Argumentation and Linguistics

Adam Wyner

Department of Computing Science
University of Aberdeen, Scotland

July 1, 2013
Argumentation in Artificial Intelligence
King's College London, London, UK
Outline

• Introduction.
• Context with respect to Argumentation Theory.
• Language Issues, methodology, computational linguistics.
• Tool introduction.
• Examples.
• Parsing and semantic representation.
• Other approaches.
Where Arguments Appear

- Consumer websites.
- Law: policy making, Supreme Court transcripts, case based reasoning, regulations.
- BBC's *Have Your Say* and *Moral Maze*.
- Medical diagnosis.
- Making plans.
- Debatepedia, Wikipedia, meeting annotations, web-forums, ....
Argument Fragment for a Camera

⭐⭐⭐⭐⭐ Stunning camera, 27 Aug 2011
By A. Reader
This review is from: Canon PowerShot SX220 HS Digital Camera - Grey (12.1MP, 14x Optical Zoom) 3.0 inch LCD (Electronics)
I adore this camera. Why? Because I get a higher percentage of good shots out of it than from any other camera I have, and that includes a few DSLRs, and the video is ridiculously good. Is the image quality as good as a DSLR or even a micro 4/3 system camera? No, but it's not far off, and a shot that's perfectly exposed and focused and free of camera shake on a smaller sensor like this beats not getting a picture at all by a long way. The SX220HS is my constant companion and lives in a pouch on my belt.

I've spent 30 years in photography, since buying my first Russian Zenit E around 1979, including being a semi-pro at one point. I know all about aperture/shutter speed/ISO ratios, exposure values, incident light metering, parallax error, Hurter & Driffield scales, flash guide numbers, the difference between depth of field and depth of focus, Ansel Adams's zone system, blah blah blah. When I'm taking a photo, I don't want to think about any of that stuff, I want to be enchanted by a scene and have a decent reproduction of it when I get home. With this camera, I get that. When I pull it out, I know I'm almost certain to get the picture I want. Canon have done an amazing job with it.

The video is astounding. When I upload it, I view it on a large screen, not a little laptop, so any flaws would be immediately apparent. I uploaded a few to Youtube for a friend to see, and she emailed me back, 'They are so clear!'
Pro and Con

Customer Reviews
Canon PowerShot SX220 HS Digital Camera - Grey (12.1MP, 14x Optical Zoom) 3.0 inch LCD

85 Reviews
5 star: (46)
4 star: (27)
3 star: (5)
2 star: (6)
1 star: (1)

Average Customer Review
(85 customer reviews)

Search Customer Reviews

Share your thoughts with other customers
Only search this product's reviews

Create your own review

The most helpful favourable review
Stunning camera
I adore this camera. Why? Because I get a higher percentage of good shots out of it than from any other camera I have, and that includes a few DSLRs, and the video is ridiculously good. Is the image quality as good as a DSLR or even a micro 4/3 system camera? No, but it’s not far off, and a shot that’s perfectly exposed and focused and free of camera shake on a smaller...

Read the full review>
Published 7 months ago by A. Reader

The most helpful critical review
Canon Warranty
The Powershot SX220 HS does take good pictures, but for me the Canon Warranty lets it down. I had the camera for about 6 weeks when I noticed the lens cover would not fully open or close, and because I’ve had the camera over 30 days amazon did not want to know, so I contacted canon support which was rubbish, the first e-mail I got back was telling me to clean the CF card...

Read the full review>
Published 5 months ago by Gmick
Comments on Comments

Comments
Track comments by e-mail
Tracked by 2 customers

Showing 1-7 of 7 posts in this discussion

Sort: Oldest first | Newest first

Initial post: 12 Nov 2011 23:32:50 GMT

Ian Beckett says:

hi,

you're not limited to 30 days under consumer legislation (6 weeks is still a "reasonable period") .. your issue is that you let them repair it which can always try and get you to do .. you should asked canon for a replacement or a refund ... contact canon and tell them you will give them one last chance to put it back to 100% condition otherwise you will be taking a small claims court action against them ... its really easy to do and costs about £17 and you can do it all online.

Reply to this post

4 of 4 people think this post adds to the discussion. Do you? Yes | No
- Question 9. Should the law be clarified with respect to whether the scanning of works held in libraries for the purpose of making their content searchable on the Internet goes beyond the scope of current exceptions to copyright?

