A closer look at learning relations from text
Katrenko, S.

Citation for published version (APA):
Katrenko, S. (2009). A closer look at learning relations from text

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
When two objects, qualities, classes, or attributes, viewed together by the mind, are seen under some connexion, that connexion is called a relation.

Augustus De Morgan. “On the syllogism, part 3”.

Abstract

This chapter reviews various aspects of semantic relations starting from how they can be represented, what logical and linguistic properties they may have to what applications they are used for. In particular, we discuss reasoning with (about) relations and highlight recent developments in such fields as question answering, information retrieval and ontology construction.

2.1 Introduction

Study of semantic relations has a long tradition in several fields, such as philosophy, linguistics, and more recently, computational linguistics. Research in the former two seeks for the explanation of what the relations are, under which conditions they hold and what criteria have to be met to be able to determine a given relation type. The researchers in the latter area are most often interested in designing methods which would allow to extract relation mentions from text corpora automatically. We first review how relations have been treated in philosophy and linguistics over the years and then turn to practical applications and show what role semantic relations play there.

2.2 Representation

To define relations, one needs to decide on what nature they have, where they are placed in the lexicon (if at all) and what properties they need to have.

When talking about relations, we distinguish between their extension and intension. For the n-ary relation its extension is determined by the set of ordered entities (of size n) that satisfy it. For example, for the relation Part - Whole this set would have a member <professor, faculty> but not <faculty, professor>. The intension of a relation is defined by what it means (what does it mean that x is part of y?). More generally, one
can represent an \( n \)-ary relation by the Cartesian product of the sets \( S_1, \ldots, S_n \) where each set corresponds to the particular argument of the relation (Def. 1). If two entities \( x \) and \( y \) are in the binary relation \( R \), we write \( xRy \) or \( R(x, y) \). A set of ordered entities that satisfy \( R \) (for instance, \( \langle x, y \rangle \)) is referred to as instances of \( R \) or its mentions. Note that the same extensions do not necessarily mean the same intension. Two relations such as \textit{study-in} and \textit{live-in} may have the same extension but their intension is different. 

**Definition 1.** A relation \( R \) over sets \( S_1, \ldots, S_n \) is a subset of their Cartesian product, \( R \subseteq S_1 \times \ldots \times S_n \).

Some relations can be also interpreted in terms of partial and total functions. Partial functions associate an element of a set (domain) with at most one element of another set (codomain) while in the total function each element from one set is associated with exactly one element from the other. Partiality indicates that for some elements from domain, there are no corresponding elements in the codomain. Following these definitions, the relation \textit{mother-of} is a total function while the relation \textit{brother-of} is a partial function.

There exist other, more complex, approaches to formalizing a binary relations, for instance the ones based on Galois lattices. In this case one considers a concept system that is defined via extension and intension of concepts (which are the basic elements of a lattice).

Relation representation is a subject of study in cognitive science as well. The main question that concerns researchers there is whether (and how) relations are stored in memory and whether there is any difference between semantic relations if one considers how they are processed in the mind. In particular, they consider semantic memory which is defined to be ‘the mental store’ ([132], p. 75). The proponents of one direction claim semantic relations to be stored in semantic memory (along with word-concept associations) while the other group advocates the view of relation derivation from semantic knowledge. Yet another view says that semantic relations are thought to fall in two groups, intraconcept relations and interconcept relations. Intraconcepts, on the contrary to the interconcepts, are stored in the memory and they exemplify relations between concepts that are associated with each other. Interconcept relations are rather defined on the basis of the features two concepts share and are treated as something that can be computed and are not necessarily stored in memory ([88]).

No matter which position (stance) one would accept, the relationships among concepts are determined by how these concepts are represented. In the classical approach one may use semantic features to represent

---

1 For the sake of uniformness, all relation types in the thesis are given in the typewriter typeface.
a concept as in the well-known example for ‘bachelor’ = [+MALE, -MARRIED] where a plus stands for the presence of a feature and a minus for its absence, correspondingly. The obvious limitations of this proposal lie in the need to define features. Rosch and colleagues took another position by promoting categorization as a process of comparison against a prototype. They emphasized that a concept cannot be completely defined by the necessary features that all its instances have to share. Instead of determining which features are necessary and sufficient, Rosch proposes to select a prototype and to compare a concept’s instances against it by some means of similarity. This approach allows accounting for such cases as a penguin being a bird because it has something in common with the typical birds. Yet another development in componential theories combined the strengths of both theories described above and proposed to divide a set of features in the ‘core’ features (necessary conditions) and others.

Semantic memory is usually represented by means of a network whose nodes correspond to the concepts and are connected by relationships. It is believed that activation of one node leads to the activation of nodes that are linked to it. However, the strength of activation of other nodes varies and is stronger if these concepts are highly associated with a given concept.

Semantic memory and relations in particular were studied by psychologists as well. Most studies have been done with subjects that were known to have a certain disorder and with other subjects that did not exhibit it. Some hypotheses proposed over the years state that impaired performance of patients on various linguistic tasks can be explained by difficulties in searching in semantic memory. It is suggested that patients may have problems with systematic search of semantic memory or with selecting strategies to do it. In particular, if disorders as Huntington’s disease effect subcortical structures, they may also lead to aphasia-like deficits. To investigate this matter, researchers employed a number of tests including a naming exercise (when a subject is shown a picture of an object and is asked to name it) and a free association experiment. The latter is particularly interesting because it involves semantic relations such as is-a and function-action (i.e., <soup, spoon>). The tests that were carried out by Smith et al. corroborated the theory about a breakdown in the organization of the lexico-semantic system in the patients with Huntington’s disease. This was clearly shown by the results of the naming exercise. It was however questionable to what extent semantic relations are affected. Based on the findings, it was concluded that Huntington’s disease “may involve disruption in a dynamic system of interactive activation (and possibly inhibition) of relations and concepts in a semantic network” but there was no clear indication that it results in a loss of semantic relations.
Rapid advances in functional magnetic resonance imaging (fMRI) research allow for studying semantic relations by other means, such as neural activation. Sachs et al. [159] examined thematically related, taxonomically related and unrelated pairs of words and observed that humans react faster to the first two types of input. It was also found that there is a difference in the size of the priming effect between taxonomic and thematic relations. In particular, thematic relations (i.e., \(<\text{car}, \text{garage}>\)) appear to cause a greater priming effect than taxonomic relations (i.e., \(<\text{car}, \text{bus}>\)). The researchers suggested that examples of thematic relations may have a more salient relationship than those of taxonomic ones. Interestingly, previous research revealed that when provided a word and asked for an association, children usually give responses that are thematically related rather than taxonomically (so for instance, ‘cat’ and not ‘white’ to ‘black’). With age people tend to switch to taxonomic responses [132]. fMRI and Event Related Potentials (ERP) studies have also shown activation of certain brain areas that largely depend on relation types. For instance, categorically related words increase activity in the right hemisphere [64], or more precisely in the right precuneus. This area is thought to have several functions such as contextual associations and personal semantic memory. Increased activation of the precuneus provides additional support for the suggestion that the categorial relation is less salient than the thematic one. Similarly to the taxonomic relation, thematic priming effects were also localized in the right hemisphere but in such brain areas as the right middle frontal gyrus and anterior cingulate. In sum, the research in this area by means of fMRI and ERP has provided support to the hypotheses about semantic memory that were introduced before, but it also leaves many open questions. To this end, most attention was paid to categorial and thematic relations, it may be as well be the case that other semantic relations are processed in other ways.

