Hopf algebra reading seminar

András Kornai

June 28 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_ThoSBOw9xH7yIw
- 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

Plan for today

1. HA preparations
2. Typical data structures used by linguists: (planar) trees, wild trees, finite functions, DAGs, codescriptive stuff,...
3. TP preparations
4. Analog/digital debate (with Zsolt in the analog and me in the digital corner)

Hopf Algebras

- Blanka found an excellent resource: click here
- These are Federico Ardila's 2012 lectures at SFSU. He is an excellent speaker: clear, energetic, didactic. Highly recommended to students who already have a first class in algebra (groups, rings, fields, polynomials) and linear algebra (vector spaces, linear mappings)
- We will use self-assessment (based on the homeworks/survey from the last meeting) to figure out how fast we can go
- Diego attended a lecture by Marcolli, I put her slides on the Resources page. This will be super helpful in understanding her papers! She is very terse on the linguistic background, but we will not be
- First courses in linear algebra rarely cover tensor products, but see Chapter V of Prasolov (in Resources). Hungarian version (Praszolov: Lineáris Algebra, Typotex 2005), and I recommend Wettl 2023 (brief but illuminating discussion of tensors)

Data structures used by linguists

- Generally *trees*, but what kind? Rooted/unrooted, labeled/unlabeled, planar/nonplanar, connected or not (forests), directed or not, binary or more branching, unary nodes permitted or not, empty nodes permited or not.
- The big dividing line: weighted or unweighted
- Usually probability weights, but can be taken from any semiring
- A very relevant semiring is the tropical semiring approximating log probabilities: addition is max, multiplication is plus

Tsujii (1988)

General Comments

Computational linguistics (CL) has borrowed a lot of ideas from *Theoretical Linguistics (TL)*. We could not have developed even a simple parser without the research results in *TL*. It is obviously nonsense to claim that we, computational linguists, do not care research results in *TL*.

However, the researchers in *TL*, it seems to me, are very fond of *fighting*, especially, those who are called *Syntacticians*. They always fight with each other by asserting that their *grammar formalisms* are superior to the others'. They are oversensitive and tend to distinguish people into two groups, *the ally* and *the enemy*.

A computational linguist using LFG (or *pseudo* LFG) as a small part in his total system is taken as the ally of LFG, and is certainly accused by the other groups. They promptly demonstrate that LFG is wrong, by showing a lot of peculiar sentences which rarely appear in real texts.

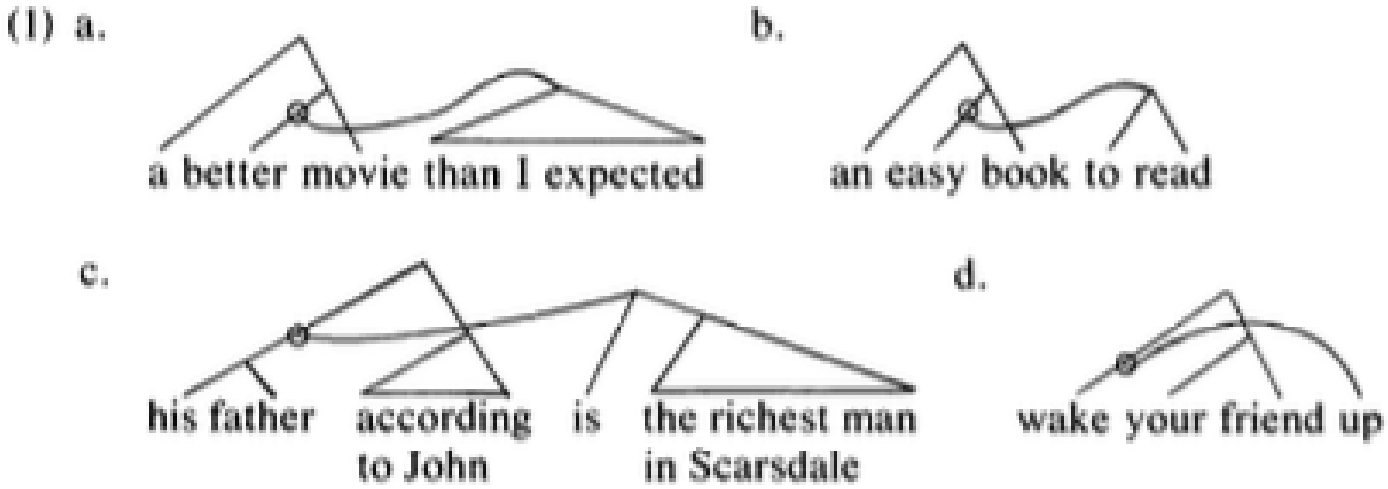
We are tired of listening to such discussions.

