THE LANGUAGE OF WORD MEANING

Edited by

Pierrette Bouillon
and Federica Busa

The Language of Word Meaning

Edited by

Pierrette Bouillon
University of Geneva

Federica Busa
Lexeme, Inc.

The Language of Word Meaning

This volume is a collection of original contributions from outstanding scholars in linguistics, philosophy, and computational linguistics exploring the relation between word meaning and human linguistic creativity. The chapters present different aspects surrounding the question of what is word meaning, a problem that has been the center of heated debate in all those disciplines that are concerned directly or indirectly with the study of language and of human cognition.

The discussions are centered around the newly emerging view of the mental lexicon, as outlined in the Generative Lexicon theory, which proposes a unified model for defining word meaning. The individual contributors present their evidence for a generative approach as well as critical perspectives, which provides for a volume where word meaning is viewed not only from a particular angle or from a particular concern, but from a wide variety of topics, each introduced and explained by the editors.

Pierrette Bouillon received a licence in classical philology and a DEA in information science from the University of Brussels, and a doctorate in linguistics from the University of Paris VII. She has been a researcher at ISSCO/TIM, University of Geneva for the past ten years, with particular interest in lexical semantics and machine translation. She has numerous refereed publications in international conferences, journals, and books, and she has served on the programme committees of several international conferences.

Federica Busa is director of knowledge engineering at Lexeme, a company developing natural language technology, in Cambridge, Massachusetts. She has a Ph.D. in computer science from Brandeis University and a Degree in Translation from the University of Geneva School of Translation and Interpretation (ETI), where she initially developed a strong interest in machine translation and lexicography. She is the author of a number of articles on computational semantics for natural language and has served on the committee of both the American and the European chapters of the Association for Computational Linguistics.

Studies in Natural Language Processing

Series Editor: Branimir Boguraev, IBM

Also in the series

List of Contributors

Nicholas Asher
DEPARTMENT OF PHILOSOPHY, UNIVERSITY OF TEXAS, AUSTIN, TX
  78712, USA
nasher@bertie.la.utexas.edu

Pierrette Bouillon
ISSCO/ETI, UNIVERSITÉ DE GENÈVE, 54 ROUTE DES ACACIAS, 1227
  GENÈVE, SWITZERLAND
Pierrette.Bouillon@issco.unige.ch

Federica Busa
LEXEME, INC. (LINGA MOTORS, INC.) 585 MASSACHUSETTS AVE.,
  CAMBRIDGE, MA 02139, USA
fede@lexeme.com

Nicoletta Calzolari
ISTITUTO DI LINGUISTICA COMPUTAZIONALE, VIA DELLA FAGGIOLA,
  32, 56126 PISA, ITALY
glottolo@ilc.pi.cnr.it

Salvador Climent
LABORATORI DE RECERCA EN LINGÜÍSTICA COMPUTACIONAL,
  UNIVERSITAT DE BARCELONA, SPAIN, GRAN VIA DE LES CORTS
  CATALANES 585, BARCELONA 08007, SPAIN
climent@fil.ub.es

Laurence Danlos
TALANA & LORIA, BP 239, 54506 VANDŒUVRE LÈS NANCY,
  FRANCE
laurence.danlos@linguist.jussieu.fr

Jerry A. Fodor
CENTER FOR COGNITIVE SCIENCE, RUTGERS UNIVERSITY, PSYCHOLOGY
  (ANNEX), BUSCH CAMPUS, NEW BRUNSWICK, NJ 08903, USA
fodor@ruccs.rutgers.edu

Elisabetta Gola
LEXEME, INC. (LINGA MOTORS, INC.), 585 MASSACHUSETTS AVE.,
  CAMBRIDGE, MA 02139, USA
gola@Lexeme.com

Jerry Hobbs
ARTIFICIAL INTELLIGENCE CENTER, SRI INTERNATIONAL, MENLO
  PARK, CA 94025, USA
HOBBS@ai.sri.com

Jacques Jayez
EHESS-CELITH, 54, BLD RASPAIL, 75006 PARIS, FRANCE
jjayez@dial.oleane.com

Adam Kilgarriff
ITRI, UNIVERSITY OF BRIGHTON, LEWES ROAD, BRIGHTON BN2 4GJ,
  UK
Adam.Kilgarriff@itri.bton.ac.uk

Alex Lascarides
CENTRE FOR COGNITIVE SCIENCE, UNIVERSITY OF EDINBURGH,
  2, BUCCLEUCH PLACE, EDINBURGH EH8 9LW, SCOTLAND,
  UK
alex@cogsci.ed.ac.uk

