Words in context: a reference perspective on the lexicon

Piek Vossen, Tommaso Caselli, Filip Ilievski, Ruben Izquierdo, Alessandro Lopopolo, Emiel van Miltenburg, Roser Morante, Minh Ngoc Lé, and Marten Postma
Faculty of Humanities, VU University Amsterdam / De Boelelaan 1105, Amsterdam
cltl@vu.nl

Abstract

In this paper, we present a rich contextual perspective on the lexicon and background knowledge for the purpose of deep semantic parsing. In the project *Understanding Language By machine*[1], we address various aspects of semantics in relation to i.) reference to entities and event instances, ii.) modeling of author and reader perspectives. Lexical resources and even resources with world-knowledge such as Wikipedia do not provide the episodic knowledge that is needed to determine reference and eventually meaning. Most resources and also the Natural Language Processing that uses these resources focus too much on semantic knowledge and local context. We argue that we need richer and more complex context models that integrate episodic knowledge, discourse structure and reader/writer perspectives to be able to correctly process text. We outline the directions of research that our project follows and the different aspects that we will study.

1 Introduction

Lexicons are considered as databases of knowledge on words and therefore also of language. Traditionally, these lexical databases are built from lexicographical intuition and capture a very condensed form of expert knowledge about the properties of words. Since the rise of corpora and the Internet as archives of language use, researchers have massively applied statistical techniques to derive properties of words. More specifically, distributional semantic models (Turney et al., 2010) are adopted by many scholars as a more empirical way to learn their semantic properties. Learning such semantic properties is based on the (mostly direct) context of the words. Simple contextual models (e.g. one-word-to-the-left, one-word-to-the-right) turn out to be not only powerful but also relatively easy to implement. Furthermore, many problems in NLP are casted as isolated problems not considering other types of information. For example, word-sense-disambiguation (WSD) is defined as a task to determine the lexical meaning of a word such as *winner* in just the sentence as context, where the meanings are defined in abstract (relational) terms in some lexicon. However, language is hardly used to refer to the abstract meaning of words but mostly it refers to specific instances in the world. Additionally, language almost never uses a single word for that, but a combinations of words (phrase, sentence, discourse). Accordingly, the noun phrase of which *winner* is the head is likely to be (co-)referential with a named entity that is known to be the winner of a specific tournament at a specific time in a specific location. To compute this actual reference, it is not sufficient to just consider the abstract meanings of words or their distributional properties. Even stronger, knowing the referential meaning may help recovering the lexical meaning of *winner*. Sense detection, coreference and reference are thus highly interconnected problems that should not be considered in isolation and require a richer and more complex notion of context.

Most NLP modules and lexicons do not address so-called episodic knowledge (Tulving, 1972) or reference to the external world and restrict themselves to semantic knowledge and abstract annotations. Partly, this is the result of the fact that a wider notion of context and the episodic aspects of meaning are difficult to define and model. There is thus one aspect that is missing in lexical resources, whether built manually or learned from corpora: they lack an episodic component. We argue that experience-based knowledge about the

[1] http://www.understandinglanguagebymachines.org

world plays an important role in the understanding of expressions in language, and in fact, language structure and vocabulary should be seen in function of making reference to these episodic experiences.

A wider conception of context that accounts for language use involves many different aspects among which the physical context of the world, the topic of the discourse, the perspective and intention of the writer and the reader, their background knowledge. This position paper describes our basic ideas and plans for research addressing these aspects. Through this paper, we argue that lexical research should be extended to richer and more complex contexts that are formally and probabilistically modeled and include episodic data as well as semantic data.

In Section 2, we first explain the referential perspective on the lexicon. Given this perspective, we explore in section 3 the interaction between episodic and semantic knowledge for NLP tasks such as WSD and Entity detection to show that both tasks cannot be seen in isolation and require the use of both types of background knowledge. In section 4, we look at the interaction within the context of the discourse, i.e. readers and writers introduce and anticipate background knowledge and can make reference with different words (lexical choice) in relation to what is known or given. Next in section 5, we discuss the relation between perceptual and language representation in terms of knowledge interaction and lexical choice, to conclude and look forward in section 7. Throughout the paper, we formulate the requirements for lexical knowledge in relation to episodic knowledge for the task of semantic parsing of text.

