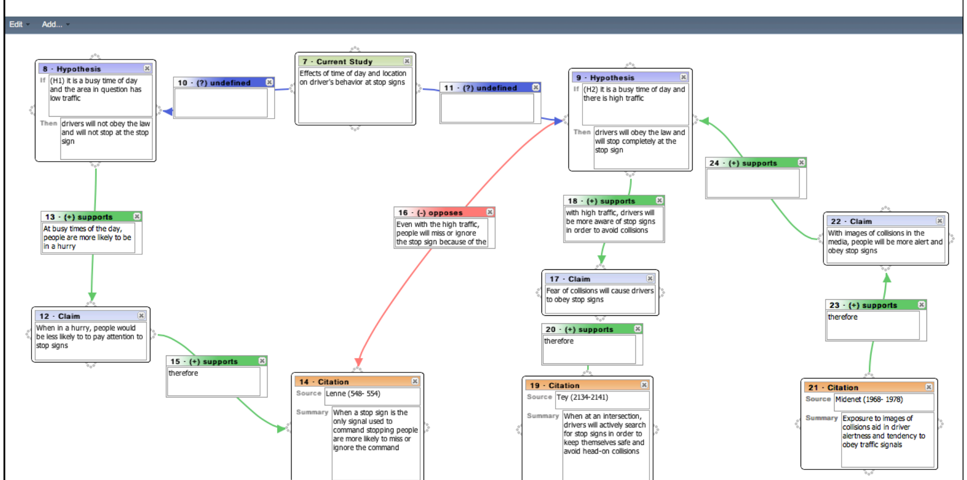
ArgumentPeer Project



Joint work with Kevin Ashley and Chris Schunn

Example Student Argument Diagram

(input via the LASAD system [Pinkwart et al.])



Ontology-Based Argument Mining and Automatic Essay Scoring
[Ong, Litman, Brusilovsky, 2014]

- System recognizes diagram ontology in essays
- System scores essays using recognized ontology

Ontology-Based Argument Mining and Automatic Essay Scoring
[Ong, Litman, Brusilovsky, 2014]

- System recognizes diagram ontology in essays
- System scores essays using recognized ontology

- How can we formalize this research in terms of argument mining subtasks?
  - Segmentation
  - Segment classification
  - Relation identification
  - Argument completion

Ontology-Based Argument Mining and Automatic Essay Scoring
[Ong, Litman, Brusilovsky, 2014]

- System recognizes diagram ontology in essays
- System scores essays using recognized ontology

- What corpus resources do we need?
- How do we evaluate them?

Ontology-Based Argument Mining and Automatic Essay Scoring
[Ong, Litman, Brusilovsky, 2014]

- System recognizes diagram ontology in essays
- System scores essays using recognized ontology

- What NLP theories/tools could we use?
- How does our data or application add constraints?

Ontology-Based Argument Mining and Automatic Essay Scoring

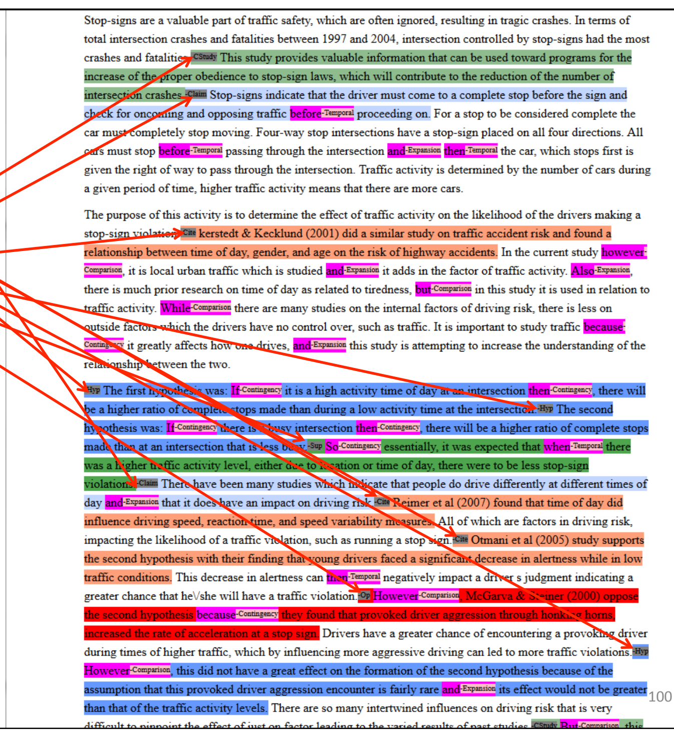
[Ong, Litman, Brusilovsky, 2014]

