Semantic Relations Between Nominals

Second Edition

Synthesis Lectures on Human Language Technologies

Editor
Graeme Hirst, ***University of Toronto***

Synthesis Lectures on Human Language Technologies is edited by Graeme Hirst of the University of Toronto. The series consists of 50- to 150-page monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

Semantic Relations Between Nominals, Second Edition
Vivi Nastase, Stan Szpakowicz, Preslav Nakov, and Diarmuid Ó Séagdha
2021

Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning
Mohammad Taher Pilehvar and Jose Camacho-Collados
2020

Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots
Michael McTear
2020

Natural Language Processing for Social Media, Third Edition
Anna Atefeh Farzindar and Diana Inkpen
2020

Statistical Significance Testing for Natural Language Processing
Rotem Dror, Lotem Peled, Segev Shlomov, and Roi Reichart
2020

Deep Learning Approaches to Text Production
Shashi Narayan and Claire Gardent
2020

Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics
Emily M. Bender and Alex Lascarides
2019

Cross-Lingual Word Embeddings
Anders Søgaard, Ivan Vulić, Sebastian Ruder, Manaal Faruqui
2019

Bayesian Analysis in Natural Language Processing, Second Edition
Shay Cohen
2019

Argumentation Mining
Manfred Stede and Jodi Schneider
2018

Quality Estimation for Machine Translation
Lucia Specia, Carolina Scarton, and Gustavo Henrique Paetzold
2018

Natural Language Processing for Social Media, Second Edition
Atefeh Farzindar and Diana Inkpen
2017

Automatic Text Simplification
Horacio Saggion
2017

Neural Network Methods for Natural Language Processing
Yoav Goldberg
2017

Syntax-based Statistical Machine Translation
Philip Williams, Rico Sennrich, Matt Post, and Philipp Koehn
2016

Domain-Sensitive Temporal Tagging
Jannik Strötgen and Michael Gertz
2016

Linked Lexical Knowledge Bases: Foundations and Applications
Iryna Gurevych, Judith Eckle-Kohler, and Michael Matuschek
2016

Bayesian Analysis in Natural Language Processing
Shay Cohen
2016

Metaphor: A Computational Perspective
Tony Veale, Ekaterina Shutova, and Beata Beigman Klebanov
2016

Grammatical Inference for Computational Linguistics
Jeffrey Heinz, Colin de la Higuera, and Menno van Zaanen
2015

Automatic Detection of Verbal Deception
Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari
2015

Natural Language Processing for Social Media
Atefeh Farzindar and Diana Inkpen
2015

Semantic Similarity from Natural Language and Ontology Analysis
Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain
2015

Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition
Hang Li
2014

Ontology-Based Interpretation of Natural Language
Philipp Cimiano, Christina Unger, and John McCrae
2014

Automated Grammatical Error Detection for Language Learners, Second Edition
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2014

Web Corpus Construction
Roland Schäfer and Felix Bildhauer
2013

Recognizing Textual Entailment: Models and Applications
Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto
2013

Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax
Emily M. Bender
2013

Semi-Supervised Learning and Domain Adaptation in Natural Language Processing
Anders Søgaard
2013

Semantic Relations Between Nominals
Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz
2013

Computational Modeling of Narrative
Inderjeet Mani
2012

Natural Language Processing for Historical Texts
Michael Piotrowski
2012

Sentiment Analysis and Opinion Mining
Bing Liu
2012

Discourse Processing
Manfred Stede
2011

Bitext Alignment
Jörg Tiedemann
2011

Linguistic Structure Prediction
Noah A. Smith
2011

Learning to Rank for Information Retrieval and Natural Language Processing
Hang Li
2011

Computational Modeling of Human Language Acquisition
Afra Alishahi
2010

Introduction to Arabic Natural Language Processing
Nizar Y. Habash
2010

Cross-Language Information Retrieval
Jian-Yun Nie
2010

Automated Grammatical Error Detection for Language Learners
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2010

Data-Intensive Text Processing with MapReduce
Jimmy Lin and Chris Dyer
2010

Semantic Role Labeling
Martha Palmer, Daniel Gildea, and Nianwen Xue
2010

Spoken Dialogue Systems
Kristiina Jokinen and Michael McTear
2009

Introduction to Chinese Natural Language Processing
Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang
2009

Introduction to Linguistic Annotation and Text Analytics
Graham Wilcock
2009

Dependency Parsing
Sandra Kübler, Ryan McDonald, and Joakim Nivre
2009

Statistical Language Models for Information Retrieval
ChengXiang Zhai
2008

Semantic Relations Between Nominals, Second Edition

Vivi Nastase, Stan Szpakowicz, Preslav Nakov, and Diarmuid Ó Séagdha

www.morganclaypool.com

ISBN: 9781636390864 paperback
ISBN: 9781636390871 ebook
ISBN: 9781636390888 hardcover

DOI 10.2200/S01078ED2V01Y202002HLT049

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES

Lecture #49
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Print 1947-4040 Electronic 1947-4059

Semantic Relations Between Nominals

Second Edition

Vivi Nastase
Institute for Natural Language Processing, University of Stuttgart

Stan Szpakowicz
School of Electrical Engineering and Computer Science, University of Ottawa

Preslav Nakov
Qatar Computing Research Institute, Hamad Bin Khalifa University

Diarmuid Ó Séagdha
Computer Laboratory, University of Cambridge

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #49

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

`Opportunity and Curiosity find similar rocks on Mars`. One can generally understand this statement if one knows that Opportunity and Curiosity *are instances of the class of* Mars rovers, and recognizes that, as signalled by the word on, rocks *are located on* Mars. Two mental operations contribute to understanding: recognize how entities/concepts mentioned in a text interact and recall already known facts (which often themselves consist of relations between entities/concepts). Concept interactions one identifies in the text can be added to the repository of known facts, and aid the processing of future texts. The amassed knowledge can assist many advanced language-processing tasks, including summarization, question answering and machine translation.

Semantic relations are the connections we perceive between things which interact. The book explores two, now intertwined, threads in semantic relations: how they are expressed in texts and what role they play in knowledge repositories. A historical perspective takes us back more than 2000 years to their beginnings, and then to developments much closer to our time: various attempts at producing lists of semantic relations, necessary and sufficient to express the interaction between entities/concepts. A look at relations outside context, then in general texts, and then in texts in specialized domains, has gradually brought new insights, and led to essential adjustments in how the relations are seen. At the same time, datasets which encompass these phenomena have become available. They started small, then grew somewhat, then became truly large. The large resources are inevitably noisy because they are constructed automatically. The available corpora—to be analyzed, or used to gather relational evidence—have also grown, and some systems now operate at the Web scale. The learning of semantic relations has proceeded in parallel, in adherence to supervised, unsupervised or distantly supervised paradigms. Detailed analyses of annotated datasets in supervised learning have granted insights useful in developing unsupervised and distantly supervised methods. These in turn have contributed to the understanding of what relations are and how to find them, and that has led to methods scalable to Web-sized textual data. The size and redundancy of information in very large corpora, which at first seemed problematic, have been harnessed to improve the process of relation extraction/learning. The newest technology, deep learning, supplies innovative and surprising solutions to a variety of problems in relation learning. This book aims to paint a big picture and to offer interesting details.

KEYWORDS

natural language processing, computational linguistics, lexical semantics, semantic relations, nominals, noun compounds, information extraction, machine learning, deep learning

Contents

Preface to the Second Edition

RELATIONS AND TEXTS

Every non-trivial text describes interactions and relations: between people, between other entities or concepts, between events. What we know about the world comprises, in large part, similar relations between concepts representing people, other entities, events, and so on. Such knowledge contributes to the understanding of relations which occur in texts. Newly found relations can in turn become part of the knowledge people store.

If an automatic system is to grasp a text's semantic content, it must be able to recognize, and reason about, relations in texts, possibly by applying and updating previously acquired knowledge. We focus here in particular on semantic relations which describe the interactions among nouns and compact noun phrases, and we present such relations from both a theoretical and a practical perspective. The theoretical exploration shows the historical path which has brought us to the current interpretation and view of semantic relations, and the wide range of proposals of relation inventories; such inventories vary according to domain, granularity and suitability for downstream applications.