- Yes.

- Not all the material digitised by publishers is scanned with OCR (Optical Character Recognition) with the purpose of making the resulting content searchable. If the rights holders will not do this, libraries should be able to offer this service. It would have a transformative effect on research, learning and teaching by opening up a mass of content to users which can be searched using search engines. The interests of copyright holders will not be harmed, because the resulting output will act as marketing material for their materials.
Context with Respect to Argumentation Theory
Locating the Problem and Engineering a Solution

- The knowledge acquisition bottleneck from NL to some formal representation.
- Relationship to other parts of the argumentation processing pipeline.
- Look at some other NL problems.
Three Stages Graph – Structured or Instantiated AFs

Knowledge Base → Argumentation Framework → Extensions of Arguments → Extensions of Conclusions

Step 1: construct arguments and attacks
Step 2: identify sets of accepted arguments
Step 3: identify sets of accepted conclusions

Three Stages - Caminada and Wu 2011

Knowledge Acquisition Bottleneck: time, labour, expertise to construct a KB at scale.
Logic-based Instantiated Argumentation
Besnard and Hunter

• An argument is an ordered pair \(<\psi, \alpha>\); \(\psi\) is a subset of a given KB and \(\alpha\) is an atomic proposition from the KB; \(\psi\) is a minimal set of formulae such that \(\psi\) implies \(\alpha\), and \(\psi\) does not imply a contradiction. \(\psi\) is said to support the claim \(\alpha\).

• Where \(p\) and \(q\) are atoms, and where the KB is comprised of \(p\) and \(p \rightarrow q\), then \(<\{p, p \rightarrow q\}, q>\) is an argument.

• We could have a KB from which we can form an argument which supports \(\neg q\), \(<\{p, p \rightarrow \neg q\}, \neg q>\). In addition and with respect to this argument, suppose we can form an undercutter \(<\{r, r \rightarrow \neg p\}, \neg p>\) and a rebuttal \(<\{r, r \rightarrow \neg p, \neg p \rightarrow q\}, q>\).

• KBs (even relatively small ones) generate lots of arguments and attack relationships which can be structured in a tree.
Abstract Argumentation

An abstract argument framework, as introduced by Dung, is a pair $AF = (A, attack)$, where $A$ is a set of arguments and $attack$ a binary relation on $A$. A subset $B$ of $A$ is said to be conflict-free if no argument in $B$ attacks another argument in $B$. $B$ is said to be admissible if: it is conflict-free; and it defends itself against any attack, i.e., suppose arguments $A_1$ and $A_3$ are in $B$, some argument $A_2$ is in $A$ but not in $B$, and $A_2$ attacks $A_1$, $A_3$ attacks $A_2$, but no argument attacks $A_3$. A preferred extension is then a maximal (with respect to set inclusion) admissible set. Several other types of extensions are defined

Preferred extension:
$\{a, c, d, h, i, k\}$
Zeroing In

- Source text
- Knowledge base & argumentation schemes
- Generated arguments (abstract or instantiated)
Current Tools to Extract and Structure Arguments from Text

• Rationale, Araucaria, Carneades (Gordon 2007), IMPACT Project, Legal Apprentice, ....

• All manual. No NLP.
Argumentation Schemes

• Patterns of presumptive (defeasible) reasoning (Walton 1996)
• Practical Reasoning with values:
  – Do action (transition) because:
    • Current circumstances - a list of literals.
    • Consequences – a list of literals.
    • Values (promoted, demoted, neutral wrt actions) – a list of terms.
• Credible Source:
  – Z is accepted because:
    • X is an expert in domain Y.
    • X stated literal Z
    • Z is about domain Y.
Proposal

• Normalise natural language source material into argumentation schemes.
• Formalise argumentation schemes in terms of roles of propositions in the scheme and internal structure of propositions (predicates and typed variables).
• Connect argumentation schemes to abstract arguments.
• Relate one scheme to another in terms of contrariness.
• Extract scheme relevant information from the source.
• Create a knowledge base to instantiate variables.
Caveat

- Low level automation, using high level structures as guides.
- For example, no automatic search for scheme filling, grounding of variables, contrast identification.
- Progress can be made on these (and for contrast identification, there is significant work already).
Language Issues
Problems with Language I

• Identification, implicit information, multiple forms with the same meaning, the same form with multiple meanings:
  
  • Entity ID: Jane Smith, for plaintiff.
  
