Erik Cambria
School of Computer Engineering, Nanyang Technological University

Bebo White
SLAC National Accelerator Laboratory, Stanford University

Jumping NLP Curves: A Review of Natural Language Processing Research

Natural language processing (NLP) is a theory-motivated range of computational techniques for the automatic analysis and representation of human language. NLP research has evolved from the era of punch cards and batch processing (in which the analysis of a sentence could take up to 7 minutes) to the era of Google and the likes of it (in which millions of webpages can be processed in less than a second). This review paper draws on recent developments in NLP research to look at the past, present, and future of NLP technology in a new light. Borrowing the paradigm of 'jumping curves' from the field of business management and marketing prediction, this survey article reinterprets the evolution of NLP research as the intersection of three overlapping curves-namely Syntactics, Semantics, and Pragmatics Curves- which will eventually lead NLP research to evolve into natural language understanding.

I. Introduction

Between the birth of the Internet and 2003, year of birth of social networks such as MySpace, Delicious, LinkedIn, and Facebook, there were just a few dozen exabytes of information on the Web. Today, that same amount of information is created weekly. The advent of the Social Web has provided people with new content-sharing services that allow them to create and share their

© BRAND X PICTURES

own contents, ideas, and opinions, in a time- and cost-efficient way, with virtually millions of other people connected to the World Wide Web. This huge amount of information, however, is mainly unstructured (because it is specifically produced for human consumption) and hence not directly machine-processable. The automatic analysis of text involves a deep understanding of natural language by machines, a reality from which we are still very far off.

Hitherto, online information retrieval, aggregation, and processing have mainly been based on algorithms relying on the textual representation of web pages. Such algorithms are very good at retrieving texts, splitting them into parts, checking the spelling and counting the number of words. When it comes to interpreting sentences and extracting meaningful information, however, their capabilities are known to be very limited. Natural language processing (NLP), in fact, requires high-level symbolic capabilities (Dyer, 1994), including:

❏ creation and propagation of dynamic bindings;

❏ manipulation of recursive, constituent structures;
❏ acquisition and access of lexical, semantic, and episodic memories;
❏ control of multiple learning/processing modules and routing of information among such modules;
❏ grounding of basic-level language constructs (e.g., objects and actions) in perceptual/motor experiences;
❏ representation of abstract concepts.

All such capabilities are required to shift from mere NLP to what is usually referred to as natural language understanding (Allen, 1987). Today, most of the existing approaches are still based on the syntactic representation of text, a method that relies mainly on word co-occurrence frequencies. Such algorithms are limited by the fact that they can process only the information that they can 'see'. As human text processors, we do not have such limitations as every word we see activates a cascade of semantically related concepts, relevant episodes, and sensory experiences, all of which enable the completion of complex NLP tasks—such as word-sense disambiguation, textual entailment, and semantic role labeling—in a quick and effortless way.

Computational models attempt to bridge such a cognitive gap by emulating the way the human brain processes natural language, e.g., by leveraging on semantic features that are not explicitly expressed in text. Computational models are useful both for scientific purposes (such as exploring the nature of linguistic communication), as well as for

practical purposes (such as enabling effective human-machine communication). Traditional research disciplines do not have the tools to completely address the problem of how language comprehension and production work. Even if you combine all the approaches, a comprehensive theory would be too complex to be studied using traditional methods. However, we may be able to realize such complex theories as computer programs and then test them by observing how well they perform. By seeing where they fail, we can incrementally improve them. Computational models may provide very specific predictions about human behaviors that can then be explored by the psycholinguist. By continuing this process, we may eventually acquire a deeper understanding of how human language processing occurs. To realize such a dream will take the combined efforts of forward-thinking psycholinguists, neuroscientists, anthropologists, philosophers, and computer scientists.

Unlike previous surveys focusing on specific aspects or applications of NLP research (e.g., evaluation criteria (Jones & Galliers, 1995), knowledge-based systems (Mahesh, Nirenburg, & Tucker, 1997), text retrieval (Jackson & Moulinier, 1997), and connectionist models (Christiansen & Chater, 1999)), this review paper focuses on the evolution of NLP research according to three different paradigms, namely: the bag-of-words, bag-of-concepts, and bag-of-narratives models. Borrowing the concept of 'jumping curves' from the field of business management, this survey article explains how and why NLP research has been gradually shifting from lexical semantics to compositional semantics and offers insights on next-generation narrative-based NLP technology.

The rest of the paper is organized as follows: Section 2 presents the historical background and the different schools of thought of NLP research; Section 3 discusses past, present, and future evolution of NLP technologies; Section 4 describes traditional syntax-centered NLP methodologies; Section 5 illustrates emerging semantics-based NLP approaches; Section 6 introduces pioneering works on narrative understanding; Section 7 proposes further insights on the evolution of current NLP technologies and suggests near future research directions; finally, Section 8 concludes the paper and outlines future areas of NLP research.

2. Background

Since its inception in 1950s, NLP research has been focusing on tasks such as machine translation, information retrieval, text summarization, question answering, information extraction, topic modeling, and more recently, opinion mining. Most NLP research carried out in the early days focused on syntax, partly because syntactic processing was manifestly necessary, and partly through implicit or explicit endorsement of the idea of syntax-driven processing.

Although the semantic problems and needs of NLP were clear from the very beginning, the strategy adopted by the research community was to tackle syntax first, for the more direct applicability of machine learning techniques. However, there were some researchers who concentrated on semantics because they saw it as the really challenging problem or assumed that semantically-driven processing be a better approach. Thus, Masterman's and Ceccato's groups, for example, exploited semantic pattern matching using semantic categories and semantic case frames, and in Ceccato's work (Ceccato, 1967) particularly, world knowledge was used to extend linguistic semantics, along with semantic networks as a device for knowledge representation. Later works recognized the need for external knowledge in interpreting and responding to language input (Minsky, 1968) and explicitly emphasized semantics in the form of general-purpose semantics with case structures for representation and semantically-driven processing (Schank, 1975).

One of the most popular representation strategies since then has been first order logic (FOL), a deductive system that consists of axioms and rules of inferences and can be used to formalize relationally-rich predicates and quantifica-

tion (Barwise, 1977). FOL supports syntactic, semantic and, to a certain degree, pragmatic expressions. Syntax specifies the way groups of symbols are to be arranged, so that the group of symbols is considered properly formed. Semantics specifies what well-formed expressions are supposed to mean. Pragmatics specifies how contextual information can be leveraged to provide better correlations between different semantics, which is essential for tasks such as word sense disambiguation. Logic, however, is known to have the problem of monotonicity. The set of entailed sentences will only increase as information is added to the knowledge base, but this runs the risk of violating a common property of human reasoning—the freedom and flexibility to change one's mind. Solutions such as default and linear logic serve to address parts of these issues. Default logic is proposed by Raymond Reiter to formalize default assumptions, e.g., "all birds fly" (Reiter, 1980). However, issues arise when default logic formalizes facts that are true in the majority of cases but are false with regards to exceptions to these 'general rules', e.g., "penguins do not fly".

