Онтологическая семантика и абдукция: обработка эллипсиса

Ontological semantics and abduction: parsing ellipsis

Petrenko M. (mpetrenk@gmail.com)

Московский Гуманитарный Институт имени Е. Р. Дашковой, Москва, Россия

В работе рассматриваются возможности абдуктивного (инференционного) анализа естественных текстов с эллиптическими сегментами в рамках Отнологической Семантики на основе инференционных правил установления зависимости между семантическими ролями, а также правил зависимости классов событий и значений скалярных атрибутов.

1. Paper goals

The paper explores a promising yet currently understudied area of application of Ontological Semantics — ellipsis processing. After a brief outline of the Ontological Semantics framework and the mechanism of abduction, i.e. inference-based form of reasoning, it will be demonstrated how an Ontosem-informed NLP application handles elliptic input abductively, i.e. in a two-step fashion similarly to an abducing human agent. Two directions for developing abductive NLP module are explored. Pertinent examples are provided.

2. Direct Meaning Access: theory, structure and applications

In the current range of largely non-semantic, method-driven and domain-restricted computational NLP systems, Ontological Semantics, or Direct Meaning Access (its current incarnation), offers a semantics-informed, formalism-independent, problem-driven and cross-domain toolbox for describing and modeling human language competence in its complexity and dynamicity. A rapidly growing list of publications provides a detailed description of the methodology [14, 28], structure [9] and application domains [24, 25, 26, 28]. Below follows a brief and generalized overview of the system.

Stemming from the fundamental tenet — the unavoidability of semantics in designing any computational NLP system — DMA incorporates a large and richly structured hierarchy of ontological concepts — the Ontology. In addition to the basic ALL → (EVENT, OBJECT, ATTRIBUTE) branching (see Figure 1), each of ca. 8,000 concepts is also defined through a large set of properties (both unique and inherited) of a slot-filler structure, whose fillers are other concepts (see Figure 2). This results in:

- a highly complex nature of the ontology, which is constrained, on the one hand, by the general principle of parsimony of its acquisition, and on the other, by the natural organization of objects, events and properties in the world, which the ontology models;
- a highly versatile nature of the ontology; the highly entangled (hypero-hyponymic, mereological, causal, etc.) conceptual network enables the ontology to emulate human semantic competence;

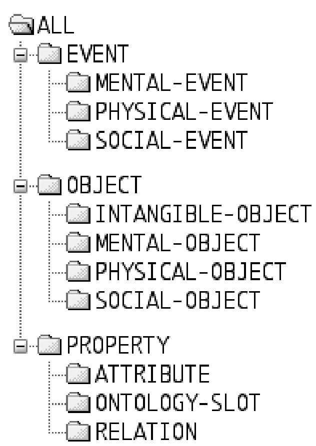

Figure 1. Basic Ontology branching: the root branch ALL breaks into OBJECT, EVENT, and PROPERTY

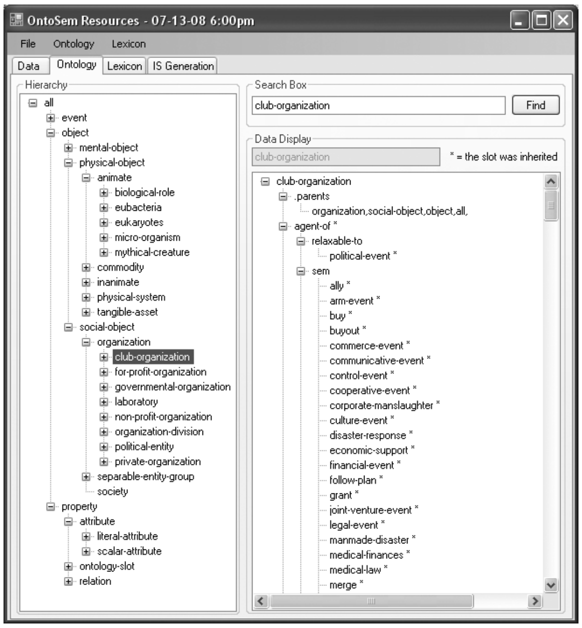

Figure 2. A snapshot of the ontology browser with the concept hierarchy and a detailed description for the CLUB-ORGANIZATION concept