  • Relation ID: Edgar Wilson disclosed the formula to Mary Hays.
  
  • Bill drove the car into Phil at 60 MPH. (agent, instrument, killing)
  
  • Jane Smith, Jane R. Smith, Smith, Attorney Smith....
  
  • Jane Smith in one case decision need not be the same Jane Smith in another case decision.
Problems with Language II

• Concepts, dispersed meanings, rules, diathesis:
  
  • Plaintiff, judge, attorney.
  
  • Jane Smith represented Jones Inc. She is a partner at Dewey, Chetum, and Howe. To contact her, write to j.smith@dch.com.
  
  • If a woman is over 62 years old and lives in the UK, she is a pensioner.
  
  • Diathesis: alternative sentence forms with (almost) synonymous meaning: Bill pushed Jill; Jill was pushed by Bill.
Problems with Language III

• Ambiguity, vagueness, underspecification:
  • The assembly passed the bill on the bank.
  • It is illegal to leave a heap of shoes on the sidewalk.
  • Vehicles may not be driven in the park.
• Sarcasm, irony.
• Interpretation.
• Context dependence, subjectivity, arbitrary meaning, when I was at school, I know language....
Problems with Language IV

• Complexity, length, and layout (see our Camera example).

• Intersentential connections:
  • Bill left the house. He drove home.
  • Bill left the house. He didn't feel comfortable there.
  • Bill left the house. It was an old house, once owned by a wealthy merchant.

• Synonymy, antonyms, meronyms (finger part of hand), etc.

• Repetition.
Problems for Annotation

• Annotate large legacy corpora.

• Address growth of corpora.

• Reduce number of human annotators and tedious work.

• Make annotation systematic, automatic, and consistent.

• Annotate fine-grained information:
  • Names, locations, addresses, web links, organisations, actions, argument structures, relations between entities.

• Map from well-drafted documents in NL to RDF/OWL/XML.
Addressing the Problems

• Decompose big problems down to smaller problems.
• Address the smaller problems.
• Compose solutions from parts.
• Identify (set aside, address, assign to someone else) remaining issues.
Methodology
Approaches


• Can do either. Where textual traceability (justification) is essential, knowledge heavy is important.
Overall Approach

• Decompose large complex problems into smaller, manageable problems for which we can create solutions.

• Software engineering approach.

Development Caveat

• Developing working prototypes (much less public and/or commercial tools) takes resources.
  • Tool development
  • Corpus development
  • Language analysis
• It is a slow, painstaking, and gradual process of constructing modules to do the small tasks you need to build the large applications you want.
  • Not a simple phone app.
Whazza Methodology?

- Rule-based and unweighted: the analysis makes use of lookup lists (gazetteers) that ascribe base annotations to strings along with unweighted rules (using the Java Annotation Pattern Engine (JAPE)). The output is not probabilistic.
- Bottom-up from simpler to complex: rules make use of simpler annotations to construct more complex annotations.
- Linguistically well-founded: rules make use of syntactic analyses and lexical-semantic information.
- Semi-automatic, interactive, and collaborative: where several alternative analyses are possible, these are provided to a human user to highlight and select; the methodology supports collaborative annotation (e.g. GATE TeamWare).
- Modular, iterative, incremental development: the analysis takes into account a continuum of textual complexity and richness; modules are designed to address simpler data, tested, then revised to account for more complex data; where data is not accounted for, the reasons are pinpointed and used to revise the rules or otherwise left for future work; the lists and rules can be augmented and adapted.
- Transparent and reproducible: all materials can be examined and reused.
- Traceability: results can be fully analysed in terms of lists and rules.
Linguistic Processing
Computational Linguistic Cascade I

- Sentence segmentation - divide text into sentences.
- Tokenisation - words identified by spaces between them.
- Part of speech tagging - noun, verb, adjective....
- Morphological analysis - singular/plural, tense, nominalisation, ...
- Shallow syntactic parsing/chunking - noun phrase, verb phrase, subordinate clause, ....
- Named entity recognition - the entities in the text.
Computational Linguistic Cascade II

• Dependency analysis – sentence subject, subordinate clauses, pronominal anaphora,…

• Relationship recognition – X is president of Y; A hit B with a car and killed B.