Another popular model for the description of natural language is production rule (Chomsky, 1956). A production rule system keeps a working memory of on-going memory assertions. This working memory is volatile and in turn keeps a set of production rules. A production rule comprises of an antecedent set of conditions and a consequent set of actions (i.e., IF <conditions> THEN <actions>). The basic operation for a production rule system involves a cycle of three steps ('recognize', 'resolve conflict', and 'act') that repeats until no more rules are applicable to the working memory. The step 'recognize' identifies the rules whose antecedent conditions are satisfied by the current working memory. The set of rules identified is also called the conflict set. The step 'resolve conflict' looks into the conflict set and selects a set of suitable rules to execute. The step 'act' simply executes the actions and updates the working memory. Production rules are modular.

Each rule is independent from the others, allowing rules to be added and deleted easily. Production rule systems have a simple control structure and the rules are easily understood by humans. This is because rules are usually derived from the observation of expert behavior or expert knowledge, thus the terminology used in encoding the rules tends to resonate with human understanding. However, there are issues with scalability when production rule systems become larger; a significant amount of maintenance is required to maintain a system with thousands of rules.

Another instance of a prominent NLP model is the ontology Web language (OWL) (McGuinness & Van Harmelen, 2004), an XML-based vocabulary that extends the resource description framework (RDF) to provide a more comprehensive set for ontology representation, such as the definition of classes, relationships between classes, properties of classes, and constraints on relationships between classes and their properties. RDF supports the subject-predicate-object model that makes assertions about a resource. RDF-based reasoning engines have been developed to check for semantic consistency which then helps to improve ontology classification. In general, OWL requires the strict definition of static structures, and therefore is not suitable for representing knowledge that contains subjective degrees of confidence. Instead, it is more suited for representing declarative knowledge. Furthermore, yet another problem of OWL is that it does not allow for an easy representation of temporal-dependent knowledge.

Networks are yet another well-known way to do NLP. For example, Bayesian networks (Pearl, 1985) (also known as belief networks) provide a means of expressing joint probability distributions over many interrelated hypotheses. All variables are represented using directed acyclic graph (DAG). Arcs are causal connections between two variables where the truth of the former directly affects the truth of the latter. A Bayesian network is able to represent subjective degrees of confidence. The representation explicitly explores the role of prior knowledge and combines pieces of evidence of the likelihood of events. In order to compute the joint distribution of the belief network, there is a need to know $\Pr(P|\mathrm{parents}(P))$ for each variable P. It is difficult to determine the probability of each variable P in the belief network. Hence, it is also difficult to enhance and maintain the statistical table for large-scale information processing problems. Bayesian networks also have limited expressiveness, which is only equivalent to the expressiveness of proposition logic. For this reason, semantic networks are more often used in NLP research.

A semantic network (Sowa, 1987) is a graphical notation for representing knowledge in patterns of interconnected nodes and arcs. Definitional networks focus on *IsA* relationships between a concept and a newly defined sub-type. The result of such a structure is called a generalization, which in turn supports the rule of inheritance for copying properties defined for a super-type to all of its sub-types. The information in definitional networks is often assumed to be true. Yet another kind of semantic networks is the assertional network, which is meant to assert propositions and the information it contains is assumed to be contingently true. Contingent truth is not reached with the application of default logic; instead, it is based more on Man's application of common-sense. The proposition also has sufficient reason in which the reason entails the proposition, e.g., "the stone is warm" with the sufficient reasons being "the sun is shining on the stone" and "whatever the sun shines on is warm".

The idea of semantic networks arose in the early 1960s from Simmons (Simmons, 1963) and Quillian (Quillian, 1963) and was further developed in the late 1980s by Marvin Minsky within his Society of Mind theory (Minsky, 1986), according to which the magic of human intelligence stems from our vast diversity—and not from any single, perfect principle. Minsky theorized that the mind is made of many little parts that he termed 'agents', each mindless by itself but able to lead to true intelligence when working together. These groups of agents, or 'agencies', are responsible for performing some type of function, such as remembering, comparing, generalizing, exemplifying, analogizing, simplifying, predicting, etc. Minsky's theory of human cognition, in particular, was welcomed with great enthusiasm by the artificial intelligence (AI) community and gave birth to many attempts to build common-sense knowledge bases for NLP tasks. The most representative projects are: (a) Cyc (Lenat & Guha, 1989), Doug Lenat's logic-based repository of common-sense knowledge; (b) WordNet (Fellbaum, 1998), Christiane Fellbaum's universal database of word senses; (c) Thought-Treasure (Mueller, 1998), Erik Mueller's story understanding system; and (d) the Open Mind Common Sense project (Singh, 2002), a second-generation common-sense database. The last project stands out because knowledge is represented in natural

TABLE 1 Most popular schools of thought in knowledge representation and NLP research.

APPROACH	CHARACTERISTIC FEATURES	REFERENCE
PRODUCTION RULE	CYCLES OF `RECOGNIZE', `RESOLVE CONFLICT', `ACT' STEPS	(CHOMSKY, 1956)
SEMANTIC PATTERN MATCHING	SEMANTIC CATEGORIES AND SEMANTIC CASE FRAMES	(CECCATO, 1967)
FIRST ORDER LOGIC (FOL)	AXIOMS AND RULES OF INFERENCES	(BARWISE, 1977)
BAYESIAN NETWORKS	VARIABLES REPRESENTED BY A PROBABILISTIC DIRECTED ACYCLIC GRAPH	(PEARL, 1985)
SEMANTIC NETWORKS	PATTERNS OF INTERCONNECTED NODES AND ARCS	(SOWA, 1987)
ONTOLOGY WEB LANGUAGE (OWL)	HIERARCHICAL CLASSES AND RELATIONSHIPS BETWEEN THEM	(MCGUINNESS & VAN HARMELEN, 2004)

language (rather than being based upon a formal logical structure), and information is not hand-crafted by expert engineers but spontaneously inserted by online volunteers. Today, the common-sense knowledge collected by the Open Mind Common Sense project is being exploited for many different NLP tasks such as textual affect sensing (H. Liu, Lieberman, & Selker, 2003), casual conversation understanding (Eagle, Singh, & Pentland, 2003), opinion mining (Cambria & Hussain, 2012), story telling (Hayden et al., 2013), and more.

3. Overlapping NLP Curves

With the dawn of the Internet Age, civilization has undergone profound, rapid-fire changes that we are experiencing more than ever today. Even technologies that are adapting, growing, and innovating have the gnawing sense that obsolescence is right around the corner. NLP research, in particular, has not evolved at the same pace as other technologies in the past 15 years.

While NLP research has made great strides in producing artificially intelligent behaviors, e.g., Google, IBM's Watson, and Apple's Siri, none of such NLP frameworks actually understand what they are doing—making them no different from a parrot that learns to repeat words without any clear understanding of what it is saying. Today, even the most popular NLP technologies view text analysis as a word or pattern matching task. Trying to ascertain the meaning of a piece of text by processing it at word-level, however, is no different from attempting to understand a picture by analyzing it at pixel-level.

In a Web where user-generated content (UGC) is drowning in its own output, NLP researchers are faced with the same challenge: the need to jump the curve (Imparato & Harari, 1996) to make significant, discontinuous leaps in their thinking, whether it is about information retrieval, aggregation, or processing. Relying on arbitrary keywords, punctuation, and word co-occurrence frequencies has worked fairly well so far, but the explosion of UGCs and the outbreak of deceptive

phenomena such as web-trolling and opinion spam, are causing standard NLP algorithms to be increasing less efficient. In order to properly extract and manipulate text meanings, a NLP system must have access to a significant amount of knowledge about the world and the domain of discourse.