- System recognizes diagram ontology in essays
- System scores essays using recognized ontology

- How can we evaluate argument mining technology?
  - Intrinsic
  - Extrinsic

Our Argument Mining I/O

- Current Study
- Claim
- Citation
- Hypothesis
- Supports
- Opposes



Stop-signs are a valuable part of traffic safety, which are often ignored, resulting in tragic crashes. In terms of total intersection crashes and fatalities between 1997 and 2004, intersection controlled by stop-signs had the most crashes and fatalities. CStudy This study provides valuable information that can be used toward programs for the increase of the proper obedience to stop-sign laws, which will contribute to the reduction of the number of intersection crashes. -Claim Stop-signs indicate that the driver must come to a complete stop before the sign and check for oncoming and opposing traffic before-Temporal proceeding on. For a stop to be considered complete the car must completely stop moving. Four-way stop intersections have a stop-sign placed on all four directions. All cars must stop before-Temporal passing through the intersection and-Expansion then-Temporal the car, which stops first is given the right of way to pass through the intersection. Traffic activity is determined by the number of cars during a given period of time, higher traffic activity means that there are more cars.

The purpose of this activity is to determine the effect of traffic activity on the likelihood of the drivers making a stop-sign violation. Cite kerstedt & Kecklund (2001) did a similar study on traffic accident risk and found a relationship between time of day, gender, and age on the risk of highway accidents. In the current study however-Comparison, it is local urban traffic which is studied and-Expansion it adds in the factor of traffic activity. Also-Expansion, there is much prior research on time of day as related to tiredness, but-Comparison in this study it is used in relation to traffic activity. While-Comparison there are many studies on the internal factors of driving risk, there is less on outside factors which the drivers have no control over, such as traffic. It is important to study traffic because-Contingency it greatly affects how one drives, and-Expansion this study is attempting to increase the understanding of the relationship between the two.

Hyp The first hypothesis was: If-Contingency it is a high activity time of day at an intersection then-Contingency, there will be a higher ratio of complete stops made than during a low activity time at the intersection. -Hyp The second hypothesis was: If-Contingency there is a busy intersection then-Contingency, there will be a higher ratio of complete stops made than at an intersection that is less busy. -Sup So-Contingency essentially, it was expected that when-Temporal there was a higher traffic activity level, either due to location or time of day, there were to be less stop-sign violations. -Claim There have been many studies which indicate that people do drive differently at different times of day and-Expansion that it does have an impact on driving risk. -Cite Reimer et al (2007) found that time of day did influence driving speed, reaction time, and speed variability measures. All of which are factors in driving risk, impacting the likelihood of a traffic violation, such as running a stop sign. -Cite Otmani et al (2005) study supports the second hypothesis with their finding that young drivers faced a significant decrease in alertness while in low traffic conditions. This decrease in alertness can then-Temporal negatively impact a driver s judgment indicating a greater chance that he\/she will have a traffic violation. -Op However-Comparison, McGarva & Steiner (2000) oppose the second hypothesis because-Contingency they found that provoked driver aggression through honking horns, increased the rate of acceleration at a stop sign. Drivers have a greater chance of encountering a provoking driver during times of higher traffic, which by influencing more aggressive driving can led to more traffic violations. -Hyp However-Comparison, this did not have a great effect on the formation of the second hypothesis because of the assumption that this provoked driver aggression encounter is fairly rare and-Expansion its effect would not be greater than that of the traffic activity levels. There are so many intertwined influences on driving risk that is very difficult to pinpoint the effect of just on factor leading to the varied results of past studies. -CStudy But-Comparison, this