• Enrichment - add lexical semantic information to verbs or nouns.

• Supertagging – adding conceptual annotations to text.

• Translation to logic for reasoning.

• Each step guided by pattern matching and rule application.
Overall Processing Strategy

- Make *implicit* information explicit by adding *machine readable* annotations.
A Tool
• General Architecture for Text Engineering (GATE) - open source framework which supports plug-in NLP components to process a corpus of text.
• **GATE Training Courses**
  https://gate.ac.uk/
• A GUI to work with the tools.
• A Java library to develop further applications.
• Components and sequences of processes, each process feeding the next in a “pipeline”.
• Annotated text output or other sorts of output.
GATE Basic Process Flow

Can add further processing components to pipeline, e.g. NER, co-reference, other other annotations,...
GATE - Gazetteers

• Gazetteers are lookup lists that add features - when a string in the text is located in a lookup list, annotate the string in the text with the feature. *Conceptual covers*.

• Feature: list of items...

• Obligation: ought, must, obliged, obligation....

• Exception: unless, except, but, apart from....

• Verbs according to thematic roles: lists of verbs and their associated roles, e.g. *run* has an agent (*Bill ran*), *rise* has a theme (*The wind blew*).
GATE – JAPE Rules

• JAPE Rules (finite state transduction rules) create overt annotations and reuse other annotations (e.g. Parser Output):

```java
( {Exception}
  ( {SyntaxTreeNode.cat==PP} | {SyntaxTreeNode.cat==NP} | {SyntaxTreeNode.cat==NN} | {SyntaxTreeNode.cat==S} )) :temp
  -> :temp.ExceptionClause2 = { rule = "ExceptionClause2" }
```
GATE – Building an Application

• Have Gazetteer lists and JAPE rules for:
  
  • lists in various forms;
  • exception phrases in various forms;
  • conditionals in various forms;
  • deontic terms;
  • associating grammatical roles (e.g. subject and object) with thematic roles (agent and theme) in various forms.
Examples
Example - Rules

• Rule identification in regulations; what one can 'argue' for and against.

• Using previous modules.

• Wyner and Peters (2011)
Sample Outputs

You may use human blood from a donor with a previous record of a reactive screening test for evidence of infection due to a communicable disease agent that is designated in paragraph a of this section, if:

(1) At the time of donation, the donor is shown to be suitable by a requalification method; and

(2) tests performed under paragraphs a are nonreactive.

Consequence, list structure, and conjuncts of the antecedent.

Except as specified in paragraphs c, you, an establishment that collects blood, must test each donation of human blood that is intended for use in preparing a product for evidence of infection due to the following communicable disease agents:

(1) Human immunodeficiency virus, type 1;

(2) Human T-lymphotropic virus, type 1; and

(3) Human T-lymphotropic virus, type II.

Exception, agent NP, deontic concept, active main verb, theme.
Sample Output

Theme, deontic modal, passive verb, agent with complex relative clause.
Sample Output - Overall

(1) If you allow any autologous donation to be used for allogeneic transfusion, you must assure that all autologous donations are tested under this section;

(2) If you ship autologous donations to another establishment that allows autologous donations to be used for allogeneic transfusion, you must assure that all autologous donations shipped to that establishment are tested under this section; and

(3) If you ship autologous donations to another establishment that does not allow autologous donations to be used for allogeneic transfusion, you must assure that, at a minimum, the first donation in each 30-day period is tested under this section.

You must further test each donation, including autologous donations, found to be reactive by a screening test performed under paragraph a whenever a supplemental test has been approved for such use by FDA, except:

(1) For autologous donations, you must further test under this paragraph, at a minimum, the first reactive donation in each 30-day period; or

(2) If you have a record for that donor of a positive result on a supplemental test approved for such use by FDA, you do not have to further test an autologous donation.