To this end, NLP systems will gradually stop relying too much on word-based techniques while starting to exploit semantics more consistently and, hence, make a leap from the Syntactics Curve to the Semantics Curve (Figure 1). NLP research has been interspersed with word-level approaches because, at first glance, the most basic unit of linguistic structure appears to be the word. Single-word expressions, however, are just a subset of concepts, multi-word expressions that carry specific *semantics and sentics* (Cambria & Hussain, 2012), that is, the denotative and connotative information commonly associated with real-world objects, actions, events, and people. Sentics, in particular, specifies the affective information associated with such real-world entities, which is key for common-sense reasoning and decision-making.

Semantics and sentics include common-sense knowledge (which humans normally acquire during the formative years of their lives) and common knowl-

edge (which people continue to accrue in their everyday life) in a re-usable knowledge base for machines. Common knowledge includes general knowledge about the world, e.g., *a chair is a type of furniture*, while common-sense knowledge comprises obvious or widely accepted things that people normally know about the world but which are usually left unstated in discourse, e.g., that *things fall downwards (and not upwards)* and *people smile when they are happy*. The difference between common and common-sense knowledge can be expressed as the difference between knowing the name of an object and understanding the same object's purpose. For example, you can know the name of all the different kinds or brands of 'pipe', but not its purpose nor the method of usage. In other words, a 'pipe' is not a pipe unless it can be used (Magritte, 1929) (Figure 2).

It is through the combined use of common and common-sense knowledge that we can have a grip on both high- and low-level concepts as well as nuances in natural language understanding and therefore effectively communicate with other people without having to continuously ask for definitions and explanations. Common-sense, in particular, is key in properly deconstructing natural language text into sentiments according to different contexts—for
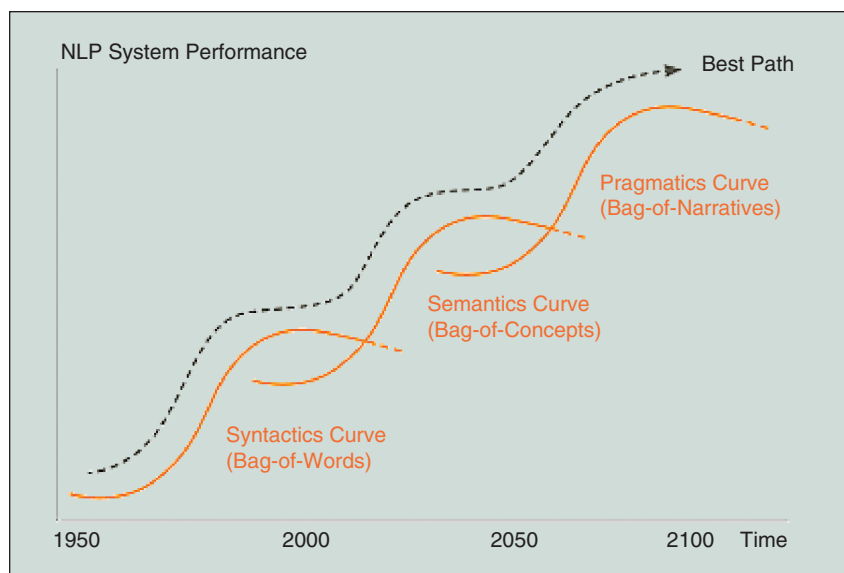

FIGURE 1 Envisioned evolution of NLP research through three different eras or curves.

example, in appraising the concept 'small room' as negative for a hotel review and 'small queue' as positive for a post office, or the concept 'go read the book' as positive for a book review but negative for a movie review.

Semantics, however, is just one layer up in the scale that separates NLP from natural language understanding. In order to achieve the ability to accurately and sensibly process information, computational models will also need to be able to project semantics and sentics in time, compare them in a parallel and dynamic way, according to different contexts and with respect to different actors and their intentions (Howard & Cambria, 2013). This will mean jumping from the Semantics Curve to the Pragmatics Curve, which will enable NLP to be more adaptive and, hence, open-domain, context-aware, and intent-driven. Intent, in particular, will be key for tasks such as sentiment analysis—a concept that generally has a negative connotation, e.g., *small seat*, might turn out to be positive, e.g., if the intent is for an infant to be safely seated in it.

While the paradigm of the Syntactics Curve is the bag-of-words model (Zellig, 1954) and the Semantics Curve is characterized by a bag-of-concepts model (Cambria & Hussain, 2012), the paradigm of the Pragmatics Curve will be the bag-of-narratives model. In this last model, each piece of text will be represented by mini-stories or interconnected episodes, leading to a more detailed level of text comprehension and sensible computation. While the bag-of-concepts model helps to overcome problems such as word-sense disambiguation and semantic role labeling, the bag-of-narratives model will enable tackling NLP issues such as co-reference resolution and textual entailment.

4. Poising on the Syntactics Curve

Today, syntax-centered NLP is still the most popular way to manage tasks such as information retrieval and extraction, auto-categorization, topic modeling, etc. Despite semantics enthusiasts having argued the importance and inevitability of a shift away from syntax for years, the vast majority of NLP researchers nowadays are still trying to keep their balance on the Syntactics Curve. Syntax-centered NLP can be broadly grouped into three main categories: keyword spotting, lexical affinity, and statistical methods.

4.1. Keyword Spotting

Keyword Spotting is the most naïve approach and probably also the most popular because of its accessibility and economy. Text is classified into categories based on the presence of fairly unambiguous words. Popular projects include: (a) Ortony's Affective Lexicon (Ortony, Clore, & Collins, 1988), which groups words into affective categories; (b) Penn Treebank (Marcus, Santorini, & Marcinkiewicz, 1994), a corpus consisting of over 4.5 million words of American English annotated for part-of-speech (POS) information; (c) PageRank (Page, Brin, Motwani, & Winograd, 1999), the famous ranking algorithm of Google; (d) LexRank (Günes & Radev, 2004), a stochastic graph-based method for computing relative importance of textual units for NLP; finally, (e) TextRank (Mihalcea & Tarau, 2004), a graph-based ranking model for text processing, based on two unsupervised methods for keyword and sentence extraction. The major weakness of keyword spotting lies in its reliance upon the presence of obvious words which are only surface features of the prose. A text document about dogs where the word 'dog' is never mentioned, e.g., because dogs are addressed according to the specific breeds they belong to, might never be retrieved by a keyword-based search engine.

4.2. Lexical Affinity

Lexical Affinity is slightly more sophisticated than keyword spotting as, rather than simply detecting obvious words, it assigns to arbitrary words a probabilistic 'affinity' for a particular category (Bush, 1999; Bybee & Scheibman, 1999; Krug, 1998; Church & Hanks, 1989; Jurafsky et al., 2000). For example, 'accident' might be assigned a 75% probability of indicating a negative event, as in 'car accident' or 'hurt in an accident'. These probabilities are usually gleaned from linguistic corpora (Kucera & Francis, 1969; Godfrey, Holliman, & McDaniel, 1992; Stevenson, Mikels, & James, 2007). Although this approach often outperforms pure keyword spotting, there are two main problems with it. First, lexical affinity operating solely on the word-level can easily be tricked by sentences such as "I avoided an accident" (negation) and "I met my girlfriend by accident" (connotation of unplanned but lovely surprise). Second, lexical affinity probabilities are often biased toward text of a particular genre, dictated by the source of the linguistic corpora. This makes it difficult to develop a re-usable, domain-independent model.

