Computing Meaning: Annotation, Representation, and Inference

Harry Bunt, Johan Bos, and Stephen Pulman

Abstract This chapter introduces the subsequent chapters in the book and how they are related, against the background of a discussion of the nature and the complexity of processes that compute the meanings of natural language expressions. The discussion focuses on three aspects of the computation of meanings that play an important part in later chapters: (1) the nature of meaning representations; (2) the integration of inferencing with compositional interpretation; and (3) the construction of semantically annotated corpora and their use in machine learning of meaning computation.

1 Introduction

While computers are very good at computing in general, they are not very good at computing meaning. There are at least three reasons why this may be so: (R1) the very notion of meaning, as expressed in natural language, is something extremely complex, and therefore difficult to compute; (R2) the process of computing meanings is extremely complex, because it requires the effective use of a variety of extremely rich information sources (linguistic knowledge, general knowledge of the world, specific knowledge of the domain of discourse, knowledge of interactive settings,...); and (R3) the very notion

Harry Bunt
Tilburg Center for Cognition and Communication (TiCC) and Department of Philosophy, Tilburg University, Netherlands, e-mail: `harry.bunt@uvt.nl`

Johan Bos
Center for Language and Cognition (CLCG), University of Groningen, Netherlands, e-mail: `johan.bos@rug.nl`

Stephen Pulman
Department of Computer Science, Oxford University, United Kingdom, e-mail: `stephen.pulman@cs.ox.ac.uk`

of meaning is not well enough understood to effectively program and/or teach computers what it is and how it can be computed for a given natural language expression, occurring in a given context.

Most of the work in formal as well as in computational semantics tacitly assumes, different from (R3), that we do have a clear understanding of what we mean by meaning, and different from (R1), that natural language meanings are simple enough to be represented by very simple structures, such as formulas in first-order logic (or, equivalently, Discourse Representation Structures). Assuming that such structures are adequate representations of meaning, computing the meaning of a given natural language expression comes down to syntactic parsing of it and composing the semantic representations of the parts to form the meaning representation, which itself has a semantics defined by the representation formalism.

Since computational semantics started to develop, in the last two decades of the twentieth century (see Blackburn and Bos, 2005), it has become clear that the dream of computing meaning representations by syntactic/semantic (de-)composition, made popular especially through the work of Richard Montague (see Thomason, 1974), cannot become reality, simply because natural language expressions much of the time do not contain sufficiently much information to construct such a representation. Other information sources are indispensable. This insight has inspired the introduction of the notion of an underspecified meaning representation, which represents the semantic information that is present in the sentence without disambiguating those aspects for which the sentence does not contain sufficient information. It also became very clear that relying solely on linguistic information for computing meanings would lead to impossibly complex interpretation processes, due to the astronomical number of readings that ordinary sentences have when considered in isolation (see Bunt and Muskens, 1999). Again, underspecified meaning representations offer solace here, as they obviate the need to fully disambiguate. Several of the chapters in this book, in particular in Part I, witness the ongoing search for appropriate forms of meaning representation and for methods of exploiting linguistic as well as other information in their computation.

A problematic aspect of the use of underspecified semantic representations is that they do not allow straightforward application of logic-based inference methods, since different resolutions of underspecifications may result in interpretations that allow different inferences (see e.g. van Deemter, 1996; Blackburn et al., 2001). This is indeed problematic on the traditional view of meaning representations as unambiguously supporting a specific set of inferences, thereby explaining differences in meaning and relations between different meanings. One way to deal with this problem is to move away from strictly deductive approaches to inferencing, and instead turn to abductive methods (Hobbs et al., 1993) or to textual entailment, where inferencing is performed directly on natural language expressions, rather than on their interpretations, and to replace logical proof by psychological plausibility (see e.g. Dagan et al., 2008, and Bos, this volume). One way or another, the use of inference processes involving natural language expressions and/or their interpretations is needed, since nonlinguistic information must be exploited in order to arrive at intended and contextually appropriate interpretations; methods for combining pieces of information therefore have to be applied in order to arrive at a appropriate interpretations. The chapters in Part II of this book are all concerned with forms of inferencing (or combining pieces of information) in the computation of meanings.