Required testing must be performed by a laboratory registered in accordance with part 607 of this chapter and either certified to perform such testing on human specimens under the Clinical Laboratory Improvement Amendments of 1988.
This is an inline representation, and not 'pure' XML as tags can overlap. There is also offset, which can be modified easily.
Sample Output – ANNIC Search

{AgentNP}{Obligation}{VP}

Context: specified in paragraph c, you, an establishment that collects blood, must ... of infection due to the

AgentNP

Obligation

VP

Page 1 (20 results) Export

<table>
<thead>
<tr>
<th>Left context</th>
<th>Match</th>
<th>Right context</th>
</tr>
</thead>
<tbody>
<tr>
<td>specified in paragraph c, you, an establishment</td>
<td>you, an establishment...cable disease agents (1) Human</td>
<td></td>
</tr>
<tr>
<td>transmission of communicable disease.</td>
<td>You must test donati...nt under paragraph a</td>
<td></td>
</tr>
<tr>
<td>test an autologous donation.</td>
<td>Required testing...Amendments of 1988. Human blood that is</td>
<td></td>
</tr>
</tbody>
</table>
Example - Camera
Argument Fragment for a Camera

⭐⭐⭐⭐⭐ Stunning camera, 27 Aug 2011
By A. Reader
This review is from: Canon PowerShot SX220 HS Digital Camera - Grey (12.1MP, 14x Optical Zoom) 3.0 inch LCD (Electronics)

I adore this camera. Why? Because I get a higher percentage of good shots out of it than from any other camera I have, and that includes a few DSLRs, and the video is ridiculously good. Is the image quality as good as a DSLR or even a micro 4/3 system camera? No, but it's not far off, and a shot that's perfectly exposed and focused and free of camera shake on a smaller sensor like this beats not getting a picture at all by a long way. The SX220HS is my constant companion and lives in a pouch on my belt.

I've spent 30 years in photography, since buying my first Russian Zenit E around 1979, including being a semi-pro at one point. I know all about aperture/shutter speed/ISO ratios, exposure values, incident light metering, parallax error, Hurter & Driffield scales, flash guide numbers, the difference between depth of field and depth of focus, Ansel Adams's zone system, blah blah blah. When I'm taking a photo, I don't want to think about any of that stuff, I want to be enchanted by a scene and have a decent reproduction of it when I get home. With this camera, I get that. When I pull it out, I know I'm almost certain to get the picture I want. Canon have done an amazing job with it.

The video is astounding. When I upload it, I view it on a large screen, not a little laptop, so any flaws would be immediately apparent. I uploaded a few to Youtube for a friend to see, and she emailed me back, 'They are so clear!'
Pro and Con

Customer Reviews
Canon PowerShot SX220 HS Digital Camera - Grey (12.1MP, 14x Optical Zoom) 3.0 inch LCD

85 Reviews

5 star: 46
4 star: 27
3 star: 5
2 star: 6
1 star: 1

Average Customer Review: 4.5 stars

Search Customer Reviews

Share your thoughts with other customers

Only search this product's reviews

Create your own review

The most helpful favourable review

Stunning camera
I adore this camera. Why? Because I get a higher percentage of good shots out of it than from any other camera I have, and that includes a few DSLRs, and the video is ridiculously good. Is the image quality as good as a DSLR or even a micro 4/3 system camera? No, but it's not far off, and a shot that's perfectly exposed and focused and free of camera shake on a smaller... Read the full review

Published 7 months ago by A. Reader

The most helpful critical review

Canon Warranty
The Powershot SX220 HS does take good pictures, but for me the Canon Warranty lets it down. I had the camera for about 6 weeks when i noticed the lens cover would not fully open or close, and because i've had the camera over 30 days amazon did not want to know, so i contacted canon support which was rubbish, the first e-mail i got back was telling me to clean the CF card...

Read the full review

Published 5 months ago by Gmick
Comments on Comments

Comments
Track comments by e-mail
Tracked by 2 customers

Showing 1-7 of 7 posts in this discussion

Sort: Oldest first | Newest first

Initial post: 22 Nov 2011 23:32:56 GMT

Ian Beckett says:

hi,

you're not limited to 30 days under consumer legislation (6 weeks is still a "reasonable period") .. your issue is that you let them repair it which canon alwayws try and get you to do .. you should asked canon for a replacement or a refund ... contact canon and tell them you will give them one last chance to put it back to 100% condition otherwise you will be taking a small claims court action against them ... its really easy to to do and costs about £17 and you can do it all online.

Reply to this post

4 of 4 people think this post adds to the discussion. Do you? [Yes] [No]
Goals

• *Extract* arguments distributed across a corpora and *evaluate* them with formal, automated tools.

• *Speed* the work of human analysts.

• Provide *semi-automatic* support.

• Use aspects of NLP to *incrementally* address a range of problems (ambiguity, structure, contrasts,....)

