

HOPF ALGEBRA READING SEMINAR

András Kornai

Aug 23 2023 2PM CET

ORGANIZATION

- 2pm zoom

<https://us02web.zoom.us/j/84045659802?pwd=L3grbWtqREE4OE>

- 11pm zoom

<https://us02web.zoom.us/j/89203668566?pwd=M1dRL2ozOWxBT>

- Slack https://join.slack.com/t/slack-qyx1689/shared_invite/zt-1xppi4d00-WnJhAvg_ThoSBOw9xH7ylw

- Course webpage

<https://nessie.ilab.sztaki.hu/~kornai/2023/Hopf>

Also reachable as kornai.com → 2023 → Hopf

- Attendance sheet

https://docs.google.com/spreadsheets/d/17cK-cl3_xdbo73_kHWCIAvwgkdg6qz44J4D6tyFfAc/edit?usp=sharing

PLAN FOR TODAY: MINIMALISM (VERY)

SUBJECTIVE RECAP

- 1 What are the key ideas (as far as I understand them)? Review in the style of Polya's *How to solve it*
- 2 Overview of Nemecek, 2023 by Michael Bukatin, see `nemecek_highlighted.pdf` in Resources
- 3 Plans for fall semester

REMEMBER FROM JULY 12 (HOPF4)

HOW TO SOLVE IT

XVI

UNDERSTANDING THE PROBLEM

- First.** *What is the unknown? What are the data? What is the condition?*
You have to *understand* the problem.
Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?
Draw a figure. Introduce suitable notation.
Separate the various parts of the condition. Can you write them down?

DEVISING A PLAN

- Second.** *Have you seen it before? Or have you seen the same problem in a slightly different form?*
Find the connection between the data and the unknown.
You may be obliged to consider auxiliary problems if an immediate connection cannot be found.
You should obtain eventually a *plan* of the solution.
Do you know a related problem? Do you know a theorem that could be useful?
Look at the unknown! And try to think of a familiar problem having the same or a similar unknown.
Here is a problem related to yours and solved before. Could you use it?
Could you use its result? Could you use its method? Should you introduce some auxiliary element in order to make its use possible?
Could you restate the problem? Could you restate it still differently?
Go back to definitions.

How to Solve It

WHAT IS THE UNKNOWN?

- The internal workings of the human mind/brain in linguistic generation (trying to tell someone something) and parsing (trying to understand what is being said)
- Actual production is not limited to single sentences, we often try to express more complex thoughts by more complex texts, but sentences are a good start. Such simplifications are a reasonable heuristic strategy, but they always build **technical debt**: How do we deal with intrasentential phenomena from sluicing (Merchant, 2001) to discourse anaphora (Chafe, 1994) and the recognition of suprasentential patterns (Schegloff, 1979) and so on?
- We need not burden syntax with these questions, but we need to understand how our syntax model interfaces with theories handling these

WHAT ARE THE DATA?

- The best accessible data are *sequences of forms* (remember *sign* = *form* + *meaning* from Aug 4 2pm (hopf7a))
- These often come naturally with sentence-level punctuation (in written corpora)
- In recorded corpora, we can get high precision labeling for speaker (who said it) and timing of pauses
- Sometimes, with a great deal of effort, we can recover some aspects of the *meaning* of signs, e.g. by paraphrase, by introspection, carefully set up experiments. *Not suitable for massive data collection*
- **Technical debt:** Actual production can include paralinguistic phenomena (expressive intonation, gestures, etc) and artistic production, poetry in particular. Note that LLMs can produce poetry of reasonable artistic quality, but so far they are weak on paralinguage

WHAT IS THE CONDITION?

- Whatever system we come up with, it needs to be *learnable* in the sense that it can be selected from alternative models based largely on positive data (very little negative feedback)
- It also needs to be *descriptively adequate* i.e. capable of dealing with the actual variety of human languages
- It also needs to be *implementable* in neural hardware
- LLMs by and large seem to meet these requirements
- In addition to the non-negotiable conditions, there are also some *nice-to-haves* such as *extensibility* to systems that handle the technical debt
- For now, the greatest missing part for both LLMs and linguistic structure is *phonology*. But NNs are making progress, see <https://theaisummer.com/speech-recognition>

DEVISING A PLAN

- Have you seen it before? Or have you seen the same problem in a slightly different form? Do you know a related problem? Look at the unknown!
- Unknowns come in two forms: static (memory engrams) and dynamic (formed in response to input data)
- How do we implement engrams in neural hardware? This is where we started on June 21st (hopf1a) with (McCulloch and Pitts, 1943), for whom memory amounts to recurrent neural net activity (cycles that just don't die down) and Little, 1974, who takes a more abstract view and identifies long-term material with vectors belonging to the eigenspace of the 2nd largest eigenvalue of the RNN transition matrix
- Smolensky, 1990 (see Gerald Penn's July 5 presentation hopf3a.mp4 and slides hopf3g.pdf) is the most significant attempt to see how symbolic computation could operate in RNNs

GRAND PLAN








- Provide a model of static representations (e.g. graphs, vectors, logic formulas)
- Provide a model of dynamic computations using these (e.g. graph unification, vector operations, logic operations)
- Understand how HAs figure in this
- Check the non-negotiable conditions (system is learnable, descriptively adequate, and implementable)
- How about extensibility? How about obtaining learning data?

CARRYING OUT THE PLAN

- Static elements (stored in the lexicon) are *roots* and *features*
- In Minimalism, they seem to differ only in multiplicity, but in traditional grammar we distinguish between *content* and *function* morphemes
- Some static elements are clearly contentful, others clearly functional, but they get fused early on e.g. noun stem + case marking
- Instead of (form, meaning) pairs, we may want to work on (form, category, meaning) triples – extended signs as in Kracht, 2003

PLAN FOR THE FALL

- We will start the week of September 11, exact time TBD
- Three-pronged attack: HA, Minimalism, LLMs
- HA (will partly depend on fall enrollment)
- MCB/MBC (pick up #textual-criticism-of-mcb, continue with what learned from AA and DK so far, possibly getting in other Minimalism experts)
- LLMs will continue with what we learn from MB, starting with work by Elhage et al., 2021; Nanda et al., 2023
- New research questions: are attention matrices positive definite? What do workspaces buy us?

-  Chafe, Wallace (1994). *Discourse, consciousness, and time; the flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.
-  Elhage, Nelson et al. (2021). *A Mathematical Framework for Transformer Circuits*. URL: <https://transformer-circuits.pub/2021/framework/index.html>.
-  Kracht, Marcus (2003). *The Mathematics of Language*. Berlin: Mouton de Gruyter.
-  Little, W. A. (1974). "The existence of persistent states in the brain". In: *Mathematical Biosciences* 19, pp. 101–120.
-  McCulloch, W.S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *Bulletin of mathematical biophysics* 5, pp. 115–133.
-  Merchant, Jason (2001). *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*. Oxford University Press.
-  Nanda, Neel et al. (2023). *Progress measures for grokking via mechanistic interpretability*. arXiv: 2301.05217 [cs.LG].



Nemecek, Adam (2023). *Coinductive guide to inductive transformer heads*. [arXiv: 2302.01834](https://arxiv.org/abs/2302.01834) [cs.LG].



Schegloff, E.A. (1979). “Identification and Recognition in Telephone Conversation Openings”. In: *Everyday Language: Studies in Ethnomethodology*. Ed. by G. Psathas. New York: Irvington, pp. 23–78.



Smolensky, Paul (1990). “Tensor product variable binding and the representation of symbolic structures in connectionist systems”. In: *Artificial intelligence* 46.1, pp. 159–216.