Introduction: Cognitive Issues in Natural Language Processing

Thierry Poibeau, Aline Villavicencio

Introduction: Cognitive Issues in Natural Language Processing

Thierry Poibeau[1] and Aline Villavicencio[2]

[1] LATTICE-CNRS, Ecole Normale Supérieure, France
[2] Institute of Informatics, Federal University of Rio Grande do Sul, Brazil
thierry.poibeau@ens.fr,avillavicencio@inf.ufrgs.br

Abstract

This introduction aims at giving an overview of the questions and problems addressed jointly in Natural Language Processing and Cognitive Science. More precisely, the idea of this introduction, and more generally of this book, is to address how these fields can fertilize each other, bringing recent advances to produce richer studies.

Natural Language Processing is fundamentally dealing with semantics and more generally with knowledge. Cognitive Science is also mostly dealing with knowledge: how knowledge is acquired and processed in the brain. The two domains have developed largely independently, as we will see in the introduction, but there are obvious links between the two, and a large number of researchers have investigated problems involving the two fields, either in the data or the methods used.

1 On the Relationships between Natural Language Processing and Cognitive Sciences

1.1 A quick historical overview

The landscape of Natural Language Processing (NLP) has dramatically changed in the last decades. Until recently, it was generally assumed that one first needs to adequately formalize an information context (for example information contained in a text) in order to be able to subsequently develop applications dealing with semantics (see e.g. [81, 4, 67]). This initial step involved manipulating large knowledge bases of manually hand-crafted rules, has resulted in the new field of "knowledge engineering" [18].

Knowledge can be seen as the result of the confrontation of our a priori ideas with the reality of the outside world. This leads to several difficulties: i) the task is potentially infinite since people constantly perceive a multiplicity of things; ii) perception interferes with information already registered in the brain,

leading to complex inferences with common sense knowledge; iii) additionally, very little is known about how information is processed in the brain, which make things even harder to formalize.

To answer some of these issues, a common assumption is that knowledge could be disconnected from perception, which led to projects aiming at developing large static databases of "common sense knowledge", from CYC [52] to more recent general domain ontologies like ConceptNet [56]. However, these projects have always led to databases that, despite their sizes, were never enough to completely and accurately formalize a given domain, and domain-independent applications were thus even more unattainable. Moreover, very quickly different problems appeared since contradicting facts, variable point of views and subjective information cannot be directly formalized in a static database aiming to provide a general and multi-purpose source of "ground-truth" information.

Despite these issues, a formalization of the textual content has often been the basis of most treatments for more than 50 years, since the beginning of NLP as a field in the late 1940s, with the creation of the first computers, to the late 1990s [46]. Things have gradually changed in the last 20 to 25 years, for two main reasons: i) the power of modern computers, capable of providing extensive calculation capacities and storing amazingly large collections of data and ii) the availability of data through the Web, which provides an unseen and constantly expanding collection of text, image and videos that goes far beyond anything people imagined before. Current corpora contain several billion words, leading to new discoveries "just" by finding patterns revealed by automatic means. As for the text processing domain, machine learning approaches [59] are capable of discovering rare word configurations and rare correlations, leading to constant progress and better performances, even for rare events.

Machine learning approaches are now prominent and achieve the best results on most tasks, including when semantics is at stake. Empirical approaches are generally geared towards practical tasks (e.g. parsing or machine translation) and most of the time do not implement any specific theory, let alone any cognitive considerations.

As a consequence, cognitive science and NLP have both evolved quite independently in last decades. On the one hand, the latter has made impressive progress in most tasks. Performance of complex systems can now be considered as satisfactory for some tasks and some specific needs, even if the results are still far from perfect. One example is the IBM Watson system that won Jeopardy[1] a few years ago [33]. On the other hand, cognitive science has also made much progress, leading to new insights in our knowledge of language processing in the brain.

So, why should we try to establish any link between these two domains if they have largely evolved apart from each other? Is it even relevant to try to establish links? We think we should answer positively to these questions since new connexions can be established between the two fields. NLP now widely uses

[1] Jeopardy is an American TV game similar to a question answering task, except that candidates have to find questions corresponding to answers, rather than answers corresponding to questions

new methods based on machine learning techniques. Even if these methods are not linked to any specific cognitive theory, they provide the means for extracting relevant information from very large masses of data. This could be compared to the activity of the human brain extracting information all the time from the different channels of perception

Ongoing research in both domains can also be enlightening for the other domain. For example distributional models are relevant for semantic theory, as well as for cognitive theories. These results can be used in studies related to lexical structure and lexical storage in the brain, and more applied fields such as language acquisition and and language pathology studies.

The opposite is also true: cognitive science have largely adopted computational models as a way to formalize, test and validate exiting theories, especially in the field of language comprehension. One of the main goals of language comprehension is to elaborate predictive models of "language complexity", often through what is called "surprisal effect": a linguistic sequence is more or less complex depending on its structure and on the frequency of lexical items used to form a sentence. The next word of a sequence can be predicted more or less accurately (with more or less "surprise") and traditional hypotheses can now be tested and validated with computational models. These models complement traditional methods, including real tests and neuro-imaging experiments (especially based on EEG) by providing a sound and formal basis for these previous proposals.

1.2 Artificial and natural systems

There is a long tradition of reflecting on the relation between natural and artificial systems [41]. In Artificial Intelligence, often the goal is not to directly reproduce how information is processed in the brain, since there is a clear lack of knowledge on how the brain works. Rather, scientists and philosophers are more interested in the results obtained by artificial systems. For example, we can ask ourselves: to what extent can a machine produce valuable results for tasks such as text production or, more specifically, translation or dialogue? Is it possible to dialogue with a machine without noticing the interlocutor is a machine and not a human? Is it possible to get a translation produced by a machine and not realizing the translation has not been made by a human being? In other words, to what extent is it possible to reproduce on a machine tasks involving some kind of "intelligence"?