On the practical side, we investigate the recognition and acquisition of relations from texts. We look at supervised learning methods. We present the available datasets. We discuss the variety of features which can describe relation instances, and learning algorithms successfully applied thus far. The overview of weakly supervised and unsupervised learning looks in detail at problems and solutions related to the acquisition of relations from large corpora with little or no previously annotated data. We show how enduring the bootstrapping algorithms based on seed examples or patterns have proved to be, and how they have been adapted to tackle Web-scale text collections. We also present a few machine learning techniques which can take advantage of data redundancy and variability for fast and reliable relation extraction.

Semantic relations play a fundamental role in ontology-based learning and information extraction from documents. They can also provide valuable information for higher-level language-processing tasks, including summarization, question answering and machine translation.

THE AUDIENCE

We expect that this book will appeal to graduate students, researchers and practitioners interested in computational semantics, information extraction and, more generally, modern natural language processing technology. We have tried to make the presentation broadly accessible to anyone with a little background in artificial intelligence. Even so, it helps to have some familiarity with computational linguistics and a modicum of tolerance for mathematical formulae. A basic understanding of machine learning is useful but not strictly necessary.

A NOTE ON THE SECOND EDITION

Two meaty chapters, 3 and 4, were the heart of the first edition. Most of the new material in this edition appears in an even more substantial Chapter 5. We have reorganized and edited all other parts of the book—Chapters 3 and 4 most thoroughly—to bring the facts, statistics and links seven years forward, and to correct previous omissions. The enlarged and restructured Chapter 1 delineates the topic of the book better, and explains what we will not discuss and why. There is also a brand new conclusion, now in Chapter 6. Chapters 3 and 4 aged well, although inevitably not all material we discussed seven years ago has survived intact on the Web. On the other hand, time flies in natural language processing. In the years since the book first appeared, deep learning has taken our discipline by storm. This edition brings the semantic relation research up to date with the new developments.

The substance of this edition owes its existence to Vivi, and the form to Stan. In particular, the new Chapter 5 is Vivi's brainchild; Stan helped whip it into shape. We are both grateful to Deniz Yuret for his constructive comments on much of Chapter 5, and to Preslav Nakov for a few incisive observations on its draft. Two anonymous reviewers made a number of most useful suggestions on the book: thank you.

Vivi Nastase and Stan Szpakowicz
January 2021

CHAPTER 1
Introduction

1.1 MOTIVATION

> *The connection is indispensable to the expression of thought. Without the connection, we would not be able to express any continuous thought, and we could only list a succession of images and ideas isolated from each other and without any link between them* [Tesnière, 1959].

The connection is indispensable. Any non-trivial text describes a group of entities and the ways in which they interact or interrelate. To identify these entities and the relations between them is a fundamental step in understanding the text. It is a step which human language users perform rapidly and reliably, assisted by their language skills and their world knowledge about entities and relations. If natural language processing[1] systems are to reach the goal of producing meaningful representations of text, they too must attain the ability to detect entities and extract the relations which hold between them. If a language understanding system is to adapt to new information just as people do, it must also use and then update existing repositories of knowledge about entities and about the way they interact.

Entity recognition (identify tokens which correspond to entity mentions) and entity resolution (identify the real-world entities or entity classes mentioned) are well-studied problems in NLP, with a voluminous associated literature. In this book, we will generally make the simplifying assumption that these steps have already been completed before the relation-processing stage—our main concern here—begins.

When a human reader interprets the relational content of a text, she draws on a spectrum of knowledge acquired from past experience, and on explicit and implicit signals in the text itself. Consider an example:

> NASA flew its final three space shuttle missions—one per orbiter remaining in the fleet—earlier this year.
>
> Discovery, one of the space shuttles in NASA's fleet, bound next year for the Smithsonian's Steven F. Udvar-Hazy Center in northern Virginia, was retired first in March. Endeavour landed June 1 and is now being prepared for display at the California Science Center in Los Angeles.

[1] The term *natural language processing* will henceforth be abbreviated to NLP.

> Atlantis flew the 135th and final shuttle mission, STS-135, last month. It will be exhibited near where it and all the other shuttles launched and most landed, at the Kennedy Space Center Visitor Complex in Florida.[2]

The first entity which this text mentions, NASA, must be recognized as referring to the U.S. space agency, not the National Auto Sports Association or the Nasa people of Colombia. As noted, we assume that this can be done before attempting to extract relational content. The recognition may trigger associations with a number of entities, such as SPACE, SPACE SHUTTLE, ORBITER, KENNEDY SPACE CENTER, DISCOVERY and ATLANTIS, some of which may appear in the text. Such associations—and the specific relations between the trigger concept and the triggered one—may help interpret the text.

The second entity mention refers to the *final three space shuttle missions*. *Space shuttle mission* is a compound of three nouns. It can only be fully interpreted by unpacking the semantic relations which hold between the concepts referred to as *space*, *shuttle* and *mission*—we already need relational processing! Roughly stated, *space shuttle mission* denotes a MISSION fulfilled (performed? aided?) by a SPACE SHUTTLE.[3] Even if the term *space shuttle mission* is unfamiliar, its meaning can be understood if we know enough about what *space shuttles* and *missions* are, and how they usually interact. Next question: what is a *space shuttle*? This term may be familiar, or can be looked up in a dictionary. While it is a relatively opaque compound, the reader can make an informed guess that such a thing moves around in space, if only she knows other uses of the word *shuttle* and sees the context. The parenthetical comment *one per orbiter remaining in the fleet* can be interpreted as implying that a SPACE SHUTTLE is a kind of ORBITER.

The second sentence in the example explicitly states that the entity referred to as *Discovery* is an instance of SPACE SHUTTLE. It does so by means of the construction `"X, one of Y"`. This information is very useful if we do not already know what DISCOVERY is. And so on; a full explanation of the relational content of a text tends to be much longer than the text itself.

Like a human reader, an NLP system must avail itself of both new information present in the text and pre-acquired knowledge about the world. The latter is often construed as a *knowledge base*, and its content may be either fixed or dynamically updated as more texts are read. So, there is a clear interaction between the tasks of knowledge acquisition and text understanding. The interest in the organization of knowledge and the principles behind the understanding of utterances can be traced back to classical antiquity. In modern times, the two tasks have been the object of computational research throughout the history of artificial intelligence (AI), even though they have often been treated as separate problems. Their tight interaction came into clear focus rather recently. It was noted when people began to develop text understanding systems—the task requires lexical, world and common-sense knowledge. At the same time, such automated text-understanding systems could produce formal representations of the knowledge in texts, and those representations could be used to build, or to add to, knowledge bases. This book surveys

[2] www.space.com/12804-nasa-space-shuttle-program-officially-ends.html

[3] This is not the only interpretation. The compound may also denote a *shuttle mission* performed in *space*.

the research landscape and the state of the art in these often intertwined tasks. Many questions arise when one deals with relations between entities. Here are some questions worth considering.

- Can one design a parsimonious and general representation of the semantic relations which appear in text?
- What linguistic signals in text can help identify semantic relations?
- Can background knowledge enhance the understanding of relations in text?
- Can the understanding of relations help acquire new world knowledge and distinguish relational information of lasting value (Discovery is a space shuttle) from contingent information with a one-off benefit (Endeavour is now prepared for display)?

1.2 APPLICATIONS

There is much promise in the ability to identify semantic relations in texts and to locate relations in structured knowledge repositories. Swanson [1987] demonstrated the potential of using relations from text for knowledge discovery. He combined relations extracted from articles in different scientific domains, and discovered previously unknown but ultimately important connections between, e.g., fish oil and blood circulation or magnesium and migraines. The biomedical literature has been growing at a double-exponential pace [Hunter and Cohen, 2006], so researchers find it impossible to keep track of everything that is being published. Yet, billions of dollars can be saved by finding out what costly experiments have already been done, and what their outcomes were, and by discovering likely new interactions between known concepts. There are simply no practical alternatives to the automatic relation extraction from text in biomedicine.