100

Essay Processing Pipeline

1. Discourse Processing
   – Tag essays with discourse connectives software

2. Argument Ontology Mining
   – Tag essays with diagram ontology elements
      • Rule-based algorithm

3. Ontology-Based Scoring
   – Use the mined argument to score the essays
      • Rule-based algorithm

Extrinsic Evaluation

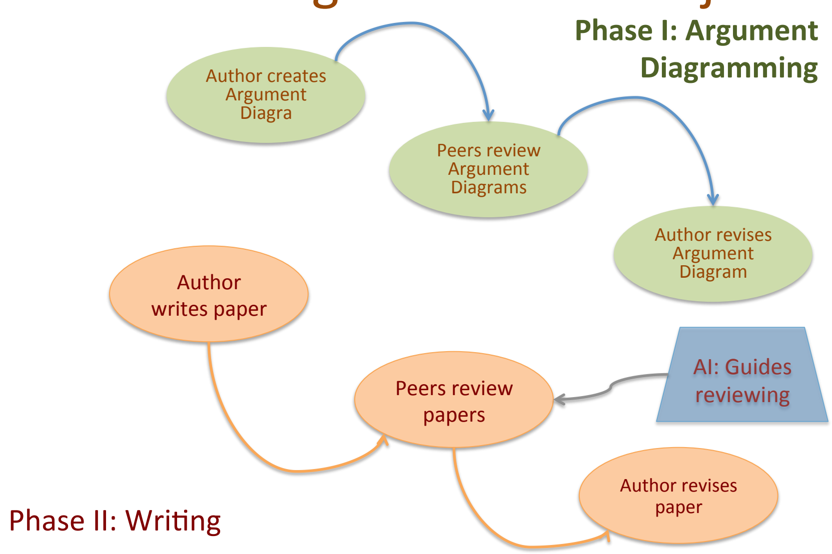
• Do automatically generated scores correlate with expert scores ?
   – Yes, for ranking: number of automatically generated tags for diagram elements are positively correlated with expert score

102

Current Argument Mining Directions

- More discourse (e.g. PDTB, RST parsers)
- Intrinsic evaluation
  - 1533 annotated sentences from 60 essays

ArgumentPeer Project



Joint work with Kevin Ashley and Chris Schunn

SWoRD: A web-based peer review system [Cho & Schunn, 2007]

- Authors submit papers (or diagrams)
- Peers submit reviews
  - *Problem:* no discussion of thesis statements
  - *Our Approach:* detect and scaffold

Identifying Thesis and Conclusion Statements in Student Essays to Scaffold Peer Review [Falakmasir, Ashley, Schunn, Litman 2014]

- Can natural language processing (NLP) detect presence/absence of thesis and conclusion sentences in essays in order to guide peer review?

Identifying Thesis and Conclusion Statements in Student Essays to Scaffold Peer Review [Falakmasir, Ashley, Schunn, Litman 2014]

- Can natural language processing (NLP) detect presence/absence of thesis and conclusion sentences in essays in order to guide peer review?
- How can we formalize this research in terms of argument mining subtasks?
  - Segmentation
  - Segment classification
  - Relation identification
  - Argument completion

Identifying Thesis and Conclusion Statements in Student Essays to Scaffold Peer Review [Falakmasir, Ashley, Schunn, Litman 2014]

- Can natural language processing (NLP) detect presence/absence of thesis and conclusion sentences in essays in order to guide peer review?
- How can we formalize this research in terms of argument mining subtasks?
  - Segmentation: sentences
  - Segment classification: probability of being thesis…
  - Relation identification
  - Argument completion

Identifying Thesis and Conclusion Statements in Student Essays to Scaffold Peer Review [Falakmasir, Ashley, Schunn, Litman 2014]