4.3. Statistical NLP

Statistical NLP has been the mainstream NLP research direction since late 1990s. It relies on language models (Manning & Schütze, 1999; Hofmann, 1999; Nigam, McCallum, Thrun, & Mitchell, 2000) based on popular machine-learning algorithms such as maximum-likelihood (Berger, Della Pietra, & Della Pietra, 1996), expectation maximization (Nigam et al., 2000), conditional random fields (Lafferty, McCallum, & Pereira, 2001), and support vector machines (Joachims, 2002). By feeding a large training corpus of annotated texts to a machine-learning algorithm, it is possible for the system to not only learn the valence of keywords (as in the keyword spotting approach), but also to take into account the valence of other arbitrary keywords (like lexical affinity),

FIGURE 2 A 'pipe' is not a pipe, unless we know how to use it.

punctuation, and word co-occurrence frequencies. However, statistical methods are generally semantically weak, meaning that, with the exception of obvious keywords, other lexical or co-occurrence elements in a statistical model have little predictive value individually. As a result, statistical text classifiers only work with acceptable accuracy when given a sufficiently large text input. So, while these methods may be able to classify text on the page- or paragraph-level, they do not work well on smaller text units such as sentences or clauses.

5. Surfing the Semantics Curve

Semantics-based NLP focuses on the intrinsic meaning associated with natural language text. Rather than simply processing documents at syntax-level, semantics-based approaches rely on implicit denotative features associated with natural language text, hence stepping away from the blind usage of keywords and word co-occurrence count. Unlike purely syntactical techniques, concept-based approaches are also able to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so. Semantics-based NLP approaches can be broadly grouped into two main categories: techniques that leverage on external knowledge, e.g., ontologies (taxonomic NLP) or semantic knowledge bases (noetic NLP), and methods that exploit only intrinsic semantics of documents (endogenous NLP).

5.1. Endogenous NLP

Endogenous NLP involves the use of machine-learning techniques to perform semantic analysis of a corpus by building structures that approximate concepts from a large set of documents. It does not involve prior semantic understanding of documents; instead, it relies only on the endogenous knowledge of these (rather than on external knowledge bases). The advantages of this approach over the knowledge engineering approach are effectiveness, consider-

able savings in terms of expert manpower, and straightforward portability to different domains (Sebastiani, 2002).

Endogenous NLP includes methods based either on lexical semantics, which focuses on the meanings of individual words, or compositional semantics, which looks at the meanings of sentences and longer utterances. The vast majority of endogenous NLP approaches is based on lexical semantics and includes well-known machine-learning techniques. Some examples of this are: (a) latent semantic analysis (Hofmann, 2001), where documents are represented as vectors in a term space; (b) latent Dirichlet allocation (Porteous et al., 2008), which involves attributing document terms to topics; (c) MapReduce (C. Liu, Qi, Wang, & Yu, 2012), a framework that has proved to be very efficient for data-intensive tasks, e.g., large scale RDFS/OWL reasoning and (d) genetic algorithms (D. Goldberg, 1989), probabilistic search procedures designed to work on large spaces involving states that can be represented by strings.

Works leveraging on compositional semantics, instead, mainly include approaches based on Hidden Markov Models (Denoyer, Zaragoza, & Gallinari, 2001; Frasconi, Soda, & Vullo, 2001), association rule learning (Cohen, 1995; Cohen & Singer, 1999), feature ensembles (Xia, Zong, Hu, & Cambria, 2013; Poria, Gelbukh, Hussain, Das, & Bandyopadhyay, 2013) and probabilistic generative models (Lau, Xia, & Ye, 2014).

5.2. Taxonomic NLP

Taxonomic NLP includes initiatives that aim to build universal taxonomies or Web ontologies for grasping the subsumptive or hierarchical semantics associated with natural language expressions. Such taxonomies usually consist of concepts (e.g., *painter*), instances (e.g., "Leonardo da Vinci"), attributes and values (e.g., "Leonardo's birthday is April 15, 1452"), and relationships (e.g., "Mona Lisa is painted by Leonardo"). In particular, subsumptive knowledge representations build upon *IsA* relationships, which are usually extracted

through syntactic patterns for automatic hypernym discovery (Hearst, 1992) able to infer triples such as <Pablo Picasso-IsA-artist> from stretches of text like "...artists such as Pablo Picasso..." or "...Pablo Picasso and other artists...".

In general, attempts to build taxonomic resources are countless and include both resources crafted by human experts or community efforts, such as WordNet and Freebase (Bollacker, Evans, Paritosh, Sturge, & Taylor, 2008), and automatically built knowledge bases. Examples of such knowledge bases include: (a) WikiTaxonomy (Ponzetto & Strube, 2007), a taxonomy extracted from Wikipedia's category links; (b) YAGO (Suchanek, Kasneci, & Weikum, 2007), a semantic knowledge base derived from WordNet, Wikipedia, and GeoNames; (c) NELL (Carlson et al., 2010) (Never-Ending Language Learning), a semantic machine-learning system that is acquiring knowledge from the Web every day; finally, (d) Probase (Wu, Li, Wang, & Zhu, 2012), a research prototype that aims to build a unified taxonomy of worldly facts from 1.68 billion webpages in Bing repository.

Other popular Semantic Web projects include: (a) SHOE (Heflin & Hendler, 1999) (Simple HTML Ontology Extensions), a knowledge representation language that allows webpages to be annotated with semantics; (b) Annotea (Kahan, 2002), an open RDF infrastructure for shared Web annotations; (c) SIOC (Breslin, Harth, Bojars, & Decker, 2005) (Semantically Interlinked Online Communities), an ontology combining terms from vocabularies that already exist with new terms needed to describe the relationships between concepts in the realm of online community sites; (d) SKOS (Miles & Bechhofer, 2009) (Simple Knowledge Organization System), an area of work developing specifications and standards to support the use of knowledge organization systems such as thesauri, classification schemes, subject heading lists and taxonomies; (e) FOAF (Brickley & Miller, 2010) (Friend Of A Friend), a project devoted

to linking people and information using the Web; (f) ISOS (Ding, Jin, Ren, & Hao, 2013) (Intelligent Self-Organizing Scheme), a scheme for the Internet of Things inspired by the endocrine regulating mechanism; finally, (g) FRED (Gangemi, Presutti, & Reforgiato, 2014), a tool that produces an event-based RDF/OWL representation of natural language text. The main weakness of taxonomic NLP is in the typicality of their knowledge bases. The way knowledge is represented in taxonomies and Web ontologies is usually strictly defined and does not allow for the combined handling of differing nuanced concepts, as the inference of semantic features associated with concepts is bound by the fixed, flat representation. The concept of 'book', for example, is typically associated to concepts such as 'newspaper' or 'magazine', as it contains knowledge, has pages, etc. In a different context, however, a book could be used as paperweight, doorstop, or even as a weapon. Another key weakness of Semantic Web projects is that they are not easily scalable and, hence, not widely adopted (Gueret, Schlobach, Dentler, Schut, & Eiben, 2012). This increases the amount of time that has to pass before the initial customer feedback is even possible, and also slows down feedback loop iterations, ultimately putting Semantic Web applications at a user-experience and agility disadvantage as compared to their Web 2.0 counterparts, because their usability inadvertently takes a back seat to the number of other complex problems that have to be solved before clients even see the application.