2 A reference perspective on the lexicon

The lexicon is traditionally addressed as what cognitive scientist call the semantic memory (Tulving, 1972): knowledge that abstracts from the individual experiences that are stored in the episodic memory. A lexicon such as WordNet (Fellbaum, 1998) thus lists the following 3 meanings of *winner*:

1. S: (n) winner, victor (the contestant who wins the contest)
2. S: (n) winner (a gambler who wins a bet)
3. S: (n) achiever, winner, success, succeeder (a person with a record of successes) "his son would never be the achiever that his father was"; "only winners need apply"; "if you want to be a success you have to dress like a success"

The glosses and the relations in WordNet represent meaning in a Fregean sense and not reference. Likewise, WordNet does not list the names of all winners of matches, people ever winning a bet in gambling nor the people that you would call typical winners. The examples shown for the third meaning of *winner* are what comes closest to episodic knowledge. Episodic knowledge is what you expect to find in Wikipedia, although it also includes semantic knowledge. Partially this knowledge is structured and partially it is free text. Efforts in the field of Word Sense Alignments (WSA) aim at providing semantic interoperability between different language resources and at enriching existing lexica. However to provide knowledge on all entities that are winners in sense 1), 2) or 3) in WordNet, we need more than linking entries across these resources but also linking the descriptions of the knowledge on these entities. For example, the Wikipedia article of *Tiger Woods* states that "by April 1997 he had already won his first major". To connect this to the semantic knowledge in WordNet, we need to infer that "he" is Tiger Woods, that "won" makes him a winner in sense 1, that "major" is a golf tournament and what tournament it was in 1997. In fact all the biographical information and news on Tiger Woods counts as episodic knowledge and WikiPedia contains just a fraction of this in textual form. To the best of our knowledge, no existing lexica nor lexical resource obtained through WSA techniques are tackling the issue of context of reference[2] and the interaction between abstract sense definitions and episodic knowledge, i.e. the actual use of a specific lexical item to refer to something in the world.

Still the way people use language is mostly to make reference to episodic situations. In those cases, it becomes less relevant what the semantic meaning of the word *winner* is but more important to whom the word may refer.

We thus propose a new layer to the lexicon that defines the referential domain of words. We define the referential domain as the typical and observed range of entities or situations to which a

[2] Notice with context of reference we are not referring to common sense knowledge.

word may refer. We employ a pragmatic and practical approach in the definition of "reference" and "referential domain" by adopting a simple extensional model of meaning. Such a range of referential values can be defined by providing a list of actual (real-world) entities that people have referred to by an expression involving the word (e.g. simply all winners sense 1 and the contest they won) or it could be defined by providing the contextual constraints to resolve the reference. The semantic definition of the word can be seen as a conceptual constraint on the type of entities (the domain) but it does not provide information on the actual referred entities. As such the semantic knowledge may help to recover the referent but it is often not sufficient. The difference can be made clear by comparing a corpus that is annotated with senses with a corpus that is annotated with entities and coreference relations and provides access to all the further knowledge on these entities. The former provides knowledge and information on the semantic meaning of a word, whereas the latter provides access to the episodic knowledge. Note that a corpus with disambiguated named entities is not the same either, since all coreferential expressions need to be included and the knowledge on these entities.

Given such an episodic layer, which can be seen as a link from the lexicon to episodic data, we can now define lexical phenomena in referential terms:

polysemy multiple disjoint referential domains are associated with the same word; i.e. individuals belong to one of the domains but never to more than one.[3]

synonymy different words or expressions have overlapping referential domains

underspecification words that can refer to the union of non-overlapping or disjoint referential domains and thus can quantify types as well as tokens

overspecification words that refer to subsets of referential domains within specific contexts only

In the case of polysemy, the referential problem is complicated by the number of referential domains that are associated, put simply: the set of referential candidates is bigger. However, solving the referential problem is not the same as solving the semantic problem of finding the appropriate meaning. In some cases it may be easier to resolve the referent than resolving the meaning. In fact, you could infer the lexical meaning from the episodic knowledge that is given on the known entity that the expression refers to.

