

Finite automata as multimodal semantic representations, probably

A visual narrative presented as a string $\alpha_1\alpha_2\cdots\alpha_n$ of pictures α_i may, to a first approximation, be regarded as a sampling at times $t_1 < t_2 < \cdots < t_n$ of a function \mathbf{q} that maps t_i to $\mathbf{q}(t_i) = \alpha_i$ and other times t to states $\mathbf{q}(t)$ of affairs left out of $\alpha_1\alpha_2\cdots\alpha_n$. A moment’s thought reveals gaps not only between α_i and α_{i+1} but also between α_i and some $\mathbf{q}(t_i)$ imperfectly portrayed by α_i . In vision, occlusion is all about bits of $\mathbf{q}(t_i)$ missing from α_i and as for natural language, the chasm between what is uttered and what is meant is a recurring theme behind logical investigations such as (AL03). But if we are to explain how these gaps are filled, is a string $\alpha_1\alpha_2\cdots\alpha_n$ a reasonable representation on which to build such an explanation? Not according to Ray Jackendoff, who rejects the filmstrip conceptualization “on the grounds that it misrepresents the essential continuity of events of motion” (J96, page 316).¹

Continuous motion in physics is specified by a state trajectory \mathbf{q} (with time derivative $\dot{\mathbf{q}}$) from times 0 to δ that minimizes an integral $\mathcal{S}(\mathbf{q}) = \int_0^\delta \mathcal{L}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) dt$ between fixed initial and final states $\mathbf{q}(0)$ and $\mathbf{q}(\delta)$, in accordance with the *least action principle* (F63). The present work ties this minimization to the search for most probable strings in large language models (LLMs), drawing on discretizations \mathcal{S}_n of $\mathcal{S}(\mathbf{q})$ from (MW01) that reduce continuous state trajectories \mathbf{q} to $n + 1$ samples $\mathbf{q}(0)\mathbf{q}(\delta/n)\cdots\mathbf{q}(\delta)$ between 0 and δ . Increasing n may bring \mathcal{S}_n arbitrarily close to $\mathcal{S}(\mathbf{q})$, but unacceptable approximations \mathcal{S}_n arising from insufficiently large n invite the question: why discretize? Because, as (SD06) points out, analytical solutions to the differential equations (of motion) associated with $\mathcal{S}(\mathbf{q})$ are not always known, and “we can only hope to approximate the solution” (page 76). From (AG), it is clear that such approximations are useful in computer graphics — part of the multimodal swathe disrupted by LLMs. Furthermore, a link between \mathcal{S}_n and string probabilities behind LLMs can be forged from the observation that the function from a string $a_1a_2\cdots a_n$ of n tokens a_i to probability

$$P(a_1a_2\cdots a_n) = \prod_{i=0}^{n-1} P(a_{i+1}|q_i) \quad \text{where } q_i = a_1a_2\cdots a_i \quad (1)$$

is at its maximum when *surprisal* from (S48)

$$-\log P(a_1a_2\cdots a_n) = \sum_{i=0}^{n-1} c_i \quad \text{where } c_i = -\log P(a_{i+1}|q_i)$$

is at its minimum, with arc costs c_i (also surprisals) adding up along successive transitions

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \cdots \xrightarrow{a_n} q_n. \quad (2)$$

The present work relaxes the assumption $q_i = a_1a_2\cdots a_i$, extracting strings \mathbf{s} other than $a_1a_2\cdots a_n$ or $q_0q_1\cdots q_n$ (over which \mathcal{S}_n is a sum, for $q_i = \mathbf{q}(i\delta/n)$) from paths (2), fragments of which appear in \mathbf{s} . These fragments can be identified with *input cells* and *inner cells* from (K56)’s finite-state representation of McCulloch-Pitts nerve nets, in service of the claim that *a nascent notion of state in (1) can be refined to understand the problem of finding a maximally probable string as a minimum cost search over paths deformed and observed partially as strings \mathbf{s} , on which finite-state constraints (hard and soft) are imposed*. This claim bears directly on David Dowty’s influential hypothesis that

- (†) the different aspectual properties of the various kinds of verbs can be explained by postulating a single homogeneous class of predicates — stative predicates — plus three or four sentential operators and connectives (D79, page 71).

¹Similar reservations are expressed by Emmon Bach and Manfred Krifka in (B86, page 587), (K98, page 98).