Many of the techniques we discuss in this book, already quite mature, have worked in some of these NLP applications. For example, Nakov's [2008b] work helps improve *Statistical Machine Translation*. Suitable paraphrases make explicit the hidden relations between the nouns in a noun compound. This makes it easier to translate English noun compounds into languages in which phrases are not as compact, and enables the recognition of different variants of the same phrase. Suppose that the phrase *oil price hikes* is interpreted as *hikes in oil prices* and *hikes in the prices of oil*. A system may find it easier to translate the more syntactically parallel *hikes in the prices of oil*—rather than the very compact *oil price hikes*—into Spanish as *alzas en los precios del petróleo.*

Another case in point: information retrieval and question answering. Suppose that one wants to ask a search engine what causes cancer. Many causes are possible, so one might want to pose this query: "list all x such that x causes cancer". This can be seen as a special kind of search, called *relational search* [Cafarella et al., 2006]. It asks for a list of things in a given relation with a given entity. Relational search can find components of objects (every x which *is part of* an automobile engine), materials with specific properties (every x which *is material for* making a submarine's hull), or types of entities (every x which *is a type of* transportation). Naturally, these examples are just the tip of the iceberg.

1.3 WHAT THIS BOOK IS ABOUT

1.3.1 RELATIONS BETWEEN NOMINALS

The book talks about relations between entities mentioned in the same sentence, and expressed linguistically as nominals. Relations are the connections we perceive among concepts or entities. A connection may come from general knowledge about the world (CHOCOLATE *is-a-kind-of* FOOD), or from a text fragment (`Chocolate is a psychoactive food`).[4] When one talks casually about a relation, one refers either to its type, such as *part-of* or *is-a*, or to its instance in which arguments accompany the relation name, such as `chocolate contains caffeine`. Throughout the book, we write simply "relation" if the context makes it clear which of the two usages is intended; otherwise we write "relation type" or "relation instance".[5]

The term *nominal* usually refers to a phrase which behaves syntactically like a noun or a noun phrase [Quirk et al., 1985, p. 335]. For our book, we have adopted a narrower definition. A *nominal* can be a *common noun* (`chocolate`, `food`), a *proper noun* (`Godiva`, `Belgium`), a *multi-word proper name* (`United Nations`), a *deverbal noun* (`cultivation`, `roasting`), a *deadjectival noun* ([the] `rich`), a *base noun phrase* built of a head noun with optional premodifiers (`processed food`, `delicious milk chocolate`), and recursively a sequence of nominals (`cacao tree`, `cacao tree growing conditions`).

The relation itself can be signalled by a phrase which links the entity mentions in a sentence (`Chocolate` *is a* `raw or processed food produced from the seed of the tropical Theobroma cacao tree.`), or it can be only implied, e.g., when the entity mentions are compressed into a noun compound (consider `cacao tree` and `cacao tree growing conditions` again).

Superficially, it seems easier to learn or detect relations when some linguistic clues exist than when the relation is only implied by the adjoining of terms. We will see, however, that, in order to rely on the linguistic expression of relations in texts, one must deal with *ambiguity*. For example, the word `in` may indicate a temporal relation (`chocolate` *in* `the 20th century`) or a spatial relation (`chocolate` *in* `Belgium`). Another difficulty is *over-specification*. Consider, for example, the ornate relation between chocolate and cultures in `Chocolate` *was prized as a health food and a divine gift by* `the Mayan and Aztec cultures.` When there are no surface indicators, the clues about the type of relation will come from knowledge about the entities (`milk chocolate`: CHOCOLATE, which is a kind of FOOD made with lots of MILK, which is an INGREDIENT of many FOODSTUFFS).

A special situation arises when an entity is actually an occurrence—event, activity or state—expressed by a deverbal noun such as `cultivation`. Relations between a deverbal noun and its modifiers mirror the relations between the underlying verb and its arguments. For example, in the clause `the ancient Mayans cultivated chocolate`, chocolate is the *theme*. So, one

[4] www.cacao-chocolate.com

[5] Chapters 1–4 will observe the following font conventions: *relation*, ENTITY/CONCEPT, `text/example/pattern`.

can also discern a *theme* in `chocolate cultivation`. We do not single out such relations for separate discussion, because the methods do not differ significantly from what is required to deal with any other relations and any other types of nominals.[6]

1.3.2 RELATIONS IN KNOWLEDGE REPOSITORIES

This book is also about relations between entities, stored for use outside a specific textual context. The relations describe the same connections as those found in a textual context, which explain how these entities interact. The difference is the absence of context. Depending on the knowledge repository, one would associate a universal or an existential quantifier with a particular relation instance. For example, the universal quantifier applies to a WordNet-style lexical semantic relation such as APPLE *is-a* FRUIT; this is always true (unless our knowledge of botany becomes revised). On the other hand, an existential quantifier should be associated with each relation instance in ConceptNet, where there are such relations as CHARACTER *located_in* NOVEL or CHARACTER *located_in* A PLAY IN A THEATRE. In repositories like Freebase or Cyc,[7] there are relations which are, or have been, true for a specific time interval, e.g., BARACK OBAMA *president_of* UNITED STATES OF AMERICA. While some systems give additional attributes to relation instances (including time span), our book will not focus on that.

By *knowledge repository*, we mean a variety of resources which store semantic relations. They have somewhat different properties and applications. When it is relevant, we will refer to a particular type.

A **taxonomy** categorizes things or concepts. Taxonomies are often based on the *is-a* relation, or on hyponymy/hypernymy when organizing linguistic information. The relation brings about the hierarchical structure of the taxonomy.

An **ontology** captures general knowledge. It covers a variety of relations one needs to express such knowledge. Ontologies often include taxonomies: *is-a* and *part_of* relations usually belong to the inventory. An ontology is built for a specific domain, where the pre-specified types of relevant concepts and relations—the ontology schema—are the scaffolding for the resource.

A **knowledge base (KB)** is a collection of facts which express knowledge, either general or in a narrower domain. Like ontologies, KBs contain a variety of relation types and concepts. Unlike ontologies, they are not structured around pre-specified schemata; new facts—maybe with new nodes or new types of edges—may be added *ad hoc*.

A **knowledge graph (KG)** is any of the knowledge repositories noted above if it is perceived as an interconnected network of entities—as opposed to separate relation instances. The graph is a powerful and expressive mathematical construct. This view of knowledge repositories makes possible a variety of solutions to issues in relation extraction and classification, as the book will show. In the context of KGs, the task of relation classification becomes *link prediction*: build a

[6] Even so, nominalization has been treated differently in some linguistic theories [Levi, 1978] and in some computational linguistic work [Lapata, 2002].

[7] The book will revisit both of them repeatedly.

model based on existing information in the graph, and use it to predict additional links (i.e., relations) between nodes in the graph.

1.4 WHAT THIS BOOK IS NOT ABOUT

1.4.1 ARGUMENT IDENTIFICATION – ENTITY RECOGNITION – WORD SENSE DISAMBIGUATION

We will not deal separately with *argument* identification, although it will be discussed in the context of the task of simultaneous argument identification and relation classification. A simplifying assumption is often made: argument identification (including possibly entity identification/disambiguation) is a task separate from relation extraction. That is why one can employ a pipeline system which first identifies, and possibly disambiguates, the entities of interest, and only then moves on to identifying semantic relations. In evaluation settings such as the ACE or SemEval relation classification tasks,[8] gold-standard entity annotations are usually part of the dataset. In some cases, such annotations may also be linked to WordNet or to another semantic network, and so mimic the output of a sense-disambiguation step.

1.4.2 DISCOURSE RELATIONS

Apart from relations in a noun compound and between entities in a sentence, there are other relations between nominals in a text, notably discourse relations. Coreference relations in particular are not included in this survey. They usually cross sentence boundaries, their arguments may be complex noun phrases (`the girl next door`) or pronouns, or they may not be explicitly expressed if ellipsis is at work. In the text below, for example, `Angela Merkel` and `German chancellor` are co-referents; the ellided noun `meeting` is marked with square brackets.

> Angela Merkel's spokesman has insisted that the German chancellor's first meeting with François Hollande, France's president-elect, will be a "getting to know you" exercise, and not [a] "decision making" [meeting].

