Rikki after a decade

I started to work on the Rikki project sometime in 2008. Here I'm commenting on the last (2014) README in the interest of continuity. The original text is quoted in blue.

Proj/Rikki is a knowledge representation project with all the attendant difficulties. AK started to work on this around 2008, his students got involved in 2010. Readers interested in placing Rikki in the tradition of linguistic semantics should read http://www.kornai.com/Papers/toq.pdf, which presents the justification from the perspective of Montague Grammar, and readers interested in placing it the tradition of lexicography should read http://www.kornai.com/Papers/mol11.pdf. Much of what is referred to only in a shorthand form here (esp. in general.note, design.tex, and DOIT) has been moved into these two papers, which, unlike the working notes, are smoother reading. The papers have duly appeared (Kornai, 2010a; Kornai, 2010b) but only my students cited it until the Webology scandal.

Relating Rikki to other strands of linguistic semantics, in particular to the work of Wierzbicka, Fauconnier, Langacker, Talmy, Jackendoff, and others, is an ongoing project. In brief, our goals are closer to the goals of this 'cognitive' school, whose approach to data is the one we like, This has actually happened, with the publication of (Kornai, 2019; Kornai, 2023), and I have met Jackendoff, with whose work my own career is intertwined at least since (Kornai, 1983; Kornai and Pullum, 1990). He continues to be a great guy, but one whose interest is exclusively with human cognition (about which he has interesting things to say to this day, check out Jackendoff and Audring, 2020)while our methods are closer to that of the MG school, whose insistence on rigorous formalization we share. The fact that we think that the methods of the cognitive semanticists are generally childish and the data of Montague Grammarians is largely irrelevant will no doubt earn us the love and respect of both groups. In fact the MG crowd begins to come around to the main criticism (hyperintensionals) that I used to make, and I am beginning to do far more logic than I ever wanted to, see Kornai, 2024b; Kornai, 2024a, and the ever-growing Mechanical Causation draft, now cut in two parts, mech.pdf and cau.pdf.

Placing the work in the AI/KR tradition is also a complex matter, in part because we feel quite comfortable sitting on the fence between the two main camps, the 'rationalists' who pursue symbol manipulation goals and techniques and use logic as their primary tool, and the 'empiricists' who pursue machine learning goals and techniques and use statistics as their primary tool. There should be a third paper devoted to presenting Rikki to those already familiar with the tradition of associative networks that starts with Quillian, 1967 and with the higher order logic approach pioneered by McCarthy, 1959; McCarthy and Hayes, 1969; McCarthy, 1980. Again the bridges are slowly getting built, especially with default logic, which needs to get out of the purely speculative realm and should actually inform itself by the linguistic data put forth on defaults by `4lang`. For a very condensed discussion see Kornai, 2022.

As far as the overall strategy pursued here the following needs to be said. First, that we do not aim at immediate neurological realism, not that it would be a bad thing, but we consider it to be an overly ambitious goal given the current (2014) state of the art. All we retain from this goal is the requirement to implement the system from resource-bound components (small finite state automata with limited fanin and fanout). 'Small' is in the eye of the beholder. Clearly, the L in LLMs is justified, especially if we conceive of these as finite automata, maybe quantized to 4 bits, though 3 should be sufficient, see Chapter 5 of https://www.kornai.com/Books/VectorSemantics/sem.pdf or its earlier incarnation as (Gyenis and Kornai, 2019). The decomposition of LLMs into smaller, more transparent, and symbolically interpretable FS devices remains a central goal, see https://lebadus.ai/Pdf/crfp.pdf. Second, we do not aim at immediate applicability even though some applications, in particular a 'block world' demo and a 'robot ticket clerk' have been created, and others such as a call routing systems have been planned out in some detail, but not implemented for lack of funding. Some more recent work by the students is summarized in Ch 9 of Kornai, 2023, for more on the ticket clerk see Nemeskey et al., 2013.

The system is expected to serve both as a mathematical model and as a practical reasoning engine, but there is no litmus test of the form "if it can't do X, it's not real" that it needs to pass rightaway. The Winograd Schema Challenge (Levesque, Davis, and Morgenstern, 2011) is actually along the right lines for what I had in mind, and the rich subsequent work such as (Sakaguchi et al., 2020) and the broader SuperGLUE etc. benchmarks remain central to the development of LLMs, in spite of the misgivings of Kocijan et al., 2023. Similarly, Rikki need not immediately show signs of evolutionary behavior: it need not itself come about as a result of (natural or artificial) selection, and it may not come with a clear mechanism of how to improve itself (or have descendants which are, at least with some probability, better). Ultimately, these are quite reasonable goals, especially when it comes to the system's ability to learn and generalize, but again we do not put a premium on these rightaway. As a matter of fact, there is work on symbolic computation that relies on self-modifying code (Bukatin and Anthony, 2017), and this clearly remains a central issue, but we have restricted our efforts to 'teachability' in the sense of Quillian, 1969. Related to this, we do not expect sophisticated social behavior from the system at the get-go, but we expect it to have enough capacity to carry Gewirth's argument to the Principle of Generic Consistency (see Kornai, 2014a). We have continued this line of research, see (Kornai, Bukatin, and Zombori, 2023).

The immediate goal is to manually craft or 'knowledge-engineer' an algebraic model that contains what we take to be all core knowledge required for (i) the representation of the meaning of natural language utterances; This is the central aim of the `4lang` project, see (Kornai, 2023). (ii) drawing elementary inferences from such utterances; Extended finite state mechanism, in particular Euclidean automata (Kornai, 2014b; Kornai, 2014c) and Eilenberg machines, as described in Kornai, 2019, still seem sufficient for this goal. (iii) exhibit some form of entelecheia (see Kornai, 2008). The tack we take is strongly reductionst, with the ultimate goal of reducing everything to a small set of primitives, see general.note for details. The file design.pdf gives the 2009 state of the design.