Alessandro Lenci
ISTITUTO DI LINGUISTICA COMPUTAZIONALE, VIA DELLA FAGGIOLA,
  32, 56126 PISA, ITALY
lenci@ilc.pi.cnr.it

Ernie Lepore
CENTER FOR COGNITIVE SCIENCE, RUTGERS UNIVERSITY, PSYCHOLOGY
  (ANNEX), BUSCH CAMPUS, NEW BRUNSWICK, NJ 08903, USA
lepore@ruccs.rutgers.edu

James McGilvray
DEPARTMENT OF PHILOSOPHY, MCGILL UNIVERSITY, LEACOCK
  BUILDING, 855 SHERBROOKE ST. W., MONTRÉAL, QUÉBEC H3A 2T7,
  CANADA
jim@philo.mcgill.ca

Monica Monachini
ISTITUTO DI LINGUISTICA COMPUTAZIONALE, VIA DELLA FAGGIOLA,
  32, 56126 PISA, ITALY
monica@ilc.pi.cnr.it

Julius M. Moravcsik
DEPARTMENT OF PHILOSOPHY, STANFORD UNIVERSITY, STANFORD,
  CA 94305, USA
julius@csli.Stanford.EDU

James Pustejovsky
COMPUTER SCIENCE DEPARTMENT, 258 VOLEN CENTER FOR COMPLEX
  SYSTEMS, BRANDEIS UNIVERSITY, WALTHAM, MA 02254, USA
jamesp@cs.brandeis.edu

Nilda Ruimy
ISTITUTO DI LINGUISTICA COMPUTAZIONALE, VIA DELLA FAGGIOLA,
  32, 56126 PISA, ITALY
nilda@ilc.pi.cnr.it

Patrick Saint-Dizier
IRIT-CNRS, 118, ROUTE DE NARBONNE, 31062 TOULOUSE CEDEX,
  FRANCE
stdizier@irit.fr

Piek Vossen
COMPUTER LINGUISTIEK, UNIVERSITEIT VAN AMSTERDAM, SPUISTRAAT
  134, 1012 AMSTERDAM, THE NETHERLANDS
piek.vossen@hum.uva.nl

Yorick Wilks
AI AND NN RESEARCH GROUP, DEPARTMENT OF COMPUTER SCIENCE,
  UNIVERSITY OF SHEFFIELD, REGENT COURT, 211 PORTOBELLO ST.,
  SHEFFIELD S1 4DP, UK
y.wilks@dcs.shef.ac.uk

Preface

What is the meaning of a word? How can the few hundreds of thousands of words we know be used to construct the many millions of utterances we make and understand in a lifetime? It would appear we need more words than we have available to us, if classical wisdom on this subject is to be believed. The subject, of course, is lexical semantics, and classical wisdom can often be wrong. This field has undergone a radical shift in recent years, in large part because of two developments. First, formal frameworks for word meaning have been developed that greatly simplify the description of lexical classes and their properties. Second, we have at our disposal new compositional techniques that allow us to view word meaning as an integral part of the overall process of semantic interpretation. These and other factors have made the issues relating to "the meaning of a word" some of the most central questions being addressed in the field of linguistics today. In fact, some classic issues have resurfaced with new data and arguments, such as the debate over analyticity and semantic knowledge, as well as the evidence of a distinction (or nondistinction) between lexical and world knowledge.

Waismann (1951) argued for what he called the "open texturedness" of terms. Although he was mainly interested in how the notion applies to the nonexhaustive nature of material object statements and the absence of conclusive verification conditions, there is another sense in which this is an interesting property of language and language use; the infinite variability of reference in language is the direct product of the *essential incompleteness* of terms and their composition. I would like to adopt this image as a characterization of what generative approaches to lexical and compositional semantics are attempting to model. From this perspective, word meanings are malleable and almost actively take on new shapes and forms in novel contexts of use.

By building a notion of "open texture" directly into word meaning, the formal mechanisms that give rise to sentence meanings will themselves ensure that both analytic and contextual aspects of interpretation are available in the model. In other words, it is the very functional nature of how words are modeled that allows them to exhibit their contextual variance, and with this enables the creative use of language.

This view of word meaning leads to an interesting comparison of two very different philosophical traditions in the study of language, namely, ordinary-language philosophy and analytic semantics. These two schools chose very different

linguistic units to analyze and from which to start their theorizing concerning language. The ordinary-language philosophers studied words and the functions of words in usage situations. Analytic semantics, on the other hand, derives from the study of sentences, sentence structures, and their propositional content, what has been called the logical syntax of the language (Carnap, 1932). The apparent incompatibility of these two approaches continues to dog certain philosophers, such as Fodor (1998), and has led to a rather pessimistic lack of concern over matters lexical.