These are exactly the kinds of questions Turing was dealing with in the 1940s and which led him to propose the famous Turing Test. The test states that if a machine is able to dialogue with a human without him noticing he is speaking with a machine, then this is the sign that the computer has some form of intelligence [85, 63].

There have been numerous discussions on the validity of the Turing test, with the key point being whether a dialogue is enough to prove the existence of intelligence. Although even very simple systems are capable of generating seemingly interesting interactions, the levels of real "understanding" shown by

these machines are extremely limited. Let's take the very famous example of Eliza, the dialoging system developed by Weizenbaum in 1966 [94]. This system was able to simulate a dialogue between a psychotherapist and a patient. Eliza was in fact just a series of regular expressions to derive questions from the patient's utterances. It was able to produce, for example, the question "why are you afraid of X?" from the sentence "I am afraid of X". The system also included a series of ready-made sentences that were used when no predefined patterns were applicable (for example "could you specify what you have in mind?", "tell me more" or "really?"). Despite its simplicity, Eliza enjoyed great success, and some patients really thought they were conversing with a real doctor through a computer.

One important point is that artificial systems do not need to directly reproduce human strategy to perform different tasks involving knowledge and reasoning. They do not even have to perform the task, only give the illusion that it is performed. Eliza has proved that very simple systems are enough to deceive human beings to some extent.[2] On the other hand, although humans may listen and answer superficially to their interlocutor in a conversation, in general they do perform these tasks using their brains and keeping track of the information expressed throughout the dialogue, and not just producing ready-made sentences without taking most of the previous exchanges into consideration.

Additionally, given the lack of knowledge about processes in the brain, the idea of drawing parallels between artificial and natural systems has never been seriously pursued. On the contrary, it has always been largely criticized. At best, artificial systems implement a specific theory but most often they just use a pragmatic approach, where what matters is the final result, not the way it is obtained (like with Eliza).

The 1990s popularized the use of machine learning methods for language processing [62, 59]. In this approach most of the knowledge comes from corpora, with the assumption that they are now large enough to provide sufficient evidence for further processing. In this case, the goal consists of identifying regularities in data that are likely to also be found in new data. The initial corpus (the training data) must be representative of the domain and should provide enough evidence so as to allow a good coverage of the data to be analyzed in the future. All NLP tasks have been affected by this new approach which currently dominates the field. For example, there are enough bilingual corpora available to directly train machine translation systems that use this data to translate new texts. There are also large treebanks with syntactically annotated texts that can be used to automatically construct statistical parsers trained on the data.

In this context, it is interesting to look at the role of semantics in the most recent generation of NLP systems. Generally, semantics and statistics are seen as opposites: on the one hand, content representation, on the other hand, computation. However, this opposition is far too simplistic. For example, statistics allows one to accurately characterize degrees of similarity between words or

[2] This was at a time when machines were rare and people more technologically naive than today, less used to interacting with computers. However, with some adaptation the same kind of experiment could probably still be reproduced nowadays.

between documents [70, 86], capturing some semantics. Statistics also offer powerful ways of representing word senses through a precise analysis of the usage of words in context, thanks to distributional models or more recently to the highly popular word embeddings [35]. One issue with these models is that there is no clear definition of what semantics, or even a word sense or a definition are. When we look at traditional dictionaries, it is immediately apparent that the number of senses per word differs: some dictionaries are more detailed, some more general, depending on their goal and their audience. However, word senses are not always mutually exclusive, and often different senses could be used to explain a word in context. The notion of graded word sense has been proposed to characterize this state of affairs: several word senses could apply to the same occurrence of a word, with different degrees of accuracy. In fact, Erk and McCarthy [30] proved that different definitions of ambiguous words can be more or less relevant at the same time. For example, "paper" could refer to"report", "publication", or "medium for writing", which are all supposed to be different word senses according to a reference dictionary. Therefore, word senses are not disjunct but largely overlap and often more than one word sense would be simultaneously activated for a given context. Statistics, by characterizing word meanings from usage and context gives complex representations that are very different from traditional ones, but can nevertheless be compared with cognitive models and may give more relevant results than previously.

Statistics are also relevant to characterize idioms, frozen expressions and even equivalencies between languages. In the last case, bilingual corpora can be aligned at the sentence or even at the word level [84]. It is possible to observe more or less regular equivalencies between words in context, and compute similarities between sequences of variable length (m-n equivalencies, like between "potatoe" and "pomme de terre" or "kick the bucket" and "passer l'arme-gauche"', to take famous examples in French and English). Because computers can automatically process very large bilingual corpora (several millions or even billion words), it is now possible to get richer and infinitely more precise bilingual dictionaries than before. The structure of these dictionaries should be studied from a cognitive point of view in mind: they are interesting since they are obtained from raw data given as input, without a priori knowledge and without any pre-defined theory. It is in this sense that we can say that artificial models based on machine learning methods encode some form of meaning that may make sense form a cognitive point of view.

The last decade has also seen an impressive amount of research aiming at linking text and image [25], the more general idea being that the acquisition of word meaning is a multimodal process in which vision plays a major role. The issue is then to provide rich enough representations of multimodal input, since it is already difficult to provide a relevant and cognitively plausible representation of an image or a video. Large online collection of images along with their metadata have been used as way to develop models of acquisition of lexical meaning. One particularly interesting point in these experiments is that meaning is not fixed and is gradually associated to specific objects through the observation of regularities in the environments. However, despite their interest,

metadata contributes for these results providing very specific information and the relation between this kind of experiment and human perception should be questioned.