5.3. Noetic NLP

Noetic NLP embraces all the mind-inspired approaches to NLP that attempt to compensate for the lack of domain adaptivity and implicit semantic feature inference of traditional algorithms, e.g., first principles modeling or explicit statistical modeling. Noetic NLP differs from taxonomic NLP in which it does not focus on encoding subsumption knowledge, but rather attempts to collect idiosyncratic knowledge about objects, actions, events, and people. Noetic NLP, moreover, performs reasoning in an adaptive and dynamic way, e.g., by generating context-dependent results or by discovering new semantic patterns that are not explicitly encoded in the knowledge base. Examples of noetic NLP include paradigms such as connectionist NLP (Christiansen & Chater, 1999), which models mental phenomena as emergent processes of interconnected networks of simple units, e.g., neural networks (Collobert et al., 2011); deep learning (Martinez, Bengio, & Yannakakis, 2013); sentic computing (Cambria & Hussain, 2012), an approach to concept-level sentiment analysis based on an ensemble of graph-mining and dimensionality-reduction techniques; and energy-based knowledge representation (Olsher, 2013), a novel framework for nuanced common-sense reasoning.

Besides knowledge representation and reasoning, a key aspect of noetic NLP is also semantic parsing. Most current NLP technologies rely on part-of-speech (POS) tagging, but that is unlike the way the human mind extracts meaning from text. Instead, just as the human mind does, a construction-based semantic parser (CBSP) (Cambria, Rajagopal, Olsher, & Das, 2013) quickly identifies meaningful stretches of text without requiring time-consuming phrase structure analysis. The use of *constructions*, defined as "stored pairings of form and function" (A. Goldberg, 2003) makes it possible to link distributed linguistic components to one another, easing extraction of semantics from linguistic structures. Constructions are composed of fixed lexical items and category-based *slots*, or 'spaces' that are filled in by lexical items during text processing. An interesting example from the relevant literature would be the construction [<ACTION> <OBJECT> <DIRECTION> <OBJECT>]. Instances of this include the phrases 'sneeze the napkin across the table' or 'hit the ball over the fence'. Constructions not only help understand how various lexical items work together to create the whole meaning, but also give the parser a sense of what categories of words are used together and thus where to expect different words.

CBSP uses this knowledge to determine constructions, their matching lexical terms, and how good each match is. Each of CBSP's constructions contributes its own unique semantics and carries a unique name. In order to choose the best possible construction for each span of text, CBSP uses knowledge about the lexical items found in text. This knowledge is obtained from looking individual lexical terms up in the knowledge bases so as to obtain information about the basic category membership of that word.

It then efficiently compares these potential memberships with the categories specified for each construction in the corpus, finding the best matches so that CBSP can extract a concept from a sentence. An example would be the extraction of the concept 'buy christmas present' from the sentence "today I bought a lot of very nice Christmas gifts". Constructions are typically *nested* within one another: CBSP is capable of finding only those construction overlaps that are semantically sensible, based on the overall semantics of constructions and construction slot categories, thus greatly reducing the time taken to process large numbers of texts. In the big data environment, a key benefit of construction-based parsing is that only small sections of text are required in order to extract meaning; word category information and the generally small size of constructions mean that the parser can still make use of error-filled or conventionally unparseable text.

6. Foreseeing the Pragmatics Curve

Narrative understanding and generation are central for reasoning, decision-making, and 'sensemaking'. Besides being a key part of human-to-human communication, narratives are the means by which reality is constructed and planning is conducted. Decoding how narratives are generated and processed by the human brain might eventually lead us to truly understand and explain human intelligence and consciousness.

Computational modeling is a powerful and effective way to investigate narrative understanding. A lot of the cognitive processes that lead humans to understand or generate narratives have traditionally been of interest to AI researchers under the umbrella of knowledge representation, common-sense reasoning, social cognition, learning, and NLP. Once NLP research can grasp semantics at a level comparable to human text processing, the jump to the Pragmatics Curve will be necessary, in the same way as semantic machine learning is now gradually evolving from lexical to compositional semantics. There are already a few pioneering works that attempt to understand narratives by leveraging on discourse structure (Asher & Lascarides, 2003), argument-support hierarchies (Bex, Prakken, & Verheij, 2007), plan graphs (Young, 2007), and common-sense reasoning (Mueller, 2007). One of the most representative initiatives in this context is Patrick Winston's work on computational models of narrative (Winston, 2011; Richards, Finlayson, & Winston, 2009), which is based on five key hypotheses:

❏ The inner language hypothesis: we have an inner symbolic language that enables event description.
❏ The strong story hypothesis: we can assemble event descriptions into stories.
❏ The directed perception hypothesis: we can direct the resources of our perceptual faculties to answer questions using real and imagined situations.
❏ The social animal hypothesis: we have a powerful reason to express the thought in our inner language in an external communication language.
❏ The exotic engineering hypothesis: our brains are unlike standard left-to-right engineered systems.

Essentially, Patrick Winston believes that human intelligence stems from our unique abilities for storytelling and understanding (Finlayson & Winston, 2011). Accordingly, his recent work has focused on developing a computational system that is able to analyze narrative texts to infer non-obvious answers to questions about these texts. This has resulted in the Genesis System. Working with short story summaries provided in English, together with low-level common-sense rules and higher-level reflection patterns that are also expressed in English, Genesis has been successful in demonstrating several story understanding capabilities. One instance of this is its ability to determine that both Macbeth and the 2007 Russia-Estonia Cyberwar involve revenge, even though neither the word 'revenge' nor any of its synonyms are mentioned in accounts describing those texts.

7. Discussion

Word- and concept-level approaches to NLP are just a first step towards natural language understanding. The future of NLP lies in biologically and linguistically motivated computational paradigms that enable narrative understanding and, hence, 'sensemaking'. Computational intelligence potentially has a large future possibility to play an important role in NLP research. Fuzzy logic, for example, has a direct relation to NLP (Carvalho, Batista, & Coheur, 2012) for tasks such as sentiment analysis (Subasic & Huettner, 2001), linguistic summarization (Kacprzyk & Zadrozny, 2010), knowledge representation (Lai, Wu, Lin, & Huang, 2011), and word meaning inference (Kazemzadeh, Lee, & Narayanan, 2013). Artificial neural networks can aid the completion of NLP tasks such as ambiguity resolution (Chan & Franklin, 1998; Costa, Frasconi, Lombardo, & Soda, 2005), grammatical inference (Lawrence, Giles, & Fong, 2000), word representation (Luong, Socher, & Manning, 2013), and emotion recognition (Cambria, Gastaldo, Bisio, & Zunino, 2014). Evolutionary computation can be exploited for tasks such as grammatical evolution (O'Neill & Ryan, 2001), knowledge discovery (Atkinson-Abutridy, Mellish, & Aitken, 2003), text categorization (Araujo, 2004), and rule learning (Ghandar, Michalewicz, Schmidt, To, & Zurbruegg, 2009).