In the case of synonymy, the referential domain of different expressions or words is the same or very similar. From a language generation or translation point of view, this can be seen as a problem of lexical choice: which expression is more appropriate in what context? The problem of lexical choice also applies to under- and overspecification relations between words and expressions. If we consider an underspecification and overspecification relation between words, we can say that the referential domains are in a subset relation. For example, there is a subset relation between the referential domain of *animal* and *dog*, which means that we can refer to dogs with the word *animal*. Compared to *dog*, the word *animal* underspecifies. On the other hand there is an overspecification relation from *lapdog* to *dog*, since the former can only refer to dog entities in certain contexts. A word such as *pet* will then both have an underspecification relation to dog entities (*pet* can also refer to cat entities), as an overspecification relation since not all dog entities can or should be called pets.

The above referential perspective on the meaning of words in the lexicon opens up the possibility of investigating what contextual knowledge is needed to determine the reference of words in text. In the next sections we explore the interaction between episodic and semantic knowledge in more detail.

3 A chicken and egg problem for word senses and entities

Semantic knowledge defines what is possible in some world (according to our cultural, experience-based, educational and scientific insights), whereas episodic knowledge defines what is the case. Not everything that is possible will also be the case. The fact that Tiger Woods as an instance of a human being could win a chess tournament does not mean he will or did. During semantic processing of text, we need to consider continuously what can be the case, what can be

[3]For the time being, we abstract from metonymic cases of polysemy or complex types for the sake of the discussion

expected and what is actually the case.

There are two classical NLP tasks that address the problem of using semantic knowledge and episodic knowledge for semantic parsing of text: Word Sense Disambiguation (WSD) and Named Entity Recognition and Linking (NERL). In the case of WSD, NLP modules try to decide on the lexical meaning of target words in sentences. The task relates to the semantic knowledge for which episodic knowledge may play a role but is not the goal. In the case of NERL, on the other hand, we want to find an expression that refers to an individual and recover the individual by selecting the correct one from a repository such as Wikipedia or Dbpedia. Very often the words in context of a named entity expression help to decide on which entity it is referring to. A third NLP task, coreference resolution, provides a bridge between WSD and NERL.

By applying the distinction of episodic and semantic knowledge, we can identify two different types of problems:

1. the semantic meaning of a word is undecided (WSD) and episodic knowledge plays a role to determine it

2. the episodic interpretation is unclear and semantic knowledge helps resolving it

Example (1) illustrates the first problem type:

(1) "The **winner** walked away with $1.5-million."[4]

In example (1), the wider context of the sentence points to episodic information that the **winner** is coreferential with Thomas Bjørn and that Wikipedia tells you that he was the winner (sense 1 in WordNet) of the Nedbank Golf Challenge in 2013). The strategy here is to first find the referent of the expression with the lemma *winner* and then to get the **episodic knowledge** on the referent to resolve the **semantic meaning** represented by the gloss 'the contestant who wins the contest'.

On the other hand we run into the second problem, when the goal is to identify *Ford* in the sentence:

(2) President Woodrow Wilson asked Ford to run as a Democrat for the United States Senate from Michigan in 1918.

Some of the possible interpretations are:

- Gerald Ford (person, born in 1917)
- Ford, the motor company
- Henry Ford, the founder of Ford Motor Company.