2 Recent Issues in Cognitive Aspects of Language Modeling

2.1 NLP and Language comprehension

Language comprehension has always been an important research issue both in linguistics and in cognitive science. For example, the 1970s and 1980s have seen a large body of research on grammar formalisms [34]. One issue was to show the plausibility of these models in terms of language comprehension [8]. These formalisms also provided information on language complexity, for example by highlighting equivalencies between linguistic phenomena and the formal devices needed to represent them [45].

Recent research have shown the convergence of models coming from both sides, linguistics and cognitive science. For example, the very popular ACT-R theory (Adaptive Control of Thought–Rational) has been developed since the early 1970s by John Anderson [5, 6]. The theory aims to explain how the brain works in a modular way and how these modules interact to make comprehension possible. Applied to language, it means that different parts of a sentence are analyzed by different modules and calculating the meaning of the sentence corresponds to assembling the various pieces of information stored by these different modules [36, 7]. Partly independently, in NLP, since the 1990s chunks also play a major role. As defined by Abney [1], they generally consist of "a single content word surrounded by a constellation of function words, matching a fixed template" (e.g. a noun chunk). These simple units can be described by a context free grammar (CFG), whereas the structure of sentences (or, in other words, the relations between chunks) correspond to more complex schemas that cannot be described by simple CFGs. Blache [15] proposes to apply this theory to parsing. The author proposes an adapted version of the activation function which takes advantage of the representation of linguistic information in terms of low levels features including frequency information. This model subsequently simplifies the analysis of sentences, in accordance with cognitive models.

A related line of research also based on the ACT-R theory investigate the notion of comprehension by focusing on comprehension difficulty [53]. There is a consensus in the comprehension community to explain comprehension with the support of two main notions: memory and expectation. Memory refers to the ability to store information and bring it to the front whenever necessary. Expectation refers to the pre-determination of a lexical category depending on the context. Levy [53] discusses memory considerations using the following examples:

1. *This is the malt that the rat that the cat that the dog worried killed ate*

2. *This is the malt that was eaten by the rat that was killed by the cat that was worried by the dog*

Whereas the first sentence is hard to understand, the second is a lot easier even if formed by the same grammar rule applied iteratively. Different hypotheses have been provided to explain this phenomenon, like the large number of incomplete and nested syntactic relationships that must be stored in the brain. Levy shows that this phenomenon is more complex than it seems, and complexity also depends on the kind of structure used, on the arguments of the sentence and on the distance between elements in memory.

Levy [53] discusses expectation in terms of a traditional task in cognitive science that consists in completing the end of unfinished sentences and measure to what extent continuing the sentence can be predicted by humans. Let us take the two following sentences:

3. *The boat passed easily under the —*

4. *Rita slowly walked down the shaky —*

The first sentence provides a strongly predictive context (leading the speaker to generally propose "bridge" to complete the sentence) whereas the second sentence is more open, which is shown by a greater variety of answers and a longer time on average to complete the sentence.

Recently, several researchers have proposed models combining memory and expectation in order to measure comprehension difficulty [17]. One of the main ideas is to measure complexity through large scale PCFG (probabilistic context-free grammar) that can capture both dimensions of the problem [38]. These researchers have thus established a direct link between computational models and cognitive issues in language processing, providing a sound basis for empirical research. However, it should be pointed out that the results obtained are not always consistent with evidence from other kinds of studies, especially EEG (electroencephalogram) or eye-tracking studies [28]. The link between these different models deserves more attention.

2.2 Language acquisition

One of the big puzzles in science is how children learn their native languages reliably and in a short period of time. Languages are complex systems, with very large vocabularies with morphologically and derivationally inflected forms. Moreover, words from this vocabulary can be combined with a diverse inventory of syntactic constructions specific to the target language to convey some meaning in a particular context. In short, children have to learn to segment sounds associating forms and meanings to individual lexical items, a processing system to generate and comprehend sentences, along with pragmatic and social skills to use language in an acceptable manner in different contexts [43]. Yet children are typically exposed to sentences that are "*propositionally simple, limited in vocabulary, slowly and carefully enunciated, repetitive, deictic, and usually*

referring to the here and now" [92]. So how can children based on this data arrive at a mature state that is so sophisticated? What are the precise mechanisms involved in acquiring a language? Are they specific to language or are they general-purpose learning mechanisms? How much do learners know about languages prior to exposure to a specific language? How much exposure to language is needed for successful learning? There are many questions related to language acquisition, and computational modelling has been used as a methodology for addressing some of them [31, 95, 3, 82, 48]. Computational models usually include five components whose degree of complexity varies according to the particular focus of the research [10]:

(i) The first is a definition of what is being learned, in this case a language, and any specific subtasks, such as word segmentation [19, 54, 29], morphology [79, 51] or syntax [11, 20, 88, 96, 89, 48, 97].

(ii) The second component defines the available hypotheses that the learning model can formulate, that is, the hypothesis space that needs to be considered for learning [23, 13, 96]. The trajectory of the learner in this space is driven by the input data towards the target language.

(iii) Additionally it is necessary to define the learning environment that the model is exposed to. This may include the order and the frequency with which the data occurs in the environment, along with any (correct or incorrect) clues about whether the data belonging to the target language or not [13, 50, 73].

(iv) The fourth component is a definition of the updating procedure for the learner's hypotheses along with any restrictions involved. This procedure determines how conservative the learner is in changing the current hypothesis [68, 21, 13].