1.4.3 TEMPORAL RELATIONS

Temporal relations between nominals, such as *morning exercise* or *afternoon snack*, are quite frequent. While they belong to the inventory of semantic relations we review in Chapters 2 and 3, they most commonly hold between two events, or between an event and a time indicator. When temporal relations are studied separately, not as part of a more general analysis of semantic relations between nominals, the emphasis is on events, which are often expressed by verbs. That is why we will single out neither the datasets for work with temporal relations,[9] nor the methods designed to detect such relations.

[8] We will return to both in Chapter 3.

[9] Consider the TempEval task (paperswithcode.com/sota/temporal-information-extraction-on-tempeval-3) or the clinical TempEval task (competitions.codalab.org/competitions/15621).

1.4.4 ONTOLOGY BUILDING / KNOWLEDGE BASE POPULATION

The book concentrates on relations between nominals in texts and in knowledge repositories: what they are and how one can identify or recognize them. We will discuss relations in existing ontologies, KBs and other repositories only insofar as they are relevant to semantic relations between nominals. While building such resources may also require interesting methods or techniques, that is not our focus. They are, however, of particular interest as sources of seed examples or training data for learning or for developing other techniques of extracting relation instances. Also, the purpose of relation extraction often is the enrichment of knowledge repositories; this *will* be discussed in the book.

There is a similar task: knowledge base population (KBP), run annually since 2009 as part of the Text Analysis Conferences (TAC);[10] Ji and Grishman [2011] describe the task, and discuss the systems which participated in the 2010 edition. KBP does not, strictly speaking, perform relation extraction or classification but the actual tasks—entity linking and slot filling—are relevant to relation extraction. *Entity linking* detects mentions of entities of predefined types, and links them to the entities in a knowledge repository provided. *Slot filling* takes an entity and relation types (e.g., Barack Obama, *birthPlace*, *birthDate* and *marriedTo*), and fills in the missing arguments. The book will not look at the particular data and systems for KBP but it will review some of its methods which touch on relation classification or extraction.

1.4.5 DATABASES AND SOCIAL NETWORKS

Databases and social networks do capture relations between entities but this book will not discuss them.

It is not always obvious where the boundary between KBs and databases lies. According to Brodie and Mylopoulos [1986], "an important difference between KBs and databases is that the former require a semantic theory for the interpretation of their contents, while the latter require a computational theory for their efficient implementation on physical machines". Our guiding principle is this: the knowledge repositories we consider here (taxonomies, ontologies, KBs and KGs) contain concepts and relations between them relevant to a wide range of applications. As positive examples, consider WordNet's lexical-semantic knowledge and Freebase's world knowledge focused on relations which connect such varying entities as people, places, artifacts, and so on. Negative examples include a network of interconnected publications, authors and citations, and a database of movie reviews: movies, users and ratings. The latter has been popular in statistical relational learning, and was among the first to undergo collective matrix factorization. (Matrix factorization is a successful technique of link prediction in KGs; Section 5.3.2 will discuss these matters at length.)

A social network is by its nature a graph, and one can interpret some databases as graphs. Knowledge repositories are often processed as graphs. The structure thus perceived benefits sev-

[10] tac.nist.gov/tracks/

eral typical operations on a repository, which rely on the interconnectedness of relation instances: derive entity and relation representations, find links, weigh entities and relations, and so on.

Because of the commonalities between databases and KBs, and between social networks and KBs, there is an overlap in the methods which can be applied to these structures to create new information or new models. Some techniques—e.g., link prediction by matrix factorization or statistical relational learning, or link prediction in social networks—have been first applied to such data. We acknowledge the proponents of such new methods but our focus is the discussion of research which explicitly applies such methods to semantic relations in texts and KBs. On the other hand, some of the techniques developed for databases and social networks are suitable to knowledge repositories but have not been applied yet. We leave it to the reader to discover such innovative applications of established methods, and to bring them into the field of semantic relations.

1.4.6 RESULTS AND COMPARISONS

The focus of this book is the interplay between linguistic information and clues on one hand, and formal organization of such information and learning methods on the other. A reader who needs a method for a concrete purpose and a concrete dataset may desire a ranking of the suitable methods by their appropriateness to the task. The variety of datasets and performance measures precludes a neat summary in a few tables, which would rank by their results the methods surveyed here. The level of performance is affected by a number of factors which interact in intricate ways: the annotation procedure for the training data, the amount of training data available, the number and nature of relations to be classified, the distribution of those relations in data, the source of the data for training and testing, the corpora and methods used for additional information (e.g., when building word representations or obtaining relational features), and so forth. Each domain—biology, medicine and so on—adds its own peculiarities.

The MUC and ACE shared evaluation tasks (which will be discussed in Section 3.2.1), as well as the tasks at 2007 and 2010 SemEval (Section 3.2.2), were meant to provide a benchmark for relation classification. They have been instrumental in advancing research on relation classification and extraction, but their small-scale datasets and their methods have long been superseded. Many conclusions drawn from the results published earlier have been rendered obsolete by more recent developments.

There is also a variety of evaluation measures, depending on the task. Precision, recall, F-score and accuracy are used in traditional relation classification tasks. In link prediction, in particular, the most commonly used measures are mean reciprocal rank (MRR), HITS@k (a precision measure computed on the first k ranked predictions), and mean average precision (MAP). Furthermore, summary results on a dataset hide the behavior on individual relation types. Good performance on one type does not ensure high performance overall.

To be blunt: a synoptic view of such heterogeneous results might be misleading rather than informative. An intrinsic evaluation of the extracted knowledge may also not be desirable. Relation extraction is not an end task. Its purpose is to build resources for use in other NLP and AI applications.

On the other hand, the NLP community (much like other scientific communities) exhibits a strong bias toward publishing positive results. This ensures that any published method has in some way improved on those it builds upon. It could be relatively easy to decide which methods perform better.

There is a more positive trend in the NLP community's publishing habits. The experimental data *and* the code are more and more often made public. That offers up a wide variety of handy methods which should be relatively easy to try out, especially on new but similar data. And even more helpfully, ML and NLP methods have been employed to gather and extract results from scientific publications. Papers with Code,[11] one of those wonderfully useful initiatives, group papers, code and results for specific tasks, and show the state-of-the-art for each of them.[12]

We encourage the reader to absorb the explanations about features and learning methods presented in the book, and then—guided by their own requirements—choose the most appropriate starting point and innovate from there.

1.5 ORGANIZATION OF THE BOOK

This brief chapter has explained why we found it worthwhile to write a book about recognizing semantic relations between nominals, what applications those recognized relations can facilitate, what the book does and what it does not discuss.

Chapter 2 recalls the history of the evolution of semantic relations in two separate but eventually intertwined threads: as relations in knowledge repositories, serving to organize our knowledge, and as connections we perceive between concepts and ideas expressed in texts. We show the progress of research on the design of lists of semantic relations, the change in understanding what a semantic relation is, and the lessons learned from that long-lasting enterprise.

Chapter 3 presents the supervised learning perspective: from annotated data sets, to features, to machine-learning formalisms employed to make the most of the available, and sometimes complex, features. We focus on methods of relation learning which can successfully build models from small, annotated datasets.

Chapter 4 goes to the other end of the learning and data spectrum. It surveys the unsupervised and distantly supervised learning of semantic relations, and shows how they can take advantage of large (unstructured) textual data. The early unsupervised methods have proved surprisingly robust and resilient. They have led to variations applied successfully even now. Distant supervision, which links unsupervised and supervised methods, has been the source of numerous ideas for producing large-scale training data from sometimes surprising sources.

[11] paperswithcode.com

[12] See paperswithcode.com/area/natural-language-processing/relation-extraction for the state-of-the-art in several benchmarking datasets.

Chapter 5, new to this edition, restates—in the context of deep learning—the matters discussed in Chapters 3 and 4. Neural networks date back to the 1940s but they have only quite recently shown their huge potential in NLP tasks, once their powerful mathematical basis could be backed up by an equally powerful (soft and hard) computational support. The adoption of this paradigm has opened up new avenues of research in semantic relation learning.