A language-dependent Lexicon (120,000 entries for English, fewer for Spanish, Russian and Turkish) constitutes another static knowledge resource within DMA. The Lexicon features semantic (linking to a concept and its properties) and syntactic (case roles, selection restrictions) information for each entry (see Figure 3). Proper names are stored in a 25,000 entries-large Onomasticon.

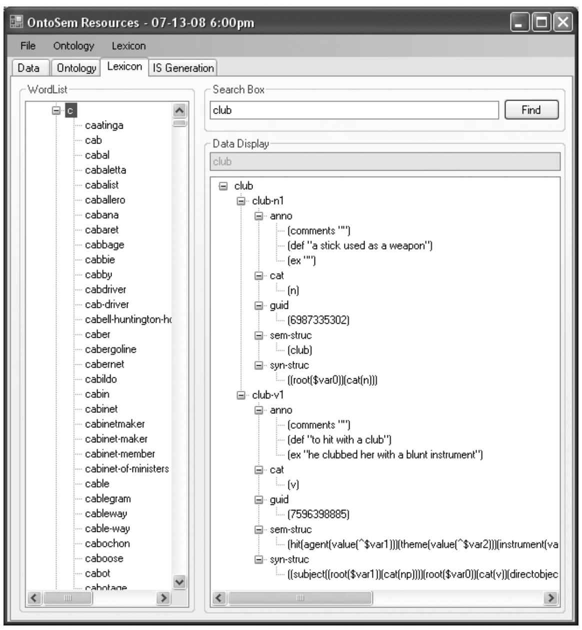

Figure 3. Lexical entries of club-n1, and club-v1 for the "club" super-entry in the Lexicon

The Fact Repository stores instances of concepts (head concepts, constraining properties and case-role fillers) from the immediate input. This allows the parsing module to operate across clauses and reconstruct elliptic and contradictory segments from recent slot fillers (see [25] for details on the contradiction-detection application of DMA).

The OntoParser, a dynamic processing module, utilizes static knowledge resources and proceeds in a step-by-step fashion from clause-breaking to lexical instantiations of concepts, syntactic constituents (e.g. multiple NP resolution), events, their case-role fillers and, ultimately clause merging, temporal, modality features, resulting in the text-meaning representation (TMR). Figure 4 illustrates the working of the OntoParser:

TMR is the final output of a DMA-informed NLP application. It constitutes lock, stock and barrel of any NLP enterprise and serves as a foothold for further machine-based applications: (information assurance, security, search, retrieval, etc.). Within DMA, a typical TMR features an event-driven description of a clause (from sentential level up) with the head event(s) and its case-role fillers. For example, the processing of the input

(1) The outlaws ran cocaine into the U.S.

would yield the following TMR:

(smuggle
(agent(sem(criminal)))
(theme(sem(cocaine)))
(destination(sem(country(has-name(value("united-states"))))))
)

, where the concept SMUGGLE in the Ontology is disambiguated through the lexical verbal entry run-v6 in the Lexicon, whose theme is the concept CONTROLLED-DRUG, which, in turn, is the ontological parent for the concept COCAINE (a more in-depth analysis of the example can be found in [28]).

Table 1 illustrates a wide range of application of Ontological Semantics and DMA with various degree of implementation (adapted from Raskin et al. 2004).