Why trees?

- Classic example Wells (1947): *The King of England opened Parliament*.
- We want to cut this in two parts that enjoy large combinatorial freedom: the best cutpoint is between the subject and the rest: (The K of E)(opened Parliament). (i) Both parts occur pretty freely elsewhere 'The K of E X(=did something)' and 'Y opened Parliament' (ii) both can be substituted by simpler (ideally, one word) material: (Joe)(slept).
- We do this recursively: *opened Parliament* is further analyzed as (opened)(Parliament) and *the King of England* as (the)(King of England), the latter as (king)(of England)
- This naturally gives trees
- Notice that the (important for grammar) notion *subject* is not used anywhere in the process

Now for some more complex examples

- *Úristen, mondtam, ez az én fiam, azonnal megismertem!*
  Lord god, I said, this is my son, I recognized him immediately
- The commas (which don't require a written form, they can be detected in the acoustic signal, (*comma intonation*, see Hetzron (1980) and Kornai and Kálmán (1988)) segment the material in four parts ABCD.
- There is no good cutpoint, the best parse is B+A..CD. Here B is called a *parenthetical*, and A..CD is called a *discontinuous constituent*. Examples from Wells (1947) via McCawley (1982):



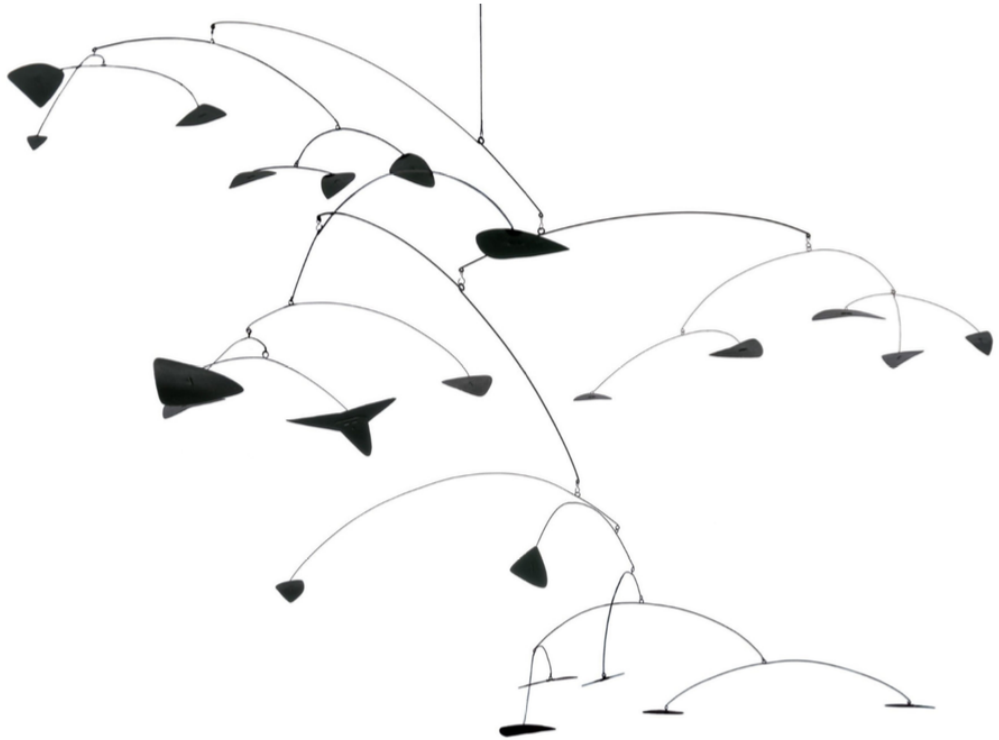
Labels: positional or functional

- What is the replacement class of *lord god*?
- This is not a direct invocation of god (vocative) but an expression of the speaker's emotive state
- We can substitute *aw fuck* just as easily as *Jesus* but not *holy mackarel*
- Labels have complex internal structure (seen e.g. from agreement phenomena): *ez az én lányom, azonnal megismertem* 'this is my daughter, I recognized *him/her/*it immediately'

- what languages appear to look like

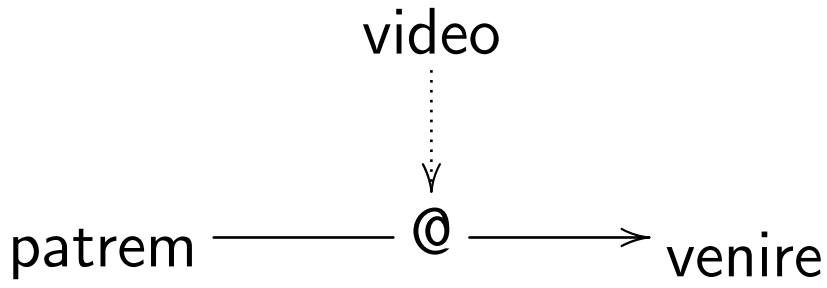
$$00220100211120001212100220000200211\ldots$$

- what languages actually look like



Hypernode graphs

- Assume 'standard' SVO order (Subject-Verb-Object) as in *Kim chose Sandy*
- *Video patrem venire*



- ([Ego] video (patrem venire [null]))
