

HOPF ALGEBRA READING SEMINAR

András Kornai

October 30 2023 6:30pm CET

ORGANIZATION

- Will (likely) alternate between 6:30PM and 10PM Monday zooms
- 6:30pm zoom
<https://us02web.zoom.us/j/85255485089?pwd=bFk5VWhva1RTQ>
- 10pm zoom
<https://us02web.zoom.us/j/84731595225?pwd=ZVdzZTFmaXJZO>
- 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_kHWCIAvwgkd-G6qz44J4D6tyFfAc/edit?usp=sharing Please fill in also if you are watching the recordings (without being present on zoom)

PLAN FOR TODAY

- ① Impromptu presentation by András on roots and features
- ② Some burning research questions for linguists
- ③ Some burning research questions for CS/AI/hacker types
- ④ Some burning research questions for mathematicians
- ⑤ Plans for the rest of the fall semester

ROOTS

- In linguistics *roots* are generally considered the smallest morphemes (meaning/form pairs)
- Typical examples include Pāṇinian roots (the Dhatupatha lists 1943, Whitney, 1885 lists only 885, triconsonental roots in Semitic languages, etc.
- Sometimes the meanings are very clear, Sanskrit *smi* 'smile', *vadh/badh* 'slay'; Hebrew *t.l.p.n* 'telephone'
- But often the meanings are more hazy, as in Skt *aNh* 'narrow, distrssing', English *be* (am, are, is, was, were, would)
- Historical depth just adds to the bleaching of the meaning, consider English *mit* (commit, demit, remit), *pose* (compose, depose, repose) etc
- Harley, 2014 departs from this tradition somewhat, paying little attention to morphological and phonological structure
- Subroot units? "phonestemes", "ideophones", "phonosemantics"

IDEOPHONES



Jess Tauber

September 10 · 🌐

I've now finished with noninitial /r/ in Khalkha (the Mongolian national language) ideophones (there are no r-initial ideophones in any case). After working out how // works in non-initial positions within ideophones, I predicted that /r/ would contrast with it, and it does. The vast majority of forms with non-initial /r/ make reference to standing or sticking out (as away from the substratum, the way the hairs on your arm would when a gentle breeze runs over your skin). // on the other hand has the opposite sense, that of relaxation to parallelism with the substratum. Also the relaxation sense of // here also extends to the material texture of the object itself, while /r/ has a stiffening, hardening sense to it. This latter I refer to as the 'you can't shave with a wet noodle' principle (for want of a better name). All sorts of interactions we make with tools, with parts of constructions, etc. depend on the relative material properties of the pieces, especially if you want to force something to happen. One yields, the other prevails.

FEATURES

- In linguistics *features* are also minimal (typically bound) morphemes, but they come in contrasting sets, and are seen as dependent on the stems. The consensus list:
- 1A Person (1st, 2nd, ...)
- 1B Number (singular, dual, ...)
- 2A Location (here, there, near, ...)
- 2B Direction (to, from, ...)
- 3A Gender (feminine, definite, animate, valuable, round-shaped, ...)
- 3B Topic (familiar, known, ...)
- 4A Tense (past, present, ...)
- 4B Aspect (perfect, habitual, ...)
- 5A Case (subject, object, ...)
- 5B Voice (active, benefactive, ...)
- 6A Degree (comparative, superlative, ...)
- 6B Mood (interrogative, negative, ...)

FEATURES IN GRAMMAR

- Long history in modern grammar: features were critical in phonological theory Trubetskoi, 1939; Jakobson, Fant, and Halle, 1952; Chomsky and Halle, 1968
- They were also critical in lexical semantics Katz and Fodor, 1963, anthropology, social science, . . .
- ϕ -features are the ones participating in agreement (Chomsky since 1990s)
- Traditional distinction between inflection and derivation
- As is typical for ling, distinctions are reasonably sharp, but not entirely clear-cut like in chemistry
- This goes both for the inflection/derivation distinction but also for the larger root/feature or syntax/semantics issue

BROAD RESEARCH QUESTIONS

- Linguistics clearly has a lot to do with symbol manipulation. What are the symbols? Are they morphemes, features, what?
- Linguistics clearly has a lot to do with structure. What is the structure? Strings, autosegmental representations, trees, (hyper)graphs, what?
- We have gold data, mountains of it, and LLMs can be taught to provide analyses in terms of the theory, e.g. output a dependency parse. Find these in the black box (or show that they are not there)
- What is the appropriate vocabulary?

NARROW RESEARCH QUESTIONS








- features and roots: is reflexive *self* a feature or a root? How about comparative *-er*?
- *unbelievable*, *unlockable*, *nonrefundable*, *unhappier* (see Laura Kalin's morphology GLOW talk)
- *cleanize, *greenize (see Kornai, 2023 Ch 1.2, Ch 2.2)
- Find (D)FSA in LLMs (see Omlin and Giles, 1996)

PLANS FOR THE REST OF THE FALL

- Remember our motto (stolen from the Hungarian beer company): “The world is as much as you pour into it”
- Students need to formulate goals, find papers they want to present in class, or solve more complex research problems
- Suggestion: we start with MBC rather than MCB
- They essentially use bialgebras, what is the full story about the antipode?
- What is the (weak and strong) generative capacity?
- Continue with applying what we learn to understand LLMs

REVISED SCHEDULE

Sept 25	6:30pm	Adam Nemecek: Machine Learning with HAs
Oct 2	10pm	Richard Luo: Old Georgian and semilinearity
Oct 9	6:30pm	Isabelle Senturia: Learning CCGs
Oct 16	10pm	Facundo Calabró: Multiple CFGs
Oct 23		No meeting (Hungarian National Holiday)
Oct 30	6:30pm	András Kornai: Roots, features, symbols
Nov 6	10pm	Stepan Shabalin: probing simple models
Nov 13	6:30pm	Blanka Kövér: Bialgebras without an antipode
Nov 20	10pm	Avery Andrews: Decyphering MCB
Nov 27	6:30pm	Michael Bukatin: Mathematical foundations of transformer analysis
Dec 4	10pm	TBA

-  Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper and Row.
-  Harley, Heidi (2014). “On the identity of roots”. In: *Theoretical Linguistics* 40.3/4, pp. 225–276.
-  Jakobson, Roman, Gunnar Fant, and Morris Halle (1952). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. MIT Press.
-  Katz, Jerrold J. and Jerry A. Fodor (1963). “The structure of a semantic theory”. In: *Language* 39, pp. 170–210.
-  Kornai, András (2023). *Vector semantics*. Springer Verlag. URL: <http://kornai.com/Drafts/advsem.pdf>.
-  Omlin, C.W. and C. Lee Giles (1996). “Constructing deterministic finite-state automata in recurrent neural networks”. In: *Journal of the ACM* 43, pp. 937–972.
-  Trubetskoi, Nikolai Sergeevich (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck and Ruprecht.



Whitney, William Dwight (1885). “The roots of the Sanskrit language”. In: *Transactions of the American Philological Association (1869–1896)* 16, pp. 5–29.