2.3 RELATION TYPES

With all the variety of semantic relations that are known, there exist several ways to classify them. One possibility would be grouping relations according to the logical properties they have (see the next section). Hjørland [73] proposes the following typological criteria:

- query/situation specific
- universal
- “deep semantics” common to all languages (cognitive structures)
- specific to some empirical languages
- domain- or discourse-specific
The first criterion stems from the information need and implies that, to be fully understood, semantic relations have to be placed in the context of a given situation. Naturally, situations vary from one to another and one has to look for what is called “typified practices”. Universality on this list is seen in the Platonic view, namely that a semantic relation is universal and given. Discussing “deep semantics”, Hjørland [73] argues that the theory of semantic primitives strongly relates to it. In a nutshell, semantic primitives are introduced as “semantemes”, the smallest semantic units that cannot be decomposed. Recall representation based on semantic features discussed earlier in Section 2.2. If one could define concepts and relations in terms of semantemes (or features), it should be possible to distinguish between them on this basis.

In our view, some criteria proposed by Hjørland [73] can be combined. For instance, the difference between situation specific and domain-specific relations is small. One can argue that they are about the same unless a situation is taken on the more fine-grained scale as such that exists within a given domain/discourse.

Let us briefly review the relationships that have been studied for a long time by various philosophers, psychologists and linguists.

The relation types that have gotten the most attention are hypernymy and mereology (part-of). Hypernymy is usually presented as is-a or kind-of relation which is used by humans when they relate an unknown object to a known one: “table is furniture”. Here, ‘furniture’ is a hypernym for ‘table’ (and ‘table’ is a hyponym of ‘furniture’). In terms of extensions, the extension of a hyponym is included in the extension of its hypernym. Murphy [132] noted that this set inclusion is unidirectional and bidirectional inclusion corresponds to synonymy. Note that there are two possibilities, either there is a new concept which can be subsumed by the other (as in the <table, furniture> example), or there is a new instance of a concept (which is rather a naming relation as in “p53 is a protein”). Besides the role that hypernymy plays for constructing thesauri [74], this relation is also taken into account while collecting selectional preferences and entailing statements. In case of selectional preferences one can replace a noun by its hyponyms. Consider, for instance, a verb ‘to eat’. Its selectional preference is ‘food’ which can be easily substituted by different food subtypes as ‘an apple’, ‘a fish’, and others. This also holds for entailments. Given a statement “an apple is on the table”, one can entail that “food is on the table”.

It is argued that there are several types of hyponymy. On the most general level hyponyms fall into taxonomic and functional categories. The examples we have mentioned so far deal with the taxonomic category. From the functional point of view we might consider ‘drink’ to be a kind of ‘poison’ but it is easy to see that not every drink is poisonous. Other accounts distinguish among geographical (<Rotterdam, city>), activity
(<football, game>), state (<fear, emotion>) and action (<walk, move>) hyponyms [25].

Another well-studied relation type is the mereological relation. Discussions on the part-whole relation can already be found in the works of Aristotle. In particular, he considers four modes of explaining things such as efficient causality (how things came into being), formal causality (what distinguishes one thing from another), final causality (what is the purpose of a thing) and material causality. The latter corresponds to what a thing consists of and thus is what we call now the part-whole relation. Formal causality can be seen as hypernymy.

Among the approaches to formalize mereology the most prominent are by Winston et al. [203], Keet [87], Smith [173], and Gerstl and Pribbenow [51]. The questions they investigated relate to what can be a part and how parts can be combined into a whole. Winston et al. [203] distinguishes 6 different subtypes of the Part-Whole relation: component-integral object (<ingredient, substance>), member-collection (<professor, faculty>), portion-mass (<meter, kilometer>), stuff-object (<silk, dress>), feature-activity, place-area (<county, state>). Their classification is based on the criteria of functionality, homeomerousity and separability. For instance, stuff-object subtype does not have any of the listed properties: in this case, the parts are neither positioned in a certain way to support functionality of a whole (functionality), nor are they similar to each other (homeomerousity), nor can they be separated/disconnected from a whole (separability). Further, Winston et al. [203] emphasize that there are certain relations meronomy can be confused with. Such relations are usually of the type topological inclusion, class inclusion, attribution, attachment or ownership.

Gerstl and Pribbenow [51] focus on the nature of the parts. They define structure-dependent parts (which are either components of an object or elements of a collection object), temporarily constructed parts (portions and spatial segments of an object), and arbitrary parts.

Yet another semantic relation whose studies can be found in the works of various philosophers is causation. Hume [75] has discussed seven kinds of philosophical relation such as resemblance, identity, relations of time and space, proportion in number or quantity, degrees in any quality, contrariety and causation. The latter is the only relation “that can be trac’d beyond our senses, and informs us of existences and objects, which we do not see or feel” (Hume [75], p. 74). He further notes that there are several necessary conditions for causation. These are contiguity in time and place, priority in time and constant conjunction. The latter condition was criticized later on by many philosophers as such that is not sufficient for causation to take place. Indeed, a simple succession of events does not necessarily means that one of them causes the other. Mill [122] proposed various methods to determine causation such as the method of difference, the method of residues, the method of agreement.
and the method of concomitant variations. The first method relies on comparing two similar instances such that one of them has an event X and the other does not. If in the instance where X occur it is followed by Y but in the instance where it does not happen Y does not happen either, one can conclude that X causes Y. The other questions that were risen by philosophers were whether effect can co-occur with a cause and what properties causes might have (i.e., sufficiency and necessity).

Nowadays causation is often studied by contrasting it to such concepts as enabling and preventing. All three concepts are thought to reflect a relation between two entities, an affector and a patient. For instance, in the phrase “a virus causes the flu”, ‘virus’ is an affector and ‘flu’ is a patient. There are commonalities between causation and enabling which can be seen by the fact that in both cases there is progress towards the end-state. Naturally, preventing differs from these two concepts because it does not allow any progress towards the end-state. It is usually assumed that the way the concept of causation is reflected in causal reasoning is equivalent to the way the causation concept is encoded in language.

