Linguistics for the Age of AI

The title-level DOI for this work is:

doi:10.7551/mitpress/13618.001.0001

Setting the Stage

Remember HAL, the "sentient computer" from Stanley Kubrick's and Arthur C. Clarke's *2001: A Space Odyssey*? To refresh your memory, here is a sample dialog between HAL and Dave, the astronaut:

Dave: Open the pod bay doors, HAL.

HAL: I'm sorry, Dave, I'm afraid I can't do that.

Dave: What's the problem?

HAL: I think you know what the problem is just as well as I do.

Dave: What are you talking about, HAL?

HAL: This mission is too important for me to allow you to jeopardize it.

Dave: I don't know what you're talking about, HAL.

HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

HAL clearly exhibits many facets of human-level intelligence—context-sensitive language understanding, reasoning about Dave's plans and goals, developing its own plans based on its own goals, fluent language generation, and even a modicum of emotional and social intelligence (note the politeness).

The movie came out over fifty years ago, so one might expect HAL to be a reality by now, like smartphones. But nothing could be further from the truth. Alas, despite the lure of AI in the public imagination and recurring waves of enthusiasm in R&D circles, re-creating human-level intelligence in a machine has proved more difficult than expected. In response, the AI community has, by and large, opted to change course, focusing on simpler, "low-hanging fruit" tasks and silo applications, such as beating the best human players in games like chess, *Go*, and *Jeopardy!* Such systems are not humanlike: they do not know what they are doing and why, their approach to problem solving does not resemble a person's, and they do not rely on models of the world, language, or agency. Instead, they largely rely on applying generic machine learning algorithms to ever larger datasets, supported by the spectacular speed and storage capacity of modern computers. Such systems

are the best that the field can offer in the short term, as they outperform handcrafted AI systems on the kinds of tasks at which they excel, while requiring much less human labor to create. But those successes have blunted the impetus to work on achieving human-level AI and have led to the concomitant avoidance of applications that require it. The upshot is that the goal of developing artificial intelligent agents with language and reasoning capabilities like those of HAL remains underexplored (Stork, 2000).

In this book, we not only explain why this goal must remain on agenda but also provide a roadmap for endowing artificial intelligent agents with the language processing capabilities demonstrated by HAL. In the most general terms, our approach can be described as follows. An artificial intelligent agent with human-level language ability—what we call a *language-endowed intelligent agent* (LEIA)—must do far more than manipulate the words of language: it must be able to *understand*, *explain*, and *learn*.

- The LEIA must *understand* the context-sensitive, intended meaning of utterances, which can require such capabilities as reconstructing elided meanings, interpreting indirect speech acts, and connecting linguistic references not only to elements of memory but also, in some cases, to perceived objects in the physical world.
- It must be capable of *explaining* its thoughts, actions, and decisions to its human collaborators in terms that are meaningful to us. It is such transparency that will earn our trust as we come to rely on intelligent systems for assistance in tasks more sensitive and critical than finding a restaurant with Siri's help or getting the gist of a foreign-language blog using Google Translate.
- It must forever be in a mode of *lifelong learning*, just like its human counterparts. Lifelong learning subsumes learning new words and phrases, new elements of the world model (ontology), and new ways of carrying out tasks. It requires remembering and learning from past actions, interactions, and simulated thought processes. And it involves perceiving and interpreting the dynamically changing properties and preferences of the agent's human collaborators.

To achieve this level of prowess, the agent must be equipped with a descriptively adequate model of language and world, which fuels heuristic algorithms that support decision-making for perception and action.

This book concentrates on describing the language understanding capabilities of LEIAs from the perspectives of cognitive modeling and system building. A prominent design feature is the orientation around *actionability*—that is, seeking an interpretation of a language input that is sufficiently deep, precise, and confident to support reasoning about action. Targeting actionability rather than perfection is far more than an expedient engineering solution—it models what people do. Real language use is messy, and no small amount of what we say to each other is unimportant and/or not completely comprehensible. So, we understand what we can and then, if needed, enhance that interpretation by

drawing on background knowledge, expectations, and judgments about the (un)importance of whatever remains unclear. How much effort we devote to understanding a particular utterance is guided by the principle of least effort, which serves to manage cognitive load. Modeling this strategy in machines is the best hope for achieving robustness in systems that integrate language understanding into an agent's overall operation.

To implement the above strategy, this book introduces two orthogonal conceptions of *incrementality*. *Horizontal* incrementality refers to processing inputs phrase by phrase, building up an interpretation as a person would upon perceiving a language stream. *Vertical* incrementality involves applying increasingly sophisticated (and resource-intensive) analysis methods to input "chunks" with the goal of achieving an actionable interpretation at the lowest possible cost. The control flow of language understanding—that is, deciding how deeply to process each chunk before consuming the next one—is handled by a cognitive architecture functionality that bridges language understanding and goal-oriented reasoning.