2.1. Abductive reasoning and Ontological Semantics: theory

The research on abductive reasoning, first defined in [17], is developing rapidly [1, 4, 5, 10, 11, 13, 15, 18, 29]. Generally, abduction is resorted to when an explanation of a fact is required, and no ultimately definitive theories are available. Most sources agree on a two-phased structure of abduction. It involves delineating a set of hypotheses, at which point a "leap of faith" is done by an abducing human when from the set of generated hypotheses the most plausible yet potentially defeasible candidate is selected. Selection criteria and algorithms are subject to debates in the literature (see [11] for a detailed overview).

Interesting parallels between abduction and the OntoParser can be drawn. Similar to abducing human

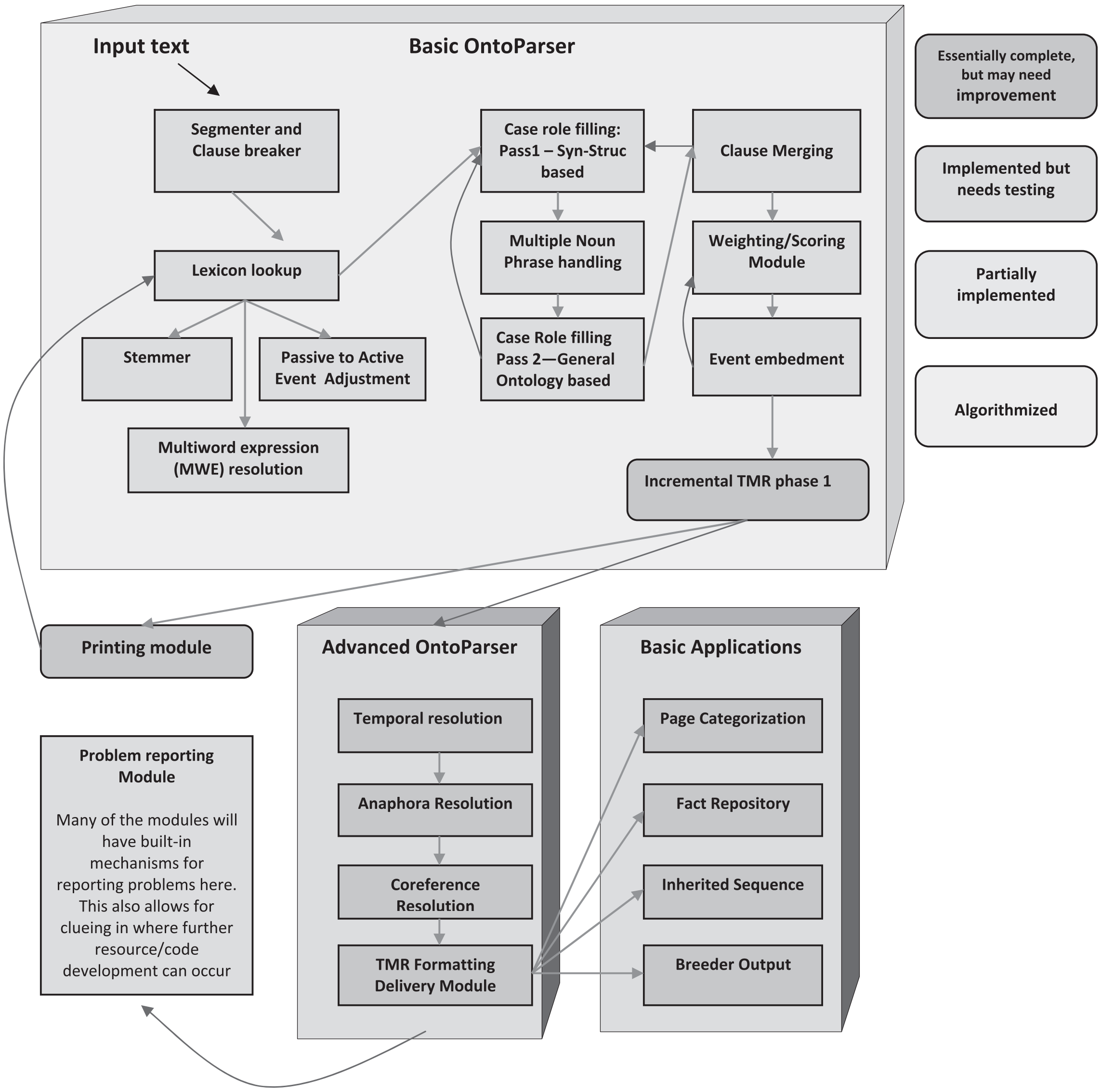

Figure 4. The flowchart of the OntoParser with sub-modules and steps of processing