Despite its potential, however, the use of computational intelligence techniques till date has not been so active in the field of NLP. The first reason is that NLP is a huge field currently tackling dozens of different problems for which specific evaluation metrics exist, and it is not possible to reduce the whole field into a specific problem, as it was done in early works (Novak, 1992). The second reason may be that powerful techniques such as support vector machines (Drucker, Wu, & Vapnik, 1999), kernel principal component analysis (Schölkopf et al., 1999), and latent Dirichlet allocation (Mukherjee & Blei, 2009) have achieved remarkable results on widely used NLP datasets, which are not yet met by computational intelligence techniques. All such word-based algorithms, however, are limited by the fact that they can process only the information that they can 'see' and, hence, will sooner or later reach saturation. Computational intelligence techniques, instead, can go beyond the syntactic representation of documents by emulating the way the human brain processes natural language (e.g., by leveraging on semantic features that are not explicitly expressed in text) and, hence, have higher potential to tackle complementary NLP tasks. An ensemble of computational intelligence techniques, for example, could be exploited within the same NLP model for online learning of natural language concepts (through neural networks), concept classification and semantic feature generalization (through fuzzy sets), and concept meaning evolution and continuous system optimization (through evolutionary computation).

8. Conclusion

In a Web where user-generated content has already hit critical mass, the need for sensible computation and information aggregation is increasing exponentially, as demonstrated by the 'mad rush' in the industry for 'big data experts' and the growth of a new 'Data Science' discipline. The democratization of online content creation has led to the increase of Web debris, which is inevitably and negatively affecting information retrieval and extraction. To analyze this negative trend and propose possible solutions, this

review paper focused on the evolution of NLP research according to three different paradigms, namely: the bag-of-words, bag-of-concepts, and bag-of-narratives models. Borrowing the concept of 'jumping curves' from the field of business management, this survey article explained how and why NLP research is gradually shifting from lexical semantics to compositional semantics and offered insights on next-generation narrative-based NLP technology.

Jumping the curve, however, is not an easy task: the origins of human language has sometimes been called the hardest problem of science (Christiansen & Kirby, 2003). NLP technologies evolved from the era of punch cards and batch processing (in which the analysis of a natural language sentence could take up to 7 minutes (Plath, 1967)) to the era of Google and the likes of it (in which millions of webpages can be processed in less than a second). Even the most efficient word-based algorithms, however, perform very poorly, if not properly trained or when contexts and domains change. Such algorithms are limited by the fact that they can process only information that they can 'see'. Language, however, is a system where all terms are interdependent and where the value of one is the result of the simultaneous presence of the others (De Saussure, 1916). As human text processors, we 'see more than what we see' (Davidson, 1997) in which every word activates a cascade of semantically-related concepts that enable the completion of complex NLP tasks, such as word-sense disambiguation, textual entailment, and semantic role labeling, in a quick and effortless way.

Concepts are the glue that holds our mental world together (Murphy, 2004). Without concepts, there would be no mental world in the first place (Bloom, 2003). Needless to say, the ability to organize knowledge into concepts is one of the defining characteristics of the human mind. A truly intelligent system needs physical knowledge of how objects behave, social knowledge of how people interact, sensory knowledge of how things look and taste, psychological

knowledge about the way people think, and so on. Having a database of millions of common-sense facts, however, is not enough for computational natural language understanding: we will need to teach NLP systems how to handle this knowledge (IQ), but also interpret emotions (EQ) and cultural nuances (CQ).

References

[1] J. Allen, *Natural Language Understanding*. Redwood City, CA: Benjamin/Cummings, 1987.
[2] L. Araujo, "Symbiosis of evolutionary techniques and statistical natural language processing," *IEEE Trans. Evol. Comput.*, vol. 8, no. 1, pp. 14–27, 2004.
[3] N. Asher and A. Lascarides, *Logics of Conversation*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[4] J. Atkinson-Abutridy, C. Mellish, and S. Aitken, "A semantically guided and domain independent evolutionary model for knowledge discovery from texts," *IEEE Trans. Evol. Comput.*, vol. 7, no. 6, pp. 546–560, 2003.
[5] J. Barwise, "An introduction to first-order logic," in *Handbook of Mathematical Logic*. (Studies in Logic and the Foundations of Mathematics). Amsterdam, The Netherlands: North-Holland, 1977.
[6] A. Berger, V. D. Pietra, and S. D. Pietra, "A maximum entropy approach to natural language processing," *Comput. Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.
[7] F. Bex, H. Prakken, and B. Verheij, "Formalizing argumentative story-based analysis of evidence," in *Proc. Int. Conf. Artificial Intelligence Law*, 2007, pp. 1-10.
[8] P. Bloom, "Glue for the mental world," *Nature*, vol. 421, pp. 212–213, Jan. 2003.
[9] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," in *Proc. ACM SIGMOD Int. Conf. Management Data*, 2008, pp. 1247–1250.
[10] J. Breslin, A. Harth, U. Bojars, and S. Decker, "Towards semantically-interlinked online communities," in *The Semantic Web: Research and Applications*. Berlin Heidelberg: Springer-Verlag, 2005, pp. 500–514.
[11] D. Brickley and L. Miller. (2010). FOAF vocabulary specification 0.98. Namespace Document [Online]. Available: http://xmlns.com/foaf/spec/
[12] N. Bush, "The predictive value of transitional probability for word-boundary palatalization in English," Unpublished M.S .thesis, Univ. New Mexico, Albuquerque, NM, 1999.
[13] J. Bybee and J. Scheibman, "The effect of usage on degrees of constituency: The reduction of don't in English," *Linguistics*, vol. 37, no. 4, pp. 575–596, 1999.
[14] E. Cambria, P. Gastaldo, F. Bisio, and R. Zunino, "An ELM-based model for affective analogical reasoning," *Neurocomputing*, Special Issue on Extreme Learning Machines, 2014.
[15] E. Cambria and A. Hussain, *Sentic Computing: Techniques, Tools, and Applications*. Dordrecht, The Netherlands: Springer-Verlag, 2012.
[16] E. Cambria, D. Rajagopal, D. Olsher, and D. Das, "Big social data analysis," in *Big Data Computing*, R. Akerkar, Ed. London: Chapman and Hall, 2013, pp. 401–414.
[17] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. Hruschka, and T. Mitchell, "Toward an architecture for never-ending language learning," in *Proc. Conf. Artificial Intelligence AAAI*, Atlanta, GA, 2010, pp. 1306–1313.
[18] J. Carvalho, F. Batista, and L. Coheur, "A critical survey on the use of fuzzy sets in speech and natural language processing," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2012, pp. 270–277.
[19] S. Ceccato, "Correlational analysis and mechanical translation," in *Machine Translation*, A. D. Booth, Ed. Amsterdam, The Netherlands: North Holland, 1967, pp. 77–135.