Chapter 6 wraps up the book with a look back at the landscape we have sought to describe. We re-emphasize the connections between the various ideas and techniques presented here. The goal is to leave the reader with a coherent, clear and informative picture of semantic relations between nominals.

CHAPTER 2

Relations Between Nominals, Relations Between Concepts

2.1 INTEGRATION OF KNOWLEDGE AND TEXTS IN TWO THOUSAND YEARS

Semantic relations describe interactions. A relation may connect nominals in a text or concepts in knowledge representation, depending on the level at which an interaction is perceived. It seems artificial to separate relations just because of this distinction, especially if one notes that the knowledge is ultimately expressed by words, and entities in particular are expressed by nominals. Indeed, we are at a stage in NLP when large amounts of textual data can help identify pairs of interacting entities. In the end, this information can be gathered and formalized to build large-scale knowledge repositories. Early work on understanding the nature of knowledge relied on contemplating the world and the objects within it, and on applying certain organizational principles to the structuring of the insights about the relations between objects in the world. Work on the study of language and the way it conveys meaning has been evolving separately until the 20th century.

The attempts of philosophers to capture and describe our knowledge about the world go back to the Antiquity. Artistotle's *Organon*, a posthumous collection assembled by his students, includes a treatise on *Categories* which presents criteria for organizing objects [Studtmann, 2008]. This endeavor must inevitably deal with language, given that world knowledge and language are intertwined. Objects in the natural world are put into categories called τὰ λεγόμενα (*ta legomena*, things which are said), and their organization is based on the relation of class inclusion.

For two millennia following Aristotle, contemplation on the nature of knowledge and on the principles of knowledge organization has been the domain of philosophers—and occasionally botanists or zoologists who would put living things into taxonomies. Let us fast-forward over centuries of hot philosophical debate, coloured by changing ideas about what concepts are and how they relate to the real world [Margolis and Laurence, 1999]. In the 1970s came a realization that a robust AI system needs the same kind of knowledge as what humans have. This revelation has spurred a concerted effort to capture and represent knowledge in a machine-friendly format, and at that point the intermingling with language became inevitable.

We now return to the Antiquity to pick up the language analysis thread and follow it to this intermingling point. During the second half of the first millennium BCE, scholars of the

Indian linguistic tradition (*vyākaraṇa*) developed a highly refined theory of language. It covered what we would now describe as morphology, syntax and semantics. The seminal document of this tradition was written by the celebrated scholar Pāṇini, often considered "the father of linguistics". The *Aṣṭādhyāyī* is an eight-volume collection of aphoristic rules which describe the process of generating a Sanskrit sentence from what would now be called a semantic representation. The latter is primarily conceptualized in terms of *kārakas*, semantic relations between events and participants—that is now studied under the name of *semantic roles*. The *Aṣṭādhyāyī* covers noun-noun compounds comprehensively from the perspective of word formation but only refers in passing to the semantics of such compounds. Subsequent commentators such as Kātyāyana and Patañjali expand on these semantic issues; they point out, for example, that compounding is only supported by the presence of a semantic relation between entities [Joshi, 1968].

Much closer to the modern day, early in the 20th century, Ferdinand de Saussure proposed a hugely influential *Course in General Linguistics* [de Saussure, 1959]. He distinguished between syntagmatic and associative relations, which *"correspond to two different forms of mental activity, both indispensable to the workings of language"*. A syntagmatic relation holds between two or more terms in a sequence *in praesentia*, in a particular context: *"words as used in discourse, strung together one after the other, enter into relations based on the linear character of languages—words must be arranged consecutively in spoken sequence. Combinations based on sequentiality may be called syntagmas."* Associative (paradigmatic) relations, on the other hand, come from accumulated experience and hold *in absentia*: *"Outside the context of discourse, words having something in common are associated withgether in the memory. [...] All these words have something or other linking them. This kind of connection is not based on linear sequence. It is a connection in the brain. Such connections are part of that accumulated store which is the form the language takes in an individual's brain."* Word associations can be morphological, phonological, grammatical or semantic.

Syntagmatic and associative relations interact in the understanding of text: word associations are summoned for the interpretation of a *syntagma*. Without the associations, a *syntagma* would have no meaning of its own. Interestingly, de Saussure's *Course* did not propose any list of relations. Harris [1987] observed that frequently occurring instances of syntagmatic relations may become part of our memory, and so turn paradigmatic. This parallels Gardin's [1965] proposal: that instances of paradigmatic relations be derived from accumulated syntagmatic data. By the way, this reflects current thinking on relation extraction from open texts.

From the point of view of structure, ontologies and texts sit at the two ends of the spectrum. The rise of formal semantics at the end of the 19th century began to bridge this gap. Starting with the work of Gottlob Frege [Frege, 1879], predicate logic and its extensions [L. T. F. Gamut, 1991] have been the standard analytical toolkit for philosophers of language. Predicate logic is an inherently relational formalism. A predicate takes one or more arguments; when modelling language, predicates which take multiple arguments usually encode semantic relations. Here is how a simple logical representation of the sentence `Google buys YouTube` might look:

$$\mathit{buy}(\texttt{Google}, \texttt{YouTube})$$

A representation like this is still commonly used in presentations of computational models. In another representation, sometimes called *neo-Davidsonian* after the philosopher Donald Davidson, additional variables represent the event or relation as something that can be explicitly modified and subject to quantification, for example:

$$\exists \mathrm{e}\ \mathit{InstanceOfBuying}(\mathrm{e}) \wedge \mathit{agent}(\mathrm{e}, \texttt{Google}) \wedge \mathit{patient}(\mathrm{e}, \texttt{YouTube})$$

or perhaps

$$\exists \mathrm{e}\ \mathit{InstanceOf}(\mathrm{e}, \mathrm{Buying}) \wedge \mathit{agent}(\mathrm{e}, \texttt{Google}) \wedge \mathit{patient}(\mathrm{e}, \texttt{YouTube})$$

Charles Sanders Peirce was one of the great thinkers to whom people could turn for inspiration for representing knowledge and relations. His accomplishments include the *existential graphs*, which rely on the notion of an *existential relation* R [Peirce, 1909]: "anything that is R to x (where x is some particular kind of object) is nonexistent in case x is nonexistent. Thus, lovers of women with bright green complexions are nonexistent in case there are no such women."

From a mathematical point of view, relations had a dual nature. In logic, they served as predicates; in graphs (later known as semantic networks), they labelled arcs between vertices which represented concepts. AI chose the representation in logic to support knowledge-based agents and inference; the idea of a graph of concepts has been adopted to represent factual knowledge, prevalent in NLP [Russell and Norvig, 2020, chapter 10]. The latter has had a strong effect on the type of relations represented. In graphs, it is quite natural to represent binary relations. Binary relations have become the *de facto* norm in ontologies built from texts, and are by far the majority of relations which NLP targets for extraction.

The advent of the computer has brought about an interest in putting this new tool to the kind of tasks which people do, and thus the field of AI began. John McCarthy was the first to describe a complete, even if hypothetical, AI program [McCarthy, 1958]. Designed to apply general world knowledge in search of solutions to a problem, the program relied on two separate components, one for knowledge represented as rules and one for reasoning mechanisms. As a logic-based system, it already built upon relational information, but it was not directly concerned with language.

Linguistically oriented AI systems soon followed. Work such as Winograd's [1972] groundbreaking interactive English dialogue system or Charniak's [1972] study on understanding children's stories demonstrated that semantic knowledge about a variety of topics is essential to computational language comprehension. That was a conceptual shift from the "shallow" architecture of primitive conversation systems such as ELIZA [Weizenbaum, 1966] and first-generation machine translation systems. The need for storehouses of background knowledge to support reasoning systems has led in several directions; one of them was the creation of large-scale hand-crafted ontologies such as Cyc [Lenat and Guha, 1990]. A more recent, and perhaps more fruitful, direction was the acquisition of collections of propositional facts about the world via volunteer contributions over the Web; this trend began with OpenMind Common Sense [Singh et al., 2002], and MindPixel,[1] and has reached truly large scale with Freebase.[2]

At the crossroads between knowledge and language, we encounter interconnected systems in which knowledge about words and their various meanings is expressed in terms of their relations to other words. The idea of defining the meaning of a word by its connections to other words is familiar to any user of a dictionary. Spärck Jones [1964] suggested that the kind of lexical relations found in a dictionary could be formalized and learned automatically from text. Around the same time, Quillian [1962] proposed the *semantic network*, a graph in which meaning is modelled by labelled associations between words. The vertices of the network are concepts onto which one maps the words in a text, and then connections—relations between concepts—are established on arcs linking some of the words.