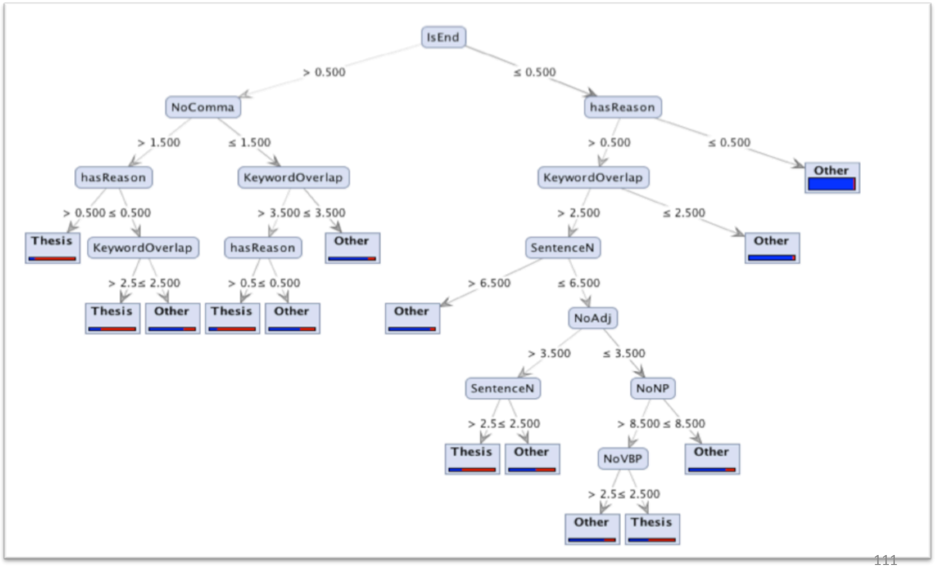
- Can natural language processing (NLP) detect presence/absence of thesis and conclusion sentences in essays in order to guide peer review?
- What corpus resources do we need?
- How do we evaluate them?
- What NLP/theories tools could we use?
- How does our data or application add constraints?
- How can we do intrinsic or extrinsic evaluations?

Features for Machine Learning

- 3 feature sets inspired by [Burstein et al. 2003]
  - **Positional:** paragraph number, sentence number in the paragraph, type of paragraph (first, body, last)
  - **Sentence Level**: syntactic, semantic, frequent words
  - **Essay Level**: number of keywords among the most frequent words of the essay, number of words overlapping with the assignment prompt, and a sentence importance score based on Rhetorical Structure Theory

110

Example Thesis Prediction Model



111

Intrinsic Evaluation

- Unseen test set

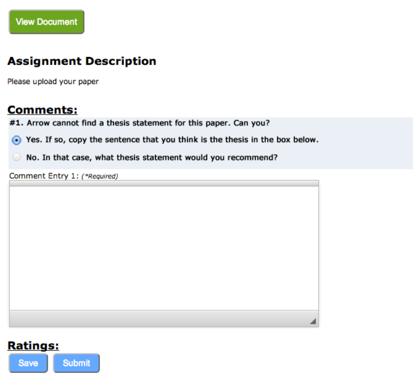
	Thesis			Conclusion			Essay		
Classifier	P	R	F	P	R	F	P	R	F
Positional Baseline	0.58	0.88	0.57	0.58	0.84	0.55	0.58	0.84	0.55
Naïve Bayes	0.70	0.79	0.74	0.65	0.69	0.67	0.63	0.65	0.64
Decision Tree	0.82	0.84	0.83	0.49	0.75	0.59	0.75	0.73	0.74
SVM	0.82	0.65	0.72	0.60	0.54	0.56	0.62	0.58	0.60

- Even with a small training corpus, NLP features can classify core argument segments of essays better than the baseline