Research in the past years has led to studying similarity effects that can be imposed on relations. For instance, Chaffin and Herrmann distinguish between item similarity and relation similarity whereby the first measure, in contrast to the second, is constant and does not depend on the relation at hand. More precisely, if one is given a word pair \(<x, y>\), item similarity is defined between arguments \(x\) and \(y\) while relation similarity measures how close \(<x, y>\) is to the target relation. However, item similarity plays a significant role only for some relations like synonymy or antonymy where such similarity effects are clearly involved. Chaffin and Herrmann studied yet another relation, part-whole, and showed that even though a part is not necessarily similar to a whole, there are still effects similar to those reported on other semantic relations. The authors concluded that relation similarity facilitates relation recognition and impedes negative response. As expected, item similarity did not contribute much to the recognition task and when held constant, relation similarity still affected the resulting performance.

2.4 Properties

The properties of semantic relations have been extensively studied by linguists and computer scientists. This section presents linguistic as well as logical properties.
2.4.1 Linguistic view

Generally speaking, linguists distinguish between relations among various linguistic entities such as sentences (e.g., entailment) or words (lexical relations). In this chapter we focus on relations among words that can be viewed as either syntagmatic or paradigmatic. Syntagmatic relations take place among the words that co-occur in the syntactic structure of a sentence, while paradigmatic relations are those that relate the words in a certain paradigm. In linguistics a paradigm is understood as a set (or a class) of linguistic units that can be used interchangeably in some linguistic environment. The paradigm is not necessarily semantic, for instance, there is a paradigmatic relation between ‘table’ and ‘tables’ (determined by number) but it is morphological and not semantic. From now on, by semantic relation we refer to paradigmatic semantic relation unless stated otherwise.

The list of linguistic properties that we present below can be further extended by adding other properties. For instance, Murphy [132] discussed binarity of semantic relations but we excluded it from consideration because it does not apply to all semantic relations.

Universality

One of the questions raised in the linguistic community has been whether relationships are universal or, in other words, whether they are shared among different languages. The notion of universality can be considered from two angles; in the broader view (do all the languages have the same relationships?) and in the more specific view (if a relation is present in two languages, do all its instances or mentions occur in both of them? Are all of them equally prototypical?). More generally, we may assume that relations that are found in a certain language reflect nature (perceptions) which would lead to the assumption that the same relations have to be found in various languages. But the well-known example of many Eskimo words for ‘snow’ in the combination with unbounded number of relations (productivity criterion) suggests that this assumption may be misleading. There is the number of semantic relationships which can be found in many languages (causation, meronomy and others) but there is little evidence to assume that all languages share the same semantic relations.

A speedy increase in multilingual resources has led to the debate on universality of semantic relations. In particular, if semantic relations were mostly universal, it would make a construction of such resources much easier. It has already been recognized that such relations as causation and mereology tend to be treated as universal ([74], p.75):

In the fields of anthropology and cross-cultural psychology, applied research has led to the conclusion that there is con-
siderable cross-cultural agreement on the meaning and use of semantic relations (Herrmann & Raybeck, 1981; Chaffin & Herrmann, 1984; Romney, Moore & Rusch, 1997). In experiments using concrete (e.g., animals) as well as more abstract (e.g., emotions) concepts, evidence points to specific types of relationships that are recognized equally easily and used with equal frequency and accuracy by diverse groups of people; the relation between opposites, the part-whole relation, and the relation between cause and effect are cited as the relationships most strongly agreed upon (Raybeck & Herrmann, 1990).

**Productivity**

Similarly to neologisms that enter a language every day, there are novel semantic relations appearing on the language landscape. If one accepts the division of relations into generic and domain-dependent, the novel semantic relations would most likely join the latter group. This happens due to developments in science and research questions that are subsequently raised. It is easy to imagine that one would not consider relationships between proteins in the 16th century while it is easier to accept that there must have been such relations as causation or hyponymy.

Productivity also suggests that there must be some mechanism to create new instances of a given relation. A natural question would be whether there exists a single universal mechanism for all semantic relations or whether it depends on the relation at hand. An attempt to explain productivity by a single rule was taken by Murphy [132] in her book where she proposed the relation by contrast (RC) principle which can be stated as follows. Given a set $S$, the contrast relation holds among its members iff “they have all the same contextually relevant properties but one” (Murphy [132], p.44). This criterion is very generic and can be applied to derive many relation types. For instance, synonomy would be defined as a relation among words such that meaning and syntactic category are the same while the word form differs.

Predictability is another property often discussed in the literature. Here, we do not tear productivity and predictability apart because, in our view, these two properties are closely related. If one is able to understand a relation, one can devise a principle that accounts for new relation instances making at the same moment this relation predictable.

**Prototypicality**

Given two examples of the same relation, it is relatively easy to conclude which of them is more prototypical. This has been done by free association tests where human subjects are given one word and the type of relation (e.g., ‘cold’ and antonymy) and have to react with the most
suitable word (e.g., ‘warm’ or ‘hot’). However, it has been demonstrated that some pairs of words can be judged as highly related but they are not the most frequently evoked pairs if one is asked to give an example of a relation. So, for instance, ‘cruel’-‘kind’ is considered antonymous by many speakers although it most likely would not be a response if they are asked to give an example of antonymy.

2.4.2 Logical properties

Semantic relations can be characterized by a number of logical properties such as reflexivity, (a)symmetry and transitivity.

**Definition 2** (Reflexivity). A relation \( R \) is reflexive if \((x, x) \in R\).

**Definition 3** (Symmetricity). A relation \( R \) is symmetric if for any \((x, y) \in R\), \((y, x) \in R\). If for any \((x, y) \in R\), \((y, x) \notin R\), a relation \( R \) is said to be antisymmetric.

**Definition 4** (Transitivity). A relation \( R \) is transitive if \((x, y) \in R \) and \((y, z) \in R\) implies that \((x, z) \in R\).

Most semantic relations are not symmetric and not reflexive (Table 1).

<table>
<thead>
<tr>
<th>relation type</th>
<th>reflexivity</th>
<th>symmetry</th>
<th>transitivity</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>synonym</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>&lt;coach, sofa&gt;</td>
</tr>
<tr>
<td>hypernymy</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>&lt;table, furniture&gt;</td>
</tr>
<tr>
<td>meronymy</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>&lt;object, whole&gt;</td>
</tr>
<tr>
<td>causation</td>
<td>no</td>
<td>no</td>
<td>yes/no</td>
<td>&lt;virus, flu&gt;</td>
</tr>
<tr>
<td>content-container</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>&lt;book, box&gt;</td>
</tr>
</tbody>
</table>

Table 1: Some semantic relations and their logical properties

Logical properties of meronymy have been discussed by many researchers in various fields. While most of them agree upon reflexivity and antisymmetry, transitivity seems to not apply always \([202, 203]\). It is argued that most merological syllogisms arise when different types of part-whole relation are considered as premises. This can be clearly seen in example (2.1). Mixing up the member-collection types in (2.1b) with the component-object meronymy type in (2.1a) leads to an invalid conclusion in (2.1c). In other cases, Cruse \([34]\) suggested that some meronymic arguments can be rephrased using ‘has’ (as in ‘the house has a door’) and therefore cannot lead to valid inferences.

