

Finite automata as multimodal semantic representations, probably



The University of Dublin

Tim Fernando

Tim.Fernando@tcd.ie
Dublin, December 2025

ABSTRACT

Visual narratives are analyzed as strings recording state changes, amenable to finite-state methods. David Dowty's hypothesis that stative predicates provide the basis for an aspectual calculus is revisited from the perspective of more recent computational work on language models and discrete mechanics, addressing objections (by Emmon Bach, Ray Jackendoff and Manfred Krifka) to a filmstrip analysis of state change.

Linguists against filmstrips

BACH 1986 (Natural Language Metaphysics, page 587)

There is apparently a strong tendency to think that states are somehow basic, a sort of filmstrip view of reality which I do not share. If anything quite the opposite seems to be true. It took about two millennia to come up with satisfactory ways of coping with Zeno's questions about what it could possibly mean to be in a state of motion at an instant or how you could possibly add together dimensionless instants to get changes (you can't).

JACKENDOFF 1996 (page 316)

I wish to reject this 'snapshot' conceptualization, on the grounds that it misrepresents the essential continuity of events of motion. For one thing, aside from the beginning and end points, the choice of a finite set of subevents is altogether arbitrary. How many subevents are there, and how is one to choose them? Notice that to stipulate the subevents as equally spaced, for instance one second or 3.5 milliseconds apart, is as arbitrary and unmotivated as any other choice. Another difficulty with a 'snapshot' conceptualization concerns the representation of nonbounded events (activities) such as John ran along the river (for hours). A finite sequence of subevents necessarily has a specified beginning and ending, so it cannot encode the absence of endpoints. And excluding the specified endpoints simply exposes other specified subevents, which thereby become new endpoints. Thus encoding nonbounded events requires major surgery in the semantic representation.

KRIFKA 1998 (page 98)

the "filmstrip" model of change . . . arguably is not the way movement and change is conceptualized

Computers against linguists

AI has arrived: fears that AI will outsmart humans are widespread.

How did this happen? Against the best efforts of linguists. Ouch.



*The bitter lesson is based on the historical observations that 1) AI researchers have often tried to build knowledge into their agents, 2) this always helps in the short term, and is personally satisfying to the researcher, but 3) in the long run it plateaus and even inhibits further progress, and 4) breakthrough progress eventually arrives by an opposing approach based on **scaling** computation by **search and learning**.*

RICHARD SUTTON (Turing Prize, 2024)

Nature (July 2023, page 677)

The development of large language models is mainly a feat of engineering and so far has been largely disconnected from the field of linguistics. Exploring links between the two directions is reopening long-standing debates in the study of language.

Filmstrips for action

State $\mathbf{q}(0)$ at time 0 changes to state $\mathbf{q}(\hat{t})$ at \hat{t} along a path that minimizes **action**, the time integral

$$S(\mathbf{q}) = \int_0^{\hat{t}} \mathcal{L}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) dt$$

with Lagrangian $\mathcal{L}(q, \dot{q})$.

Vol 2, Ch 19, The Principle of Least Action

State sequences $\mathbf{q}(t_0)\mathbf{q}(t_1) \cdots \mathbf{q}(t_n)$ approximate $S(\mathbf{q})$ via methods from *Discrete Variational Mechanics*

$$S(\mathbf{q}) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} L_{n,i}(\mathbf{q}) \quad \text{reducing } \mathbf{q} \text{ to } \mathbf{q}(t_0)\mathbf{q}(t_1) \cdots \mathbf{q}(t_n) \quad (1)$$

$$L_{n,i}(\mathbf{q}) = h \cdot \mathcal{L}\left(\frac{\mathbf{q}(t_i) + \mathbf{q}(t_{i+1})}{2}, \frac{\mathbf{q}(t_{i+1}) - \mathbf{q}(t_i)}{h}\right) \quad \text{where } h = \frac{\hat{t}}{n} \text{ and } t_i = ih$$

applied fruitfully in computer graphics (e.g., Stern & Desbrun 2006).