The crux of natural language understanding is semantic and pragmatic analysis. We treat the large number of semantic and pragmatic problems (from lexical disambiguation to coreference resolution to indirect speech acts and beyond) using *microtheories*. Each microtheory sketches a complete problem space, classifies component problems, and details methods of solving the subset of problems that can be treated fully automatically. Full automaticity is important, as it is nonsensical to develop agent systems that rely on unfulfillable prerequisites. The heuristics brought to bear on language analysis tasks are drawn from many sources: static knowledge bases, such as semantic lexicons and the associated ontological (world) model; the system's situation model, which contains the agent's active goal and plan agenda; results of processing prior inputs within the same dialog, task, or application; and any other machine-tractable resource that we can render useful, whether it was developed in-house or imported. In other words, the approach to language understanding described in this book operationalizes the many facets of situational context that we, as humans, bring to bear when participating in everyday language interactions.

Just as building intelligent agents requires a combination of science and craft, so does writing a book of this genre. The biggest challenge is balancing the generic with the specific. Our solution was to write most of the book (chapters 2–7) in relatively generic terms, without undue focus on implementation details but, rather, with an emphasis on the decision-making involved in modeling. Still, we devote two chapters to specifics: namely, systems developed using the described microtheories (chapter 8) and evaluations of a LEIA's language understanding capabilities (chapter 9). Both of these chapters (a) validate that we are working on real AI in real systems, with all their expected and unexpected hurdles, and (b) emphasize the need for holistic—rather than isolated or strictly modularized—approaches to agent modeling. For example, our Maryland Virtual Patient prototype system was designed to train physicians via interactions with simulated LEIAs playing the role of virtual patients. The virtual-patient LEIAs integrated cognitive capabilities (e.g.,

language processing and reasoning) with a physiological simulation that not only produced medically valid outcomes but also provided the agent with interoception of its own symptoms, which contributed to its health-oriented decision-making. This holistic approach to agent modeling stands in contrast to the currently more popular research methodology of carving out small problems to be solved individually. However, solving small problems does not help solve big problems, such as building LEIAs. Instead, we must take on the big problems in their totality and prepare agents to do the best they can with what they've got—just like people do.

We think that this book will be informative, thought-provoking, and accessible to a wide variety of readers interested in the fields of linguistics, cognitive science, natural language processing (NLP), and AI. This includes professionals, students, and anyone motivated enough to dig into the science-oriented offerings of the popular press. The book suggests ways in which linguists can make essential contributions to AI beyond corpus annotation and building wordnets, it offers students a choice of many topics for research and dissertations, it reminds practitioners of knowledge-lean NLP just how far we still have to go, and it shows developers of artificial intelligent agent systems what it will take to make agents truly language-endowed. To promote readability, we have divided the chapters, when applicable, into the main body followed by deep dives that will likely be of interest primarily to specialists. We also include pointers to online appendixes.

As a point of reference, we have successfully incorporated the book into undergraduate and graduate courses at Rensselaer Polytechnic Institute with the sole prerequisite of an introductory course in linguistics. As a tool for stimulating students to think about and discuss hard issues, the book has proven quite valuable. The suggested chapter-end exercises offer hands-on practice as well as a break from reading. But although the book nicely serves pedagogical goals, it is not a textbook. Textbooks tend to look backward, presenting a neat picture of work already accomplished and striving for extensive and balanced coverage of the literature. By contrast, this book looks forward, laying out a particular program of work that, in the historical perspective, is still in its early days—with all the unknowns commensurate with that status.

As mentioned earlier, this genre of exploration is far removed from the current mainstream, so it would be natural for some readers to come to the table with expectations that, in fact, will not be fulfilled. For example, we will say nothing about neural networks, which is the approach to machine learning that is receiving the most buzz at the time of writing. As for machine learning more broadly understood, we believe that the most promising path toward integrating machine learning–based and knowledge-based methods is to integrate the results of machine learning into primarily knowledge-based systems, rather than the other way around—though, of course, this is a wide-open research issue.

The genre of this book is very different from that of compendia such as Quirk et al.'s *A Comprehensive Grammar of the English Language* (1985). We do not offer either a complete description of the phenomena or the final word on any of the topics we discuss. While

the microtheories we describe are sufficiently mature to be implemented in working systems and presented in print, they all address active research areas, and they will certainly continue to evolve in planned and unplanned directions.

There has been significant competition for space in the book. We expect that some readers will find themselves wanting more detail about one or another phenomenon than what the book contains. We took our best shot at estimating the levels of interest that particular material would elicit in readers.

In short, this book is about giving readers new things to think about in a new way. Ultimately, we hope that it will serve as a reminder of just how complex human languages are, and how remarkable—verging on miraculous—it is that we can use them with such ease.