(v) The last is a definition of success in the task. The model needs to be evaluated according to a definition of successful learning that indicates when the target language has been successfully acquired. Success criteria include those defined by learning paradigms like Identification in the Limit [37], Probably Approximately Correct learning [87] and the Minimum Description Length principle [78].

One of the challenges with research in this area is that in general we have only very limited and often indirect access to the neural regions involved in language production and understanding, especially during the language acquisition period, and this is usually restricted to the output product. Corpora containing naturalistic language acquisition data from transcripts of child-directed and child-produced speech have been used as the basis for research in the area. They include data from longitudinal studies, following the same child for several years and allowing the investigation of different developmental stages. There are also latitudinal studies that include various children of particular age groups, and

these may help to avoid any individual bias from personal language traits. Initiatives like CHILDES [57] have provided repositories for language acquisition data for over 25 languages, some with additional information from part-of-speech taggers and parsers [80, 91], others providing audio and video recordings with the transcripts.[3]

The availability of language acquisition data brings an enormous potential for the "in-vitro" testing of different theories of acquisition via simulations in computational models [88, 75, 48, 82]. For instance, there has been much interest in Hierarchical Bayesian models and their application to child language acquisition [76, 42, 71, 72]. Much of their appeal comes from being able to handle noise and ambiguity in the input data, while also accounting for known pre-existing language biases via prior probabilities. They have been applied to the acquisition of word segmentation [74, 77], verb alternations [76, 72, 90], argument structure [2], multiword expressions [66], among other tasks.

2.3 Clinical Conditions

As many clinical conditions have an impact on language abilities, computational methods have also been used as tools to investigate possible language changes associated with these pathologies. For example, Alzheimer's Disease, which affects millions of people around the world, has from its early stages a noticeable impact on lexical search and retrieval processes. The use of computational methods that allow the creation of synthetic simulations compatible with this condition may contribute to an early diagnosis by helping to distinguish changes that are triggered by the disease from those that arise as a natural consequence of aging.

One promising line of investigation uses concepts from graph theory [93] to model the lexicon as a complex network where words or concepts correspond to nodes and are connected to one another by specific relations, such as proximity in a sentence or synonymy [83, 27]. Measures like the clustering coefficient of the network, the number of connected components and the average length of the shortest path between pairs of nodes have been used to determine characteristics of networks in healthy and clinical cases [22, 9], in studies related to semantic storage and the mechanisms that operate on it [27, 64] and in investigations that use simulations of changes that lead from healthy to clinical networks [16]. For example, starting from networks of semantic priming in healthy subjects Borge-Holthoefer et al. [16] simulated the evolution to a clinical condition through changes in network connectivity that led to a progressive degradation of the network structure that has qualitative agreement with real observations of clinical patients with Alzheimers Disease.

Resources such as MRC Psycholinguistic Dataset [26], WordNet [32], or the University of South Florida Free Association Norms [65] provide additional information for analyzing language data. They include characteristics like the familiarity, concreteness and age of acquisition of the words as well as the se-

[3] `http://childes.psy.cmu.edu`

mantic similarity or association strength among them. However, since these resources have limited coverage and are not available for all languages, some alternatives are to also use data-driven methods to automatically extract relevant information from corpora, and to employ crowdsourcing for additional judgements [60, 69, 40, 47]. For instance, distributional semantic models [55, 49] capture semantic relatedness among words, and they have been found to successfully explain human performance in semantic priming tasks. Moreover, the more recent models based on neural networks [61] provided a better fit to the behavioral data [58].

Extensive, accurate, and fast detection of patients in early stages of pathological conditions has enormous potential for maximizing the effectiveness of treatments while minimizing their costs, and computational methods can contribute toward this goal.

2.4 The origins of language and language evolution

Investigations on the origins of language have long been the focus of much interest for centuries. They have examined questions that range from determining the biological mechanisms involved with language production and understanding, to how languages evolved into their contemporary variants [14, 39, 12]. These problems have been considered some of the hardest in science [24], given the uniqueness and complexity of the human language and the lack of empirical evidence.[4] According to Hauser et al. [39], the way forward for empirical work involves combined efforts from:

- comparative animal behavior, looking at natural communication and artificial languages,
- paleontology and archaeology, examining structural characteristics of skulls and bones that can be linked to brain functions,
- molecular biology, mapping genes to complex behavior, and
- mathematical modeling of computations and representations.

In particular, the latter allow the definition of complex simulations involving populations of linguistic agents which interact with one another to try to approximate possible scenarios for the emergence of language. Language is seen as a complex adaptive system that may be affected by variables like the learning algorithms adopted and the communicative efficiency of competing alternatives, in addition to factors like language contact between different populations and population size. These need to be realistic models of how phonological, syntactic and semantic representations arose and were selected for in populations, with the possibility of testing their assumptions with regards to plausibility [39].

[4] Discussions on the topic were even banned for a while by the Linguistic Society of Paris in the 19th century [44].

3 Content and Structure of the book

The chapters in this collection present different aspects of research on computational models of language and cognition. They display a cross-section of recent research in the area, covering the spectrum from theoretical considerations and formalizations to more applied models and the construction of applications and resources.

The chapters in the first part of the book describe works that analyze psycholinguistic data using neural and cognitive language processing. Recordings of brain activity data are one of the most direct reflections of the states and processes involved in language processing, and when analyzed in the light of cognitive and linguistic theories they can provide insights about functions and architecture of the language faculty. The first chapter "Decoding Language from the Brain" by Brian Murphy, Leila Wehbe and Alona Fyshe provides an overview of recent works that use computational modeling of language and of brain activity. They start with a discussion of how patterns of electrophysiological activity have been associated with sentence processing difficulties and language complexity. For words, they describe how computational models can distinguish relevant aspects of brain activity for word meaning from noise using distributional semantic theory. Looking at syntax, they examine how lexical units are combined to form short phrases and how existing theories of language characterize the representations produced by compositional processes. They finish with a discussion of experiments that use more natural language understanding tasks for a more holistic and realistic language processing.