• Wyner, Schneider, Atkinson, and Bench-Capon (2012).
Variables in schemes as targets for extraction.

Premises:
• Camera X has property P.
• Property P promotes value V for agent A.

Conclusion:
• Agent A should Action Camera X.
Analyst’s Goal: Instantiate

Premises:

• The Canon SX220 has good video quality.
• Good video quality promotes image quality for casual photographers.

Conclusion:

• Casual photographers should buy the Canon SX220.
Annotating Text

- **Annotate** text:
  - Simple or complex annotations.
  - Highlight annotations with **colours**.
  - Search for and extract text by annotation.
- GATE “General Architecture for Text Engineering”.
  - Works with large corpora of text.
  - Rule-based or machine-learning approaches.
To Find Argument Passages

- Use:
  - Indicators of **premise**
    - *after, as, because, for, since, when, ....*
  - Indicators of **conclusion**
    - *therefore, in conclusion, consequently, ....*
The paper manual supplied with the camera is a simplified version that does not explain the complex functions of the camera. The electronic version of the manual, supplied on the CD and downloadable from canon's website, is more comprehensive but not definitive. Some experimentation is required to get to grips with some of the camera's features.

If you are looking for a compact point-and-shoot camera that will also allow a high level of creative control then this camera is well worth consideration.
To Find What is Being Discussed

- Use **domain terminology**:
  - Has a flash
  - Number of megapixels
  - Scope of the zoom
  - Lens size
  - The warranty
I think the issue is one of canon screwing up the logic in auto/program modes making it useless indoors or in dull conditions and though you can set a low iso in program mode if you leave it and move indoors or in to dull conditions you run the risk of getting horribly under exposed shots.

The many reviews you read look at the image quality at a SET ISO so dont really pick up on the poor auto and program mode.

I have decided to keep mine (for outside travel shots) and also keep my old ixus to guarantee good point and shoot shots indoors at weddings, parties etc which it always produces.
To Find Attacks Between Arguments

• Use contrast terminology:
  – Indicators
    *but, except, not, never, no, ....*
  – Contrasting sentiment
    The flash worked *poorly*.
    The flash worked *flawlessly*. 
I think the issue is one of canon screwing up the logic in auto/program modes making it useless indoors or in dull conditions and though you can set a low iso in program mode if you leave it and move indoors or in to dull conditions you run the risk of getting horribly under exposed shots.

The many reviews you read look at the image quality at a SET ISO so don’t really pick up on the poor auto and program mode.

I have decided to keep mine (for outside travel shots) and also keep my old ixus to guarantee good point and shoot shots indoors at weddings, parties etc which it always produces.
I adore this camera. Why? Because I get a higher percentage of good shots out of it than from any other camera I have, and that includes a few DSLRs, and the video is ridiculously good. Is the image quality as good as a DSLR or even a micro 4/3 system camera? No, but it's not far off, and a shot that's perfectly exposed and focused and free of camera shake on a smaller sensor like this beats not getting a picture at all by a long way. The SX220HS is my constant companion and lives in a pouch on my belt.
Query for Patterns

\{PremiseIndicator\}(\{Token\})^{10}\{Positive\}(\{Token\})^{10}\{CameraProperty\}

Context:

Canon did it this way because a bigger battery would have increased the size of the camera, and

<table>
<thead>
<tr>
<th>Context</th>
<th>Token</th>
<th>PremiseIndicator</th>
<th>Positive</th>
<th>CameraProperty</th>
</tr>
</thead>
</table>
An Argument for Buying the Camera

Premises:

The pictures are perfectly exposed.
The pictures are well-focused.
No camera shake.
Good video quality.
Each of these properties promotes image quality.

Conclusion:

(You, the reader,) should buy the Canon SX220.
An Argument for NOT Buying the Camera

Premises:
- The colour is poor when using the flash.
- The images are not crisp when using the flash.
- The flash causes a shadow.
- Each of these properties demotes image quality.

Conclusion:
(You, the reader,) should not buy the CanonSX220.
Counterarguments to the Premises of “Don’t buy”

- The colour is poor when using the flash.
- For good colour, use the colour setting, not the flash.

- The images are not crisp when using the flash.
- No need to use flash even in low light.