- The first zero [ego] is substantive – we *know* from the conjugation not just that it is present tense, active, indicative but also that it is 1st person singular (cf. *tu vides, is videt, nos videmus, . . .*
- The second zero [null] is technical
- Left for later: DAGs, finite functions, codescription

Theorem proving: the main goals

- Understanding SOTA in Hopf algebras
- Understanding SOTA in 'physics-style' arguments
- Linking HAs to standard NLP tasks such as
- Computing structure from surface string
- Computing surface string from structure
- Already very hard at the word level: conjugating video

Analog or digital?

- Analog superiority claim: "If neural outputs were 0/1 this would invalidate current learning methods"
- Digital equivalence claim: "Whatever can be learned by analog networks can be learned just as well by digital ones"
- Approximation claim: "Whatever A can do D can approximate to arbitrary degree of precision"
- What would constitute a counterexample to digital equivalence? An analog network A that cannot be approximated by a digital one, so:
- Some ϵ such that $\nexists\delta$ such that specifying inputs within δ would produce the same output within ϵ
- Does learning make a difference?

Approximation

- Standard definition of analog computation: General Purpose Analog Computer (GPAC) model
- Is this applicable to NNs?
- There are digitally computable functions, such as Γ, which are not GPAC computable
- But maybe we should be satisfied with approximation over time?
- Brings in Zeno paradox issues (infinitely many computation steps in finite time)
- Church-Turing Thesis: no mathematical algorithm/computation method escapes the Turing bound
- *Physical Church-Turing Thesis*: no real-device implementable algorithm/computation method escapes the Turing bound

- Hetzron, Robert (1980). “Hungarian Tonosyntax”. In: *Nyelvtudományi Értekezések* 104.
- Kornai, András and László Kálmán (1988). “Hungarian sentence intonation”. In: *Autosegmental studies on pitch accent*. Ed. by Harry van der Hulst and Norval Smith. Dordrecht: Foris, pp. 183–195.
- McCawley, James D. (1982). “Parantheticals and discontinuous constituents”. In: *Linqistic Inqiury* 13.1, pp. 91–106.
- Wells, Roulon S. (1947). “Immediate constituents”. In: *Language* 23, pp. 321–343.