The network-based style of representation has remained very influential. It informs large-scale lexical resources such as WordNet [Fellbaum, 1998], a network whose latest active version—now over a decade old—had over 155,000 words (nouns, verbs, adjectives, adverbs) and over 117,000 groups of near-synonyms called *synsets*.[3] WordNet represents over twenty semantic relations between synsets, including synonymy, antonymy, hypernymy (a *sandwich* is a kind of *snack food*), hyponymy (*snack food* has a *sandwich* among its kinds), meronymy (*bread* is part of a *sandwich*) and holonymy (a *sandwich* has *bread* as a part). Section 2.2.2 will revisit WordNet's relations as examples of relations between concepts.

The early work on manual knowledge acquisition has quickly made it apparent that the process must be automated.[4] Luckily, work on text analysis has revealed that much of the knowledge we wish to extract is contained in texts. In parallel, methods of finding the structure in free-form texts—in the shape of part-of-speech and grammatical parsing—have been developed to fill in more of the gap between structured ontologies and unstructured texts. Work on the automatic construction of KBs from text collections took off with Hearst's [1992] pioneering research. At the beginning, the focus was on the relations which are the backbone of ontologies, Hearst's *is-a* and Berland and Charniak's [1999] *part-of*. The bootstrapping techniques developed for these relations were then applied to other relations; see for example Ravichandran and Hovy [2002] and Patwardhan and Riloff [2007].

One can observe that having specific targets for relation extraction may cause the omission of a wealth of information in texts. This has led to open information extraction. In this paradigm, one begins by hypothesizing about how a relation may be expressed, e.g., as a pattern over parts of speech [Fader et al., 2011], a path in a syntactic parse tree [Ciaramita et al., 2005], or a sequence of high-frequency words [Davidov and Rappoport, 2008b]. Next, all matching instances are

[1]The project has been dormant since 2005. See en.wikipedia.org/wiki/Mindpixel for a bit of history.

[2]Freebase was frozen in 2016.

[3]See wordnet.princeton.edu, in particular the WordNet 3.0 statistics at wordnet.princeton.edu/documentation/wnstats7wn.

[4]We accept the descriptor "manual", prevalent when the NLP literature talks about people creating language resources. But: the word means something *worked or done by hand and not by machine* (www.merriam-webster.com/dictionary/manual). Ontology creation, rule design, knowledge acquisition and text annotation are *intellectual* activities, not handiwork.

extracted as candidate relation instances. The downside of such methods is the high variability in relation expressions; a mapping onto a set of "canonical" relation expressions is the subject of ongoing work.

2.2 A MENAGERIE OF RELATION SCHEMATA

The two threads of parallel work, on the organization of knowledge and on texts, have led to two perspectives on relations. A relation manifests itself in text at the word level, and arises from the particular context in which it appears; we look at how relations between nominals have attained prominence in NLP research. In ontologies and other taxonomies, relations connect concepts, expressing facts considered (believed) to be true in view of the current state of our collective understanding of the world; we look at relations in a few of the knowledge repositories which have been applied in NLP, and at the trouble semantic relations may cause when they are used in ontologies. The next two sections briefly survey these two perspectives, and the different ways in which they have been turned into practice.

2.2.1 RELATIONS BETWEEN NOMINALS

Standard lexical-semantic literature discusses semantic relations at great length. We recommend that the interested reader consult Geeraerts's [2010] comprehensive monograph and the citations therein to all the classic publications. This section shows the evolution of work on determining, or building, a list of relations with coverage wide enough for text analysis—complete coverage, if at all possible. The work has first concentrated on people, trying to find out what kind of connections they perceive between various word combinations. Later, the focus shifted to data in attempts to design a list of relations which covers all the connections perceived in the texts under consideration. Then the nature of the texts changed—from general-purpose news texts or literature to texts in specialized domains like biology or medicine—and that caused another shift in perspective.

Casagrande and Hale [1967] attempted to build a *list* of semantic relations by asking native speakers of an exotic language to give definitions for a predetermined list of words. They analyzed the definitions into declarative sentences which state simple facts, and determined the relations expressed in those sentences. The result, a list of 13 relations not only between nominals, appears in Table 2.1.

Chaffin and Herrmann [1984] presented an exercise in the analysis of relations themselves, and of their distinguishing properties. They explored human perception of similarities and differences between relations via an exercise in grouping instances of 31 semantic relations. The results of the experiment showed that the subjects perceived five classes of semantic relations—see Table 2.2. Instances of these five classes can be distinguished by three properties: contrasting/non-contrasting, logical/pragmatic, and inclusion/non-inclusion.

Much of the debate on the correct representation of semantic relations has played out with regard to characterizing the interpretation of noun compounds. A noun compound is a sequence

Table 2.1: Casagrande and Hale's [1967] relations

Relation	Example	Relation	Example
attributive	toad–small	*contingency*	lightning–rain
function	ear–hearing	*spatial*	tongue–mouth
operational	shirt–wear	*comparison*	wolf–coyote
exemplification	circular–wheel	*class inclusion*	bee–insect
synonymy	thousand–ten hundred	*antonymy*	low–high
provenience	milk–cow	*grading*	Monday–Sunday
circularity	X is defined as X		

Table 2.2: Chaffin and Herrmann's [1984] relations

Relation	Example
constrasts	night–day
similars	car–auto
class inclusion	vehicle–car
part-whole	airplane–wing
case relations—agent, instrument	

of two or more nouns which functions as a single noun, e.g., `space shuttle` or `space shuttle mission`. Compounding is a frequent and productive process in English:[5] any text will contain numerous compounds, and many of these will be infrequent. Semantic interest in the special case of two-word noun compounds, or noun-noun compounds, is due not just to their ubiquity but also to the fact that they can encode a variety of relations. For example, a *taxi driver* is a driver who drives a taxi, while an *embassy driver* is a driver who is employed by / drives for an embassy, and an *embassy building* is a building which houses, or belongs to, an embassy.

The main questions about representation which arise in the study of compounds reflect broader questions relevant to any attempt at formalizing semantic relations in general. Noun compounds can therefore be viewed as an informative case study or microcosm. There is a voluminous literature on the semantics of compounds, from the perspective of both linguistics and NLP.[6] In linguistics, the primary aim is to find the most comprehensively explanatory representation. In NLP, it is to select the most useful representation for a particular application: this should have the right trade-off between generality and specificity to be computationally tractable and to give informative output to downstream systems. The two perspectives are complementary.

[5] Compounding is a feature of many other languages; see [Bauer, 2001] for a comprehensive cross-linguistic overview.
[6] See www.cl.cam.ac.uk/~do242/Resources/compound_bibliography.html for a long list.

Table 2.3: Warren's [1978] major semantic relations

Relation	Example
Possession	family estate
Location	water polo
Purpose	water bucket
Activity-Actor	crime syndicate
Resemblance	cherry bomb
Constitute	clay bird

Here is an important question: can the relational semantics of compounding be explained by a concise listing of possible semantic relations? Or is the set of distinguishable relations in practice boundlessly large? In linguistics, the former assumption has led to the compilation of relation inventories, starting with early descriptive work [Grimm, 1826, Jespersen, 1942, Noreen, 1904] and continuing through to the age of generative linguistics [Levi, 1978, Li, 1971, Warren, 1978].

For example, Warren proposed an inventory of relations informed by a comprehensive study of the Brown Corpus [Kučera and Francis, 1967]. The inventory consists of six *major* semantic relations, each of them further subdivided according to a hierarchy of up to four levels. Table 2.3 shows the major relations. As an example of further division, the relation *Time*—a direct child of the major relation *Location*—is specialized into *Time-Animate Entity* (`weekend guests`), *Time-Concrete, Inanimate Entity* (`Sunday paper`) and *Time-Abstract Entity* (`fall colors`).