Related to the limitations of effectively following strictly logic- and rule-based methods in the computation of meaning is the exploration of statistical and machine learning techniques that have been successfully applied in other areas of computational linguistics. These techniques presuppose the availability of large corpora, and can benefit in particular from semantically annotated resources. The development of such corpora (e.g.Basile et al. (2012)), and of well-founded semantic annotation methodologies (see Bunt, 2013), have supported the use of these new methods in computational semantics research (see e.g. Clark and Pulman, 2007), as reflected in several of the chapters in this book, both in Part I and in Part III.

2 About this book

The chapters in this book are organized into three parts. A first cluster of four chapters is focused on aspects of the representation of meaning and the computation of these representations. A second group of four chapters is concerned with issues of inferencing and its role in language understanding. The chapters in the third and final cluster of four deal with resources for meaning computation and their use.

2.1 Semantic Representation and Compositionality

In the opening chapter of this part of the book, entitled **Deterministic Statistical Mapping of Sentences to Underspecified Semantics**, the authors Hiyan Alshawi, Pi-Chuan Chang and Michael Ringgaard present a method for training a statistical model for mapping natural language sentences to semantic expressions. The semantics are expressions of an underspecified logical form that has properties making it particularly suitable for statistical mapping from text. An encoding of the semantic expressions into dependency trees with automatically generated labels allows application of existing methods for statistical dependency parsing to the mapping task (without the need for separate traditional dependency labels or parts of speech). The encoding also results in a natural per-word semantic-mapping accuracy measure.

The authors report on the results of training and testing statistical models for mapping sentences of the Penn Treebank into the semantic expressions, for which per-word semantic mapping accuracy ranges between 79% and 86% depending on the experimental conditions.

The particular choice of algorithms used also means that the trained mapping is deterministic (in the sense of deterministic parsing), paving the way for large-scale text-to-semantics mapping.

In the next chapter, **A Formal Approach to Linking Logical Form and Vector-Space Lexical Semantics**, the authors Dan Garrette, Katrin Erk and Raymond Mooney argue that first-order logic provides a powerful and flexible mechanism for representing natural language semantics, but that it is an open question of how best to integrate it with uncertain, weighted knowledge, for example regarding word meaning. They describe a mapping between predicates of logical form and points in a vector space. This mapping is used to project distributional inferences to inference rules in logical form. The authors then describe the first steps of an approach that uses this mapping to recast first-order semantics into the probabilistic models that are part of Statistical Relational AI. Specifically, they show how Discourse Representation Structures can be combined with distributional models for word meaning inside a Markov Logic Network and used to successfully perform inferences that take advantage of logical concepts such as negation and factivity, as well as weighted information on word meaning in context.

In the chapter **Annotations that Effectively Contribute to Semantic Interpretation** Harry Bunt presents a new perspective on the use of semantic annotations. He argues that semantic annotations should capture semantic information that is supplementary to the information that is expressed in the source text, and should have a formal semantics. If the latter condition is satisfied then the information in semantic annotations can be effectively combined with information extracted by a compositional semantic analysis. This can be used (1) for making semantic relations explicit which are not expressed in the text as such, such as coreference relations and implicit discourse relations, and (2) for specializing an interpretation to one that is contextually appropriate.

Bunt shows how such uses of semantic annotations can be optimally facilitated by defining a semantics of annotations in the form of a compositional translation of annotations into a formalism that is also suitable for underspecified semantic representations as commonly built by compositional semantic analyzers, allowing a unification-like combination of pieces of information from different sources. He shows that slightly modified Discourse Representation Structures, where discourse referents are paired with annotation markables, are particularly convenient for this purpose.

The approach is illustrated with examples from recent efforts concerning the annotation of information about time and events, about coreference, about semantic roles, and about discourse relations.