Table 1. Areas of DMA application

Application	Function	Implementation	Reference
Syntactic NL Watermarking	Embeds the watermark in the syntactic tree of a sentence	Pilot/demo	[2]
Semantic NL Watermarking	Embeds the watermark in the TMR tree of a sentence	Pilot	[3]
NL Tamperproofing	Embeds a brittle watermark to detect any changes to the text	Pilot	[3]
NL Sanitization	Seamlessly removes and replaces sensitive information	Proof of concept	[12]
Automatic Terminology Standardizer	Translates different terminological dialects in IAS into TMRs	Proof of concept	[23]
Perimeter Protection	Sanitizes outgoing e-mail online	Proof of concept	[22]
NL Streaming Processor	Interprets incoming information before it is complete	Research	[23]
Ontosem-based humor research	Processing, modeling and generation of humorous texts based on Ontosem and General Theory if Verbal Humor frameworks	Research	[6, 8, 19]
Ontosem-driven Internet search engine	Semantic search, classification, QA applications	Pilot/demo	[9]
Abductive reasoning modeling	Processing elliptic input segments	Research	[20]
Ontosem-based (de)classification	Controls information sharing across security levels	Research	[27]

agents handling ambiguity, the OntoParser processes elliptic input by:

- projecting its static knowledge resources (Ontology, Lexicon, Onomasticon) onto immediate input;
- constraining the resources by selecting a pool of possible slot-fillers, and, ultimately,
- detecting the most plausible candidate (or a minimal set of them) within that pool.

Abduction thus lies in the core of OntoParser's processing of ambiguity. Both human agent and OntoParser operate on same principles. What the "pool of possible explanations" and the "best explanation" are to the human competence is what the "multiplicity of slot-fillers (mainly case-role fillers)" and the "constrained filler" are to the Ontoparser (Table 2).

Table 2. Similarity of two-staged abductive reasoning by human agent and OntoParser

Phase	Human competence	OntoParser
Hypotheses set	Pool of plausible hypotheses selected based on general knowledge	Set of case-role fillers delineated based on ontological properties and data from lexical entries in the input
Candidate selection	Best plausible explanation generated based on immediate goals	Most plausible case-role filler identified based on Fact Repository data and other explicit slot fillers

2.2. Abductive reasoning and Ontological Semantics: practice

The section will discuss two cases of ellipsis and strategies for their processing by the OntoParser.

Modeling abduction within Ontological Semantics is an ongoing study. Currently, two routes are being explored for enabling the parser to effectively process ambiguous (in particular, elliptic) input. Each route is defined by the initial input conditions, i.e. what segments of the text are missing and need restoring.

- For input with elliptic case roles (and explicit clause-forming events), inference rules can be designed that capture dependencies between case role fillers across clauses; if the dependency is there, the elliptic case role filler can be reconstructed based on the rule and the already present fillers from the recent input (stored in the Fact Repository);
- For input with elliptic clause-forming events (and explicit case roles), inference rules can be designed that capture correlations between scalar attributes and the epistemic modality values; when applied, the rule allows to narrow down the pool of potential elliptic events;

Examples will illustrate each case.

