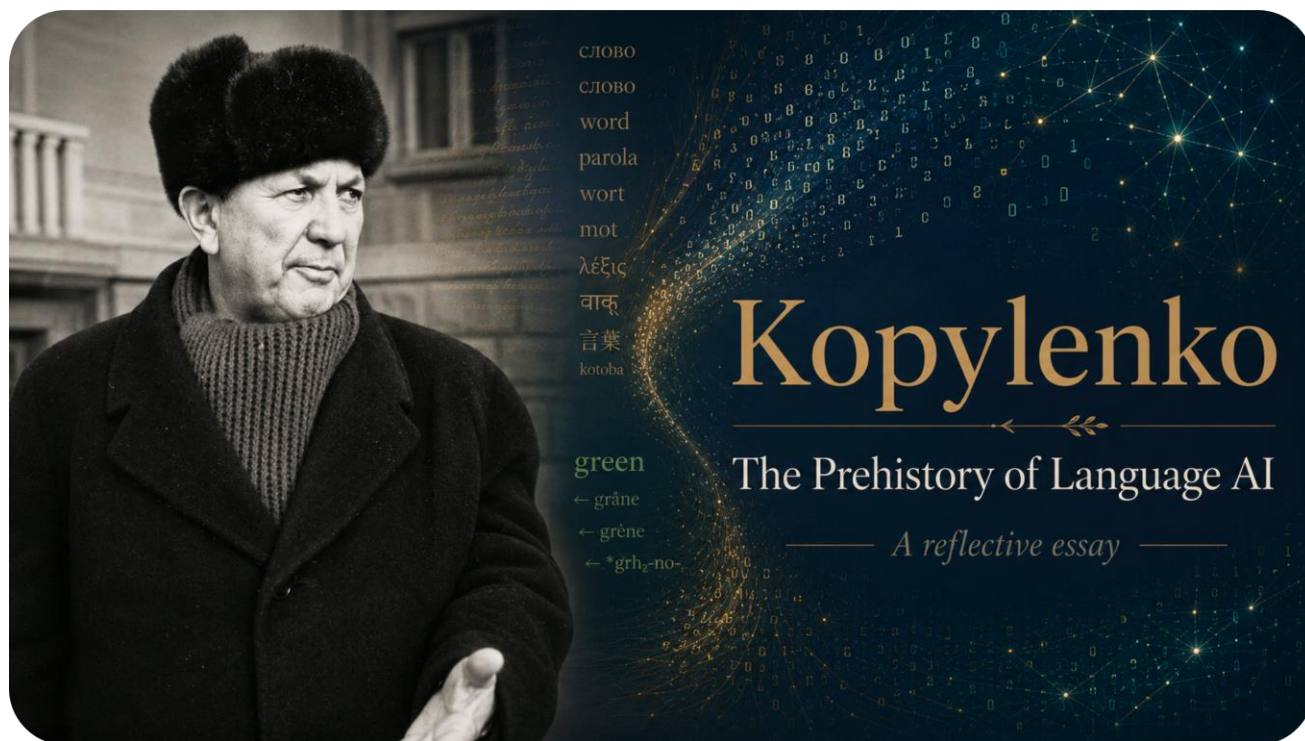




The Green Word

Kopylenko, Combinability, and the Prehistory of Language AI

A personal reflection by Lev Neymotin, PhD
Uncompressed Thinking

On the Way to Penn Station

There are conversations that begin almost accidentally and then, without warning, open a door into a much larger room. This one began on the way to Penn Station, after a visit to the Raphael exhibition at the Metropolitan Museum. The exhibition itself had nothing to do with the subject that followed, except perhaps in one hidden sense: when the mind is still full of color, form, and human presence, it becomes more willing to notice connections that ordinarily pass in silence.

I was walking with Professor Kopylenko's daughter, an old friend, and I asked her a question that was meant to be casual, although I knew very well that it was not an easy question. I asked whether she could give me, in a few words, the central professional accomplishment of her father, Moisei Mikhailovich Kopylenko - something that could be said to an uninitiated audience, not as a formal biography, not as a catalog of publications, but as the kind of living snapshot that allows one human being to understand why another mattered.