In the last chapter of this part of the book, entitled **Concrete Sentence Spaces for Compositional Distributional Models of Meaning**, a group of authors consisting of Edward Grefenstette, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke, and Stephen Pulman describe a compositional model of meaning that they have developed for distributional semantics, in which each word in a sentence has a meaning vector and the distributional meaning of the sentence is a function of the tensor products of the word vectors. Abstractly speaking, this function is the morphism corresponding to the grammatical structure of the sentence in the category of finite dimensional vector spaces.

The authors provide a concrete method for implementing this linear meaning map by presenting an algorithm for computing representations for various syntactic classes which have functional types; this algorithm results in assigning concrete corpus-based vector spaces to the abstract type of 'sentence'. The construction method is based on structured vector spaces whose basis vectors are pairs of words and grammatical roles. The concrete sentence spaces only depend on the types of the verbs of sentences; the authors use an embedding of these spaces and compare meanings of sentences with different grammatical structures by simply taking the inner product of their vectors in the bigger space. The constructions are exemplified on a toy corpus.

2.2 Inference and Understanding

In the first of the four chapters forming the second part of the book, entitled **Recognising Textual Entailment and Computational Semantics**, Johan Bos notes that recognising textual entailment (RTE) — deciding whether one piece of text contains new information with respect to another piece of text — remains a big challenge in natural language processing.

One attempt to deal with this problem is combining deep semantic analysis and logical inference, as is done in the Nutcracker RTE system. In doing so, various obstacles will be met on the way: robust semantic interpretation, designing interfaces to state-of-the-art theorem provers, and acquiring relevant background knowledge. The coverage of the parser and semantic analysis component is high (nearly reaching 100%). Yet the performance on RTE examples yields high precision but low recall.

An empirical study of the output of Nutcracker reveals that the true positives are caused by sophisticated linguistic analysis such as coordination, active-passive alternation, pronoun resolution and relative clauses; the small set of false positives are caused by insufficent syntactic and semantic analyses. Most importantly, the false negatives are produced mainly by lack of background knowledge.

The next chapter, entitled **Abductive Reasoning with a Large Knowledge Base for Discourse Processing** presents a discourse processing framework based on weighted abduction. The authors, Ekaterina Ovchinnikova, Niloofar Montazeri, Theodore Alexandrov, Jerry Hobbs, Michael C. McCord, and Rutu Mulkar-Mehta, elaborate on ideas concerning abduction in language understanding described in Hobbs et al. (1993) and implement the abductive inference procedure in a system called *Mini-TACITUS*. Particular attention is paid to constructing a large and reliable knowledge base for supporting inferences. For this purpose such lexical-semantic resources are exploited as WordNet and FrameNet. English Slot Grammar (McCord, 1990) is used to parse text and produce logical forms.

The proposed procedure and the resulting knowledge base are tested on the Recognizing Textual Entailment task using the data sets from the RTE-2 challenge for evaluation. In addition, an evaluation is provided of the semantic role labeling produced by the system taking the Frame-Annotated Corpus for Textual Entailment as a gold standard.

In the chapter **Natural Logic and Natural Language Inference**, Bill MacCartney and Chris Manning propose a model of natural language inference which identifies valid inferences by their lexical and syntactic features, without full semantic interpretation. They extend past work in *natural logic*, which has focused on semantic containment and monotonicity, by incorporating both semantic exclusion and implicativity. The proposed model decomposes an inference problem into a sequence of atomic edits linking premise to hypothesis; predicts a lexical entailment relation for each edit; propagates these relations upward through a semantic composition tree according to properties of intermediate nodes; and joins the resulting entailment relations across the edit sequence.

A computational implementation of the model achieves 70% accuracy and 89% precision on the FraCaS test suite (Cooper et al., 1996). . Moreover, including this model as a component in an existing system is shown to yield significant performance gains on the Recognizing Textual Entailment challenge.

In the final chapter of this part of the book, **designing Efficient Controlled Languages for Ontologies**, the authors Raffaella Bernardi, Diego Calvanese, and Camilo Thorne describe a methodology to recognize *efficient* controlled natural languages that compositionally translate into ontology languages, and as such are suitable for use in natural language front-ends to ontology-based systems. Efficiency in this setting is defined as the tractability (in the sense of computational complexity theory) of logical reasoning in such fragments, measured in the size of the data they aim to manage.