The implementation of an inference rule establishing case-role dependencies in clauses with specific event classes was described in detail in [20]. The rule allowed reconstructing non-verbalized fillers for the case role of destination in events whose EFFECT property slot was filled by a MOVEMENT-EVENT with an explicit destination case role:

(2) A bomb was thrown at a building. No serious damage reported. (Roughly adapted from [20])

(throw		(damage
(theme(sem(bomb)))	→	(theme(sem(*building — reconstructed*)))
(destination(sem(building))))		)

Based on the rule and employing the immediate data from Fact Repository and proceeding algorithmically, the parser reconstructs missing theme case role for the event DAMAGE with BUILDING from the preceding THROW (which IS-A MOVEMENT-EVENT) event, where BUILDING fills in the destination slot.

The inference rule below stipulates identical fillers for agent and theme case roles in CHANGE-IN-QUANTITY events related through the PRECONDITION and EFFECT properties. Example (3),

(3) If the US plans to increase its troops in Afghanistan, the reduction in Iraq needs to be accelerated.

features two events, INCREASE and DECREASE (represented by increase-v1 and reduce-v1), the latter having a non-verbalized theme filler. Before the inference rule is applied, the following TMR would be produced:

```
(increase
 (agent(sem(country(has-name(value("united-states"))))))
 (theme(sem(military-unit)))
 (location(sem(country(has-name(value("afghanistan"))))))
 (volitive(value(>0.5)))
 (precondition(sem
  (decrease
   (agent(sem(nothing))))
   (theme(sem(nothing)))
   (location(sem(country(has-name(value("iraq"))))))
   (saliency(value(>0.5)))
   (velocity(value(>0.5)))
  )
 )
)
```

The italicized "nothing's" indicate that no explicit or default fillers for the slots have been found the parser. This being said, the following rule can be formulated:
For events E1 and E2,
If E1 and E2 are CHANGE-IN-QUANTITY events, which stand in the PRECONDITION or EFFECT relation, then E1 and E1 have identical fillers for agent and theme case roles, if no other explicit or default fillers are available.
A formal definition of the rule is provided below:

IF (E1(pre-condition/effect(E2)));
E1 = CHANGE-IN-QUANTITY
E2 = CHANGE-IN-QUANTITY
No explicit case role fillers for E2 available;
THEN (E1(agent)) = (E2(agent))
(E1(theme)) = (E2(theme))

As the result of the rule application, the concepts COUNTRY(has-name(value(“united-states”))) and MILITARY-UNIT will fill in the slots for agent and theme case roles in the DECREASE event.

Reconstructing elliptic events presents a challenge. In the ongoing study in this area, an interesting avenue is currently being explored, in which for cases with explicit scalar attribute values, explicit case roles and implicit events, inference rules can be designed based on the correlation between the value of epistemic modality and the value of large class of SCALAR-ATTRIBUTE’s for the non-verbalized event. In the current Ontology, it appears, some scalar attributes can determine the value of the epistemic modality of the clause-forming event. To illustrate, the zero value of INTELLIGENCE attribute seems to precondition the zero value of epistemic modality for the branch of COGNITIVE-EVENT’s:

IF (intelligence(value(0)))
THEN (cognitive-event(epistemic(value(0))))

Example (4),

(4) I tried to explain him the theory, but he turned out to be completely dumb.

contains an explicit zero value of the INTELLIGENCE attribute and a non-verbalized EVENT:

(expressive-act
(agent(sem(human)))
(theme(sem(theory)))
(beneficiary(sem(human intelligence(value(0)))))
(domain(value(*event missing*)))
(epiteuctic(value(<0.5)))
)

Naturally, in the given situation, successful understanding presupposes a higher-than-zero value of intelligence property attributed to the recipient. In the TMR, the italicized *event missing* can be reconstructed through an inference rule which would constrain the domain[1] for the INTELLIGENCE attribute to a number of event classes, COGNITIVE-EVENT being one of them. The case role filler match would then narrow the search down to a specific event. A rule can thus be formulated:

For a clause C that contains a verbalized case role CR, a verbalized scalar attribute SA preconditioning a set of event classes EC, and non-verbalized event E, the value of SA will be identical with the value of the epistemic modality of E, and CR will fill in one of the slots for E;