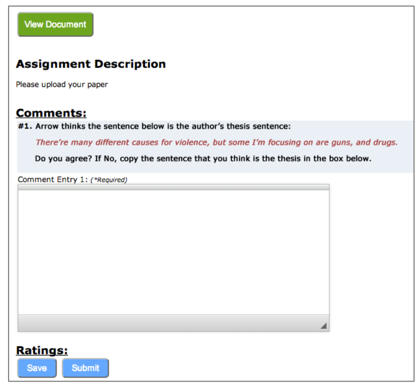
112

Extrinsic Evaluation: Impact of Scaffolding During Peer Review

When argument mining cannot find a thesis sentence, the first peer review prompt is:



When argument mining identifies a thesis sentence, the first peer review prompt is:



Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - Tree-like structures
- Resources
  - Corpora
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - **Automated Writing Assessment**
- Looking Forward

Automatic Scoring of an Analytical Response-To-Text Assessment (RTA)

[Rahimi, Litman, Correnti, Matsumura, Wang & Kisa, 2014]

- Writing assessment via interpretable features that operationalize the ***Evidence*** rubric of RTA

115

Rubric for the Evidence dimension of RTA

1	2	3	4
Features one or no pieces of evidence	Features at least 2 pieces of evidence	Features at least 3 pieces of evidence	Features at least 3 pieces of evidence
Selects inappropriate or little evidence from the text; may have serious factual errors and omissions	Selects some appropriate but general evidence from the text; may contain a factual error or omission	Selects appropriate and concrete, specific evidence from the text	Selects detailed, precise, and significant evidence from the text
Demonstrates little or no development or use of selected evidence	Demonstrates limited development or use of selected evidence	Demonstrates use of selected details from the text to support key idea	Demonstrates integral use of selected details from the text to support and extend key idea
Summarize entire text or copies heavily from text	Evidence provided may be listed in a sentence, not expanded upon	Attempts to elaborate upon Evidence	Evidence must be used to support key idea / inference(s)

Automatic Scoring of an Analytical Response-To-Text Assessment (RTA)

[Rahimi, Litman, Correnti, Matsumura, Wang & Kisa, 2014]

- Writing assessment via interpretable features that operationalize the ***Evidence*** rubric of RTA
  - Problem sormalization in terms of argument mining subtasks?
  - What corpus resources do we need?
  - How do we evaluate them?
  - What NLP theories/tools could we use?
  - How does our data or application add constraints?
  - How can we do intrinsic or extrinsic evaluations?

117

Prompt and Good Essay

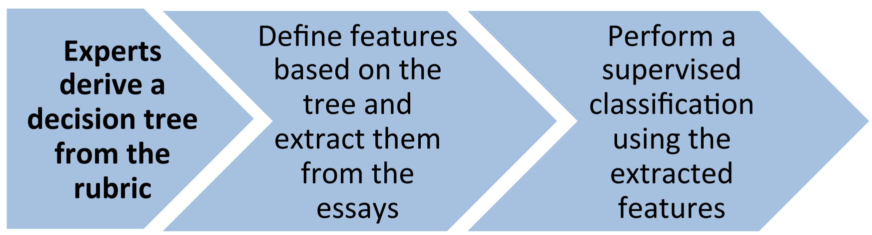
Prompt: Did the author provide a convincing argument that winning the fight against poverty is achievable in our lifetime? Explain why or why not with 3-4 examples from the text to support your answer.

Example Essay, Score of 4 on Evidence dimension:
I was convinced that winning the fight of poverty is achievable in our lifetime. Many people couldn't afford medicine or bed nets to be treated for malaria . Many children had died from this dieseuse even though it could be treated easily. But now, bed nets are used in every sleeping site . And the medicine is free of charge. Another example is that the farmers' crops are dying because they could not afford the nessacary fertilizer and irrigation . But they are now, making progess. Farmers now have fertilizer and water to give to the crops. Also with seeds and the proper tools . Third, kids in Sauri were not well educated. Many families couldn't afford school . Even at school there was no lunch . Students were exhausted from each day of school. Now, school is free . Children excited to learn now can and they do have midday meals . Finally, Sauri is making great progress. If they keep it up that city will no longer be in poverty. Then the Millennium Village project can move on to help other countries in need.