Even as I asked, I understood the unfairness of the request. Kopylenko was not a one-line scientist. He was a major linguist, a scholar of wide reach and great discipline, a professor, a builder of schools, a teacher whose students carried his work further, and a man whose professional interests moved across comparative and typological linguistics, phraseology, bilingualism, sociolinguistics, history of linguistic thought, and what we would now call the more formal and structural approaches to language. To compress such a life into a few sentences is already a kind of violence against it. And still, in conversation, we ask such questions because we are searching not for an archive, but for a point of entry.

She did not hesitate. Perhaps only a daughter can do this - not because she is simplifying, but because she carries the person and the work together. What she gave me was not a definition; it was an image. And that image has not left me.

The Daughter's Answer

She said, in essence: my father developed a mathematically grounded way of looking at how any word in a language can be connected to other words in that language. Then she gave me a simple example, so simple that at first it almost sounded like a game.

Take the word green.

What can follow it? Green frog. Green alligator. Green leaf. Green eyes. Green apple. Green envy. Green energy. Green light. Green pencil. Green beret. Green forest. Green zone. Green powder. Green tea. Green room. Green signal. Green fear, perhaps, if the context allows it. Some combinations are ordinary; some are metaphorical; some are rare; some are nearly impossible, but not absolutely forbidden if a poet or a madman finds the right sentence for them.

Then the game changes. It stops being merely playful and becomes scientific. One can search through thousands of texts and ask: how often does this word appear with that word? What follows what? Which combinations are expected, which are surprising, which are stable enough to be part of ordinary speech, and which exist only as a possibility waiting for a context? Suddenly the word is no longer a lonely item in a dictionary. It becomes a point in a field of continuations. It has neighbors, habits, permissions, attractions, resistances, and probabilities.

This, as I understood her explanation, was part of the world Kopylenko entered and helped shape: the world of lexical combinability, the world in which language is not simply a list of meanings, but a living system of possible and actual connections. And because his mind was mathematical as well as philological, the question was not only whether one combination sounded right, but how such combinations might be described, grouped, measured, and made intellectually visible.

A Scientist Who Found Joy in Structure

What touched me most in her telling was not only the idea itself, but the atmosphere around it. She spoke of his enjoyment, of the pleasure he took in this work, and of his students having a wonderful time with him as they explored these patterns. I found this detail deeply important. Serious science is often presented to the public as a grim monument of difficulty, as if importance must always wear a severe face. But in the best minds, the opposite is often true. The deeper the seriousness, the more childlike the delight in seeing a hidden structure appear.

This is part of why Kopylenko's personality matters in the story. I did not know him as his students knew him, and I do not want to pretend that I did. But through his daughter, through the

traces of his life, and through the affection with which his name is spoken, one feels the presence of a scholar for whom language was not a dead object pinned under glass. It was alive. It was playful and rigorous at once. It could be loved, and precisely because it was loved, it could be studied with discipline.

There is something very moving about that combination. A man passes through the harshness of the twentieth century, war, displacement, injuries, and the heavy geography of Soviet academic life, and still builds around himself a world of books, students, words, and exact thought. One does not need to exaggerate such a life. The facts are already large enough. What matters is to feel that the scientific achievement did not float separately from the person; it came from the same inner engine - curiosity, discipline, memory, and a certain freedom of mind that refuses to treat language as obvious merely because everyone uses it.

The Shock of Recognition

Then something happened in my own mind. It was not a conclusion reached by slow analysis. It was more like a light switching on. The description I had just heard - a word surrounded by possible continuations, those continuations appearing with different degrees of likelihood across large bodies of text - suddenly sounded uncannily close to one of the deep intuitions behind modern language AI.

Of course, I want to be precise here. Modern large language models are not simple tables of which word follows which word. They are immense neural systems, built on architectures and computational methods that did not exist in Kopylenko's working world. They operate with tokens, embeddings, layers, attention mechanisms, and enormous training corpora. They do not merely count the word after green. They learn patterns across long contexts and across levels of language that no human scholar could enumerate by hand.