Minimum cost search

The function from a string $a_1 a_2 \cdots a_n$ of n tokens a_i to probability

$$P(a_1 a_2 \cdots a_n) = \prod_{i=0}^{n-1} P(a_{i+1}|q_i) \quad \text{where } q_i = a_1 a_2 \cdots a_i \quad (2)$$

is at its maximum when Shannon **surprisal**

$$-\log P(a_1 a_2 \cdots a_n) = \sum_{i=0}^{n-1} c_i \quad \text{where } c_i = -\log P(a_{i+1}|q_i) \quad (3)$$

is at its minimum, with arc **costs** c_i adding up along a path

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \cdots \xrightarrow{a_n} q_n. \quad (4)$$

Drop the assumption $q_i = a_1 a_2 \cdots a_i$, and describe paths (4) by strings s other than $a_1 a_2 \cdots a_n$, including state sequences $\mathbf{q}(t_0)\mathbf{q}(t_1) \cdots \mathbf{q}(t_n)$ used in (1) to approximate $S(\mathbf{q})$

$$\begin{aligned} S(\mathbf{q}) &\approx -\log P(a_1 a_2 \cdots a_n) \quad \text{LHS of (3)} \\ L_{n,i}(\mathbf{q}) &\approx -\log P(a_{i+1}|q_i) \quad \text{from RHS of (3)} \\ &= K_i - V_i \quad \text{where } K_i = -\log P(q_i a_{i+1}) \quad \text{and } V_i = -\log P(q_i) \end{aligned}$$

with kinetic energy $K(\dot{q})$ minus potential energy $V(q)$ as the Lagrangian.

Uncertainty about forces over an incomplete account of states q turns \mathcal{L} into a probability function P

single path (4) \rightsquigarrow multiple strings that are weighed probabilistically

From strings to language models

mathematician known for his work in algebraic geometry and then for research into vision and pattern theory. He won the Fields Medal and was a MacArthur Fellow. In 2010 he was awarded the National Medal of Science.



Mumford defines **Pattern Theory** as

*the analysis of the patterns generated by the world in any modality, with all their naturally occurring complexity and ambiguity, with the goal of **reconstructing the processes, objects and events that produced them** [1994, page 187]*

PROPOSAL: reconstruct processes, objects and events in paths (4) filmed as strings