- Word windows containing evidence are highlighted

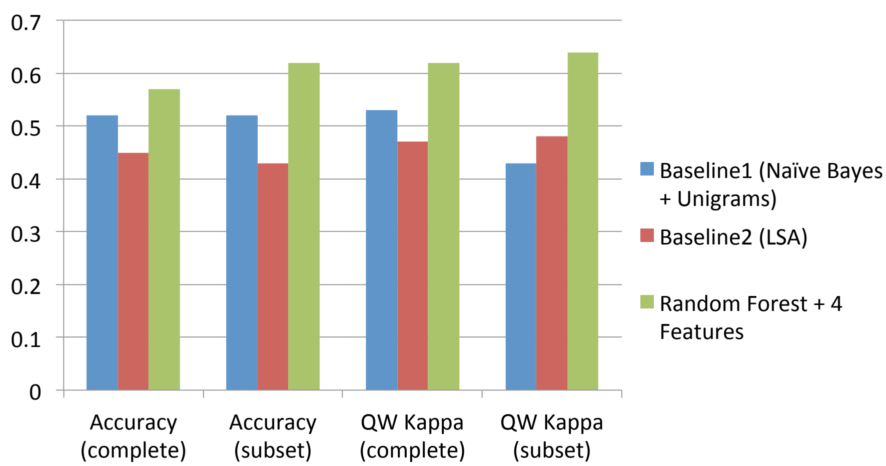
11

Automatic Scoring Approach



15

Extrinsic Evaluation



- Proposed features outperform both baselines
- Absolute performance **improves** on less noisy data

Outline

- Argument Mining (from Text)
- Computational Discourse
  - Functional structures
  - Predicate-argument structures
  - Tree-like structures
- Resources
  - Corpora
  - Software
- Evaluation
  - Intrinsic versus Extrinsic
  - Quantitative Metrics
- Educational Case Studies
  - Teaching Writing with Diagramming and Peer Review
  - Automated Writing Assessment
- **Looking Forward**

Open Challenges

- Annotation Schemes and Annotated Corpora
- Enabling NLP and Discourse Technology
- Human Annotation versus Computer Analysis
  – Genre-Neutral vs. Genre-Specific Features & Methods
  – Fine-Grained vs. Coarse-Grained Annotations