The second chapter entitled "Light-and-Deep Parsing: a Cognitive Model of Sentence Processing" by Philippe Blache provides an overview of language processing from various perspectives, including neurolinguistics with findings from electrophysiological studies. On this basis, the author argues for an alternative to the classical architectures involving modular and serial processing, that takes into account language as a whole. He proposes a new representation for linguistic information, based on properties. For basic properties that are assessable in a simple and direct manner, the default processing mechanism based on light parsing is applied. This mechanism stores words in working memory, assembles them into chunks, infers properties and activates constructions, resulting in fast even if shallow processing and direct access to interpretation. For more complex cases, a deeper processing needs to be adopted with classical strictly incremental and serial interpretation that is compositionally constructed from a syntactic structure.

The computational modeling of clinical groups in psycholinguistic tasks can also provide insights about the language faculty, by characterizing how particular conditions affect language use. In the final chapter of this section "Graph Theory applied to speech: Insights on cognitive deficit diagnosis and dream research" by Natália Bezerra Mota, Mauro Copelli and Sidarta Ribeiro, graph theory is used for a structural analysis of the language used by clinical groups, represented as networks, beyond what a lexical analysis would reveal, to help in the psychiatric diagnosis of psychoses and dementias. The first study examines

how networks representing the flow of thoughts of Bipolar and Schizophrenic patients are able to distinguish clinical from control groups based on their verbal reports of dreams or waking events. A second study looks at the use of networks representing a verbal fluency task. They model a group of clinical participants diagnosed with Alzheimer's dementia, another with Moderate Cognitive Impairment, and a third of healthy elderly participants. Based on topological analysis of the networks it was possible to distinguish these three groups.

The chapters in the second part of the book are related to the use of data-driven methods for acquiring information from large amounts of language, in tasks ranging from translation, inference about semantic roles, native language identification, and speech segmentation. The first chapter, "Putting Linguistics back into Computational Linguistics" by Martin Kay discusses the place of knowledge about languages and speakers in computational linguistics and natural language processing. For Kay communication is a collaborative task that involves the hearer guessing the speaker's intentions. He argues that it is not enough to examine large quantities of texts to discover all we need to know about languages. The referential function of language should also be taken into account, both for a better understanding of the human language ability and for language technology. Looking at the case of translation, he analyzes the advantages of doing that comparing the syntactic and pragmatic traditions of translation. The former uses information about lexical correspondences in source and target language and possibly the reordering of words, which can be learned from huge quantities of data using statistical approaches. The latter starts from what the original author wants to communicate and finds a way of expressing it in the target language sometimes independently of the words and phrases in the source text, and possibly making implicit information in the source explicit in the target, if important to convey the message.

In the second chapter, entitled"A Distributional Model of Verb-Specific Semantic Roles Inferences", Gianluca Lebani and Alessandro Lenci start with an overview of research on acquisition and representation of thematic roles, where roles describe the relation of each of the arguments of a verb in the event or situation it expresses. Adopting the view of thematic roles as clusters of properties entailed by verb arguments, they use evidence from behavioral data to define a more fine-grained characterization of the properties activated by a verb, focusing on a subset of English verbs, and examine to what extent these properties can be acquired from corpus based distributional data.

The following chapter in this section is "Native Language Identification Using Large, Longitudinal Data", by Xiao Jiang, Yan Huang, Yufan Guo, Jeroen Geertzen, Dora Alexopoulou, Lin Sun and Anna Korhonen. As mentioned in its title, this chapter deals with the automatic identification of the native language of second language learners. This has theoretical consequences, especially to determine to what extent L1 backgrounds influences L2 learning and "whether there is a significant difference between the writings of L2 learners across different L1 backgrounds". This research domain has also immediate and practical applications for example in language tutoring systems and authorship profiling. The chapter offers new insights based on a new corpus, called the EF-Cambridge

Open English Learner Database (EFCAMDAT), which is multiple times larger than previous L2 corpora and provides longitudinal data across 16 proficiency levels. The system for native language identification presented in the paper employs accurate machine learning (SVM) with a wide range of linguistic features. The authors report high overall accuracy of around 80% at low and medium proficiency levels, and 70% at advanced levels, and the detailed analysis show that the top performing features differ from one proficiency level to another, which a fine grained analysis is necessary to take into account the difference of the various learner proficiency.

The last chapter in this section is entitled "Evaluating language acquisition models: A utility-based look at Bayesian segmentation", by Lisa Pearl and Lawrence Phillips. The authors address the problem of evaluation in an unsupervised domain, especially when we have an imperfect knowledge of this domain. The problem is even more difficult when it comes to child language acquisition due to "uncertainty about the exact nature of the target linguistic knowledge and a lack of empirical evidence about children's knowledge at specific stages in development". The idea of a gold standard for language acquisition is thus not realistic. The rest of the paper investigate this issue through the study of initial stages of speech segmentation, where a fluent stream of speech is divided by the learner into useful units, such as words. The authors show that segmentation based on Bayesian models, which has proven successful for English, also obtains good results for a variety of other languages. This is particularly true if a relevant segmentation ("useful and valid non-word units") is taken into account, which can be quite different that the traditional gold standard based on written word segmentation. The authors conclude by showing that "this serves as a general methodological contribution about the definition of segmentation success, especially when we consider that useful units may vary across the world's languages".