- The flash causes a shadow.
- There is a corrective video about the flash shadow.
Gold Standards
have discovered such formulations, this is not a patent action and "obviousness" is not a benchmark. See Ferroline Corp. v. General Aniline & Film Corp., 207 F.2d 912 (7th Cir. 1953), cert. denied, 347 U.S. 953, 74 S.Ct. 678, 98 L.Ed. 1098 (1954).

4 The compositions of series 400 inks were not known to others in the industry. Steps taken to guard their secrecy—notably confidentiality agreements employees were required to sign with M & T and with Whittaker. There was no dispute as to the value of the formulas or with respect to the time and effort spent in developing them. Duplication was not sought as to deprive the series 400 inks of trade secret status. A few sophisticated competitors
Results of Annotation

• The annotators carry out their task and complete the project.
• Carry out inter-annotator agreement analysis.
• Curate the disagreements to create a Gold Standard corpus. Can use this for machine learning.
• Search the annotations using an online tool, e.g. ANNIC.
Additions
Add to Explorer (or Teamware)

• Verbs for propositional attitudes, e.g. *believe, know, hope* and speech acts, e.g. *stated, mentioned, guessed*.
• Opinion adverbials - *obviously, so far as I know, scientifically*.
• Question words and markers – *who, why, ?*
• Rhetorical connectives - *elaboration, example, contrast*.
• Others....
For the Argument Explorer tool, some additions are very useful to handle the volume of data extracted. Want to homogenise similar sentences.

**Summarisation** - extracting subset of (common or salient) text; using extracted information and NLG to create a 'natural' sounding summary by filling in a template (like what we've done with our app).

**Sentence compression** - removing words while preserving meaning, e.g. *I would really like to have a delicious tasty cup of coffee* -> *I want coffee.*
The Whole Hog
Another Approach: Parsing and Semantic Representations

• Render text into a machine-processable logic to:
  – have detailed semantic representation (e.g. predicate logic).
  – test for consistency (or subsets of the KB).
  – draw inferences from input facts.
  – give meaningful explanations for inferences/arguments.
  – answer questions.
• Attempto Controlled English
  – Expressive yet finite vocabulary and syntax.
  – Gives one parse and interpretation per input sentence.
  – Can be used as an NL interface to First Order Logic.
  – Semantics for discourse, e.g. *Bill walked in. He sat down.*

• C&C/Boxer
  – Less restricted
  – Statistically most highly ranked parses.
  – First order logic, temporal and modal operators.
  – Semantics for discourse.
Full Parsing and Semantic Representation

- ACE
  - http://attempto.ifi.uzh.ch/site/
- C&C/Boxer
- GMB Webdemo
  - https://urd.let.rug.nl/basile/gsb/webdemo/demo.php
- Groningen Meaning Bank
  - http://gmb.let.rug.nl/
- Wyner, Bos, Basile, and Quaresma (2012).
Parse - Cameron is a British citizen
DRS - Cameron is a British citizen
Cameron is a British citizen

First Order Logic

\[ \forall A : ( \text{per1cameron}(A) \land \exists B : \exists C : ( \text{a1british}(C) \land ( \text{n1citizen}(C) \land A = C ) ) ) \]

RuleML

```xml
<?xml version="1.0" standalone="no"?>
<!--
<?xml-model href="http://www.ruleml.org/1.0/relaxng/folog_relaxed.rnc"
?>
<?xml-model href="http://www.ruleml.org/1.0/xsd/folog.xsd" type="application/xml" schematypens="http://www.w3.org/2001/XMLSchema"?>
xsi:schemaLocation="http://ruleml.org/spec http://www.ruleml.org/1.0/xsd/folog.xsd">
  -->
  <RuleML>
  <Exists>
    <Var>A</Var>
  </Exists>
  <And>
    <Atom>
      <Rel>per1cameron</Rel>
    </Atom>
    <Var>A</Var>
  </And>
</RuleML>
```
• If an application is made to register as a British citizen a person who is a British overseas territories citizen, the Secretary of State may, if he thinks fit, cause the person to be so registered.
**First Antecedent**

1. A *person* (x9) is identified as a *British overseas citizen* (see x14 in P10).
2. A *making event* (e7) has an *application* (x6) as *Theme*.
3. A *registering event* (e12) has the *application* (x6) as *Agent* and x9 as *Theme*.
4. The *registering event* (e12) is in the *as* relation with a *British citizen* (x13), which is distinct from x14.