In particular, to identify efficient controlled languages, fragments are considered which correspond to the *DL-Lite* familiy of description logics, known to underpin data intensive ontologies and systems. The proposed methodology exploits the link between syntax and semantics of natural language captured by categorial grammars, controlling the use of lexical terms that introduce logical structure outside the allowed fragments. A major role is played by the control of function words introducing logical operators in first-order meaning representations.

Bernardi et al. present a preliminary analysis of semantically parsed English written corpora, which was carried out in order to show how empirical methods may be useful in identifying CLs that provide good trade-offs between coverage and efficiency.

2.3 Semantic Resources and Annotation

Part 3 of the book opens with a chapter by Harry Bunt, **A Context-Change Semantics for Dialogue Acts**, which presents an update semantic for dialogue acts, defined in terms of combinations of very simple 'elementary update functions' for updating the information state of an addressee of a dialogue act. This approach, which is rooted in Dynamic Interpretation Theory (Bunt, 1995; 2000) is motivated by the observation that related types of dialogue acts such as answers, confirmations, and disconfirmations give rise to similar but slightly different information state updates, which can be described elegantly in terms of overlapping sets of elementary update functions. This makes fine-grained distinctions between types of dialogue acts explicit and explains semantic relations like entailment and exclusion between dialogue acts.

The approach is applied to dialogue act representations as defined in the Dialogue Act Markup Language (DiAML), which forms part of the recently established ISO standard 24617-2 for dialogue annotation (), and to the varieties of dialogue act types defined in this standard and in the DIT^{++} taxonomy of dialogue acts.

The next chapter, by Susan Windisch Brown, Dmitriy Dligach, and Martha Palmer deals with the semantic classification of verb senses, and is entitled **VerbNet Class Assignment as a WSD Task**. The VerbNet lexical resource classifies English verbs based on semantic and syntactic regularities and has been used for a variety of NLP tasks, most notably, semantic role labeling. Since, in addition to thematic roles, it also provides semantic predicates, it can serve as a foundation for further inferencing. Many verbs belong to multiple VerbNet classes, with each class membership corresponding roughly to a different sense of the verb. A VerbNet token classifier is essential for current applications using the resource and could provide the basis for a deep semantic parsing system, one that made full use of VerbNet's extensive syntactic and semantic information. The authors describe their VerbNet classifier, which uses rich syntactic and semantic features to label verb instances with their appropriate VerbNet class. It is shown to achieve an accuracy of 88.67% with multiclass verbs, which is a 49% error reduction over the most frequent class behaviour as a baseline.

In the chapter **Annotation of Compositional Operations with GLML**, James Pustejovsky, Jessica Moszkowics, Olga Batiukova, and Anna Rumshisky introduce a methodology for annotating compositional operations in natural language text and describe the Generative Lexicon Mark-up Language (GLML), a mark-up language inspired by the Generative Lexicon model, for identifying such relations. While most annotation systems capture surface relationships, GLML captures the "compositional history" of the argument selection relative to the predicate. The chapter provides a brief overview of GL before moving on to the proposed methodology for annotating with GLML.

Three main tasks are described in this chapter. The first one is based on atomic semantic types and the other two exploit more fine-grained meaning parameters encoded in the Qualia Structure roles: (i) argument selection and coercion annotated for the SemEval-2010 competition; (ii) qualia in modification constructions; (iii) type selection in modification constructions and verb-noun combinations involving dot objects. The authors explain what each task comprises and include the XML format for annotated sample sentences. It is shown that, by identifying and subsequently annotating the typing and subtyping shifts in these constructions, an insight is gained into the workings of the general mechanisms of composition.

In the closing chapter of this book, entitled **Incremental Recognition and Prediction of Dialogue Acts**, by Volha Petukhova and Harry Bunt, is concerned with incremental machine-learned recognition of the communicative functions of dialogue utterances. Language use in human conversation is fundamentally incremental, and human language processing is continuously sensitive to multiple partial constraints, where contextual ones play a very important role. The question arises whether dialogue systems can be enabled to access and use various sources of information well enough and fast enough to interpret incoming spoken utterances from its users in real time. This chapter focuses on the on-line recognition of the communicative functions of user utterances, more specifically on the question of how the intended (multi-)functionality of dialogue utterances can be recognized on the basis of observable features of communicative behaviour in a data-oriented way.