A stative predicate in (†) is expressed by the verb `own` in (4), but *not* by `buy` in (3) (H20).

Facebook bought Instagram. (3)

Facebook owns Instagram. (4)

To analyze the connection between (3) and (4), (F23) treats `buy(x, y)` as a label on a transition

$$\boxed{(\text{own}(x, y), 0)} \xrightarrow{\text{buy}(x, y)} \boxed{(\text{own}(x, y), 1)} \quad (5)$$

from a state where x does *not* own y (indicated by 0) to a state where x does (indicated by 1), leaving open the possibility of further changes arising from `buy(x, y)` and/or acts simultaneous with or subsequent to any occurrence of `buy(x, y)`, including an utterance of (3) that accounts for the past tense `bought`. These possibilities are structured within a logical system meeting the conditions for an *institution* (GB92), with explicit notions of bounded (finite) but refinable granularity Σ , along which to vary satisfaction $\mathbf{s} \models_{\Sigma} \varphi$ between Σ -models \mathbf{s} and Σ -sentences φ . The transition (5) is encoded as a string $\boxed{(\text{own}(x, y), 0), \text{buy}(x, y) \mid (\text{own}(x, y), 1)}$ of Σ -boxes that can be lengthened or compressed for a Σ -model, on which constraints expressed by Σ -sentences φ are imposed, defining sets of strings accepted by finite automata.

To bring out the homogeneity of stative predicates in (†) against continuous change, we turn to *Hybrid automata* (H00) that describe the state of the environment by a d -tuple $\vec{r} \in \mathbb{R}^d$ of real numbers. A state l of a finite automaton \mathcal{A} is paired with \vec{r} to form transitions $(l, \vec{r}) \xrightarrow{\delta} (l, \vec{r}')$ labelled by $\delta > 0$, specifying change *not* in l but from \vec{r} to \vec{r}' continuously over the time-step δ . More precisely, l is assumed to come with sets $Inv_l \subseteq \mathbb{R}^d$ and $Flow_l \subseteq \mathbb{R}^d \times \mathbb{R}^d$ (for invariants and flow), relative to which a $(\delta, \vec{r}, \vec{r}', l)$ -witness is defined to be a differentiable function f on the closed interval $[0, \delta] \subseteq \mathbb{R}$ such that $f(0) = \vec{r}$, $f(\delta) = \vec{r}'$ and for all $t \in [0, \delta]$,

$$f(t) \in Inv_l \quad \text{and} \quad (f(t), \dot{f}(t)) \in Flow_l \quad \text{where } \dot{f} \text{ is the time derivative of } f.$$

A transition $(l, \vec{r}) \xrightarrow{\delta} (l', \vec{r}')$ says nothing more nor less than: $l = l'$ and a $(\delta, \vec{r}, \vec{r}', l)$ -witness exists. Following (MW01)'s discretization of the action integral $\mathcal{S}(\mathbf{q})$ by \mathcal{S}_n , we can equate the time derivative with the slope $\frac{\vec{r}' - \vec{r}}{\delta}$ in an approximation $\overset{\delta}{\rightsquigarrow}$ of $\xrightarrow{\delta}$ given by

$$(l, \vec{r}) \overset{\delta}{\rightsquigarrow} (l', \vec{r}') \iff l = l' \quad \text{and} \quad \frac{\vec{r}' + \vec{r}}{2} \in Inv_l \quad \text{and} \quad \left(\frac{\vec{r}' + \vec{r}}{2}, \frac{\vec{r}' - \vec{r}}{\delta} \right) \in Flow_l.$$

To improve the approximation $\overset{\delta}{\rightsquigarrow}$ of $\xrightarrow{\delta}$, we define, for any integer $n > 0$, transitions

$$(l, \vec{r}) \xrightarrow{\delta/n}_n (l', \vec{r}') \iff (l, \vec{r}) \text{ is connected to } (l', \vec{r}') \text{ by a path of } n \overset{\delta/n}{\rightsquigarrow} \text{-transitions}$$