The third and last part of the book deals with social issues in language evolution. Most people admit that the primary goal of languages is to make it possible for humans to communicate and easily exchange even the more complex ideas. What is not so clear is why there are so many languages around the world, how and why these languages constantly evolve, change and even disappear. This section provide theoretical as well as practical accounts and also consider how computational models can shed new light on this complex issue.

The first chapter in this section, by Anne Reboul, is "Social Evolution of public languages: between Rousseau's Eden and Hobbes' Leviathan". She observes that nearly all models of language evolution rely on social scenarios, where language is the main tool for specific purposes like hunting, sexual selection or tool making. Moreover, apart from when language is seen as primarily a mental tool, all hypotheses involve some social dimension. The question addressed in this chapter is whether "the social pressure leading to the emergence of language is due to prosocial attitudes" (the cooperative/altruistic hypothesis) "or to an arms race motivated by inside group competition and conflict". The scenarios range between altruistic scenarios (Rousseauist scenario) or more conflictual ones, where competition comes before cooperation (the Hobbesian

scenario). In the rest of her chapter, Reboul criticizes the idea that language is a communication system in the strong sense: language did not emerge primarily for communication. Instead, the author shows convincingly that negotiation and persuasion are more important. In this context, language is not only a tool for communication but also a perfect tool for implicit communication and argumentation. The chapter is based on recent theories of communication and argumentation and shed new light on a hot topic in the domain.

The following chapter, "Genetic biases affecting language: What do computer models and experimental approaches suggest?" is by Rick Janssen and Dan Dediu. The authors observe that language evolution as an area has been highly inspired by biological models of evolution. Moreover computational models have shown in practice how specific features can be amplified from generation to generation, leading to preferential selection of characteristic language feature like recursion, compositionally and other universal features. In these models the evolution of languages is based on specific biological characteristics of the human species, encoded in the human genome, but that "agents might evolve to a state of predisposed adaptability, while particularly stable language features might get assimilated into the genome via Baldwinian niche construction". Although this issue is largely controversial, it may be considered as a valid alternative to the adaptation of language specific features, "for example explaining speech perception as a possible co-option of more general learning and pattern recognition mechanisms". The authors claim that the evolution of language cannot be explained solely from a biological perspective and that social interaction must also been taken into account. Computational and agent-based models give a sound basis for this thesis that deserves to be exposed and discussed among researchers.

The last chapter in this section, by Remi Van Trijp is entitled "Linguistic Assessment Criteria for Explaining Language Change: A Case Study on Syncretism in German Definite Articles". The author addresses ambiguity in natural languages and argues that it may lead to greater efficiency in language processing. The claim is supported by a case study on the German declension system. The author proposes a formalization that shows case syncretism is "efficiently processed as long as the case forms are still in functional opposition of each other". Syncretism and ambiguity should thus be studied within the whole linguistic system or "linguistic eco-system" according to the author of the chapter.

Acknowledgments

Aline Villavicencio was partly sponsored by projects AIM-WEST (FAPERGS-INRIA 1706- 2551/13-7), CNPq 482520/2012-4, 312114/2015-0, "Simplificação Textual de Expressões Complexas", sponsored by Samsung Eletrônica da Amazônia Ltda. under the terms of Brazilian federal law No. 8.248/91.

References

[1] Steven Abney. Parsing by chunks. In Robert Berwick, Steven Abney, and Carol Tenny, editors, *Principle-Based Parsing*. Kluwer Academic Press, Dordrecht, 1991.

[2] A. Alishahi and S. Stevenson. A computational model of early argument structure acquisition. *Cognitive Science*, 32(5):789–834, 2008.

[3] Afra Alishahi. *Computational Modeling of Human Language Acquisition*. Synthesis lectures on human language technologies. Morgan & Claypool, 2011.

[4] James Allen. *Natural Language Understanding*. Pearson, New York, 1994.

[5] John R. Anderson. *Language, memory, and thought*. Lawrence Erlbaum Associates, Mahwah, NJ, 1976.

[6] John R. Anderson, Dan Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060, 2004.

[7] John R Anderson, Daniel Bothell, Michael D Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological review*, 111(4):1036–1060, 2004.

[8] Benjamin Bergen and Nancy Chang. Embodied construction grammar in simulation-based language understanding. In J.-O. Ostman and M. Fried, editors, *Construction Grammar(s): Cognitive and Cross-Language Dimensions*. Johns Benjamins, Amsterdam, 2003.

[9] Laiss Bertola, Natalia Bezerra Mota, Mauro Copelli, Thiago Rivero, Breno Satler, De Oliveira Diniz, Marco Aurelio Romano, Sidarta Ribeiro, and Leandro Fernandes Malloy-diniz. Graph analysis of verbal fluency test discriminate between patients with Alzheimer's disease, mild cognitive impairment and normal elderly controls. *Frontiers in Aging Neuroscience*, 6(185), 2014.

[10] Stefano Bertolo. A brief overview of learnability. In Stefano Bertolo, editor, *Language Acquisition and Learnability*, pages 1–14. Cambridge University Press, 2001.

[11] Robert C. Berwick. *The Acquisition of Syntactic Knowledge*. MIT Press, 1985.

[12] Robert C. Berwick and Noam Chomsky. *Why Only Us: Language and Evolution*. The MIT Press, 2015.

[13] Robert C. Berwick and Partha Niyogi. Learning from Triggers. *Linguistic Inquiry*, 27(4):605–622, 1996.

[14] Derek Bickerton. *Roots of language*. Language Science Press, 2016.

