Extra Credit

Dear Professor Kogge,

I spoke to you around Thanksgiving break regarding doing an extra credit assignment related to Context Free Grammars and the Chomsky hierarchy. One of the ideas you suggested was to do research on Chomsky's views on linguistics and write about them. I was quickly led down a variety of paths in linguistics, philosophy, artificial intelligence, mathematics, and theory of computing. I probably spent over 10 hours doing research on all of these different topics, but I found it to be absolutely fascinating and a great way to sum up this semester. It is striking to me how connected Chomsky's theory of syntax and language structure is to the content that we have been covering in class. I also found the connection of Chomsky's ideas to artificial intelligence to be particularly insightful. My essay begins by describing some important linguistic developments prior to Chomsky, then summarizes Chomsky's contributions, and finally relates them to modern developments in artificial intelligence.

Best,

Collin Bowers

Background

To begin, it is essential to discuss the ideas of Ferdinand de Saussure, an influential early 20th-century linguist whose work profoundly shaped modern linguistics and, as this essay will later show, has important implications for artificial intelligence. De Saussure is widely recognized as the founder of structuralism, a theoretical framework that examines how elements of a system derive meaning from their relationships to one another rather than from inherent qualities. One of his most important contributions is the difference between the "signifier" and the "signified." The signifier refers to the physical form of a word—whether spoken or written—such as the word *tree*. The signified, on the other hand, represents the concept or mental image associated with the signifier, such as the idea of a tree. Naturally, de Saussure makes the point that the signifier is arbitrary because it varies depending on the language (e.g., tree is *árbol* in Spanish). However, de Saussure's crucial argument is that the relationship between the signifier and the signified is arbitrary, implying that there's no natural connection between a word and its meaning—it is instead established through convention and cultural agreement. He says that "time changes all things; there is no reason why language should escape

this universal law."[1] Thus, for de Saussure, meaning is constructed from the relationship between words rather than a direct connection between words and reality. Notably, this perspective resonates well with modern transformer models, and I shall return to this later in this essay.

The field of linguistics after de Saussure experienced two major developments: the structuralist movements in Europe and the Americas. In Europe, structural linguists followed de Saussure's lead, focusing on the analysis of language based on sound systems and the relational meanings of words within sentences.[2] This approach emphasized the interplay of linguistic elements within a structured system, reflecting de Saussure's foundational ideas. In contrast, the structuralist movement in the Americas took a different trajectory, influenced not only by linguistics but also by structural psychology and behaviorism.[3] Behaviorism, as a psychological framework, prioritizes the study of observable behaviors and their interactions with the environment, largely disregarding internal mental states. Thus, American linguists mostly avoided speculation about meaning and focused on the observable, structural properties of language, such as sound patterns and sentence structures.

Noam Chomsky's Developments

Beginning in the 1950s, Noam Chomsky wrought a sea change in the field of linguistics.[4] He argued that there is a Universal Grammar, where there are constraints on what the grammar of a possible human language could be, and that humans have a Language Acquisition Device, or an innate capacity that we are born with to acquire language. This theory helps explain how children develop language remarkably quickly despite receiving limited linguistic input, a phenomenon often referred to as the "poverty of the stimulus." His ideas dealt a massive blow to behaviorism by demonstrating that language acquisition involves complex mental processes that behaviorists had largely overlooked. His work not only reshaped linguistics but also contributed

---

[1] https://genevasolutions.news/explorations/how-geneva-s-past-thinkers-can-help-us-brace-for-the-ai-upheaval/ferdinand-de-saussure-from-linguistic-patterns-to-ai

[2] Graffi, Giorgio, ' European Linguistics since Saussure', in Keith Allan (ed.), *The Oxford Handbook of the History of Linguistics* (2013; online edn, Oxford Academic, 1 July 2013), https://doi.org/10.1093/oxfordhb/9780199585847.013.0021, accessed 14 Dec. 2024.

[3] Blevins, James P., ' American Descriptivism ('Structuralism')', in Keith Allan (ed.), *The Oxford Handbook of the History of Linguistics* (2013; online edn, Oxford Academic, 1 July 2013), https://doi.org/10.1093/oxfordhb/9780199585847.013.0019, accessed 14 Dec. 2024.

[4] Freidin, Robert, ' Noam Chomsky's Contribution to Linguistics: A Sketch', in Keith Allan (ed.), *The Oxford Handbook of the History of Linguistics* (2013; online edn, Oxford Academic, 1 July 2013), https://doi.org/10.1093/oxfordhb/9780199585847.013.0020, accessed 14 Dec. 2024.

to the cognitive revolution, influencing fields as diverse as psychology, neuroscience, and computer science. By postulating a Universal Grammar and Language Acquisition Device, he suggested that something more fundamental was going on in the human brain to make possible this capacity for language. Chomsky's breakthroughs challenged de Saussure's idea that language is arbitrary and there is no natural connection between a word and its meaning.