This captures some interesting issues - the difference between different forms of predication, e.g. *as a British Citizen* and *is a British Citizen*, while leaving their exact interpretation aside. The proposition (p10) correctly predicates being a British overseas territory citizen of the person (x9) who wants to be a British citizen. There are two events - a making event and a registering event, where the theme of making (the application) is the agent of registering; this captures accurately the link between the passive and the subject of the infinitive. However, it makes an abstract noun the agent of registering, which is a conceptual ambiguity of the text. The agent of making is left unspecified, as it is in the source, though conceptually there ought to be one.
**Output Analysis**

**Subordinate Consequent**

1. The modal operator *may* is represented as a diamond over the whole box.
2. The *secretary* (x3) is in the *of* relation to *state* (x4).
3. A *causing* event (e18) has the *secretary* (x3) as *Theme*; and the *causing* event is in a *Predicate* relation to a *proposition* (p19).
4. The *proposition* (p19) is a *registering* event (e21) with the *person* (x9) as *Theme*.
5. The predicate *so* applies to the *registering* event (e21).

As a semantic representation, several aspects seem correct and clear: the modal has scope over the whole clause; the secretary and state are related; and what is caused is an event of registering the person. Several things are problematic: the agentive interpretation of cause is not used (rather using a *middle* construction); the *of* is obscure; and, very importantly, the *so* predication fails to function as an ellipptic reference to *as a British citizen*. This last point is key since the meaning of the sentence is intended to be the conditions under which someone is registered as a British citizen.
• Alethic (necessity, ....), Epistemic (know, ....), Temporal, Deontic (must, may, ....).

• Subclasses of modal operators and propositional attitudes with different inference properties. Why are these not relations rather than sentential operators?
  – Factive – know, saw, heard,...
  – Non-factive – believe, remember (?),....
  – Contrafactive – wish, imagine, ...
  – Epistemic-root interpretation – may, must, can,...
  – Scope issues with de re/de dicto.
Issues

• Representational correspondence and/or redundancy.
• Example, NL semantics uses event arguments to tie together the representation of terms with respect to a predicate, not flat, labeled arguments as in other representations.
• Different typing of expressions in various representations (data, individual, relation).
• Legal rules are defeasible – what is the linguistic representation of this?
Future work

• Close evaluation of parses, semantic translations, and conversion to XML/RuleML.
• Compare and contrast outputs.
• Extension and application of ACE and GMB. For example, extract the vocabulary and structures from SBVR or see what is/is not available in ACE or GMB.
• Corpus development.
• Gold standards for ACE or GMB.
• Applications
  – Ontologies
  – Knowledge bases
  – Business and policy rule systems

Wyner, ACAI, London, UK
(cc) by-nc-sa license
Other Approaches
Sentiment Analysis
Debate Graphs

• A presentation (series of statements) by an MP correlated to a Aye or Nay (Gold Standard).
• Classify presentations according to positive or negative sentiment terminology and with respect to the Gold Standard.
• Relate presentations according to common terminology
• Create graph of relationships between presentations.
• No internal structure to the analysis of the presentations; homogenises differences within a statement; very light language component.
Sentiment Analysis
Debate Graphs
Textual Inference

• Textual inference for argument analysis. Cabrio and Villata (2012).
• **T1**: Internet access is essential now; it must be a right. The internet is the only wire that delivers freedom of speech, freedom of assembly, and freedom of the press in a single connection.

• **T2**: Internet not as important as real rights. We may think of such trivial things as a fundamental right, but consider the truly impoverished and what is most important to them. The right to vote, the right to liberty and freedom from slavery or the right to elementary education.

• **H**: Making Internet a right only benefits society.

• **T1 entails H; T2 contradicts H.**
Textual Inference

• Analyse corpus of statements on Debatepedia.
• Identify network of statements for and against a given position.
• Issue – grounding the justifications for the results, which depends on how TI is done.
Rhetorical Structure Theory

- Intrasentential relationships: background, circumstance, concession, elaboration, motivation, restatement,....
• Explicit vs implicit information. Must be able to infer information within or relations between statements - enthymemes, presuppositions, inferences, abduction, background information.

• In particular, how to infer rules (if...then...) or argumentation schemes, where explicit markers are missing?
Thanks for your attention!

- Questions?
- Contacts:
  - Adam Wyner  
    adam@wyner.info