[15] Philippe Blache. Chunks et activation: un modle de facilitation du traitement linguistique. In *Proceedings of the conference Traitement Automatique du Langage Naturel*. ATALA, Les Sables d'Olonnes, 2013.

[16] Javier Borge-Holthoefer, Yamir Moreno, and Alex Arenas. Modeling abnormal priming in Alzheimer's patients with a free association network. *PLoS ONE*, 6(8), 2011.

[17] Marisa F. Boston, John T. Hale, Shravan Vasishth, and Reinhold Kliegl. Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26(3):301–349, 2011.

[18] Ronald Brachman and Hector Levesque. *Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[19] Michael R. Brent. An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 34(1):71–105, 1999.

[20] Ted Briscoe. Co-evolution of language and of the language acquisition device. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 418–427. Association for Computational Linguistics, 1997.

[21] Ted Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, pages 245–296, 2000.

[22] Álvaro Cabana, Juan C Valle-lisboa, Brita Elvevåg, and Eduardo Mizraji. Detecting orderdisorder transitions in discourse: Implications for schizophrenia. *Schizophrenia Research*, 2011.

[23] Noam Chomsky. *Aspects of the Theory of Syntax*. The MIT Press, Cambridge, 1965.

[24] Morten H. Christiansen and Simon Kirby. Language evolution: The hardest problem in science? In M.H. Christiansen and S. Kirby, editors, *Language Evolution: The States of the Art*. Oxford University Press, 2003.

[25] Grzegorz Chrupaa, Ákos Kádár, and Afra Alishahi. Learning language through pictures. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 112–118. ACL, Beijing, 2015.

[26] Max Coltheart. The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33A:497–505, 1981.

[27] Simon De Deyne and Gert Storms. Word associations: Network and semantic properties. *Behavior research methods*, 40(1):213–231, 2008.

[28] Vera Demberg and Frank Keller. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210, 2008.

[29] Micha Elsner, Sharon Goldwater, Naomi Feldman, and Frank Wood. A joint learning model of word segmentation, lexical acquisition, and phonetic variability. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2013.

[30] Katrin Erk and Diana McCarthy. Graded word sense assignment. In *Proceedings of the Empirical Methods in Natural Language Processing Conference*. ACL, 2009.

[31] Afsaneh Fazly, Afra Alishahi, and Suzanne Stevenson. A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34(6):1017–1063, 2010.

[32] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA, 1998.

[33] David A. Ferrucci. Introduction to "this is watson". *IBM J. Res. Dev.*, 56(3):235–249, 2012.

[34] Nissim Francez and Shuly Wintner. *Unification Grammars*. Cambridge University Press, Cambridge, 2012.

[35] Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu. Learning semantic hierarchies via word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014*, pages 1199–1209, 2014.

[36] James P. Gee and Franois Grosjean. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15(4):411–458, 2004.

[37] E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

[38] John T. Hale. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, PA, 2001.

[39] Marc D Hauser, Charles Yang, Robert C Berwick, Ian Tattersall, Michael J Ryan, Jeffrey Watumull, Noam Chomsky, and Richard C Lewontin. The mystery of language evolution. *Frontiers in Psychology*, 5(May):1–12, 2014.

[40] Felix Hill, Roi Reichart, and Anna Korhonen. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 41(4):665–695, 2015.

[41] John H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge, MA, USA, 1992.

[42] Anne S. Hsu and Nick Chater. The logical problem of language acquisition: A probabilistic perspective. *Cognitive Science*, 34(6):972–1016, 2010.

[43] Nina Hyams. The theory of parameters and syntactic development. In Thomas Roeper and Edwin Williams, editors, *Parameter Setting*, pages 1–22. Springer Netherlands, Dordrecht, 1987.

[44] Sverker Johansson. *Origins of Language: Constraints on hypotheses*. Converging Evidence in Language and Communication Research. John Benjamins Publishing Company, 2005.

[45] Aravind K. Joshi. Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes*, 5(1):1–27, 1990.

[46] Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 2000.

[47] Maximilian Köper and Sabine Schulte im Walde. Automatically generated affective norms of abstractness, arousal, imageability and valence for 350 000 german lemmas. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 2595–2598, Portoroz, Slovenia, 2016.

[48] Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer, and Mark Steedman. A probabilistic model of syntactic and semantic acquisition from child-directed utterances and their meanings. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 234–244, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[49] Quoc Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. In *31st International Conference on Machine Learning*, 2014.

[50] Julie Anne Legate and Charles Yang. The richness of the poverty of the stimulus. *On the occasion of Happy Golden Anniversary, Generative Syntax: 50 years since Logical Structure of Linguistic Theory*, 2005.

[51] Julie Anne Legate and Charles Yang. Morphosyntactic learning and the development of tense. *Language Acquisition*, 14(3):315–344, 2007.

[52] Douglas B. Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.

[53] Roger Levy. Memory and surprisal in human sentence comprehension. In Roger P. G. van Gompel, editor, *Sentence Processing*, page 78114. Hove: Psychology Press, 2013.

[54] Constantine Lignos. Modeling infant word segmentation. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, CoNLL '11, pages 29–38, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[55] Dekang Lin. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, pages 768–774, 1998.

[56] Hugo Liu and Push Singh. Conceptnet - a practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4):211–226, 2004.

[57] Brian MacWhinney. *The CHILDES Project: Tools for analyzing talk.* Lawrence Erlbaum Associates, Mahwah, NJ, 2000.

[58] Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, in press.

[59] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing.* MIT Press, Cambridge, MA, USA, 1999.