But precision should not kill recognition. Under all that machinery sits a deceptively simple linguistic intuition: language is a system of possible continuations. A word carries with it a horizon of what can come next. A phrase leans toward some completions and away from others. Meaning is not stored only inside a word; it emerges in the way words combine, collide, stabilize, and surprise us. Modern AI did not invent that intuition. It mechanized it, multiplied it, and scaled it until the whole world could suddenly speak with machines that had learned, in their own strange non-human way, the statistical texture of language.

That is why the story shook me. I am not saying that Kopylenko invented the large language model. That would be both unfair to the history of AI and too small a way to honor him. A great scholar does not need to be converted into a modern technology mascot. But I am saying something I believe is both more accurate and more interesting: Kopylenko belongs to the prehistory of language AI. He worked in the humanistic and mathematical territory where language was already being seen as a grand system of combinability before the machines arrived that could process such combinability at planetary scale.

Not Prediction Alone, but Combinability

The word prediction is now everywhere. We are told that a language model predicts the next token, and technically this is true enough as a doorway into the subject. But if we stop there, the whole matter becomes too poor. Prediction sounds mechanical, almost trivial, as if all that is happening is a guessing game. Kopylenko's lens restores something richer. A word is not just followed by another word. It participates in a culture of permissible combinations, conventional

pairings, metaphorical extensions, historical habits, professional vocabularies, jokes, idioms, and poetic violations of expectation.

Green light is not the same kind of combination as green envy. Green tea is not the same as green energy. Green beret carries history and uniform; green apple carries taste, childhood, and the kitchen table; green zone carries military and political geography; green fear sounds strange until some writer makes it necessary. Each combination opens a different semantic chamber. The statistical surface is real, but below it there are meanings, memories, contexts, and forms of life.

This is where the older linguistic vision still matters. Modern AI is powerful precisely because it absorbs unimaginably many patterns, but it does not automatically understand why those patterns matter to us. Human linguistics begins from a different place. It asks why some combinations feel natural, why others require pressure, why some become idioms, why some survive across generations, and why the smallest shift in a word can change the entire emotional temperature of a sentence. Kopylenko's world and the AI world are not the same world, but they touch each other along this edge: the edge where language becomes a system of relations.

The Other Name in the Conversation

Then Ms. Kopylenko brought up another name: Igor Melchuk, another Soviet mathematical linguist, a scholar associated with formal models of language and with the Meaning-Text tradition. The name mattered because it reminded me that this was not an isolated spark. There was a whole intellectual atmosphere in which language was being approached not only as literature, grammar, or national inheritance, but also as an object of formal modeling. The Soviet linguistic world, for all its pressures and limitations, produced people who thought about language with a technical seriousness that now looks surprisingly modern.

This does not diminish Kopylenko; it situates him. Scientific prehistories are rarely made by one person standing alone. They are made by clusters of minds, by conversations, schools, seminars, arguments, dissertations, and students. Later, when technology finally arrives, the public sees the machine and forgets the long invisible preparation of thought that made certain questions thinkable in the first place. The machine receives the spotlight. The earlier intellectual labor becomes a quiet foundation.

A Discovery After the Fact

What moved me most was the delay. Kopylenko, as his daughter told me, did not appear to think of this work as a precursor to something that would one day shake the world. Why would he? The world in which a person could sit at a laptop and ask an artificial system to write, translate, summarize, argue, compose, imitate, and respond in human language still belonged to science fiction. The machinery was not there. The data was not there. The public imagination was not there. And yet the idea, or at least one of its ancestors, was there: language as a structured universe of combinations.

This is how some scientific ideas live. They are born inside one field and only later reveal their second life inside another. The originator may never see the later consequence, and perhaps would not even agree with every way the later world uses his idea. But from the distance of time, we can recognize a family resemblance. Not identity. Not direct descent in the crude sense. But resemblance strong enough to make us pause.