For (E(SCALAR-ATTRIBUTE(value(x)))(case-role(y)))
IF E is elliptic;
x is verbalized;
y is verbalized;
THEN
(E(case-role(y));
(epistemic(value(x)))),
where $E \in \{EC\}$; $y \in \{EC(case\text{-}roles)\}$;

The tentative list of preconditioning scalar attributes includes, but not restricted to:

difficulty-attribute, feasibility-attribute, intensity, orderliness, precision-attribute, rapidity, safety, secrecy-attribute, success-attribute, survivability, treatability, stability, age, endurance, flexibility, resistance, roundness, almost all SCALAR-HUMAN-ATTRIBUTE, etc.

For each of these attributes, a limited number of events can be defined as their domain. Based on this dependency, the OntoParser would:

1) Look up the set of events preconditioned by the attribute in question;
2) From this set, select events whose case-role fillers can be found in the given input;
3) Identify the non-verbalized event;
4) Based on the given attribute value, determine the probability of the event by assigning a specific value of epistemic modality;

A tentative algorithm can thus be formulated:

1. Input contains an ATTRIBUTE?
   Yes — go to 2. No — terminate.
2. Non-verbalized EVENT in the clause?
   Yes — go to 3. No — terminate.
3. Identify events preconditioned by the attribute.
   Yes — go to 4. No — terminate.
4. Identify case-role fillers in the clause.
   Yes — go to 5. No — terminate.

[1] The notions of “attribute domain” and “attribute range” need clarification here. Within Ontological Semantics, along with OBJECTS and EVENTS, the ontology features a class of PROPERTIES, of which ATTRIBUTE is a subclass. Every PROPERTY is defined through domain and range: property (domain, range). The set of events or objects to which PROPERTY pertains is defined in its domain, whereas the set of values of PROPERTY defined in its range (see [14], but also [16: 31] for similarities with the notions of domain and range of a mathematical function). To illustrate, in the ontology, AGE is a literal attribute, whose domain is EVENT and OBJECT, and whose range is ANY-NUMBER (i.e. a numeric value of age of a particular object or event). The author is grateful to the anonymous reviewers for emphasizing the need to elaborate on the notions.

5. Match the case-role fillers with those for the events preconditioned by the attribute.
   Declare the selected set of events elliptic.
6. Identify the value of the ATTRIBUTE.
   Assign identical value of EPISTEMIC modality to the event or events declared elliptic.

Consider the examples:

(5) He saw the shore on the horizon but was too tired.

(6) He found the spade but was too tired.

For both examples, the parser would identify the following set of events preconditioned by the STRENGTH-ATTRIBUTE: dig, enclose, shift, shift-material, entwine, unwrap, wrap, fasten-together, operate-device.

In the set, the events would be selected which could have COASTAL-GEOLOGICAL-ENTITY (lexical item "shore") and SHOVEL (lexical item "spade") as case-role fillers:

```
(5) (coastal-geological-entity
      (destination-of(sem(horizontal-liquid-motion)))
    )

(6) (shovel
      (instrument-of(sem(operate device dig shift-material
      enclose fasten-together operate-device unwrap wrap)))
    )
```

Having narrowed down the search, the parser would then proceed with the STRENGTH-ATTRIBUTE and EPISTEMIC value assignment, at which point the ultimate TMR's for the two examples will be generated:

```
(5) He saw the shore on the horizon but was too tired.
(visual-event
   (agent(sem(human(gender(value(male))))))
   (theme(sem(coastal-geological-entity)))
(horizontal-liquid-motion
   (agent(sem(human
         (gender(value(male))))))
         (strength-attribute(value(<0.2)))
   (destination(sem(coastal-geological-entity)))
   (epistemic(value(<0.2)))
   )
)

(6) He found the spade but was too tired.
(find
   (agent(sem(human(gender(value(male))))))
   (theme(sem(shovel)))
(operate device dig shift-material enclose fasten-together
operate-device unwrap wrap
   (agent(sem(human
         (gender(value(male))))))
         (strength-attribute(value(<0.2)))
   (instrument(sem(shovel)))
   (epistemic(value(<0.2)))
   )
)
```

In (6), the natural ambiguity of the input would be registered even by a human agent since it is unclear for what purposes the spade would be used. This is interpreted by the OntoParser as the multiplicity (significantly reduced, after to the inference processing) of possible events.