breaking $\overset{\delta}{\rightsquigarrow}$ into n steps of $\overset{\delta/n}{\rightsquigarrow}$, with the expectation that raising n brings $\xrightarrow{\delta/n}_n$ closer to $\xrightarrow{\delta}$. Apart from labels $\delta > 0$, transitions $(l, \vec{r}) \xrightarrow{\sigma} (l', \vec{r}')$ changing both l and \vec{r} may arise from a transition $l \xrightarrow{\sigma}_{\mathcal{A}} l'$ in the automaton \mathcal{A} , with σ encoding non-stative predicates such as `buy(x, y)` in (5). That said, stative vocabularies Σ are a natural starting point for strings $\alpha_1 \alpha_2 \cdots \alpha_n$ of pictures α_i (F19). In their simplest form, the panels of a comic strip are snapshots which “have stative informational content” (A14, page 19), although, as Scott McCloud observes, “from the earliest days, the modern comic has grappled with the problem of showing motion in a static medium” leading to “motion lines” (M93, page 110) as well as multiple speech bubbles. In his attack on filmstrips, Jackendoff warns that adopting them “requires major surgery in the semantic representation” (J96, page 316). The present work embraces the challenge of major surgery as a task in *Pattern Theory* (GM07), a theory David Mumford defines 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 (M94, page 187).

Continuous or not, processes, objects and events produce signals that are sampled discretely (feeding, at sufficient scale, LLMs which produce, for all intents and purposes, language).

References

- [A14] D. Abusch. 2014. Temporal succession and aspectual type in visual narrative. In L. Crnić and U. Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 1, pages 9–29. MIT Working Papers in Linguistics 70.
- [AG] Applied Geometry group led by Mathieu Desbrun, publications webpage. <http://www.geometry.caltech.edu/pubs.html>
- [AL03] N. Asher and A. Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.
- [B86] E. Bach. 1986. Natural language metaphysics. In R. Barcan Marcus, G.J.W. Dorn, & P. Weingartner, editors, *Logic, Methodology and Philosophy of Science VII*, Elsevier, 573 — 595.
- [D79] D.R. Dowty. 1979. *Word Meaning and Montague Grammar*. Reidel.
- [F19] T. Fernando. 2019. Pictorial narratives and temporal refinement. In K. Blake, F. Davis, K. Lamp, and J. Rhyne, editors, *Proceedings of the 29th Semantics and Linguistic Theory Conference*, pages 43–62. Linguistic Society of America.
- [F23] T. Fernando. 2023. Triadic temporal representations and deformations. In *Proc. Natural Logic meets Machine Learning IV*, pages 51–61. ACL Anthology.
- [F63] R. Feynman. 1963. The principle of least action. In *The Feynman Lectures on Physics*, volume 2, chapter 19. Caltech, https://www.feynmanlectures.caltech.edu/II_19.html.
- [GB92] J.A. Goguen and R.M. Burstall. 1992. Institutions: Abstract model theory for specification and programming. *Journal of the ACM*, 39(1):95–146.
- [GM07] U. Grenander and M. Miller. 2007. *Pattern Theory: From Representation to Inference*. Oxford University Press.
- [H00] T.A. Henzinger. 2000. The theory of hybrid automata. In M.K. Inan and R.P. Kurshan, editors, *Verification of Digital and Hybrid Systems*, NATO ASI Series F: Computer and Systems Sciences 170, pages 265–292. Springer.
- [H20] M.J. Hosseini. 2020. Unsupervised learning of relational entailment graphs from text. PhD Thesis, School of Informatics, University of Edinburgh.
- [J96] R. Jackendoff. 1996. The proper treatment of measuring out, telicity, and perhaps even quantification in English. *Natural Language and Linguistic Theory* 14, 305–354.
- [K56] S.C. Kleene. 1956. Representation of events in nerve nets and finite automata. In C.E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 3–42. Princeton University Press.
- [K98] M. Krifka. 1998. The Origins of Telicity. In S. Rothstein, editor, *Events and Grammar*. Studies in Linguistics and Philosophy, vol 70. Springer.
- [MW01] J.E. Marsden and M. West. 2001. Discrete Mechanics and Variational Integrators. *Acta Numerica*, 10:357–515.
- [M93] S. McCloud. 1993. *Understanding Comics: The Invisible Art*. Harper Collins.
- [M94] D. Mumford. 1994. Pattern Theory: a unifying perspective. *First European Congress of Mathematics*, pages 187–224. Birkhäuser.
- [S48] C.E. Shannon. 1948. A Mathematical Theory of Communication. *Bell Systems Technical Journal*, 27(3):379–423.
- [SD06] A. Stern and M. Desbrun. 2006. Discrete geometric mechanics for variational time integrators. In *ACM SIGGRAPH 2006 Courses*, 75–80.