Levi [1978] proposed a set of relations (for theory-internal reasons called "recoverable deletable predicates" or RDPs), which she claimed underlie all compositional non-nominalized compounds in English. They appear in Table 2.4. The *Role* column shows the modifier's function in the corresponding paraphrasing relative clause: when the modifier is the subject of that clause, the RDP is marked with the index 2.

In Levi's theory, nominalizations such as `taxi driver` are accounted for by a separate procedure because they are assumed to be derived from a different kind of deep representation. For those who are not committed to the transformational view of syntax and semantics, this separation is unnecessary and only leads to spurious distinctions (`horse doctor` would be labelled *for* but `horse healer` would have another relation label, *agent*). Levi deems the degree of ambiguity afforded by 12 relations to be sufficiently restricted for a hearer to identify the relation intended by a speaker by recourse to lexical or encyclopaedic knowledge, while still allowing for the semantic flexibility of compounding.

Levi's relations have influenced further proposals of relation inventories. One example is Ó Séaghdha and Copestake's [2007] work. They started from Levi's set of relations, and followed a set of principles based on empirical and theoretical considerations:

Table 2.4: Levi's [1978] relations

RDP	Example	Role	Traditional Name
$CAUSE_1$	tear gas	object	causative
$CAUSE_2$	drug deaths	subject	causative
$HAVE_1$	apple cake	object	possessive/dative
$HAVE_2$	lemon peel	subject	possessive/dative
$MAKE_1$	silkworm	object	productive/composite
$MAKE_2$	snowball	subject	productive/composite
USE	steam iron	object	instrumental
BE	soldier ant	object	essive/appositional
IN	field mouse	object	locative
FOR	horse doctor	object	purposive/benefactive
FROM	olive oil	object	source/ablative
ABOUT	price war	object	topic

i. the inventory of relations should have good coverage;
ii. relations should be disjoint, and each relation should describe a coherent concept;
iii. the class distribution should not be too skewed or too sparse;
iv. the concepts underlying the relations should generalize to other linguistic phenomena;
v. the guidelines should make the annotation process as simple as possible;
vi. the categories should provide useful semantic information.

The result was an inventory of eight relations. Four of Levi's relations (*about*, *be*, *have*, *in*) were kept, and *for* was replaced with *agent* and *inst* (instrument). Ó Séaghdha and Copestake introduced *rel* for compounds which encode non-specific relations, and *lex* for compounds which are idiomatic.

In contrast with the comprehensive view, Zimmer [1971] pointed to the great variety of English compounds. He concluded that it may be simpler to categorize the semantic relations which cannot be encoded in compounds than those which can. Downing [1977] cited compounds such as *plate length* ("what your hair is when it drags in your food") in order to argue: "The existence of numerous novel compounds like these guarantees the futility of any attempt to enumerate an absolute and finite class of compounding relationships." A complementary argument holds that simple relations chosen from a discrete set do not suffice to capture the richness of relational meaning, and that the meaning of word combinations arises from the interaction between necessarily complex representations of events and entities. This view received a detailed treatment in Coulson's [2001] work on frame semantics.

While this debate has arisen from theoretical linguistic concerns, the tension between parsimony and expressiveness in semantic representation is also a fundamental concern for computational linguists. The inventory approach has been popular in NLP because it is computationally suited to both rule-based and statistical classification methods. Su [1969] was, as far as we know, the first researcher to report on noun compound interpretation from a computational perspective. He described 24 semantic categories for use in producing paraphrase analyses of compounds. These categories contain many relations familiar from linguistically motivated inventories: *Use*, *Possessor*, *Spatial Location*, *Cause*, and so on. Other inventories proposed for noun compound analysis include those of Girju et al. [2005], Leonard [1984], Vanderwende [1994] and Ó Séaghdha [2008]. A large inventory, later used by a number of researchers, appeared in Nastase and Szpakowicz [2003]: 30 relations were grouped into 5 categories—see Table 2.5.

The example inventories presented here should make it clear that these alternative accounts have much in common. They all have categories for locative relations, for possessive relations, for purposive relations, and so on. Tratz and Hovy [2010] proposed a new inventory of 43 relations in 10 categories, developed in an iterative crowd-sourcing process to find a scheme which maximizes agreement between annotators. The relations appear in Tables 2.6–2.7. Tratz and Hovy performed meta-analysis of the most notable previous proposals; it has shown that they all cover essentially the same semantic space, although they differ in how exactly they partition that space.

Every representational framework considered in this section thus far has assumed that semantic relations are abstract constructs which correspond to logical predicates rather than to lexical items. In another take on meaning, semantic relations can be expressed by paraphrases. The relation in *weather report* can be attributed to the abstract predicate named *about* or *topic*; or the same relation can be described by saying that a *weather report* is "a report about the weather" or "a report forecasting the weather". Lauer [1995] proposed a widely cited analysis of noun compounds as paraphrases. He cast the task of interpreting compounds as that of choosing a prepositional paraphrase from the following set of precisely eight prepositions: *of*, *for*, *in*, *at*, *on*, *from*, *with*, *about*. For example, `olive oil` could be analyzed as OIL *from* OLIVES, `night flight` as a FLIGHT *at* NIGHT, and `odour spray` as a SPRAY *for* ODOURS. From a computational point of view, paraphrasing is attractive because a predictive model can be built by identifying noun-preposition co-occurrences in a corpus or even on the Web [Lapata and Keller, 2004].

On the other hand, the lexical nature of Lauer's relations has disadvantages. Prepositions themselves are polysemous, and the assignment of a prepositional paraphrase to a compound does not unambiguously identify the compound's meaning. In other words, once a compound has been identified as, say, an *of*-compound, there remains a question: what kind of relation does *of* indicate? The paraphrases *school of music*, *theory of computation* and *bell of (the) church* do not describe the same kind of semantic relation. Furthermore, the assignment of different categories does not necessarily entail a difference in semantic relations. The categories *in*, *at* and *on* have a significant overlap. The lexical distinction between `prayer in (the) morning`,

Table 2.5: Nastase and Szpakowicz's [2003] relations. H stands for *head*, M stands for *modifier*.

Relation Group	Examples	Paraphrase
Causality		
Cause	flu virus	H causes M
Effect	exam anxiety	M causes H
Purpose	concert hall	H is for M
Detraction	headache pill	H opposes M
Participant		
Agent	student protest	M performs H
Beneficiary	student discount	M benefits from H
Instrument	laser printer	H uses M
Object	metal separator	M is acted upon by H
Object_Property	sunken ship	H underwent M
Part	printer tray	H is part of M
Possessor	national debt	M has H
Property	blue book	H is M
Product	plum tree	H produces M
Source	olive oil	M is the source of H
Stative	sleeping dog	H is in a state of M
Whole	daisy chain	M is part of H
Quality		
Container	film music	M contains H
Content	apple cake	M is contained in H
Equative	player coach	H is also M
Manner	stylish writing	H occurs in the way indicated by M
Material	brick house	H is made of M
Measure	expensive book	M is a measure of H
Topic	weather report	H is concerned with M
Type	oak tree	M is a type of H
Spatiality		
Direction	outgoing mail	H is directed towards M
Location	home town	H is the location of M
Location_at	desert storm	H is located at M
Location_from	foreign capital	H originates at M
Temporality		
Frequency	daily experience	H occurs every time M occurs
Time_at	morning exercise	H occurs when M occurs
Time_through	six-hour meeting	H existed for the duration of M

Table 2.6: The relations from Tratz and Hovy [2010] (part I). The *Approximate Mappings* column shows the mapping of the proposed relation to a relation from [Barker and Szpakowicz, 1998] (B), [Girju et al., 2005] (G), [Levi, 1978] (L), [Nastase and Szpakowicz, 2003] (N), [Vanderwende, 1994] (V) and [Warren, 1978] (W).