References

1. *Aliseda A.* Abductive Reasoning: Logical Investigations into Discovery and Explanation // Mexico: National Autonomous University of Mexico, 2006.
2. *Atallah M. J., Raskin V., Crogan M., Hempelmann C. F., Kerschbaum F., Mohamed D., Naik S.* Natural Language Watermarking: Design, Analysis, and a Proof-of-Concept Implementation // I. S. Moskowitz (ed.), Information Hiding: 4th International Workshop, IH 2001, Pittsburgh, PA, USA, April 2001 Proceedings. Berlin: Springer, 2001. P. 185–199.
3. *Atallah M. J., Raskin V., Hempelmann C. F., Karahan M., Sion R., Topkara U., Triezenberg K. E.* Natural Language Watermarking and Tamperproofing. // F. A. P. Petitcolas (ed.), Information Hiding: 5th International Workshop, IH 2002, Proceedings. Berlin: Springer, 2002. P.196–210.
4. *Attardo S.* On the Nature of Rationality in (Neo-Gricean) Pragmatics // International Journal of Pragmatics. (Special issue on Neo-Gricean pragmatics, edited by Ken Turner). 2003. V. 14, P. 3–20.
5. *Flach P. A., Kakas A. C.* Abduction and Induction: Essays on their Relation and Integration // Kluwer Academic Publishers, 2002.
6. *Hempelmann C. F.* Paronomasic Puns: Target Recoverability towards Automatic Generation. Unpublished Ph.D. Dissertation // Interdisciplinary Program in Linguistics. West Lafayette, IN: Purdue University, 2003.
7. *Hempelmann C. F., Raskin V., Triezenberg K. E.* Semantic Forensics: NLP Systems for Deception Detection // Proceedings of the First Annual MidwestColloquium in Computational Linguistics. Damir C., Rodriguez P.(Eds.) Bloomington, IN: Indiana University. 2004.