While for humans it is obvious that neither a car vendor company nor a one-year-old boy ran for Senate representative, it is more complicated for a machine. In order to make the correct choice, the machine first needs to known what it means *to run as a politician for a senate*. The semantic knowledge of this expression and the specific meaning of *run* define the constraints that are needed to reason over the choices for deterring the referent of *Ford*. The machine would need to be able to access (at least) first the following semantic knowledge: (i) a senator must be a human (ii) he or she must be at least 18 years old and (iii) it is highly unlikely that someone (or something) who does not meet these minimum requirements runs for a senator. This kind of knowledge belongs to the category of semantic knowledge and is usually represented and made explicit through ontologies.

In order to resolve the problems mentioned above, we need to go back and forth between episodic and semantic knowledge. In the literature there have been some attempts to bring episodic and semantic knowledge closer to each other. In (Navigli and Ponzetto, 2012), WordNet and Wikipedia were merged into BabelNet, enabling a system to handle both words, senses and entities in an unified framework. (Hovy et al., 2011; Clark and Harrison, 2009) syntactically and semantically parsed large amounts of text, extracted propositions and abstracted them into more general propositional statements. The resulting propositional store was used in NLP tasks such as semantic parsing and textual entailment. Another example is IBM Watson (Hajishirzi and Mueller, 2012) which combines statistical relations extracted from text and crowdsourced data to enhance question answering. The sources of information for Watson include encyclopedias, dictionaries, thesauri, news articles, as well as databases, taxonomies, and ontologies (DBPedia, WordNet, and YAGO).

The extraction of episodic knowledge is also related to Open Information Extraction (OIE)

[4]source: `http://www.southafrica.info/news/sport/golf-nedbank-210613.htm\#.VEAWkYusVW8` creation time: **21 June 2013**

(Etzioni et al., 2006). An OIE system makes a single data-driven pass over a corpus and extracts a large set of relational tuples that represent facts. A system like TextRunner could extract automatically 60.5 million tuples from a given corpus of 9 million Web pages (Banko et al., 2007). The extracted information is stored in a unified knowledge base. Another related approach is the Never-Ending Language Learner (NELL, (Mitchell et al., 2015))[5], which has been learning to read the web 24 hours/day for over four years. NELL creates beliefs based on what it reads, which can be considered a type of episodic knowledge. For example, beliefs about Tiger Woods, such as "tiger_woods is an athlete who beat phil_mickelson (athlete)", can be found in the following address: `http://rtw.ml.cmu.edu/rtw/kbbrowser/athlete:tiger\_woods`.

We take a step in this direction by proposing a system that parses and extracts knowledge about words, entities and events from a large amount of data that could include Wikipedia, DBPedia, news articles, etc., with the goal to introduce better text comprehension. It would be analogous to a reader who understands today's news in virtue of the knowledge she harnessed yesterday. This is significantly different from existing systems for the reason it leverages not only abstract, but also concrete knowledge – i.e. not only golf players in general, but also Thomas Bjørn as demonstrated in example (1).

The knowledge base we aim to build should contain the required information needed to perform the tasks described above, basically extracting semantic and episodic meaning. It could include resources already available such as WordNet, Wikipedia or DBpedia (Lehmann et al., 2014). Using these resources, in some cases we would be able to obtain the semantic meaning and rely on this meaning to infer episodic knowledge, or vice versa. In other cases, additional resources would need to be queried (like the news of the past months in order to extract episodic information about someone cited in today's newspaper). And finally, in some other cases (the egg or the chicken) we may find a deadlock where both types of meanings are undetermined, or not reliable, and the problem cannot be solved.

The use of such a huge knowledge base in NLP creates an interesting and challenging problem on its own. Although reasoning has always been a computationally expensive process, recently researchers have started to question large-scale reasoning and its use in NLP (Ovchinnikova et al., 2011; Inoue et al., 2012), raising hope for more practical usages.