(2.1) (a) Andrew’s head is part of Andrew.

(b) Andrew is part of the CS department.

(c) Andrew’s head is part of the CS department.
Logical properties of semantic relations have implications for automatic reasoning and learning. For instance, transitivity can be used while building taxonomies or for inferring tasks. Symmetry is often explicitly considered in learning. Results in learning theory state that it is nearly impossible to learn from positive information alone. In practice, a learning algorithm takes positive and negative examples as its input and uses both to generate a hypothesis. When only positive examples are given and it is known that the relation at hand is not symmetric, this property is used to generate negative examples (according to the closed world assumption).

Posession of certain logical properties may influence learning and recognition of semantic relations by humans. It has been a long debate in cognitive science as to whether asymmetry contributes to learning causal relations. The two main accounts are contradictory. Proponents of the associative view distinguish between causal and temporal asymmetry and claim that learning is influenced by the latter one. Even though causal asymmetry has a temporal component (where effects usually occur later in time than their causes), it differs from the temporal asymmetry because it presupposes knowledge about causation. In contrast, temporal asymmetry only means that there are differences in the strength of associative links between two words. Causal-model theory presents a contradictory view and states that learning is guided by causal asymmetry.

Furthermore, if asymmetry is important for relation learning, it may also play a role in relation retrieval. This was studied by Fenker et al. [45] whose experiments revealed that the order of arguments has a great impact on recognition of causation. In their experimental setting, participants were asked to judge whether word pairs as such that are examples of causation or whether they are not. The findings of Fenker et al. [45] lead to the conclusion that humans answer faster if they are given examples of causal relation where a cause precedes an effect. Interestingly, when the order of arguments was known in advance, the reaction times for the experimental pairs where the effect was given first and the cause followed, were still slower compared to the situation where the order was reversed.

### 2.5 Reasoning with (About) Relations

#### 2.5.1 By humans

Reasoning with relations has been of interest for a long period of time. The most attention was paid to linear syllogisms so that human subjects were given a pair of premises and were supposed to draw a conclusion or to validate it. For instance, given the premises in (2.2), one would conclude that Bob is the tallest, followed by Ann and Andrew in this order.
(2.2) (a) Ann is taller than Andrew.
(b) Bob is taller than Ann.

Several accounts have been proposed in the literature some of which claim that reasoning is facilitated by mental operations while others provide explanations in terms of linguistic observations. For (2.2), the theory of mental operations \[(2.2)\] would predict that the second premise, \((2.2b)\) will be swapped with the first premise \((2.2a)\) so that they are in a linear order. In combination with the transitivity of the taller-than relation, it would speed up reasoning. Yet another theory grounded on linguistic principles suggests that lexically marked terms make transitive inferences harder for humans \[(2.2)\]. In linguistics, two items can be distinguished by absence/presence of a distinctive feature. If this feature is present, the item is referred to as marked. Markedness is studied at different levels of linguistic analysis, from phonological to syntactical. Terms that represent the enveloping property of size, height, age such as ‘tall’, ‘big’, ‘old’ are considered unmarked whereas their antonyms are marked. Unmarked items are neutralized in questions, for instance, a common way of asking about someone’s height is “How tall is he?” rather than “How short is he?” In the example above taller-than is the unmarked comparative (in contrast to shorter-than) and this is believed to make inferences easier.

There are three main strategies that humans employ when they reason about relations \[(2.3)\]. In Goodwin and Johnson-Laird \[(2.3)\]’s study reasoning was concerned with relations between relations such as in \((2.3)\). Given two premises of this kind, a goal was to infer the complete order (e.g., Steve > Bob > Ann > Andrew).

(2.3) (a) Steve is taller than Bob to a greater extent than Ann is taller than Andrew.
(b) Bob is taller than Andrew to a greater extent than Ann is taller than Andrew.

The first strategy (and the most common one) in such a setting was composing a ternary order given the second premise (Bob > Ann > Andrew) and updating it by using information in the first clause. The second strategy relied on transitivity whereby a participant used information from both premises to build a transitive order of three individuals (e.g., Steve > Bob > Andrew) and then the fourth one was added. The third strategy (frame strategy) relies on the initial guess (e.g., Steve seems to be the tallest among the four) and integrating the remaining individuals. The frame strategy was the least popular but it is worth noting that the participants often mixed their strategies.

We should stress that the research by Goodwin and Johnson-Laird \[(2.3)\] focused on simple transitive relations and as they rightly mention other relations “may lead to more difficult inferences, because they call for
more complex models” (Goodwin and Johnson-Laird \[50\], p. 1067). So far, work in reasoning about relations only explores integration strategies and relational complexity (e.g., how many unique arguments are given) but is not concerned with the arguments of relations. The examples are deliberately chosen in a way that a reader cannot use his background knowledge to speed up reasoning. This may also explain why the first strategy is the one used most often. For instance, using proper names when studying the relation higher-than does not facilitate guessing while using common objects would most likely prompt it.

2.5.2 By machine

Automatic reasoning is often placed in the context of ontologies. A reasoner is understood here as a procedure that given a set of logical axioms can output their entailments and check satisfiability of the axioms. There exist several formalisms to make this reasoning possible with OWL being often used in practice \[118\]. As first order logic (FOL) is expressive but undecidable, OWL makes use of a subset of FOL, namely of Description Logics (DL). There are two types of assertions in DL, on concepts (located in the so-called TBox) and on individuals (located in ABox). For instance, the assertions that a concept ‘person’ is included in the concept ‘organism’ are to be find in the TBox. The reasoning mechanisms are defined for both ABox and TBox.

There are several aspects that must be considered while studying reasoning mechanisms \[58\]. Firstly, inference engines can be distinguished based on whether they offer solutions for multiple inheritance of concept attributes and relations. Secondly, to be able to take into account exceptions (such as a penguin being a bird but not being able to fly), non-monotonic reasoning should be supported. Thirdly, it should be possible to detect inconsistencies and this is what constraint checking is used for.

2.6 Learnability of Relations

Since the introduction of the probably approximately correct (PAC) learning paradigm, there have been many studies on concept learnability. In this section we consider relation learnability from the formal point of view by reviewing some existing results and discussing how they can be stated for relations.