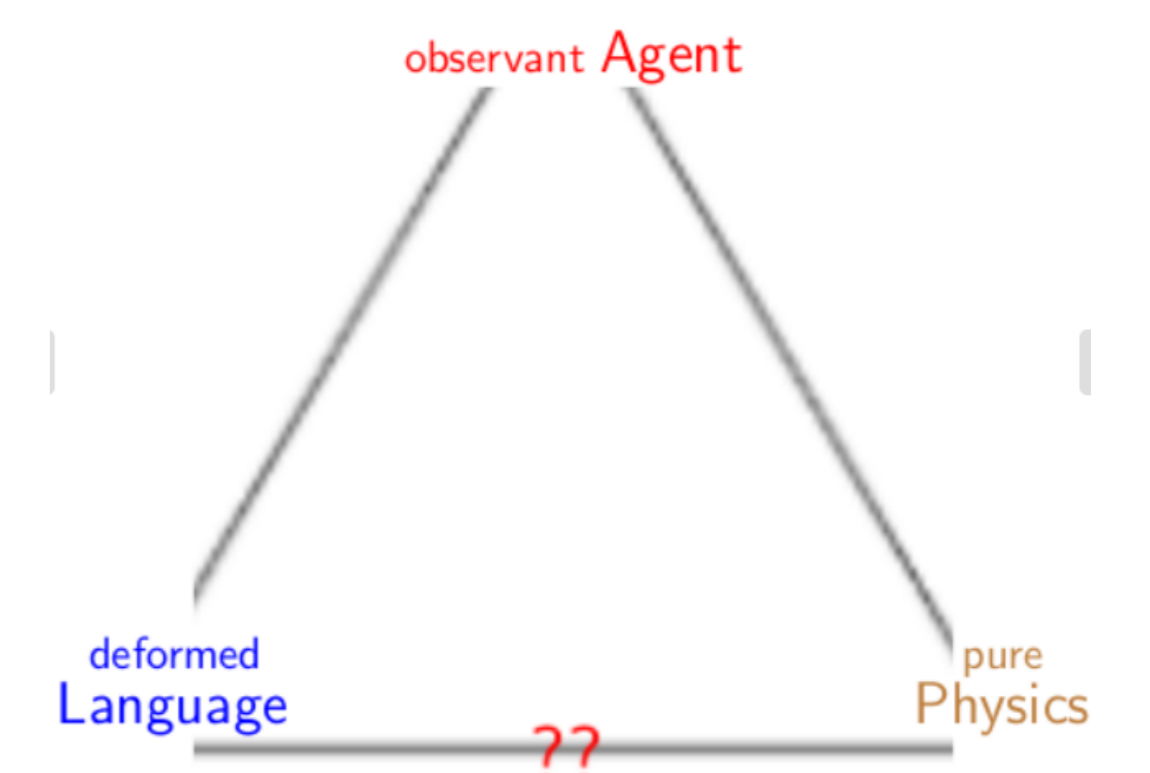
Beyond language models: triads

Saying it's true (over and over) does *not* make it true.

Updating the *Triangle of Meaning* (Ogden & Richards 1923)

*the pattern should not merely describe the 'pure' situation that underlies reality but the 'deformed' situation that is actually **observed** in which the pure pattern may be hard to recognize.*

[Mumford 2019, page 203]



The probability of a transition $q \xrightarrow{a} q'$ to q' after a **given** q is the product

$$P(aq'|q) = P(a|q)P(q'|qa)$$

of an **agent policy**

$$P(a|q) \approx \text{probability agent does } a \text{ given } q$$

and a **state transition**

$$P(q'|qa) \approx \text{probability of world at } q' \text{ given previously } a \text{ is done at } q$$

making probabilities from P **subjective** and variable.

(4) \approx string $(q_0, a_1)(q_1, a_2) \cdots (q_{n-1}, a_n)(q_n, \cdot)$ of pairs (q_i, a_{i+1}) of stative q_i and active a_{i+1}

	linguistic example	change	homogeneous	Kleene 1956 nerve net
stative	Facebook owns Instagram	no	yes	inner cells (internal state q)
active	Facebook bought Instagram	yes	no	input cells (external state a)

Learning vector embeddings by scaling

Exactly what make up q and a are for a machine to learn — or so argues Sutton's *Bitter Lesson*

*the actual contents of minds are tremendously, irredeemably complex . . . They are not what should be built in, as their complexity is endless; instead we should build in only the meta-methods that can find and capture this arbitrary complexity. Essential to these methods is that they can **find good approximations**, but the search for them should be by our methods, not by us. We want AI agents that can discover like we can, not which contain what we have discovered. Building in our discoveries only makes it harder to see how the discovering process can be done.*

Scaling (to "find good approximations") can proceed in different directions

(i) the number m of inner cells \approx dimension of a *Hybrid Automaton*

(ii) the number s_i of values the i th inner cell may take

with (i) and (ii) together giving $\prod_{i=1}^m s_i$ internal states q (Kleene 1956, §8), whereas external states a are discrete analogs of the time derivative \dot{q} in the Lagrangian $\mathcal{L}(q, \dot{q})$, with

(iii) real time discretized by choosing a minimal step size $h > 0$.

Even if we let machines learn the components of q and a , we can well ask

(*) what do/might those components **mean**?

To address (*), strings from (4) are organized in logical systems called *institutions* (Goguen & Burstall 1992)

- spelling out bounded granularities relative to which time arises from non-stative predicates (such as **buy**) and changes to the extension of stative predicates (such as **own**)

- differentiating semantic strings (weighed probabilistically) from (syntactic) strings that describe them

- compressing semantic strings according to the stativity of predicates, complicating the amalgamation of strings of different granularities and giving rise to pattern-theoretic deformations.

Deformations as sequential art

A few well chosen images convey the idea; plenty of detail (within and between images) can be left out.



P.S. Did AI make that? Can AI beat us at any Imitation Game?

CONCLUSION

A nascent notion of state in next-token prediction

$$P(a_1 a_2 \cdots a_n) = \prod_{i=0}^{n-1} P(a_{i+1}|q_i) \quad \text{where } q_i = a_1 a_2 \cdots a_i \quad (2)$$

is refined to understand the maximization of (2) as a minimum cost **search** over paths deformed and observed partially as strings on which finite-state constraints (hard and soft) can be **learned**. The **film-strip** approach to state change is **scaled** for open-ended finite-state approximations, with **action** from non-stative predicates (such as **buy**) and changes to the extension of stative predicates (such as **own**) interpreted model-theoretically across different modalities.