The authors discuss and examine an incremental approach to dialogue utterance interpretation. A token-based approach combining the use of local classifiers, which exploit local utterance features, and global classifiers which use the outputs of local classifiers applied to previous and subsequent tokens, is shown to result in excellent dialogue act recognition scores for unsegmented spoken dialogue. This can be seen as a significant step forward towards the development of fully incremental, on-line methods for computing the meaning of utterances in spoken dialogue.

References

Basile, V., J. Bos, K. Evang, and N. Venhuizen (2012). Developing a large semantically annotated corpus. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey, pp. 3196–3200.

Blackburn, P. and J. Bos (2005). *Representation and Inference for Natural Language. A First Course in Computational Semantics.* Stanford: CSLI Publications.

Blackburn, P., J. Bos, M. Kohlhase, and H. De Nivelle (2001). Inference and computational semantics. In H. Bunt, R. Muskens, and E. Thijsse (Eds.), *Computing Meaning, vol. 2*, pp. 11–28. Dordrecht: Kluwer.

Bos, J. (2013). Recognising textual entailment and computational semantics. In H. Bunt, J. Bos, and S. Pulman (Eds.), *Computing Meaning, vol. 4*, pp. 47–74. Berlin: Springer.

Bunt, H. (1995). Dynamic interpretation and dialogue theory. In *The structure of multimodal dialogue II, M. Taylor, F. Neél, and D. Bouwhuis (eds.)*, pp. 139–166. Amsterdam: John Benjamins.

Bunt, H. (2000). Dialogue pragmatics and context specification. In H. Bunt and W. Black (Eds.), *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics.*, pp. 81–150. Amsterdam: John Benjamins.

Bunt, H. (2013). A methodology for designing semantic annotations. TiCC Technical Report TR 2013-001, Tilburg University.

Bunt, H. and R. Muskens (1999). Computational semantics. In H. Bunt and R. Muskens (Eds.), *Computing Meaning, vol. 1.*, pp. 1–32. Dordrecht: Kluwer.

Clark, S. and S. Pulman (2007). Combining symbolic and distributional models of meaning. In *Proceedings AAAI Spring Symposium on Quantum Interacction.*, Stanford, CA., pp. 52–55.

Cooper, R., D. Crouch, J. van Eijck, C. Fox, J. van Genabith, J. Jaspars, H. Kamp, D. Milward, M. Pinkal, M. Poesio, and S. Pulman (1996). *The FRACAS Consortium: Using the Framework.* Project FraCaS, A Framework for Computational Semantics, Deliverable D16, University of Edinburgh.

Dagan, I., R. Bar-Haim, I. Szpektor, I. Greental, and E. Shnarch (1996). Natural language as the basis for meaning representation and inference. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Pro-*

cessing., pp. 151–170. Berlin: Lecture Notes in Computer Science 4949, Springer.

Hobbs, J., M. Stickel, D. Appelt, and P. Martin (1993). Interpretation as abduction. *Artificial Intelligence 63*, 69–142.

ISO (2012). *ISO Standard 24617-2: Semantic annotation framework Part 2: Dialogue acts. September 2010.* Geneva: International Organisation for Standardisation ISO.

McCord, M. (1990). Slot grammar: A system for simpler construction of practical natural language grammars. In *Natural Language and Logic: International Scientific Symposium.*, pp. 118–145. Berlin: Springer.

Thomason, R. (1974). *Formal Philosophy. Collected papers by Richard Montague.* New Haven, CT.: Yale University Press.

van Deemter, K. (1996). Towards a logic of ambiguous expressions. In K. van Deemter and S. Peters (Eds.), *Semantic Ambiguity and Underspecification.*, pp. 203–237. Stanford: CSLI Publications.