In the PAC model, a concept over some set $S$ is a subset of $S$, $c \subset S$. By learning $c$ the goal is to output a hypothesis concept $h$ such that it closely approximates $c$. The goodness of the approximation is measured in terms of an error between $c$ and $h$. Namely, for any fixed probability distribution $D$ over the instance space $X$ the error $\text{error}(h)$ is equal to
If one represents two concepts \( c \) and \( h \) graphically by a Venn diagram, the error would be equal to the symmetric difference between \( c \) and \( h \). A size of \( C \) is denoted by \( \text{size}(c) \) and stands for the size of the smallest representation given some representation scheme.

To learn \( c \), we assume that a learning algorithm has access to the oracle \( \text{EX} \) that returns a labeled example \( <x, c(x)> \) whenever it is called. The ideal model of learning would require not only a small error but also efficiency so that the number of calls to \( \text{EX} \) would be low and the amount of computation small. In the definition of the PAC model below, the goal is to bound an error by \( \epsilon \) (where \( \epsilon \) is seen as the error parameter) with high confidence \( \delta \).

**Definition 5** (The PAC model \([86]\)). Let \( C_n \) be a representation class over \( \mathcal{X}_n \) (where \( \mathcal{X}_n \) is either \([0, 1]^n \) or \( n \)-dimensional Euclidean space \( \mathbb{R}^n \)), and let \( \mathcal{X} = \bigcup_{n \geq 1} \mathcal{X}_n \) and \( C = \bigcup_{n \geq 1} C_n \). \( C_n \) is PAC learnable if there exists an algorithm \( L \) with the following property: for every concept \( c \in C \), for every distribution \( D \) on \( \mathcal{X} \), and for all \( 0 < \epsilon < 1/2 \) and \( 0 < \delta < 1/2 \), if \( L \) is given access to \( \text{EX}(c, D) \) and inputs \( \epsilon \) and \( \delta \), then with probability at least \( 1 - \delta \), \( L \) outputs a hypothesis concept \( h \in C \) satisfying \( \text{error}(h) \leq \epsilon \). This probability is taken over the random examples drawn by calls to \( \text{EX}(c, D) \), and any internal randomization of \( L \). \( L \) runs in time polynomial in \( n \), \( \text{size}(c) \), \( 1/\epsilon \), \( 1/\delta \).

As it can be seen from Def. 5, the PAC framework provides a very general definition of learnability. It is distribution-free and as a result the bounds that can be obtained by PAC are generally not tight. Despite the criticism of PAC, it has been shown that a number of simple concepts are not PAC learnable. But, given the generality of this framework, the negative results are strong results.

In our view, it is interesting to consider PAC learnability of relations in terms of their logical properties. In particular, we can raise questions as “Are symmetric relations PAC-learnable?”. To study learnability of relations one has to decide on their representation. It has been shown by Kearns and Vazirani \([86]\) that the choice of hypothesis representation is crucial for learnability results and if chosen inappropriately may lead to intractable learning. In general, we would represent a \( k \)-ary relation as a disjunction over \( k \)-tuples. An example of a tuple for a binary Part-Whole relation is \(<\text{artifact}, \text{collection}>\). However, if we consider a binary symmetric relation, the order can be neglected and it can be represented in \( k \)-DNF. A logical formula is in \( k \)-DNF when it is a disjunction of conjunctive clauses. Moreover, conjunctions are defined over \( k \) literals. For a binary relation \( \mathcal{R} \) this means that all its instances are combined by disjunction and the arguments of every instance are written by conjunctions as in \( (x_1 \land y_1) \lor (x_2 \land y_2) \) for two instances \(<x_1, y_1>\) and \(<x_2, y_2>\). It follows immediately that a symmetric binary relation is PAC-learnable (Theorem 1).

**Theorem 1.** Any symmetric binary relation is PAC-learnable.
Proof. This trivially follows from Kearns and Vazirani [86] if a relation \( R \) is represented by 2-DNF. To prove that k-DNF formulae are learnable, it is necessary to represent it via CNF formulae. The learning of k-DNF is then reduced to learning conjunctions which are known to be PAC-learnable.

Interesting work on PAC learning concept (is-a) hierarchies was done by Kearns [85]. Concepts in this case were represented as sets and the main focus was paid to studying inclusions. This means learning consisted of two steps: (i) learning concepts by using a learning algorithm and storing the hypothesis representations in a pool, (ii) performing inclusion tests on any two hypothesis representations \( r_1 \) and \( r_2 \) of two concepts \( c_1 \) and \( c_2 \), correspondingly. Inclusion tests determine whether \( c_2 \subseteq c_1 \), \( c_1 \subseteq c_2 \) or two concepts are incompatible.

Theorem 2. [Kearns [85]] A concept hierarchy is PAC-learnable.

Kearns [85] proves that a hierarchy of concepts is PAC learnable by showing that it is only possible in the case where concepts are learnt independently of one another (all runs of a learning algorithm are oblivious) and the algorithm produces multiple hypothesis concepts. An idea about multiple hypotheses stems from the intuition that it should be possible to infer two hypotheses per target concept would refer to its lower and upper bound. Availability of such hypotheses for two target concepts \( c_1 \) and \( c_2 \) would enable inclusion tests. Moreover, there are two other conditions to be met. These are considering a pool of hypothesis representations as a closed system (so that information necessary for the inclusion test is already contained in the hypothesis representations), and succinctness of hypotheses (so that the hypothesis representation is more succinct than the training data). From the inclusion tests it is possible to reconstruct a hierarchy of concepts.

PAC learning of relations may provide additional insights on their nature. It may as well be possible that, in order to learn a particular relation type, it is necessary to introduce additional constraints. PAC learnability results also reveal that along with constructing an efficient learning algorithm, hypothesis representation plays a significant role and cannot be neglected.

2.7 Computational Approaches to Relation Recognition

When one encounters a relation instance in text, it can be realized syntactically in many ways. Here we limit ourselves to the relation mentions that occur within a single sentence and do not consider discourse. Most solutions that were proposed to relation extraction are based on this assumption [117]. Recognizing relations from a wider scope is an interesting enterprise but it would require a more complex system that would
take into account anaphora resolution and other phenomena. In order to extract relations from text, there are several steps to be taken.

In general, the relation recognition problem can be seen as a two-step process. First, the relation arguments have to be identified. Further, it is necessary to check whether the relation holds. This setting has also been used for relation discovery in other domains [208], moreover, it is often assumed that the arguments have already been found. In this case, relation extraction is reduced to the second step which involves procedures enabling such verification. It has been shown by Bunescu et al. [19] that if the correct names of proteins are given, the accuracy of relation discovery is much higher.

In the supervised setting, a training set contains examples of a given relation (which we can see as extensional information about this relation) and a goal becomes to infer a model such that if applied to a new, unseen data set, it is able to recognize all instances of the given relation in this new data set.