4 Discourse knowledge

Adopting a discourse perspective to the problem of reference and context modelling implies awareness of the compositional mechanisms of a text and their decoding. Basically, a text is composed by juxtapposition of units (normally sentences) which hold together by means of a coherence relation. It is up to the reader to figure out what the relation is through an incremental decoding process. Failure to identify the coherence relation(s) between the text segments implies a failure of the communicative process. Different models of text/discourse processing have been proposed ((Webber, 1978; Hobbs et al., 1988; Kamp et al., 2011; Kruijff-Korbayová and Steedman, 2003), among others). In this project we are not aiming at proposing a new discourse model but rather to integrate existing frameworks with streaming contexts. This will provide a new perspective and larger notion of discourse knowledge as an important block of information to deal with the reference problem.

Texts/discourses, like the perception of entities, do not occur in isolation. In particular, they do not occur in isolation from two different perspectives: firstly, they are objects in the world; and secondly, they occur in a streaming context where information unfolds and evolves in time. The notion of streaming context can be clarified by considering a specific type of text/discourse like a news article. Every news article is a continuos interplay between episodic/world knowledge, background knowledge and shared beliefs. The more an event, or topic, is known and active in the background knowledge of the community, that is, the more a certain topic is in the streaming context, the more the writer will tend to use vague, contextual and opinionated expressions. Vice versa, the less the event, or topic, is known and active in the streaming context, the more background knowledge must be provided in order to reinforce or build the shared beliefs and attain the communicative goal. The modelling of streaming context is essential in

[5]urlhttp://rtw.ml.cmu.edu/rtw/

order to deal with background and episodic knowledge, and in the design of NLP tools. A temporally driven access to events, or topics, such as to (re-)construct timelines is the first step to facilitate the understanding of background knowledge, its interaction with the episodic knowledge, and access to stories. We cannot expect NLP modules to successfully process an isolated news article that is published in the middle of a streaming context that runs for some time and which involve a story. In the middle of a debate or discourse, writers can more freely make reference to entities given in the background knowledge, hence use more ambiguous or vague words, express opinions or judgements or frame entities that are assumed to be known by the audience. Likewise, dealing with such expressions to resolve their semantics, the discourse function should be seen in the light of the position within that discourse and the previous information.

Streaming aspects of the news and story lines of the events are the basic starting points for interpreting the news articles in the right order, given a reasonable assumption on background knowledge and given the public opinion, perspective and world view.

5 Perceptual context

The interaction between episodic and semantic knowledge is apparent when language is related to perception. Perceptual data is by definition episodic and it is always about instances. Any symbolic description in language that is related to such perception also connects it to the more abstract knowledge we have about these entities.

In the current digital era, information is not only provided in the form of written texts. Digital resources make use of multimedia modalities such as speech, text, images, and video. From a reference perspective to the lexicon we propose that multimodal information should be integrated in the lexicon in order to ground the episodic meaning of words with perceptual stimuli. For example, the word *winner* can be linked to images of winners or to videos of events in which winners participate.

This line of work has mostly been focused on images. Examples are ImageNet (Deng et al., 2009), SUMO (Pease et al., 2002), and the TinyImages database (Torralba et al., 2008). While these resources serve a clear need (they provide an ordered collection of images that is useful to many applications), the use of these resources is often misguided. The reason for this is that the problem of connecting language and perception is often being viewed as a single image recognition task. The 'solution' for this task is to take a large set of labeled images, and to train a model to recognize the visual features corresponding to a particular label. This leads to what we may call the **label specificity problem**: images get tagged with overly specific labels that do not make any sense without context. Figure 1 illustrates this problem. Depicted are 12 random images that are labeled 'pharmaceutical chemist' by the TinyImages algorithm. Such a labeling is problematic: how could we ever tell (without any context) whether someone is a pharmaceutical chemist? The answer is: we cannot and we should not. Clearly, the image labeling algorithm is overstepping a line here.