Chomsky also made syntax play a larger role in linguistic theory, which had important implications for Computer Science. In his seminal work *Syntactic Structures*, Chomsky introduced a formal theory of grammar, uncovering abstract underlying structures in language. These insights proved particularly valuable for the Theory of Computing. One of his most enduring contributions, the Chomsky hierarchy, is a formal classification system for grammars based on their generative power, connecting linguistics to computational models. The connection between what we learned in class regarding automata theory and the linguistic implications Chomsky's framework has for actual languages used by people in the real world is particularly fascinating. For starters, let's take a Type 3 language (i.e., a regular language), which can be recognized by finite automata. It can produce a sentence like "If you get an A in Theory of Computing, then you must have worked hard." However, if the sentence grows more complex, with multiple instances of "if" and "then," the structure demands the ability to match corresponding pairs. This requirement demonstrates the limitation of regular languages, which cannot "count," as we observed in class. In contrast, Type 2 languages (i.e., context-free languages) can handle some forms of nested dependencies but falter when faced with more complex patterns, such as arbitrary crossing dependencies.[5] These limitations reveal why English and most natural languages belong to the broader category of Type 1 languages (context-sensitive languages), which can be modeled by linear-bounded non-deterministic Turing Machines. Unlike the unbounded Turing Machines we studied in class, linear-bounded automata (LBAs) reflect real-world constraints, as practical computers operate with finite memory. Finally, a Type 0 language is described by an unbounded Turing Machine. It is remarkable that the rules governing both human languages and programming languages can be classified within the four language types of the Chomsky hierarchy. Equally fascinating is the historical coincidence: the

[5] See 6:30 https://www.youtube.com/watch?v=5-uOijZ5mRo. The description of the video also provides a proof using the pumping lemma for context-free languages.

tools to model these languages, such as Finite Automata and Turing Machines, were developed around the same time as Chomsky's groundbreaking work.

Artificial Intelligence

Given the seemingly deep connection between computation and language, it is fitting to turn our attention to Artificial Intelligence (AI). The relationship between the Chomsky hierarchy, human languages, and Turing Machines suggests a profound parallel between human language and the nature of computation. Just as Chomsky was developing his ideas of a Universal Grammar and a hierarchy of languages, the fields of automata theory, computer science, and AI began to grow rapidly. However, despite this theoretical connection, Chomsky's ideas do not align neatly with the direction AI has taken in recent decades. Generative AI relies on probabilistic models for predicting the next word in a sentence based on the context and previous words used. This approach resonates more closely with de Saussure's linguistic theory, which posits that meaning arises from the relationships between words rather than a direct connection to reality. Chomsky, by contrast, has maintained that human language is rooted in an innate, structured cognitive system, not statistical or probabilistic mechanisms.

Chomsky has not shied away from controversy on the issue of modern generative AI models. In the article *On Chomsky and the Two Cultures of Statistical Learning*,[6] Peter Norvig responds to several of Chomsky's claims regarding AI. Norvig argues that Chomsky has focused too much on the generative side of language and not the interpretive side of it. He says that in "interpretation (such as speech recognition) the listener receives a noisy, ambiguous signal and needs to decide which of many possible intended messages is most likely. Thus, it is obvious that this is inherently a probabilistic problem." Having written his article around 2011, Norvig's claims seem to stand up surprisingly well based on the success of ChatGPT. However, Chomsky has not let the newest generative AI tools go unchallenged, and in 2023 he wrote an article entitled *The False Promise of ChatGPT*.[7] In it, he argued that the human mind is not a massive statistical or probabilistic engine, but rather an efficient and elegant system that seeks to create explanations, even from a small amount of information. Against Norvig's view, he argues that the probabilistic nature of generative AI intrinsically limits what it can do. He says that

[6] https://norvig.com/chomsky.html
[7] https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

generative AI tools are, by design, “unlimited in what they can ‘learn’ (which is to say, memorize),” and are consequently “incapable of distinguishing the possible from the impossible… They trade merely in probabilities that change over time.” Humans, by contrast, are “limited in the kinds of explanations we can rationally conjecture,” but that plays to our advantage because our minds have an almost “mathematical elegance” to how they work, allowing us to distinguish the possible from the impossible.

Overall, Chomsky’s career has been defined by his ability to challenge paradigms that underestimate the human mind. He dealt significant blows to behaviorists and structural linguists through his groundbreaking theories, which have reshaped the study of language and cognition. Whether his critiques of AI will prove similarly transformative remains to be seen. His contributions to the Theory of Computing, particularly the Chomsky hierarchy, have been invaluable in synthesizing the core concepts of computation we’ve explored this semester. The philosophical implications of his work are immense and invite us to delve deeper into the intricate relationship between language, computation, and human intelligence.