An example from the biomedical domain is given below. In the typical scenario, one starts with the preprocessing (which includes such steps as tokenization and might require some additional analysis depending on the method used). The first step consists of named entity recognition, where all proteins occurring in the sentence are identified. There are three of them, retinoblastoma, RIZ, and E1A. The next step is to detect if there are any relations among them. The correct answer is an interaction between retinoblastoma and RIZ, while E1A does not participate in any interaction. In the parts of this thesis that are concerned with semantic relation extraction we focus on Step2 assuming that entities were already found in the previous step.

| Input: The retinoblastoma protein binds to RIZ, a zing-finger protein that shares an epitope with the adenovirus E1A protein. |
| Preprocessing: The retinoblastoma protein binds to RIZ, a zing-finger protein that shares an epitope with the adenovirus E1A protein. |
| Step1: The (prot) retinoblastoma (\prot) protein binds to (prot) RIZ (\prot), a zing-finger protein that shares an epitope with the adenovirus (prot) E1A (\prot) protein. |
| Step2: The (p1 pair="1") (prot) retinoblastoma (\prot) (\p1) protein binds to (p1 pair="1") (prot) RIZ (\prot) (\p1), a zing-finger protein that shares an epitope with the adenovirus (prot) E1A (\prot) protein. |
| Output: interaction(retinoblastoma, RIZ) |

Even within one sentence, relation arguments can be located within close proximity (i.e., in one noun phrase) or further away from each other (in different noun phrases). The following subsection describes
approaches that were taken for recognition of relations whose arguments occur in the same noun phrase, which are commonly known as ‘noun compounds’. We proceed further by reviewing more general accounts to relation extraction and discuss pattern-based and learning methods.

2.7.1 Noun compounds

Noun compounds such as ‘flu virus’ have been studied in linguistics for a long time which resulted in several definitions to what they are. Lauer [96] summarized these definitions in a list which includes noun premodifiers (any constituent that occur in front of a noun as ‘out-in-the-wilds cottage’), complex nominals (where non-predicating adjectives are taken into account too as in ‘electrical engineer’), noun-noun compounds (any sequence of nouns that functions as a noun). Lauer [96] himself and other researchers who studied relation recognition (i.e., Rosario [155]) excluded all definitions but the last one in their work.

The interest in noun compounds in theoretic and more applied research can be explained by several reasons. Firstly, such compounds occur often in text and if not taken into consideration, the coverage of any relation extraction system would be much lower. Secondly, interpretation of noun compounds is a challenging task because little context is available to do it. It is not surprising that many applications have made used of theories that were proposed in the linguistic community. Levy [105] proposed that there exists a finite set of semantic relations that compounds express such as ‘in’, ‘for’, ‘from’, ‘cause’ and some others. When given a compound ‘software engineer’, one can interprete it as ‘engineer that makes software’. Such accounts as Levy’s can be criticized as having a limited scope and not being able to cover all kinds of compounds, but they led to many practical approaches based on paraphrasing and corpus analysis. Rosario [155] has shown that noun compounds are important not only for generic texts but also for relation extraction in the biomedical field.

2.7.2 Methods

Most approaches to relation extraction fall in one of two categories, either pattern-oriented or other approaches. Pattern-oriented methods comprise hand-written patterns and learnt patterns.

Hand-written patterns

The approaches based on hand-written patterns are usually time-consuming since they often assume the use of rules (patterns) written by an expert. Consequently, when such rules are applied to unseen data, they fail to take into account relations expressed in another way. Although patterns provide a high precision, recall might be much lower [186]. Let us re-
consider the sentence from the biomedical field that we gave earlier. If one is interested in finding protein names in text, one would expect a pattern ‘the * protein’ (where the wildcard stands for a protein name) to perform well. When applied to the sentence, three candidates are extracted, ‘retinoblastoma’, ‘zing-finger’, and ‘adenovirus E1a’. Already at this stage we may notice that these extractions are not always accurate. While ‘retinoblastoma’ is indeed tagged as a protein name, the phrase ‘adenovirus E1a’ contains a protein name (‘E1A’) but there is yet another word occurring in it which was not considered as a protein name by human annotators. In addition, ‘zing-finger’ was not classified as a protein name at all. This simple example reveals that even though patterns are generally considered accurate, they may provide erroneous instances as well. Patterns have been studied not only for named entity recognition (NER) but also for relation discovery. For the same sentence, a pattern such as ‘X binds to Y’ would indicate that X and Y are related to each other and are arguments of the interaction relation. It is easy to see that such a pattern is too generic. In spirit of the two-step relation extraction procedure, the first step would detect named entities in text. The more accurate pattern would take this information into account and require that both X and Y are protein names.

In general, hand-written patterns can be of two types. The first type is sequential and based on often occurring sequences of words in a sentence. Sequential hand-written patterns (as mentioned above) were initially used for extraction of hypernymy [67], with several attempts to extend them for other relations. Patterns that account for hypernymy such as ‘such X as Y’, ‘X including Y’ are often referred to as Hearst patterns. Berland and Charniak [10] introduced a method for extracting parts in large newspaper corpora. They concluded that patterns themselves do not guarantee accurate recognition and they should be combined with statistical measures to rank the extracted parts.

The second type of patterns [89] attempts to account for the syntactic structure of a sentence. The dependency structure of a sentence can usually be represented as a tree and the patterns in this case become subtrees. Such patterns are sometimes referred to as graphical. With the aim to identify examples of causation, Khoo et al. [89] applied this type of patterns to texts in the medical domain. This study has shown that graphical patterns are sensitive to the errors made by the parsers, do not cover all examples in the test data and extract many spurious instances. We hypothesize that, compared to sequential patterns, it is also more difficult for humans to construct graphical patterns.

A simpler approach is to consider not the dependency tree as a whole but certain predefined syntactic functions. This idea has been used by Hahn and Romacker [65], Rinaldi et al. [152], to name a few. Rinaldi et al. [152] have focused on triples such as ‘predicate-subject-object’ and
Automatic pattern acquisition

The drawback of approaches using hand-written patterns is their low recall. Another way to obtain such patterns is to learn them from large corpora or from the World Wide Web. It is assumed that using large corpora will result not only in high precision patterns but also in good coverage. Most methods that explore this avenue rely on bootstrapping. In other words, a list of relation instances (seeds) is given as an input and the goal becomes to find patterns in text such that they extract seeds. The patterns are then selected according to a certain criterion and are applied to text in order to recognize more instances of the same relation types. The initial list of relation instances are expanded with the acquired ones and used again to extract more patterns.