References

- Azar, M. (1999). Argumentative text as rhetorical structure: An application of rhetorical structure theory. *Argumentation, 13*, 97-114.
- Budzynska, K., Janier, M. , Reed, C., Saint-Dizier, P., Stede, M. & Yakorska, O. (2014). A model for processing illocutionary structures and argumentation in debates. In: *Proc. of LREC*, Reykavik.
- Burstein, J., Marcu, D. & Knight, K. (2003). Finding the WRITE stuff: Automatic identification of discourse structure in student essays. *IEEE Intelligent Systems: Special Issue on Advances in NLP,* 18:32–39.
- Carlson, L., Marcu, D. & Okurowski, M.E. (2002). RST discourse treebank, LDC2002T07 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.
- Carlson, L., Marcu, D., & Okurowski, M. E. (2003). *Building a discourse-tagged corpus in the framework of rhetorical structure theory* (pp. 85-112). Springer Netherlands.
- Cho, K. & Schunn, C.D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education*, 48.
- Falakmasir, M. H, Ashley, K. D, Schunn, C. D. & Litman, D. J. (2014). Identifying thesis and conclusion statements in student essays to scaffold peer review. In *Proceedings 12th International Conference on Intelligent Tutoring Systems,* Honolulu, HI.
- Feng, V. W., & Hirst, G. (2011). Classifying arguments by scheme. In *Proceedings 49th ACL*, Portland, OR, 987-996.
- Feng, V. W., & Hirst, G. (2012). Text-level discourse parsing with rich linguistic features. In *Proceedings 50th ACL*, Jeju, Korea, 60-68.
- Green, N. L. (2010). Representation of argumentation in text with rhetorical structure theory. *Argumentation, 24*, 181-196.
- Grosz, B. J. Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics* 12, 3, 175-204.
- Hirohata, K., Okazaki, N., Ananiadou, S. & Ishizuka, M. (2008). Identifying sections in scientific abstracts using conditional random fields. In *Proc. 3rd Int'l Joint Conf. on Natural Language Processing*, 381–388.
- Lee, A., Prasad, R., Joshi, A., Dinesh, N. & Webber, B. (2006). Complexity of dependencies in discourse. In *Proc. 5th Int'l Workshop on Treebanks and Linguistic Theories*, Prague.
- Lin, Z., Ng, H. T., & Kan, M. (2014). A PDTB-Styled End-to-End Discourse Parser. *Natural Language Engineering*, 20, 51-184.
- Madnani, N., Heilman, M., Tetrault, J. & Chodorow, M. (2012). Identifying high-level organizational elements in argumentative discourse. In *Proceedings NAACL-HLT,* 20–28, Montreal, Quebec, Canada.
- Mann, W. C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text, 8* (3), 243-281.
- Ong, N., Litman, D & Brusilovsky, A. (2014). Ontology-based argument mining and automatic essay scoring. In *Proceedings of the 1st ACL Workshop on Argumentation Mining,* Baltimore, MD.
- Oza, U., Prasad, R., Kolachina, S., Sharma, D. M., & Joshi, A. (2009). The Hindi Discourse Relation Bank. In *Proc. 3rd Linguistic Annotation Workshop*, Singapore.

References

- Palau, M . & Moens, M. F. (2009). Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the ICAIL*, Barcelona, Spain, 98-109.
- Peldszus, A. & Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence*, *7(1)*, 1-31.
- Pinkwart, N., Ashley, K., Aleven, V. & Lynch, C. (2008). Graph Grammars: an ITS Technology for Diagram Representations. In *Proceedings of The 21st International FLAIRS Conference, Special Track on Intelligent Tutoring Systems*, 433-438.
- Pitler, E. & Nenkova, A. (2009). Using Syntax to Disambiguate Explicit Discourse Connectives in Text. In *Proceedings of ACL*.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. & Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proc. 6th Int'l Conference on Language Resources and Evaluation,* Marrakech, Morocco.
- Rahimi, Z., Litman, D., Correnti, R, Matsumura, L. C, Wang, E. & Kisa, Z. (2014). Automatic scoring of an analytical Response-To-Text assessment, In *Proceedings 12th International Conference on Intelligent Tutoring Systems*, 601-610, Honolulu, HI.
- Reed, C. (2005). Preliminary Results from an Argument Corpus. In *Proceedings of the IX Symposium on Social Communication*, Santiago de Cuba, pp5 76-580.
- Saint-Dizier, P. (2012). Processing natural language arguments with the TextCoop platform. *Journal of Argumentation and Computation, 3*(1), 49-82.
- Stab, C. & Gurevych, I. (2014). Annotating argument components and relations in persuasive essays. In *Proceedings of the the 25th International Conference on Computational Linguistics*, Dublin.
- Stab, C. & Gurevych, I. (2014, to appear) Identifying Argumentative Discourse Structures in Persuasive Essays. In *Proceedings Conference on Empirical Methods in Natural Language Processing.*
- Taboada, M. (2004). *Building Coherence and Cohesion: Task-Oriented Dialogue in English and Spanish*. Amsterdam and Philadelphia: John Benjamins.
- Taboada, M. & Mann, W. C. (2006a). Applications of Rhetorical Structure Theory. *Discourse Studies, 8* (4), 567-588.
- Taboada, M. & Mann, W. C. (2006b). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies, 8* (3), 423-459.
- Teufel, S. & Moens, M. (2002). Summarizing scientific articles – experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4), 409-445.

Thank You!

- Final Questions?

- Further Information
  - http://www.cs.pitt.edu/~litman