That pause is the point of this essay. It is not a claim of priority. It is an act of recognition. It says: before language AI became a product, before it became a platform, before it became an argument in every newspaper and every boardroom, there were linguists who understood that language could be studied through its possible combinations. They did this without the glamour of Silicon Valley, without industrial-scale models, without global publicity, and often without any expectation that the world would later return to their questions through machines.

Why This Felt Personal

For me, the moment was personal because it came not from a conference lecture, not from a technical paper, and not from a formal search through the history of artificial intelligence. It came from an old friend speaking about her father while we were simply moving through New York after an afternoon of art. That is not a minor detail. Ideas often reach us through human trust before they become formal arguments. I heard the story not as a historian, but as a friend, and perhaps that is why I felt its force so immediately.

I was also struck by the tenderness of the transmission. A daughter carries a father differently from a bibliography. A bibliography gives titles, dates, institutions, and numbers; a daughter can give the tone of a mind. She can say: he liked doing this. He had fun with it. His students had fun too. For a scientist, that may be one of the most beautiful forms of remembrance. Not only that he was right, not only that he was productive, not only that he was honored, but that the work itself was alive around him.

And so the word green stayed with me. It became more than an example. It became a small door through which one could see a much larger intellectual landscape: Kopylenko at his desk, students around him, texts being searched and classified, words moving in constellations, and somewhere far ahead, beyond the horizon of his time, machines learning to speak by absorbing the same great truth in a radically different form - that language is not built from isolated stones, but from relations.

The Claim I Want to Make

So here is the claim, stated as clearly and carefully as I can state it.

Moisei Mikhailovich Kopylenko did not invent artificial intelligence, and he did not create the large language model. But his work on lexical combinability belongs to the intellectual prehistory of language AI. He helped make visible a principle that later became central to computational language modeling: that words and meanings live in structured fields of possible combination, and that these fields can be studied, compared, and in some sense measured.

Modern AI turned that principle into machinery. Kopylenko approached it through linguistics, scholarship, and disciplined intuition. The two are not the same. But they are not strangers either.

That is enough. In fact, it is more than enough. It allows us to honor him without forcing him into a history he did not live to see. It allows us to say that some minds prepare the future not by predicting it, but by touching a structure so fundamental that the future, when it finally arrives, has to pass through it.

Gratitude

I am grateful to Ms. Kopylenko for giving me that doorway. She did not offer me a thesis; she offered a memory. But sometimes a memory is more powerful than a thesis because it carries the warmth of the person inside the idea. Through her few sentences, the scholar became present not as a distant academic name, but as a living mind at play with language, seeing order where most of us only hear speech passing by.

The shock I felt was real. Perhaps it was also a little naive at first, because every discovery made in excitement needs later correction and care. But the corrected version is still exciting. Maybe even more exciting. It is not the story of one man secretly inventing AI before AI existed. It is the story of a great linguist working deeply inside language and revealing, in his own terms, one of the patterns that would later become indispensable to the age of language machines.

On that walk to Penn Station, after Raphael and before the train, I thought I had asked a small question. Instead, I received a bridge: from a father's scientific joy to a daughter's memory, from Soviet mathematical linguistics to contemporary AI, from the word green to the immense machinery of modern language models. I do not know whether this connection has been recognized before. Perhaps it has; perhaps it has not. But for me it arrived with the force of discovery, and with the quieter force of gratitude.

Language remembers more than we think. Sometimes it even remembers the future.

Sources and Notes

This draft is written as a personal reflective essay. The source notes below are included to support the factual framing, not to turn the essay into a technical paper.

1. [Moisei Mikhailovich Kopylenko biography and scientific activity, memorial site](#)
2. [Kopylenko, Moisei Mikhailovich, Russian Wikipedia](#)
3. [Kopylenko, M. M. Sochetaemost leksem v russkom yazyke, bibliographic listing](#)
4. [GPT-4 Technical Report, OpenAI](#)
5. [Vaswani et al., Attention Is All You Need](#)
6. [Igor Melchuk, biography, College de France](#)
7. [Meaning-Text Theory and Igor Melchuk, RUDN Journal of Language Studies, Semiotics and Semantics](#)