Andras Kornai Fri Oct 31 19:57:54 GMT 2008 Fri Jan 2 23:19:45 EST 2009 Fri Feb 7 14:47:03 CET 2014 Fri Apr 5 13:12:57 CEST 2024

References

Bukatin, Michael and Jon Anthony (2017). "Dataflow Matrix Machines and V-values: a Bridge between Programs and Neural Nets". In: *K + K = 120: Papers dedicated to Laszlo Kalman and Andras Kornai on the occasion of their 60th birthdays*. MTA Research Institute for Linguistics, pp. 153–186.

Gyenis, Zalán and András Kornai (2019). "Naive probability". In: *ArXiv*, p. 1905.10924.

Jackendoff, Ray and Jenny Audring (2020). *The texture of the lexicon*. Oxford University Press.

Kocijan, Vid et al. (2023). *The Defeat of the Winograd Schema Challenge*. arXiv: 2201.02387. URL: https://arxiv.org/pdf/2201.02387.pdf.

Kornai, András (1983). *X-vonás nyelvtanok*. Eötvos Loránd University.

— (2008). "On the proper definition of information". In: *Living, Working and Learning beyond Technology: Conference Proceedings of ETHICOMP 2008*. Ed. by T. Bynum et al. Tipographia Commerciale, pp. 488–495. URL: https://www.kornai.com/Papers/ec08.pdf.

— (2010a). "The algebra of lexical semantics". In: *Proceedings of the 11th Mathematics of Language Workshop*. Ed. by Christian Ebert, Gerhard Jäger, and Jens Michaelis. LNAI 6149. Springer, pp. 174–199. DOI: 10.5555/1886644.1886658.

— (2010b). "The treatment of ordinary quantification in English proper". In: *Hungarian Review of Philosophy* 54.4, pp. 150–162.

— (2014a). "Bounding the impact of AGI". In: *Journal of Experimental and Theoretical Artificial Intelligence* 26.3, pp. 417–438.

Kornai, András (2014b). "Euclidean Automata". In: *Implementing Selves with Safe Motivational Systems and Self-Improvement*. Ed. by Mark Waser. Proc. AAAI Spring Symposium. AAAI Press, pp. 25–30.

— (2014c). "Finite automata with continuous input". In: *Short Papers from the Sixth Workshop on Non-Classical Models of Automata and Applications*. Ed. by S. Bensch, R. Freund, and F. Otto.
— (2019). *Semantics*. Springer Verlag. ISBN: 978-3-319-65644-1. DOI: 10.1007/978-3-319-65645-8. URL: http://kornai.com/Drafts/sem.pdf.
— (2022). "Deception by default". In: *Philosophy and Theory of Artificial Intelligence (PTAI 2021)*. Ed. by Vincent Mueller. Springer Verlag, pp. 171–177. DOI: 10.1007/978-3-031-09153-7_14. URL: http://kornai.com/Drafts/deception.pdf.
— (2023). *Vector semantics*. Springer Verlag. DOI: 10.1007/978-981-19-5607-2. URL: http://kornai.com/Drafts/advsem.pdf.
— (2024a). "Dyadic negation in natural language". In: *Acta Linguistica Academica* 71, pp. 235–257. DOI: 10.1556/2062.2024.00656. URL: https://akjournals.com/view/journals/2062/71/1-2/article-p235.xml.
Kornai, Andras (2024b). "What is the simplest semantics imaginable?" In: *From fieldwork to linguistic theory: A tribute to Dan Everett*. Ed. by Edward Gibson and Moshe Poliak. Language Science Press, pp. 247–259. URL: https://langsci-press.org/catalog/book/434.
Kornai, András, Michael Bukatin, and Zsolt Zombori (2023). "Safety without alignment". In: *ArXiv* 2303.00752.
Kornai, András and Geoffrey K. Pullum (1990). "The X-bar theory of phrase structure". In: *Language* 66, pp. 24–50.
Levesque, Hector J, Ernest Davis, and Leora Morgenstern (2011). "The Winograd schema challenge." In: *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*. Vol. 46, p. 47.
McCarthy, John (1959). "Programs with common sense". In: *Proc. Teddington conference on Mechanization of Thought Processes*. Crown Publishing, pp. 75–91.
— (1980). "Circumscription — A Form of Non-Monotonic Reasoning". In: *Artificial Intelligence* 13, pp. 27–39.
McCarthy, John and Patrick J. Hayes (1969). "Some Philosophical Problems from the Standpoint of Artificial Intelligence". In: *Machine Intelligence 4*. Ed. by B. Meltzer and D. Michie. Edinburgh University Press, pp. 463–502.
Nemeskey, Dávid et al. (2013). "Spreading activation in language understanding". In: *Proceedings of the 9th International Conference on Computer Science and Information Technologies (CSIT 2013)*. Yerevan, Armenia: Springer, pp. 140–143. URL: https://hlt.bme.hu/media/pdf/nemeskey_2013.pdf.
Quillian, M. Ross (1967). "Semantic memory". In: *Semantic information processing*. Ed. by Minsky. Cambridge: MIT Press, pp. 227–270.
— (1969). "The teachable language comprehender". In: *Communications of the ACM* 12, pp. 459–476. DOI: 10.1145/363196.363214.
Sakaguchi, Keisuke et al. (2020). "WinoGrande: An Adversarial Winograd Schema Challenge at Scale". In: *Proc. 34th, AAAI Conference on Artificial Intelligence*, pp. 8731–8738.