[20] S. Chan and J. Franklin, "Symbolic connectionism in natural language disambiguation," *IEEE Trans. Neural Netw.*, vol. 9, no. 5, pp. 739–755, 1998.
[21] N. Chomsky, "Three models for the description of language," *IRE Trans. Inform. Theory*, vol. 2, no. 3, pp. 113–124, 1956.
[22] M. Christiansen and N. Chater, "Connectionist natural language processing: The state of the art," *Cogn. Sci.*, vol. 23, no. 4, pp. 417–437, 1999.
[23] M. Christiansen and S. Kirby, "Language evolution: The hardest problem in science?" in *Language Evolution*, M. Christiansen and S. Kirby, Eds. Oxford, U.K.: Oxford Univ. Press, 2003. pp. 1–15.
[24] K. Church and P. Hanks, "Word association norms, mutual information, and lexicography," in *Proc. 27th Annu. Meeting Association Computational Linguistics*, 1989, pp. 76–83.
[25] W. Cohen, "Learning to classify English text with ILP methods," in *Advances in Inductive Logic Programming*, L. De Raedt, Ed. Amsterdam, The Netherlands: IOS Press, 1995, pp. 124–143.
[26] W. Cohen and Y. Singer, "Context-sensitive learning methods for text categorization," *ACM Trans. Inform. Syst.*, vol. 17, no. 2, pp. 141–173, 1999.
[27] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.
[28] F. Costa, P. Frasconi, V. Lombardo, P. Sturt, and G. Soda, "Ambiguity resolution analysis in incremental parsing of natural language," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 959–971, 2005.
[29] D. Davidson, "Seeing through language," in *Royal Institute of Philosophy, Supplement*. Cambridge, U.K.: Cambridge Univ. Press, 1997, vol. 42 , pp. 15–28.
[30] L. Denoyer, H. Zaragoza, and P. Gallinari, "HMM-based passage models for document classification and ranking," in *Proc. 23rd European Colloq. Information Retrieval Research*, Darmstadt, Germany, 2001.
[31] F. de Saussure, *Cours de Linguistique Générale*. Paris: Payot, 1916.
[32] Y. Ding, Y. Jin, L. Ren, and K. Hao, "An intelligent self-organization scheme for the Internet of things," *IEEE Comput. Intell. Mag.*, vol. 8, no. 3, pp. 41–53, 2013.
[33] H. Drucker, D. Wu, and V. Vapnik, "Support vector machines for spam categorization," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1048–1054, 1999.
[34] M. Dyer, "Connectionist natural language processing: A status report," in *Computational Architectures Integrating Neural and Symbolic Processes*, R. Sun and L. Bookman, Eds. Dordrecht, The Netherlands: Kluwer Academic, 1995, vol. 292, pp. 389–429.
[35] N. Eagle, P. Singh, and A. Pentland, "Common sense conversations: Understanding casual conversation using a common sense database," in *Proc. Int. Joint Conf. Artificial Intelligence*, 2003.
[36] C. Fellbaum, *WordNet: An Electronic Lexical Database* (language, speech, and communication). Cambridge, MA: The MIT Press, 1998.
[37] M. Finlayson and P. Winston, "Narrative is a key cognitive competency," in *Proc. 2nd Annu. Meeting Biologically Inspired Cognitive Architectures*, 2011, p. 110.
[38] P. Frasconi, G. Soda, and A. Vullo, "Text categorization for multi-page documents: A hybrid naive Bayes HMM approach," *J. Intell. Inform. Syst.*, vol. 18, nos. 2–3, pp. 195–217, 2001.
[39] A. Gangemi, V. Presutti, D. Reforgiato, "Frame-based detection of opinion holders and topics: A model and a tool," *IEEE Comput. Intell. Mag.*, vol. 9, no. 1, pp. 20–30, 2014.
[40] A. Ghandar, Z. Michalewicz, M. Schmidt, T. To, and R. Zurbruegg, "Computational intelligence for evolving trading rules," *IEEE Trans. Evol. Comput.*, vol. 13, no. 1, pp. 71–86, 2009.
[41] J. Godfrey, E. Holliman, and J. McDaniel, "Switchboard: Telephone speech corpus for research and development," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1992, pp. 517–520.

[42] A. Goldberg, "Constructions: A new theoretical approach to language," *Trends Cogn. Sci.*, vol. 7, no. 5, pp. 219–224, 2003.

[43] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley, 1989.

[44] C. Gueret, S. Schlobach, K. Dentler, M. Schut, and G. Eiben, "Evolutionary and swarm computing for the semantic Web," *IEEE Comput. Intell. Mag.*, vol. 7, no. 2, pp. 16–31, 2012.

[45] E. Günes and D. Radev, "LexRank: Graph-based lexical centrality as salience in text summarization," *J. Artif. Intell. Res.*, vol. 22, no. 1, pp. 457–479, 2004.

[46] K. Hayden, D. Novy, C. Havasi, M. Bove, S. Alfaro, and R. Speer, "Narratarium: An immersive storytelling environment," in *Proc. Human-Computer Interaction*, 2013, vol. 374, pp. 536–540.

[47] M. Hearst, "Automatic acquisition of hyponyms from large text corpora," in *Proc. 14th Conf. Computational Linguistics*, 1992, pp. 539–545.

[48] J. Heflin and J. Hendler, "Shoe: A knowledge representation language for internet applications," Univ. Maryland, College Park, Maryland, Tech. Rep., 1999.

[49] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. 22nd Annu. Int. ACM SIGIR Conf. Research Development Information Retrieval*, 1999, p. 50–57.

[50] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learn.*, vol. 42, nos. 1–2, pp. 177–196, 2001.

[51] N. Howard and E. Cambria, "Intention awareness: Improving upon situation awareness in human-centric environments," *Human-Centric Computing Information Sciences.* vol. 3, Cambridge, MA: Springer-Verlag, 2013. no. 9.

[52] N. Imparato and O. Harari, *Jumping the Curve: Innovation and Strategic Choice in An Age of Transition.* San Francisco, CA: Jossey-Bass, 1996.

[53] P. Jackson and I. Moulinier, *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization.* Philadelphia, PA: John Benjamins. 1997.

[54] T. Joachims, *Learning To Classify Text Using Support Vector Machines: Methods, Theory and Algorithms.* Norwell, MA: Kluwer Academic, 2002.

[55] K. Jones and J. Galliers, "Evaluating natural language processing systems: An analysis and review," *Comput. Linguistics*, vol. 24, no. 2, 1995.

[56] D. Jurafsky, A. Bell, M. Gregory, W. Raymond, J. Bybee, and P. Hopper, *Probabilistic Relations Between Words: Evidence From Reduction In Lexical Production.* Amsterdam, The Netherlands: John Benjamins, 2000.

[57] J. Kacprzyk and S. Zadrozny, "Computing with words is an implementable paradigm: Fuzzy queries, linguistic data summaries, and natural-language generation," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 461–472. 2010.

[58] J. Kahan, "Annotea: An open RDF infrastructure for shared web annotations," *Comput. Netw.*, vol. 39, no. 5, pp. 589–608, 2002.

[59] A. Kazemzadeh, S. Lee, and S. Narayanan, "Fuzzy logic models for the meaning of emotion words," *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 34–49, 2013.

[60] M. Krug, "String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change," *J. Eng. Linguistics*, vol. 26, no. 4, pp. 286–320, 1998.

[61] H. Kucera and N. Francis, "Computational analysis of present-day American English," *Int. J. Amer. Linguistics*, vol. 35, no. 1, pp. 71–75, 1969.

[62] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Machine Learning*, 2001, pp. 282–289.

[63] L. Lai, C. Wu, P. Lin, and L. Huang, "Developing a fuzzy search engine based on fuzzy ontology and semantic search," in *Proc. IEEE Int. Conf. Fuzzy Systems*, Taipei, Taiwan, 2011, pp. 2684–2689.

[64] R. Lau, Y. Xia, and Y. Ye, "A probabilistic generative model for mining cybercriminal networks from online social media," *IEEE Comput. Intell. Mag.*, vol. 9, no. 1, pp. 31–43, 2014.

[65] S. Lawrence, C. Giles, and S. Fong, "Natural language grammatical inference with recurrent neural networks," *IEEE Trans. Knowledge. Data Eng.*, vol. 12, no. 1, pp. 126–140, 2000.

[66] D. Lenat and R. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project.* Boston, MA: Addison-Wesley, 1989.