[60] Scott McDonald and Chris Brew. A distributional model of semantic context effects in lexical processing. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, 21-26 July, 2004, Barcelona, Spain.*, pages 17–24. ACL, 2004.

[61] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[62] Thomas M. Mitchell. *Machine Learning.* McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.

[63] James H. Moor. Turing test. In *Encyclopedia of Computer Science*, pages 1801–1802. John Wiley and Sons Ltd., Chichester, UK, 2003.

[64] Natalia B Mota, Nivaldo A P Vasconcelos, Nathalia Lemos, Ana C Pieretti, Osame Kinouchi, Guillermo A Cecchi, Mauro Copelli, and Sidarta Ribeiro. Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. 7(4):1–9, 2012.

[65] Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407, 2004.

[66] Aida Nematzadeh, Afsaneh Fazly, and Suzanne Stevenson. Child acquisition of multiword verbs: A computational investigation. In A. Villavicencio, T. Poibeau, A. Korhonen, and A. Alishahi, editors, *Cognitive Aspects of Computational Language Acquisition*, pages 235–256. Springer, 2013.

[67] Sergei Nirenburg and Victor Raskin. *Ontological Semantics*. MIT Press, Cambridge, MA, 2004.

[68] Partha Niyogi and Robert C. Berwick. A language learning model for finite parameter spaces. *Cognition*, 61(1-2):161–193, 1996.

[69] Sebastian Padó and Mirella Lapata. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199, 2007.

[70] Patrick Pantel and Dekang Lin. Discovering word senses from text. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'02, pages 613–619, New York, NY, USA, 2002. ACM.

[71] Christopher Parisien, Afsaneh Fazly, and Suzanne Stevenson. An incremental bayesian model for learning syntactic categories. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, CoNLL '08, pages 89–96, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[72] Christopher Parisien and Suzanne Stevenson. Learning verb alternations in a usage-based bayesian model. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 2010.

[73] Lisa Pearl. The input for syntactic acquisition: Solutions from language change modeling. In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition*, pages 1–9, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.

[74] Lisa Pearl, Sharon Goldwater, and Mark Steyvers. How ideal are we? incorporating human limitations into bayesian models of word segmentation. In *Proceedings of the 34th annual Boston University Conference on Child Language Development*, page 315326, Somerville, MA, 2010. Cascadilla Press.

[75] Amy Perfors, Joshua B Tenenbaum, and Elizabeth Wonnacott. Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of child language*, 37(03):607–642, 2010.

[76] Amy Perfors, Joshua B. Tenenbaum, and Elizabeth Wonnacott. Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, (37):607–642, 2010.

[77] Lawrence Phillips and Lisa Pearl. Bayesian inference as a viable cross-linguistic word segmentation strategy: It's all about what's useful. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, CogSci 2014, Quebec City, Canada, July 23-26, 2014*, 2014.

[78] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 of *Series in Computer Science*. World Scientific, 1989.

[79] D. E. Rumelhart and J. L. McClelland. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 2. chapter On Learning the Past Tenses of English Verbs, pages 216–271. MIT Press, Cambridge, MA, USA, 1986.

[80] Kenji Sagae, Brian MacWhinney, and Alon Lavie. Adding syntactic annotations to transcripts of parent-child dialogs. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal*. European Language Resources Association, 2004.

[81] John Sowa, editor. *Principles of Semantic Networks: Explorations in the Representation of Knowledge*. Morgan Kaufmann Series in Representation and Reasoning, (San Francisco, 1991.

[82] Mark Steedman. Probabilistic models of grammar acquisition. In *Proceedings of the Workshop on Computational Models of Language Acquisition and Loss*, pages 19–19, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[83] Mark Steyvers and Joshua B Tenenbaum. The Large-Scale Structure of Semantic Networks : Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29:41–78, 2005.

[84] Jörg Tiedemann. *Bitext Alignment*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2011.

[85] Alan Turing. Computing machinery and intelligence. *Mind*, 236:433–460, 1950.

[86] Peter D. Turney and Patrick Pantel. From frequency to meaning: Vector space models of semantics. *J. Artif. Int. Res.*, 37(1):141–188, January 2010.

[87] Leslie Valiant. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books, Inc., New York, NY, USA, 2013.

[88] Aline Villavicencio. *The acquisition of a unification-based generalised categorial grammar*. PhD thesis, University of Cambridge, 2002.

[89] Aline Villavicencio. Language acquisition with a unification-based grammar. In R. Borsley e K. Borjars, editor, *Non-transformational Syntax*. Blackwell, 2011.

[90] Aline Villavicencio, Marco Idiart, Robert Berwick, and Igor Malioutov. Language acquisition and probabilistic models: keeping it simple. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1321–1330, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[91] Aline Villavicencio, Beracah Yankama, Rodrigo Wilkens, Marco Idiart, and Robert Berwick. An annotated english child language database. In *Proceedings of the Workshop on Computational Models of Language Acquisition and Loss*, pages 23–25, Avignon, France, April 2012. Association for Computational Linguistics.

[92] Eric Wanner and Lila R. Gleitman, editors. *Language Acquisition: The State of the Art*. Cambridge University Press, Cambridge, 1982.

[93] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world' networks. *Nature*, 393(June):440–442, 1998.

[94] Joseph Weizenbaum. Eliza - a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45, January 1966.

[95] Shuly Wintner. Computational models of language acquisition. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: 11th International Conference, CICLing 2010, Iaşi, Romania, March 21-27, 2010. Proceedings*, pages 86–99. Springer Berlin Heidelberg, 2010.

[96] Charles Yang. Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10):451–456, October 2004.

[97] Charles Yang. Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 110(16):6324–6327, April 2013.