Category Name	Example	Approximate Mappings
Causal Group		
communicator of communication	court order	BGN: Agent, L: act_a+$Product_a$, V: Subj
performer of act/activity	police abuse	BGN: Agent, L: Act_a+$Product_a$, V: Subj
creator/provider/cause of	ad revenue	BGV: Cause(d-by), L: $Cause_2$, N: Effect
Purpose/Activity Group		
perform/engage in	cooking pot	BGV: Purpose, L: For, N: Purpose, W: Activity, Purpose
create/provide/sell	nicotine patch	BV: Purpose, BG: Result, G: Make-Produce, GNV: Cause(s), L: $Cause_1$, $Make_1$, For, N: Product, W: Activity, Purpose
obtain/access/seek	shrimp boat	BGNV: Purpose, L: For, W: Activity, Purpose
modify/access/change	eye surgery	BGNV: Purpose, L: For, W: Activity, Purpose
mitigate/oppose/destroy	flak jacket	BGV: Purpose, L: For, N: Detraction, W: Activity, Purpose
organize/supervise/authority	ethics board	BGNV: Purpose/Topic, L: For/$About_a$, W: Activity
propel	water gun	BGNV: Purpose, L: For, W: Activity, Purpose
protect/conserve	screen saver	BGNV: Purpose, L: For, W: Activity, Purpose
transport/transfer/trade	freight train	BGNV: Purpose, L: For, W: Activity, Purpose
traverse/visit	tree traversal	BGNV: Purpose, L: For, W: Activity, Purpose
Ownership, Experience, Employment and Use		
possessor + owned/possessed	family estate	BGNVW: Possess*, L: $Have_2$
experiencer + cognition/mental	voter concern	BNVW: Possess*, G: Experiencer, L: $Have_2$
employer + employee/volunteer	team doctor	BGNVW: Possess*, L: For/$Have_2$, BGN: Beneficiary
consumer + consumed	cat food	BGNVW: Purpose, L: For, BGN: Beneficiary
user/recipient + used/received	voter guide	BNVW: Purpose, G: Recipient, L: For, BGN: Beneficiary
owned/possessed + possession	store owner	G: Possession, L: $Have_1$, W: Belonging-Possessor
experience + experiencer	fire victim	G: Experiencer, L: $Have_1$
thing consumed + consumer	fruit fly	W: Obj-SingleBeing
thing/means used + user	faith healer	BNV: Instrument, G: Means, Instrument, L: Use, W: MotivePower-Obj
Temporal Group		
time (span) + X	night work	BNV: Time(At), G: Temporal, L: In_c, W: Time-Obj
X + time (span)	birth date	G: Temporal, W: Obj-Time
Location and Whole + Part/Member of		
location/geographic scope of X	hillside home	BGV: Locat(ion/ive), L: In_a, $From_b$, B: Source, N: Location(At/From), W: Place-Obj, PlaceOfOrigin
whole-part/member of	robot arm	B: Possess, G: Part-Whole, L: $Have_2$, N: Part, V: Whole-Part, W: Obj-Part, Group-Member

Table 2.7: The relations from Tratz and Hovy [2010] (part II). The *Approximate Mappings* column shows the mapping of the proposed relation to a relation from [Barker and Szpakowicz, 1998] (B), [Girju et al., 2005] (G), [Levi, 1978] (L), [Nastase and Szpakowicz, 2003] (N), [Vanderwende, 1994] (V) and [Warren, 1978] (W).

Category Name	Example	Approximate Mappings
Composition and Containment Group		
substance/material/ ingredient + whole	plastic bag	BNVW: Material*, GN: Source, L: From$_a$, L: Have$_1$, L: Make$_{2b}$, N: Content
part/member + collection/config/series	truck convoy	L: Make$_{2ac}$, N: Whole, V: Part-Whole, W: Parts-Whole
X + spatial container/location/ bounds	shoe box	B: Content, Located, L: For, L: Have$_1$, N: Location, W: Obj-Place
Topic Group		
topic of communication/imagery/info	travel story	BGNV: Topic, L: About$_a$, W: SubjectMatter, G: Depiction
topic of plan/deal/arrangement/rules	loan terms	BGNV: Topic, L: About$_a$, W: SubjectMatter
topic of observation/study/evaluation	job survey	BGNV: Topic, L: About$_a$, W: SubjectMatter
topic of cognition/emotion	jazz fan	BGNV: Topic, L: About$_a$, W: SubjectMatter
topic of expert	policy wonk	BGNV: Topic, L: About$_a$, W: SubjectMatter
topic of situation	oil glut	BGNV: Topic, L: About$_a$
topic of event/process	lava flow	G: Theme, V: Subj
Attribute Group		
topic/thing + attrib	street name	BNV: Possess*, G: Property, L: Have$_2$, W: Obj-Quality
topic/thing + attrib value charac of	earth tone	
Attributive and Coreferential		
coreferential	fighter plane	BV: Equative, G: Type, IS-A, L: Be$_{bcd}$, N: Type, Equality, W: Copula
partial attribute transfer	skeleton crew	W: Resemblance, G: Type
measure + whole	hour meeting	G: Measure, N: TimeThrough, Measure, W: Size-Whole
Other		
highly lexicalized/fixed pair	pig iron	
other	contact lens	

`prayer at night` and `prayer on (a) feast day` does not signal different relations. There is another problem. Many noun-noun compounds which cannot be paraphrased using prepositions (*woman driver*, *taxi driver*) are excluded from the model. Other compounds admit only unintuitive paraphrases: should *honey bee* really be analyzed as *bee for honey*?

Nakov [2008a] and Butnariu et al. [2010] depart from the assumption that a handful of phrases can characterize a semantic relation. They consider a relation to be expressed by any combination of verbs and prepositions which occur in texts: `olive oil` can now be interpreted as, e.g., OIL *that is extracted from* OLIVES or OIL *that is squeezed from* OLIVES. Such paraphrases—more informative than Lauer's OIL *from* OLIVES or Levi's *from* (OIL, OLIVES)—come closer to the richness demanded by Downing's [1977] linguistic arguments. A semantic relation is represented as a distribution over multiple paraphrases, and this allows comparisons. Two compounds may be similar in some ways (`olive oil` and `sea salt` both match the paraphrase `N1 is extracted from N2`) and different in others (salt is not *squeezed from* the sea).

2.2.2 RELATIONS BETWEEN CONCEPTS

Machine-readable knowledge is usually stored in ontologies, KBs or KGs. Relations in such repositories connect concepts rather than words or phrases. Concepts are unambiguous—a concept is represented by a unique name/label—and they refer to something particular evoked by that name or label. Relations in a knowledge repository also have specific characteristics: they should be unambiguous; different relation names/labels should refer to different types of connections; and they should capture some form of enduring knowledge, as in the example in Section 1.1: ENDEAVOUR *is a* SPACE SHUTTLE vs. ENDEAVOUR *is now being prepared* for DISPLAY.

To develop an ontology or a KB, one must choose which entities and relations between them to represent. Both these choices depend on the domain of the knowledge to be captured, and we will see further on that there is much variety. There is, however, some consensus. The backbone of an ontology is the *is-a* relation, and *part-of* is desirable as well. The consensus may break over granularity: even *is-a* and *part-of* can be further refined.

An instance of the *is-a* relation usually links a more specific and a more general concept. From the point of view of formalizing knowledge, there is a distinction between linking two generic concepts (CHOCOLATE *is-a* FOOD), and linking a concept instance and its superordinate concept (TOBLERONE *is-a* CHOCOLATE). The first formalizes *class inclusion*, while the second models *class membership*. Such a distinction was added to WordNet's hyponym/hypernym hierarchy in the form of the *instance hypernymy* relation [Hristea and Miller, 2006]. Further distinctions can be made. Wierzbicka [1984] refined *is-a* into five subrelations. The two most interesting for us are the *is-a-kind-of* relation, which she calls *taxonomic* (chicken–bird) and *is-used-as-a-kind-of*, called *functional* (adornment–decoration).

Meronymic (*part-of*) relations also can be refined, and in certain situations they should be. Winston et al. [1987] made a convincing case for six types of meronymy; they are listed along with examples in Table 2.8. This is motivated by the apparent contradictions in the transitivity