8. *Hempelmann C. F.* Computational Humor: Beyond the Pun // The Primer of Humor Research. Raskin V. (Ed.). Berlin, New York: Mouton de Gruyter, 2008. P. 335–363.
9. *Hempelmann C. F., Raskin V.* Semantic Search: Content Vs. formalism // Rome: Proceedings of Langtech 2008. http://www.langtech.it/en/technical_program/technical_program.htm (full paper).
10. *Hobbs J. R., Stickel M. E., Appelt D. E., Martin P.* Interpretation as Abduction. // Pereira, F.C.N. and Barbara J. Crosz (eds.) Natural Language Processing. MIT Press, 1994. P.69–142.
11. *Gabbay D. M., Woods J.* The Reach of Abduction. Insight and Trial // A Practical Logic of Cognitive Systems. Amsterdam: Elsevier Science, 2005. V. 2.
12. *Mohamed D.* Ontological Semantics Methods for Automatic Downgrading. An unpublished Masters' thesis // Interdisciplinary Program in Linguistics and CERIAS. West Lafayette, IN: Purdue University, 2001.
13. *Niiniluoto I.* 1998. Defending Abduction // Philosophy of Science, Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association, 1999. Part I: Contributed Papers. Vol. 66.
14. *Nirenburg S., Raskin V.* Ontological Semantics // Cambridge, MA: MIT Press, 2004. A prepublication draft, chapter by chapter, can be found on www.ontologicalsemantics.com
15. *Paavola S.* Abduction as a Logic and Methodology of Discovery: The importance of strategies // Foundations of Science. 2004. V. 8. P. 267–283. http://www.helsinki.fi/science/commens/papers/abductionstrategies.html
16. *Partee B. H., Meulen A. T., Wall R. E.* Mathematical Methods in Linguistics. Dordrecht, Boston, London: Kluwer Academic Publishers, 1990.
17. *Peirce Ch. S.* Philosophical Writings of Peirce (1955) // Collected Papers. Charles Hartshorne and Paul Weiss (eds.), New York: Dover: Harvard University Press, 1960–1966. V. 8.
18. *Peng Y., Reggia J. A.* Abductive Inference Models for Diagnostic Problem Solving // New York, Berlin: Springer-Verlag, 1990.
19. *Petrenko M.* The Narrative Joke: Conceptual Structure and Linguistic Manifestation. An unpublished Ph.D. dissertation // Interdisciplinary Program in Linguistics, West Lafayette, IN: Purdue University, 2008.
20. *Petrenko M., Raskin V.* Modeling Abduction within Ontological Semantics // Proceedings of Midwestern Computational Linguistics Colloquium 5. East Lansing: Michigan State University, May 2008.
21. *Raskin V., Atallah M. J., McDonough C. J., Nirenburg S.* Natural Language Processing for Information Assurance and Security: An Overview and Implementations // M. Schaefer (ed.), New Security Paradigm Workshop. Proceedings. New York: ACM Press. Ballycotton, County Cork Ireland: 2001. P. 51–65.
22. *Raskin V., Atallah M. J., Hempelmann C. H., Mohamed D.* Hybrid Data and Text System for Downgrading Sensitive Documents // Technical report. West Lafayette, IN: Center for Education, Research in Information Assurance and Security, 2001.
23. *Raskin V., Nirenburg S., Atallah M. J., Hempelmann C. F., Triezenberg K. E.* Why NLP should move into IAS // Steven Krauwer (ed.), Proceedings of the Workshop on a Roadmap for Computational Linguistics. Taipei, Taiwan: Academima Sinica, 2002. P. 1–7.
24. *Raskin V., Nirenburg S., Nirenburg I., Hempelmann C. F., Triezenberg K. E.* The genesis of a script for bankruptcy in ontological semantics // ACL: Edmonton, Alberta, Canada: G. Hirst and S. Nirenburg (eds.), Proceedings of the Text Meaning Workshop, HLT/NAACL 2003: Human Language Technology and North American Chapter of the Association of Computational Linguistics Conference., 2003
25. *Raskin V., Hempelmann C. F., Triezenberg K. E.* 2004. Semantic Forensics: An Application of Ontological Semantics to Information Assurance // ACL Publication**.** G. Hirst and S. Nirenburg (eds.), Proceedings of the Second Workshop on Text Meaning and Interpretation, ACL, Barcelona, Spain, Edmonton, Alberta, Canada**:** 2004.
26. *Raskin V., Hempelmann C. F., Triezenberg K. E.* Ontological Semantic Forensics: Meaning-based Deception Detection // Meaning computation, 2008. V. 1.
27. *Raskin V., Hempelmann C. F., Triezenberg K. E., Buck B., Keen A.* Accessing and Manipulating Meaning of Textual and Data Information for Information Assurance and Security and
28. *Intelligence* Information // Proceedings of the Cyber Security and Information Intelligence Research Workshop, Oak Ridge National Laboratory, Oak Ridge, Tennessee: May 12–14, 2008.
29. *Raskin V., Hempelman C. F., Taylor J. M., Petrenko M. S., Trienzenberg K. E., Buck B.* The Why's, How's, and What-of's of Natural Language Ontology // 2nd Interdisciplinary Ontology Conference. Proceedings. Tokyo: 2009 (forthcoming)
30. *Walton D.* Abductive Reasoning // University of Alabama Press, 2004.