Several relation extraction systems have made use of the idea described above. These include DIPRE [13] and Snowball [2]. Both systems examine sentential information to derive patterns and employ a pattern matching procedure to extract more relation instances. The obvious advantage of such methods lies in no need to annotate large corpora which significantly reduces the human cost. Nevertheless, there are also several limitations. For instance, Snowball exploits information provided by NER tools. This means that it is limited to the relation types that take as arguments types of entities that the tool can recognize. In addition, such semi-supervised systems often require a number of parameters to be set. The actual values of parameters may depend on various factors such as the type of data that is being used, relation types, quantitative information and others which makes setting all parameters a non-trivial decision.

While the aforementioned approaches were applied to generic data, pattern learning has also been investigated on domain-specific data. The performance of rule learning methods in the biomedical domain has been studied in detail by Bunescu et al. [19]. The authors have addressed the problems of protein identification and extraction of the protein interactions. For relation extraction, two approaches have been developed, based on the Rapier rule learning method and on the longest common subsequences. It has been shown that these two approaches outperform hand-written rules. In the food domain, van Hage et al. [193] focused on Part-Whole relation and used known relation instances to acquire patterns for food items and their ingredients.

Other approaches

Contrary to the approaches discussed above, Pustejovsky et al. [148] and Leroy and Chen [101] have employed finite state automata to learn relations. When testing their approach on the inhibit relation, Pustejovsky has shown that these patterns lead to relatively accurate extraction of biological relations.

Automatic pattern acquisition

The drawback of approaches using hand-written patterns is their low recall. Another way to obtain such patterns is to learn them from large corpora or from the World Wide Web. It is assumed that using large corpora will result not only in high precision patterns but also in good coverage. Most methods that explore this avenue rely on bootstrapping. In other words, a list of relation instances (seeds) is given as an input and the goal becomes to find patterns in text such that they extract seeds. The patterns are then selected according to a certain criterion and are applied to text in order to recognize more instances of the same relation types. The initial list of relation instances are expanded with the acquired ones and used again to extract more patterns.

Several relation extraction systems have made use of the idea described above. These include DIPRE [13] and Snowball [2]. Both systems examine sentential information to derive patterns and employ a pattern matching procedure to extract more relation instances. The obvious advantage of such methods lies in no need to annotate large corpora which significantly reduces the human cost. Nevertheless, there are also several limitations. For instance, Snowball exploits information provided by NER tools. This means that it is limited to the relation types that take as arguments types of entities that the tool can recognize. In addition, such semi-supervised systems often require a number of parameters to be set. The actual values of parameters may depend on various factors such as the type of data that is being used, relation types, quantitative information and others which makes setting all parameters a non-trivial decision.

While the aforementioned approaches were applied to generic data, pattern learning has also been investigated on domain-specific data. The performance of rule learning methods in the biomedical domain has been studied in detail by Bunescu et al. [19]. The authors have addressed the problems of protein identification and extraction of the protein interactions. For relation extraction, two approaches have been developed, based on the Rapier rule learning method and on the longest common subsequences. It has been shown that these two approaches outperform hand-written rules. In the food domain, van Hage et al. [193] focused on Part-Whole relation and used known relation instances to acquire patterns for food items and their ingredients.

Other approaches

Contrary to the approaches discussed above, Pustejovsky et al. [148] and Leroy and Chen [101] have employed finite state automata to learn relations. When testing their approach on the inhibit relation, Pustejovsky
et al. [148] got high precision and moderate recall. A particularly interesting approach has been proposed by Bunescu and Mooney [16], who have studied subsequence kernels for relation extraction. Comparative experiments on the biomedical data sets have revealed that the relation kernel outperforms the approaches based on the longest common subsequences and hand-written rules.

In general, kernel methods belong to the most popular machine learning methods that have been used for automatic relation recognition [208, 16, 57, 35]. This is not surprising given the fact that relation extraction cannot be easily cast as a machine learning problem in the commonly used attribute-value format. Kernel methods (discussed in more detail in Chapter 3) provide an alternative to this standard setting by allowing work with various complex data structures.

Approaches to automatic relation recognition do not necessarily involve learning. Some methods that are based on pure co-occurrence of terms proved useful, however, their performance depends on the type of relation [181]. In the biomedical domain, the co-occurrence of terms denoting diseases and genes is likely to provide evidence for the relation between them. In contrast to a gene-disease relation, a relation between genes is less predictable by the pure co-occurrence of genes in a sentence.

Learning with background knowledge

Learning with background (or prior) knowledge constitutes yet another direction in relation discovery. These approaches are grounded in the assumption that relevant information found in numerous knowledge resources such as ontologies should improve extraction results. With all the merits that these resources provide, there are several less attractive aspects to be taken into account. First of all, such resources are by no means complete which may be problematic if an extraction system heavily relies on them. Furthermore, there are many domains where they are not available and one is forced to explore other methods for relation extraction.

For generic relations, the most commonly used resource is WordNet [44], which is a lexical database for English. It contains words grouped in synsets based on synonymy and also provides other information on meronymy, hypernymy along with the definitions of synsets. WordNet can be employed for different purposes such as studying semantic constraints for certain relation types [54], combining it with information found in the training set [57, 139].

Since many knowledge resources have been created in the biomedical community in recent years, it has been especially valuable to test their impact on the biomedical entity extraction task. Leroy and Chen [101] have presented a hybrid system integrating linguistic parsing with existing knowledge sources, such as Gene Ontology, UMLS, and the HUGO
nomenclature. They have evaluated 549 relations from Medline abstracts containing the p53 gene. In comparison to the relations extracted by a parser, the relations provided by a co-occurrence based semantic net Concept Space are less precise and relevant. However, when adding relations containing terms found in GO and HUGO, precision increases. By approaches such as Leroy’s, it has been demonstrated that the knowledge sources can contribute to the named entity recognition and relation extraction tasks in the biomedical domain.

2.8 applications of semantic relations

The fields that can benefit from better understanding of semantic relations range from database design and management \[182\] to various applications in NLP. In this section we consider three areas: information retrieval (IE), question answering (QA) and ontology construction, and discuss the impact that semantic relations have shown to have in their context.

2.8.1 Ontology Construction

In many fields there is a growing interest to construct domain ontologies that ideally would facilitate reasoning and could be helpful for many practical applications. Creating such resources manually is a very time-consuming process and their maintenance leads to more burden. An appealing solution is constructing ontologies in a semi-automatic way whereby concepts and relations are gathered automatically but have to be verified by a human expert. From the semantic relation viewpoint, automatic ontology building benefits from using hypernymy as this is the relation that allows to group concepts into hierarchical structures. Many existing approaches to ontology learning use the well-known Hearst patterns \[67\]. The other methods employ distributional similarity measures in combination with clustering. In such cases the first step was to detect similar terms in text and the second to use them as an input for a clustering algorithm \[23\]. Another avenue that was taken by us was exploring grammatical inference (or grammar induction) as an alternative to clustering. It was shown earlier that grammatical inference leads to constituents that are not only syntactically meaningful (i.e., clusters of nouns or adjectives) but also provides a semantically biased output \[1\]. Our experiments with biomedical data and modern grammatical inference methods proved this claim \[80\]. In particular, we have shown that the resulting clusters often correspond to semantic classes such as gene names, cell types, DNA, RNA and others. Moreover, these classes are detected with relatively high precision (around 70-80%) but the coverage (compared to the an-
notations in the text) is much lower. Extracting relations by means of grammatical inference is more elaborate and may lead to detection of common patterns but does not cover all relation instances mentioned in the text.