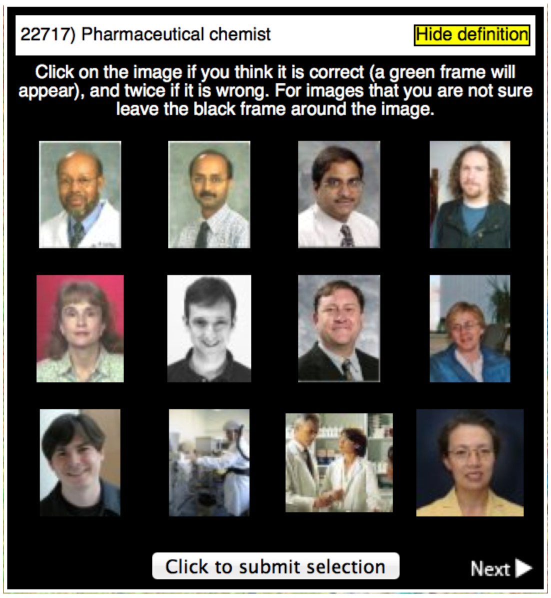

Figure 1: Images labeled 'pharmaceutical chemist' in the TinyImages database.

We need to recognize that image labeling is not a single problem, but rather consists of two separate problems:

1. The entity recognition problem: what is being depicted?

2. The problem of 'role detection': what is the role of the entities in the scene being depicted?

This division leads to new questions. For the first problem of entity recognition, we need to find out how specific the labeling of entities can be without any context. I.e. we need to establish 'perceptual basic level terms' indicating for each kind of entity to what degree we can reliably identify them. These terms might be added to WordNet, so that every branch has a marker indicating the boundaries of visual recognition. Synsets below (or above) that marker correspond to labels that can only applied with additional knowledge.

To illustrate the second problem, let's say we have a picture of Thomas Bjørn. In the context of the Nedbank Golf Challenge, if we can correctly identify him in the picture, we might label that picture with the tag *the winner*, but in an every day context (e.g. in a mall), it would make more sense to say *shopper* or *passer-by*. On the other hand if no specific circumstances apply, the best way to call him is Thomas Bjørn and if his identity is unknown *a middle-aged man*. In other words, we need an accurate model of the discourse context to provide any labeling.

6 A context model for combining lexical, semantic and episodic knowledge

Within this project, we will test context models on a data set of 1.3 million English news articles on the automotive industry spanning a period of 10 years. The articles were semantically parsed in the NewsReader project[6]. They contain the output of various WSD systems based on WordNet, NERL based on DBPedia Spotlight (Daiber et al., 2013) and Semantic Role Labeling using Mate (Björkelund et al., 2009) with mappings to FrameNet (Baker et al., 1998) and other ontologies. Furthermore, it contains coreference relations linking pronouns and noun phrases to entities and connecting events in which they participate. These systems use local contexts within isolated articles and even sentences. None of the systems uses wider background knowledge nor the fact that the news has been published as an opinionated information stream within a discourse. We will use this data set to test various aspects of contexts on interpretation:

- The impact of background information of entities on the WSD of co-referring noun phrases and the events in which they participate: knowing who they are can we tell how reference is made.
- The impact of streaming aspects and trendiness of news on the identification and linking of entities.
- The impact of discourse contexts (perspectives and opinions) on the interpretation of roles referring to these entities.

We will consider the discourse context of news as a topic that evolves in time. We assume that a topic is first grounded to the background knowledge on the entities and related events and, when time passes, new information and perspectives are added to the discourse context that was set before. We predict that traditional semantic parsing based on local context and isolated articles will get more difficult and less successful while the topic develops. We also build up episodic and semantic knowledge about the domain. This includes biographical knowledge about the major participants but also generalized knowledge what typical events they are involved in with their typical roles. Finally, we will connect textual reference to non-textual perceptual data to measure the impact on the interpretation of visual context and vice versa. Deciding on the meaning of expressions and their reference thus will be modeled within a rich knowledge and perspective context.

7 Conclusions

In this position paper, we described our perspective on lexical and knowledge resources. To semantically parse natural language text, we need to have access to both semantic and episodic knowledge. Lexical resources provide semantic knowledge that by itself is not sufficient to resolve reference and perspective. We thus propose to link the semantic knowledge to episodic data beyond the WSA linking and creating richer context models. Such data takes the form of corpora and data that is not only annotated for senses but also for entity reference, co-reference, writer perspective and background knowledge that is needed for interpretation but not expressed. Furthermore, we need to process the episodic knowledge (news, biographies) of real world entities to be able to buildup knowledge on potential reference and reason over actual references in text.

