This write-up is my opinion (more of a brain dump, hence the lack of proper coherence; I have material to read and cringe in a few months or years, though) on the connectionism vs. symbolism/linguistics debate before listening to MLST's Chomsky episode. I'll write P2 once I finish listening to all of it.

Chomsky's linguistics is to language what custom-made kernels were to image recognition. I haven't studied linguistics extensively, but what I've come across conveys this message. Billions of people use Google translate every day. The attention matrices of the transformer Google uses to accomplish this probably understand languages better than all 20th-century linguists combined by capturing a ruleset far richer (it has economic value because of this!) than whatever they conceived. The transformer may not be intelligent, but it captures languages' statistics better than any human despite not "understanding" it (hence its feats of translation). I cannot help but conclude that Chomsky's universal grammar is like a Fabergé egg. Cute, but not helpful. More pop science than science.

You have lost the game in trying to "understand" language the moment you even begin to analyze the syntactic properties of language or nuances like "subjects" and "verbs" (there are tribal languages without these things). It is more humble to recognize that GPT and scaled architectures akin to it were inevitable - that the only way to make machines speak is to let them pick up features of language by showing them human-generated signals. And those who were humble built systems that picked up on the quirks of various languages. They realized the futility of formalizing language. Language is simply some subset of all possible analog sound signals that go into the ear canal or out of the vocal chord, and get processed by the brain simultaneously, during which the signals come in contact with other information the system has to enable various modifications to itself and to the world around it. The original idea of universal grammar failed because of the enormous number of combinations in which these analog signals can be generated and processed by the brain. But DL practitioners succeeded with transformers in natural language processing (and, dare I say it, NLU - Natural Language Understanding) because they implicitly understood that the set of signals to care about is too large to do manual analysis on but not too large to not be amenable to systems (with proper inductive biases and scalability) that can learn (NNs).

What if humans communicated via radio waves instead of vibrating air molecules? Would discrete structures arise from such an intriguing mode of communication? We don't know. We only know that for the agents on Earth that communicate by vibrating air

molecules, these discretely structured representations of sound waves and associated visual symbols and concepts ("called" "as" "language" "by" "a" "subset" "of" "those" "agents" "that" "speak" "English") is an excellent local minimum.

Let's do a thought experiment. Imagine a reinforcement learning scenario in some complex environment; much more complex (but still computationally tractable) than any we have built. Imagine each agent *a_i* in the environment has a huge matrix (or tensor) *M_i* initially (call it the "communication matrix"). Agent *a_i* also has an internal state *I_i* (also a tensor) and an external state *O_i*. *O_i* is accessible to other agents (hence "external"). The agent can pick an action from a set of actions - move around, wait, pick up, throw things, and maybe mate and produce new agents. When two agents *a_i* and *a_j* come together within some distance *D* (this is a hyperparameter), *a_i* can write ("speak") to *a_j*'s matrix *M_j*, and *a_j* can write (either bits, integers, floating point values, or embeddings) to *a_i*'s matrix *M_i*. Whether an agent writes or not is also an action ("whether or not it chooses to speak") to be chosen from the set of possible actions. At each time step, an agent's action is a function of all or a sample of 1. the local states of the environment it visited, 2. it's previous internal states, 3. external states of other agents it encountered, 4. communication matrices (i.e., the memory of what other agents "said" to it), and 5. rewards.

If the agent's action at a timestep is "write" upon encountering another agent, what it writes is a function of the same (or more or less; I don't know exactly; this post is a brain dump). If many agents are in the vicinity, it could write to all their matrices. I think a "block" action is also interesting - when another agent nearby chooses "write" ("speak"), don't let it write ("don't listen"). Put GPTs, ViTs, plain transformer encoders, or ResNets wherever you see "function" in the above scenario. Define some objectives for all the agents and reward functions. It won't be straightforward to do but suppose it's tractable. Let the simulator run for a while. Come back and look at all the communication matrices. If we defined the objectives and rewards well enough, will these agents develop something akin to human language? Will the bits in communication matrices resemble symbols or words? Will these agents create a syntax for this virtual language because a semblance of syntax (or actual syntax) allows for more efficient communication? Will these agents develop a lingua franca, i.e., a commonly agreed alphabet and vocabulary? By accessing other agents' external states, can an agent implicitly approximate their internal states (a "theory of mind")?

I don't know. I don't even know if what I have proposed here is workable. But one can see how the agents' environment informs the language they could develop. If the agents do develop a language with symbols, we can use a simple pre-trained transformer to translate between that and English and vice-versa. Then you could give it English instructions. How cool would that be? We can study language this way; language as an emergent phenomenon. What I've proposed here is probably a bit silly. Still, consider the insanely parallel and powerful "physics engine" that is the Earth, the scale at which evolution played its game with its available materials (a shit ton of elements and molecules), and how long this game has been going on. I find it hard to believe that discrete algorithms whose pseudocode we can comprehend - specifically for this thing "language" that one species would use someday - came to be in brains. It's all mush and goo, electrical signals, and learning algorithms.

Humans don't use algebraic reasoning methods like dogs don't. Algebra may be an emergent mental tool, but it's not hard coded in the brain. We invented algebra only a few centuries ago. And we suck at it. We suck at discrete problems, "...except a small number of humans who've been using pen and paper, and only in the last couple of millennia," as Lecun put it. That's why we invented computers. We created them because our brain's "FLOPs" is too low, and our short-term memory is too fallible to keep track of discrete variables, which may be most likely why we took so long to invent mathematics. Were humans reasoning by operating over discrete variables a hundred thousand years ago? They spoke using grunts and calls, which - like evolution - had time to evolve and gain all of the beauty and nuance we see today. Whatever is in our brain that lets us use language is in an ape's brain, albeit in a rudimentary form. Mathematicians use math the way we use language (it's the "language of the cosmos" after all). When we learn math, we pick up on rules and patterns - a kind of pattern that is more rule-based - defined and studied extensively by mathematicians of past centuries. If saying some people have a "math gene" is silly, and anyone can learn math (which I believe strongly), then so is saying people have a "language gene" (which Chomsky said came from a random mutation millennia ago). To put a pseudocode snippet "$x \leftarrow F(x)$" here and have a reader somewhere else comprehend it is not a result of some gene.