The novelty of the generative lexical approach to language is the way in which these traditions and ideas are synthesized. What generative lexicon shares with ordinary language philosophy is a focus on words and word use. What it shares with analytic semantics is of course a concentration on the formalization of rules and types into a coherent and explicable system. The debt to the generative tradition in linguistics is almost too obvious to state: that meanings are compositional and recursive in nature (Chomsky, 1995). Furthermore, the generative devices available to the semantics are both more flexible than conventional approaches to compositional semantics such as Montague Grammar and more constrained than the view resulting from an arbitrary application of lexical rules.

The chapters in this volume constitute one of the first major collections addressing the synthesis of the issues mentioned above within this new paradigm. From the perspectives of philosophy, linguistics, computational linguistics, and lexicography, the fundamental problems of the creative use of language and the open texture of word sense are confronted. Although the contributors do not all agree with the basic principles of the generativity of lexical senses and composition, they have positioned their arguments in relation to this thesis, which is both useful and informative.

If the study of language is to bring us any closer to understanding concepts and the nature of our thoughts, it will be accomplished only by appreciating the importance of word meaning and the realization that compositional semantics is sensitive to word internal knowledge. It is, therefore, the very open texture of our language that reflects the compositionality of our thought.

James Pustejovsky

References

Carnap, R. 1932. *The Logical Structure of Syntax*, Littlefield, Adams and Company.
Chomsky, N. 1995. Language and nature. *Mind*, 104, 1–61.
Fodor, J. 1998. *Concepts*, Oxford, Oxford University Press.
Waismann, F. 1951. Verifiability. In Flew, A. (ed.), *Essays on Logic and Language*, Oxford, Basil Blackwell.

Introduction: Word Meaning and Creativity

Utterers create meanings by using words in context.
Hearers create interpretations.
Patrick Hanks, *SPARKLE Workshop*, Pisa, January 1999.

This quote from Patrick Hanks reflects very closely the spirit of this volume that tackles the relation between word meaning and human linguistic creativity. We are interested in pursuing the view that words are rich repositories of semantic information that people use to talk about the world in a potentially infinite number of ways. Our goal is to tackle the essence of words insofar as they provide a window onto human cognition and the compositional nature of thought. It is thus imperative that a theory of language addresses how lexical items contribute to the peculiar human ability that goes under the label of "linguistic creativity."

It is undeniable that words have "meanings" that go above and beyond the scope of linguistic research: They often carry the weight of a person's own experience. We are not aiming at exploring the unexplorable, but in proving that, given an appropriate set of methodological tools, a formal modeling of word meaning and linguistic creativity can be achieved. *Linguistic creativity* is a "generative" ability to extend the expressive possibilities of language in a potentially infinite number of ways. From the perspective of the lexicon (i.e., word meaning), it is the ability to give new meanings to words beyond their literal use.

As such, the overall task follows the strategy of contemporary generative syntax, which has achieved a basic understanding of how speakers produce and understand novel utterances and has brought simplicity to the great complexity underlying the syntactic structure of sentences. This became possible once the basic data set was isolated and regularities were identified in the massive variability of grammatical expressions.

From the viewpoint of the lexicon, however, creativity looks even more like a miracle. Researchers tend to disagree on what constitutes "linguistic" evidence (as opposed to context-dependent use of words) as well as what is the basic vocabulary for describing regularities in the behavior of lexical items. As a result, lexical semantics has been confined to a narrow domain of investigation, essentially dependent on a specific grammatical framework.

Broader issues surrounding the study of the meaning of words, how it is represented, and how it contributes to syntactic and semantic processing have been the

object of controversy among philosophers and the source of serious problems for researchers in computational linguistics and artificial intelligence.

This volume aims to bring together diverse contributions in philosophy, linguistics, computer science, and lexicography to highlight the breadth of the field and, at the same time, to point out how different issues converge around a common need: a framework establishing a set of principled properties of the lexicon and a unified vocabulary for representing word meaning. This cannot be achieved unless researchers isolate the data set to be studied and agree on the way in which these data should be understood.

1   Generativity in the Lexicon

One of the threads that links the contributions to this volume is the attempt to strengthen the foundations for defining the boundary between literal and extended (viz. creative) meanings. This requires two steps:

1. Narrowing and delimiting the definition of a word sense.
2. Defining the compositional processes/rules to extend the sense of a given word.

The second aspect takes advantage of a "strictly limited fragment of world knowledge" (see Asher and Lascarides in this volume) and defines the domains where components of meaning can be shifted. Thus, generativity in the lexicon is about "shifting" meaning in ways and across domains that are not random. To use the metaphors cited by Moravcsik (in this volume): If there is counterfeit money, then there must be real money. If there are meaning shifts, then there must be an underlying literal semantic representation from which the extended sense is projected.