Acknowledgments

This project is funded by the Dutch National Science Foundation from the Spinoza-2013 price.

[6] `www.newsreader-project.eu`

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

M. Banko, M. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni. 2007. Open information extraction from the web. In *International Joint Conference on Artificial Intelligence*.

Anders Björkelund, Love Hafdell, and Pierre Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of The Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 43–48, Boulder.

Peter Clark and Phil Harrison. 2009. Large-scale extraction and use of knowledge from text. In *Proceedings of the fifth international conference on Knowledge capture*, pages 153–160. ACM.

Joachim Daiber, Max Jakob, Chris Hokamp, and Pablo N. Mendes. 2013. Improving efficiency and accuracy in multilingual entity extraction. In *Proceedings of the 9th International Conference on Semantic Systems*, I-SEMANTICS '13, pages 121–124, New York, NY, USA. ACM.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE.

Oren Etzioni, Michele Banko, and Michael J Cafarella. 2006. Machine reading. In *AAAI Conference on Artificial Intelligence*.

Christiane Fellbaum. 1998. *WordNet: an electronic lexical database*. MIT Press.

Hannaneh Hajishirzi and Erik T Mueller. 2012. Question answering in natural language narratives using symbolic probabilistic reasoning. In *FLAIRS Conference*.

Jerry R Hobbs, Mark Stickel, Paul Martin, and Douglas Edwards. 1988. Interpretation as abduction. In *Proceedings of the 26th annual meeting on Association for Computational Linguistics*, pages 95–103. Association for Computational Linguistics.

Dirk Hovy, Chunliang Zhang, Eduard Hovy, and Anselmo Peñas. 2011. Unsupervised discovery of domain-specific knowledge from text. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1466–1475. Association for Computational Linguistics.

Naoya Inoue, Ekaterina Ovchinnikova, Kentaro Inui, and Jerry Hobbs. 2012. Coreference resolution with ilp-based weighted abduction. In *COLING*, pages 1291–1308.

Hans Kamp, Josef Van Genabith, and Uwe Reyle. 2011. Discourse representation theory. In *Handbook of philosophical logic*, pages 125–394. Springer.

Ivana Kruijff-Korbayová and Mark Steedman. 2003. Discourse and information structure. *Journal of Logic, Language and Information*, 12(3):249–259.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2014. DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*.

Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, Justin Betteridge, Andrew Carlson, Bhavana Dalvi, Matt Gardner, Bryan Kisiel, Jayant Krishnmurthy, Ni Lao, Kathryn Mazaitis, Tahir Mohammad, Ndapandula Nakashole, Emmanouil Antonios Platanios, Alan Ritter, Mehdi Samadi, Burr Settles, Richard Wang, Derry Wijaya, Abhinav Gupta, Xinlei Chen, Abulhair Saparov, Malcolm Greaves, and Joel Welling. 2015. Never-ending learning. In *AAAI*.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, December.

Ekaterina Ovchinnikova, Niloofar Montazeri, and Theodore Alexandrov. 2011. Abductive reasoning with a large knowledge base for discourse processing. *Computational Semantics IWCS 2011*, page 225.

Adam Pease, Ian Niles, and John Li. 2002. The suggested upper merged ontology: A large ontology for the semantic web and its applications. In *Working notes of the AAAI-2002 workshop on ontologies and the semantic web*, volume 28.

Antonio Torralba, Robert Fergus, and William T Freeman. 2008. 80 million tiny images: A large data set for nonparametric object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1958–1970.

E. Tulving, 1972. *Episodic and semantic memory*, pages 381–402. Academic Press.

Peter D Turney, Patrick Pantel, et al. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1):141–188.

Bonnie Lynn Webber. 1978. A formal approach to discourse anaphora. Technical report, DTIC Document.