Too Early: Reflections on an N-Space Model for Language and Vision (1975)

The Hardware Era (1971–1973)

In the early 1970s I was a graduate student in the EE department at North Carolina State University where, under Professor John Staudhammer, we built what was to be for many tears the fastest interactive 3D computer graphics system in the world. This system used a 32K/30-bit word Adage Graphics Terminal (AGT-30) to manipulate and display 3D *vector* images in real time. In this environment my colleagues David Wooten, Nick England and I built a modeling program written in AGT assembly language that could construct and edit 3D *polygonal* models using the keyboard, buttons and dials on the computer. We also built an interface to an 8K/16-bit word Varian 620 minicomputer that would accept the modeled polygons from the AGT-30. On this minicomputer I wrote a scan conversion and hidden surface removal (SCHSA) in machine code that could convert the 3D polygons into run-length encoded scan lines using an algorithm similar to the work of John Warnock at the University of Utah. Unlike Warnock's Fortran implementation this consisted of about 650 words of machine code in memory and was quite fast in comparison. The run-length encoded scan lines were then output to a FIFO hardware buffer that I designed and built that synchronized with the rotation of a video disk system to push the segments through D/A converters to write individual scan lines on the video disk. Because the Sony Trinitron that was connected to the disk was interlaced, the process took at least 30 seconds to write a complete image. Here's an old video of me operating this system to illustrate its interactivity: https://www.youtube.com/watch?app=desktop&v=JRM6KaBZo6w.

Linguistic Insights (1973–1975)

By 1973 I was working on an EE Ph.D. degree with a minor in Computer Science. Since the fledgling CS department didn't yet offer graduate degrees, it was a part of the Mathematics Department at NC State and my official minor was in mathematics. As a result, while I was studying Computational Linguistics and the works of Chomsky, Bobrow, Winograd, Schank and others I was also taking courses in differential equations and advanced matrix algebra. When my early linguistics class projects using symbolic language models became overly complex, I had an insight that perhaps the use of vectors to encode meaning was a more scalable approach. Initial prototyping confirmed this insight, with most of the punched IBM cards moving from the program deck to the lexicon deck. With the support of Professor Staudhammer and Professor Alan Tharp of CS, I chose this as my dissertation's topic of research.

In 1975, I completed a Ph.D. dissertation titled "*An N-Space Model for Visual and Verbal Concepts*". The central idea was simple, if ambitious: that words and their meanings could be modeled as points in a structured, multidimensional semantic space—and that natural language could be used to generate and manipulate 3D visual scenes through transformations in that space. I wrote a program called ENGRAF that demonstrated this concept, integrating natural language input, a vector-based meaning representation, and producing 3D graphics system editing commands for the AGT-30 as output.

The ACL Conference Experience (1974)

In early 1974, I submitted a version of this work to the 12th Annual Meeting of the Association for Computational Linguistics (ACL) conference, to be held July 26-27, 1974, at UMass in Amherst. The ACL conference in 1974 was a seminal event in NLP history. At the time, ACL was a small but influential academic gathering. The community heavily favored symbolic and rule-based approaches to language, and ENGRAF's vector-based approach was outside of the mainstream. Its emphasis on grounded semantics and multidimensional representation probably seemed unorthodox.

When I arrived at the conference, I was informed that although my abstract had been accepted, my paper would not be included in the proceedings. Nonetheless, I was permitted to give a live demonstration over TSO and to deliver my lecture to the audience. The demonstration itself was limited. At the time, the ENGRAF parser and semantic engine ran on the IBM 370-168 mainframe at the North Carolina Research Triangle Computer Center, while the 3D graphics rendering system lived in our lab in Raleigh. The two systems were not yet integrated; the data flow was theoretical and manual, not real-time. I had hoped to bridge that gap before the conference, but time and resources ran out.

In hindsight, that limitation likely colored the audience's perception. I was asking them to imagine a fusion that wasn't fully there yet describing a working conceptual pipeline from sentence to visual model, while only showing one half live. To me, the vision was clear; to them, perhaps it seemed incomplete or overly speculative. The reaction to my talk was harsh. During the presentation, I encountered pointed critiques, dismissive comments, and overt skepticism, not just about the feasibility of the system but the validity of my entire conceptual framework. There was little curiosity, no questions of possibility—only questions of credibility. I left feeling embarrassed and discouraged and my career moved on from academic AI afterward.

Revisiting ENGRAF through GPT (2024)

Recently I've been learning about Large Language Models via dialogs with ChatGPT and have been impressed with the similarity of that technology to my ideas from 1975. With GPT's help, I found a copy of my dissertation on ProQuest and asked it to evaluate the work. The responses were both stunning and unexpected. Here's a brief one:

"Your 1975 dissertation anticipated several foundational ideas that underpin modern AI, especially in natural language processing and multimodal systems. The use of a multidimensional vector space to represent meaning closely parallels today's embedding models like Word2Vec, GloVe, and BERT, which project words into high-dimensional semantic spaces to capture contextual relationships. Your vision of integrating language understanding with graphical transformations

foreshadows the architecture of contemporary systems such as OpenAI's CLIP and GPT-4V, which bridge language and vision through shared embedding spaces. The use of a linguistic interface to control a 3D modeling system is conceptually aligned with current research in natural language-guided image generation, scene editing, and robotics. While your implementation was constrained by the hardware of the time, the architecture and ambition of ENGRAF were strikingly ahead of their time."

Closing Reflections

I'm left wondering: what might have happened if ACL had published my paper in the proceedings after all? Would it have influenced the field? Probably not directly. But perhaps it would have planted a seed, or at least signaled to others that modeling meaning with vectors and connecting language to visual scenes was a worthy pursuit. In any case, I now look back not with regret, but with a kind of quiet affirmation: the ideas were sound. Their time has come.