Adopting and expanding Plato's metaphor, if there is a shadow then there must be an object that projects it and a source of light striking that object from a particular angle. In this volume, we take literal meaning to correspond to the object, the extended projection to the shadow, and the source of light to the specific compositional processes at work.

2   Generative Lexicon(s)

In this volume we have explored the lexicon from a particular perspective: a generative approach to word meaning as highlighted by the work in the Generative Lexicon (see Chapter 4 in this volume).

It is important to stress that this volume is not meant to be a defense of the Generative Lexicon (henceforth GL), but rather a way to focus the discussion around a concrete framework that has an important merit: It has taken an explicit stand toward establishing a methodology in lexical semantics and providing its own answer to the basic question of what is word meaning. Although its answers

are far from being uncontroversial or complete, such an explicit attempt allows for serious scientific discourse. Our goal is to present different views around a set of basic assumptions made in GL:

 a. The polymorphic nature of words can be studied in a principled way;
 b. There is a semantic vocabulary for structuring word meaning, which provides the input to the rules of semantic composition;
 c. There are generative rules of semantic compositions;
 d. The structuring of word meaning is an empirical question.

The polymorphic behavior of words is reflected in the ability of a lexical item to change meaning in different contexts. Nobody would deny this property of words: most of them appear in different contexts, with different meanings. This is what gives rise to the variability of interpretations and reflects speakers' creativity in using words in a variety of ways.

The more controversial question is how to represent and structure them. One solution is to take an *enumerative* (or monomorphic) approach, where the different senses are listed in the lexicon. This view, however, faces the problem of predicting a priori all the possible senses a word can take in context. As senses are only restricted by the number of possible contexts, it becomes questionable whether a semantics for natural language is even possible.

GL argues that compositional meaning takes place under predictable circumstances, provided there is an adequate set of principles for describing lexical structure. Senses need not be enumerated but can be derived *generatively* from richer representations of words and compositional mechanisms for combining them. The nature of individual lexical representations characterizes literal meaning, which involves two closely tied elements: analytic and conventionalized knowledge. The analytic component is the structuring itself, namely, *qualia structure*; the conventionalized component is represented by the value that fills each qualia role.

Irrespective of any particular framework, anyone studying word meaning has to address these points. We will present solutions within the generative lexicon, as well as alternative views of its critics.

3   Organization of the Volume

The volume is organized into four parts. The first part, *Linguistic Creativity and the Lexicon*, provides the philosophical foundations for the work presented in the book. James McGilvray presents a lucid discussion on the implication of Chomsky's creativity for a theory of the Lexicon. The three subsequent chapters approach the philosophical questions surrounding word meaning in the form of a debate. We have reprinted the debate between Jerry Fodor and Ernie Lepore on one hand, and James Pustejovsky on the other, which appeared in the 1998 spring issue of *Linguistic Inquiry*. In this volume, Yorick Wilks has joined in, contributing the view from three decades of research in artificial intelligence (AI).

Part two of the book, *The Syntax of Word Meaning*, opens with the contribution of James Pustejovsky, who presents new developments in GL, and then moves on to the analysis of fairly standard topics in lexical semantics: verb semantics (Pierrette Bouillon and Federica Busa; Jacques Jayez, and Patrick Saint-Dizier), partitive constructions (Salvador Climent), adjectives (Patrick Saint-Dizier), and causal relations (Laurence Danlos).

The third part of the volume, *Interfacing the Lexicon*, presents contributions on metonymy (Jerry Hobbs) and metaphor (Julius Moravcsik; Nicolas Asher and Alex Lascarides). Here the focus is on the need as well as the feasibility of appealing to structured lexical representations to study those phenomena that are on the edge between the lexicon and pragmatics. These views are challenged in the contribution by Adam Kilgarriff, who presents data against the usefulness of a generative-like approach to sense extension phenomena.

The last part of the book, *Building Resources*, contains contributions on the development of actual resources for natural language processing (NLP) using current developments in lexical semantics. The first two chapters present two aspects of the SIMPLE project: an effort sponsored by the European Commission to develop twelve computational semantic lexicons for the major European languages (Federica Busa, Nicoletta Calzolari, and Alessandro Lenci; Nilda Ruimy, Elisabetta Gola, and Monica Monachini). The last chapter, by Piek Vossen, discusses the EuroWordNet project, whose goal is to build Wordnets for a number of European languages.

Each part of the volume is introduced by a more extensive discussion of the chapters, in order to provide a roadmap for the reader, highlight the common issues raised by each set of papers, and stress their relevance toward our goals.

Pierrette Bouillon
Federica Busa

Contents

Linguistic Creativity and the Lexicon