[67] C. Liu, G. Qi, H. Wang, and Y. Yu, "Reasoning with large scale ontologies in fuzzy pD⋆ using mapreduce," *IEEE Comput. Intell. Mag.*, vol. 7, no. 27, pp. 54–66, 2012.

[68] H. Liu, H. Lieberman, and T. Selker, "A model of textual affect sensing using real-world knowledge," in *Proc. 8th Int. Conf. Intelligent User Interfaces,* 2003, pp. 125–132.

[69] M. Luong, R. Socher, and C. Manning, "Better word representations with recursive neural networks for morphology," in *Proc. Conf. Natural Language Learning*, 2013.

[70] R. Magritte, "Les mots et les images," *La Révolution surréaliste*, no. 12, 1929.

[71] K. Mahesh, S. Nirenburg, and A. Tucker, *Knowledge-Based Systems for Natural Language Processing.* Boca Raton, FL: CRC Press, 1997.

[72] C. Manning, and H. Schütze, *Foundations of Statistical Natural Language Processing.* Cambridge, MA: MIT press, 1999.

[73] M. Marcus, B. Santorini, and M. Marcinkiewicz, "Building a large annotated corpus of english: The penn treebank," *Comput. Linguistics*, vol. 19, no. 2, pp. 313–330, 1994.

[74] H. Martinez, Y. Bengio, and G. Yannakakis, "Learning deep physiological models of affect," *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 20–33, 2013.

[75] D. McGuinness and F. Van Harmelen, OWL web ontology language overview, *W3C recommendation*, 2004.

[76] R. Mihalcea, and P. Tarau, "TextRank: Bringing order into texts," in *Proc. Conf. Empirical Methods Natural Language Processing*, Barcelona, 2004.

[77] A. Miles and S. Bechhofer, "SKOS simple knowledge organization system reference," *W3C Recommendation,* Tech. Rep. 2009.

[78] M. Minsky, *Semantic Information Processing.* Cambridge, MA: MIT Press, 1968.

[79] M. Minsky, *The Society of Mind.* New York: Simon and Schuster, 1986.

[80] E. Mueller, *Natural Language Processing with Thought-Treasure.* New York: Signifonn, 1998.

[81] E. Mueller, "Modeling space and time in narratives about restaurants," *Literary Linguistic Comput.*, vol. 22, no. 1, pp. 67–84, 2007.

[82] I. Mukherjee, and D. Blei, "Relative performance guarantees for approximate inference in latent dirichlet allocation," in *Proc. Neural Information Processing Systems,* Vancouver, BC, 2009, pp. 1129–1136.

[83] G. Murphy, *The Big Book of Concepts.* Cambridge, MA: MIT Press, 2004.

[84] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using EM," *Machine Learn.*, vol. 39, nos. 2–3, pp. 103–134, 2000.

[85] V. Novak, "Fuzzy sets in natural language processing," in *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, Yager Ed. Norwell, MA: Kluwer Academic, 1992, pp. 185–200.

[86] D. Olsher, "COGVIEW & INTELNET: Nuanced energy-based knowledge representation and integrated cognitive-conceptual framework for realistic culture, values, and concept-affected systems simulation," in *Proc. 2013 IEEE Symp. Computational Intelligence Human-Like Intelligence*, Singapore, 2013, pp. 82–91.

[87] M. O'Neill and C. Ryan, "Grammatical evolution," *IEEE Trans. Evol. Comput.*, vol. 5, no. 4, pp. 349–358, 2001.

[88] A. Ortony, G. Clore, and A. Collins, "The cognitive structure of emotions," Cambridge, U.K.: Cambridge Univ. Press, 1988.

[89] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web," Stanford Univ., Stanford, CA, Tech. Rep., 1999.

[90] J. Pearl, "Bayesian networks: A model of self-activated memory for evidential reasoning," UCLA comput. Sci., Irvine, CA: Tech. Rep. CSD-850017, 1985.

[91] W. Plath, "Multiple path analysis and automatic translation," in *Machine Translation*, A. D. Booth, Ed. Amsterdam, The Netherlands: North-Holland, 1967, pp. 267–315.

[92] S. Ponzetto and M. Strube, "Deriving a large-scale taxonomy from Wikipedia," in *Proc. AAAI'07 22nd Nat. Conf. Artificial Intelligence*, Vancouver, BC, 2007, pp. 1440–1445.

[93] S. Poria, A. Gelbukh, A. Hussain, D. Das, and S. Bandyopadhyay, "Enhanced SenticNet with affective labels for concept-based opinion mining," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 31–38, 2013.

[94] I. Porteous, I. Newman, A. Ihler, A. Asuncion, P. Smyth, and M. Welling, "Fast collapsed Gibbs sampling for latent dirichlet allocation," in *Proc. 14th ACM SIG-KDD Int. Conf. Knowledge Discovery Data Mining*, 2008, pp. 569–577.

[95] R. Quillian, "A notation for representing conceptual information: An application to semantics and mechanical english paraphrasing," System Development Corp., Santa Monica, California, Tech. Rep. SP-1395, 1963.

[96] R. Reiter, "A logic for default reasoning," *Artificial Intell.*, vol. 13, pp. 81–132, 1980.

[97] W. Richards, M. Finlayson, and P. Winston, "Advancing computational models of narrative," MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, Tech. Rep. 2009-063, 2009.

[98] R. Schank, *Conceptual Information Processing.* Amsterdam, The Netherlands: Elsevier Science Inc., 1975.

[99] B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A. Smola, "Input space versus feature space in kernel-based methods," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1000–1017, 1999.

[100] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, no. 1, pp. 1–47, 2002.

[101] R. Simmons, "Synthetic language behavior," *Data Processing Manage.*, vol. 5, no. 12, pp. 11–18, 1963.

[102] P. Singh. (2002). The open mind common sense project. [Online]. Available: http://www.kurzweilai.net/

[103] J. Sowa, "Semantic networks," in *Encyclopedia of Artificial Intelligence*, S. Shapiro, Ed. New York: Wiley, 1987.

[104] R. Stevenson, J. Mikels, and T. James, "Characterization of the affective norms for english words by discrete emotional categories," *Behav. Res. Methods*, vol. 39, no. 4, pp. 1020–1024, 2007.

[105] P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 4, pp. 483–496, 2001.

[106] F. Suchanek, G. Kasneci, and G. Weikum, "Yago: A core of semantic knowledge," in *Proc. 16th Int. World Wide Web Conf.*, 2007. pp. 697–706.

[107] P. Winston, "The strong story hypothesis and the directed perception hypothesis," in *Proc. AAAi Fall Symp.: Advances Cognitive Systems*, 2011.

[108] W. Wu, H. Li, H. Wang, and K. Zhu, "Probase: A probabilistic taxonomy for text understanding," in *Proc. ACM SIGMOD Int. Conf. Management Data,* Scottsdale, AZ, 2012, pp. 481–492.

[109] R. Xia, C. Zong, X. Hu, and E. Cambria, "Feature ensemble plus sample selection: A comprehensive approach to domain adaptation for sentiment classification," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 10–18, 2013.

[110] R. Young, "Story and discourse: A bipartite model of narrative generation in virtual worlds," *Interaction Studies*, vol. 8, pp. 177–208, 2007.

[111] H. Zellig, "Distributional structure," *Word*, vol. 10, pp. 146–162, 1954.