Chomsky's mistake in the first place was to see language as being computational (programming languages inspired this idea). Language is not computational. Dolphins communicate with each other using whistles and boops; perhaps human language is just glorified dolphin speech. That makes much more sense to me than people analyzing "the dog that the cat that the boy saw chased barked." You struggle to make sense of the

sentence because it is silly, and you aren't likely to encounter it; GPT struggles with sentences like it as well because it captures the statistics of human language (as it should). Grammatically incorrect sentences may make sense in some contexts ("I can haz cheezeburger!1!"). Grammatically correct sentences may make no sense ("colorless green ideas sleep furiously"). Your brain doesn't care about grammar but about what sentences it's likely to encounter. This is an important distinction. Joseph D. Becker's "The Phrasal Lexicon" from 1975 goes into more detail. He almost precisely describes how modern language models work (the last page has the abstract) and calls out linguists.

Consider how one (sometimes) corrects sentences when writing (or speaking): I don't always check for grammatical correctness by doing "operations over variables." That would be too slow because, as we know, our brains suck at discrete problems. Instead, the brain turns it into a continuous problem by making you say the sentence aloud (or aloud "in your head") and check if it feels right - a classification problem without clear explainability for classifications, just like we see in supervised learning. This "vibe checking" is what GPT sort of does. It "feels" for correctness.

Our first goal should be - as Lecun put it in a thread - to attain cat-level or middle-aged Chimpanzee-level intelligence. Mathematics, logic, etc., can come after (but here, too, we see DL systems (e.g., Minerva) achieving SoTA results much earlier than expected). One would've expected real-world planning (say, a cat moving a few objects, perhaps downing a wooden ramp, etc., to make way for itself and catch a mouse) would've been "solved" much before language. But we have language models that even outperform humans despite language supposedly being a symbolic problem, according to some. It was a symbolic problem until the introduction of word embeddings. What will the "word embeddings" of the 2020s be? I don't know yet. We can put in systems the ability to do real-world planning and abstract reasoning without using symbolic methods. But we need better sensors and materials engineering to capture the rich information from the environment, which ML systems can use. Current systems need better priors and better long-term intents, which evolution put into organisms over epochs. The question is, can all of this be learned? I think so. Living organisms have the advantage of a vibrant and high-resolution sensor suite developed at the atomic scale over billions of years in the super parallel physics engine that is the Earth. What do our best RL agents have? Python libraries like OpenAI gym.

I agree with Gary Marcus and Chomsky that too much emphasis on NLP (and the major benchmark tasks) can shoehorn DL or general AI research for solely commercial aims,

drawing attention away from scientific understanding (for e.g., Jensen Huang uses marketing speech like "transformer engine!" to refer to FP8 matmuls in the H100 GPU). But intelligence is not a symbolic or algebraic problem. It existed in dinosaurs, and they did not use set theory, math, or algebra. The electrical mush suspended in the cerebral fluid of the skull doesn't have a symbolic algorithm whose pseudocode we can uncover and put on Wikipedia. In a way, the brain is even worse than a neural net. It did not develop in a stable environment. It has all these requirements like food, electrolytes, etc., and it's remarkable it learns anything at all (that's how I see it as a non-expert in neuroscience). We must be humble and develop better systems with connectionist architectures (perhaps even new kinds of processors) that learn ideal structures for reasoning. Nobody is smart enough to uncover a discrete master algorithm and write it on a whiteboard.

Deep learning wouldn't get the attention it has today if not for an engineer in 2012 who decided he was going to implement a blazingly fast CNN using low-level GPU code. An engineer's work led to new science and understanding by focusing the attention of theorists on something different. If symbolists want to make themselves useful, they should deliver results. Make GitHub repositories and come up with workable implementations of ideas. If they think achieving SoTA results is Goodhart's law, they should come up with something really good - like making embodied virtual agents do things that blow people's minds (DL has done this). Most symbolists' criticisms focus on feedforward models only. The critiques fall apart if it's aimed at connectionism in general. After all, the brain is just a bunch of ion channels opening and closing. Surely it cannot come up with Zermelo-Fraenkel set theory. Right?

ABSTRACT

Theoretical linguists have in recent years concentrated their attention on the productive aspect of language, wherein utterances are formed combinatorically from units the size of words or smaller. This paper will focus on the contrary aspect of language, wherein utterances are formed by repetition, modification, and concatenation of previously-known phrases consisting of more than one word. I suspect that we speak mostly by stitching together swatches of text that we have heard before; productive processes have the secondary role of adapting the old phrases to the new situation. The advantage of this point of view is that it has the potential to account for the observed linguistic behavior of native speakers, rather than discounting their actual behavior as irrelevant to their language. In particular, this point of view allows us to concede that most utterances are produced in stereotyped social situations, where the communicative and ritualistic functions of language demand not novelty, but rather an appropriate combination of formulas, cliches, idioms, allusions, slogans, and so forth. Language must have originated in such constrained social contexts, and they are still the predominant arena for language production. Therefore an understanding of the use of phrases is basic to the understanding of language as a whole.

You are currently reading a much-abridged version of a paper that will be published elsewhere later.