Wandmacher et al. [199] emphasized that extraction of semantic relations alone is not sufficient for ontology construction. Relation integration is as important as their correct identification. Firstly, not all relation mentions that are automatically extracted are equally reliable. Their reliability can be estimated by considering in how many sources they were found combined with a local confidence score provided by each resource. Secondly, integration takes place once the reliable relation instances are identified and have to be added to the already existing ontology. At this stage, it is needed to unify word senses and to resolve inconsistencies if there are any. Inconsistency constraints exploit logical properties such as transitivity and anti-reflexivity.

2.8.2 Question Answering

Question answering is a field closely related to information retrieval. It also aims at satisfying information needs of a user but, in contrast to information retrieval, a user does not obtain a collection of documents or passages he has to browse but rather an explicit answer [129]. For instance, if a user is interested in the date of birth of Mozart, a question answering engine has to return ‘1756’. To be able to deliver a correct answer, the engine needs to go through several phases such as recognition of a question type, retrieving the documents that are likely to contain an answer and actual extraction of an answer. Common pitfalls are ambiguity (are there perhaps more Mozarts a user could have had in mind?), temporal constraints (if the question is rephrased into the date of birth of a president of France, which one should it be?) and others. Given the granularity of this task, semantic relations would likely help here more than in case of information retrieval.

Semantic relations can be useful at different stages of question answering. They have to be taken into account when identifying the type of a question and they have to be considered at actual answer extraction time. It has been shown earlier that in order to enhance existing QA systems, it is necessary to be able to find various expressions of the same relation. Lin and Pantel [109] proposed a method to detect inference rules that would account for the mismatch between questions and the information given in the text. Combining a distributional hypothesis with similarity of dependency paths between two entities, the researchers were able to infer that “X solves Y” is similar to “X finds a solution to Y”, “X addresses Y” or to “Y is solved by X”. The experimental results on a subset of the TREC-8 collection revealed that this method produced
many correct paraphrases of the input sequence most of which were not identified by humans manually.

Lopez et al. [114] addressed the problem of finding expressions of the same relation in an ontology-driven way. The QA system that they built includes a so-called relation similarity service which presents an input query in the form of triples \(<\text{subject}, \text{predicate}, \text{object}>\) and maps them into the ontology’s concepts and relations. In other words, its goal is “to map the relationships in the linguistic triple into an ontology-compliant-triple” [114]. The advantages of such a system are its portability (as it allows to plug in various ontologies) and its ability to handle various expressions of relations. Its obvious limitation lies in the linguistic coverage, which has to be considerable in size in order to use this approach in real world scenarios.

In her PhD thesis, van der Plas [192] explored usefulness of several lexico-semantic relations and resources for QA. She considered three methods to acquire lexico-semantic information from corpora: a syntax-based approach, a proximity-based method, and an alignment-based method. All three methods measure similarity between words by means of various inputs. To compute similarity, the syntax-based approach uses syntactic structure while the proximity-based method discards this information and focuses solely on the words that can be found within a given context window. The third, alignment-based method, exploits multilingual parallel corpora. It turned out that the alignment-based method very often results in synonyms, the syntax-based approach provides semantically related words but mostly co-hyponyms, and the proximity-based approach usually accounts for associations. If the output of these methods is considered in the context of QA, the lexico-semantic information given by the syntactically oriented method proved most useful. The proximity-based method is helpful for query expansion and, quite surprisingly, the alignment-based method did not improve QA.

### 2.8.3 Information Retrieval

Several studies have revealed that the role of semantic relations in information retrieval depends on what type of relations are used, in what stage of retrieval they are employed and on the granularity of retrieval.

Semantic relations can be used in two ways, either to refine queries before actual retrieval is done or to manipulate output returned by a search engine (e.g. identifying whether a fragment of text contains a given relation or not). Intuitively, the first strategy is concerned with increasing recall while the second one is supposed to increase precision.

The most widely used relation types for query expansion are hypernymy (or broader terms from a thesaurus) and synonymy or, more generally, any paradigmatic relations that are believed to help retrieve more relevant information. Voorhees [198] focused on semantic relations that can be
extracted from WordNet and concluded that if a query is defined very well, the query expansion does not have a large impact on the retrieval results while if it is not, the effectiveness of retrieval improves.

Increased interest in domain-specific search has led to numerous studies in biomedical [70], chemical, legal [188] and other domains. Meij et al. [120] showed that synonymy improve results in the biomedical domain but the best performance is obtained by combining several methods such as gene name expansion, thesaurus lookup and augmenting a query with thesaurus based feedback. Zhao et al. [210] explored co-occurrence based thesaurus in the legal domain and concluded that it is effective on most queries but it could as well degrade performance on some others.

To boost precision, relations can also be used to filter out the retrieval results or to re-rank them. Relations that are usually employed here are syntagmatic [88]. However, in most cases relation matching, as it is referred to in the literature, does not appear to affect retrieval results. If existing dictionaries are used, they may not contain relations that would match a query [133]. Sometimes, an exact relation match cannot be found and one can resort to partial matching [111]. A relation can be represented by a triple ‘concept1, relation, concept2’ and partiality would require splitting it into ‘concept1, relation’ and ‘relation, concept2’. This technique provides better results for long queries.

Overall, the role of semantic relations in information retrieval depends on the granularity of the retrieval itself. If the goal is to retrieve documents, then terms proximity may already be sufficient to achieve the same performance as the one that is obtained by using more elaborate natural language processing [88]. If however one aims at retrieving passages, semantic relations may be more helpful.

In the various applications that we have discussed so far, the acquired relations can either be used immediately or they can facilitate composing resources such as thesauri and ontologies. In either case, instances of semantic relations are obtained by using similar methods, with the distinction that additional constraints on information integration have to be introduced for the latter case.

2.9 CONCLUSIONS

In this chapter we gave a definition for a relation in terms of its intension and extension and reviewed its linguistic and logical properties. We also emphasized the role semantic relations play in human reasoning and learning by humans. Further, we discussed several applications that could benefit from relation learning such as question answering and information retrieval. Natural language understanding is very complex and understanding relations among words can be considered as one of the building blocks that is needed in order to accomplish this goal.