Unifying formulaic, geometric, and algebraic theories of semantics

András Kornai
SZTAKI Computer Science Research Institute

ESSLLI, August 2 2021


ตายอด
คลำช้าง

Blind men and the elephant

A group of blind men heard that a strange animal, called an elephant, had been brought to the town, but none of them were aware of its shape and form. Out of curiosity, they said: "We must inspect and know it by touch, of which we are capable". So, they sought it out, and when they found it they groped about it. The first person, whose hand landed on the trunk, said, "This being is like a thick snake". For another one whose hand reached its ear, it seemed like a kind of fan. As for another person, whose hand was upon its leg, said, the elephant is a pillar like a tree-trunk. The blind man who placed his hand upon its side said the elephant, "is a wall". Another who felt its tail, described it as a rope. The last felt its tusk, stating the elephant is that which is hard, smooth and like a spear. (Apocryphal back-story to Johnstone, 2002)

Outline

Who is this course for?

- Semanticists, both "mainstream" and "cognitive"
- Morphologists, lexicographers
- Linguists interested in AI, KR, NLP
- People (still) interested in major questions raised 30-50 years ago
- People not afraid of formal theories
- People attracted by some of the coming attractions

Coming attractions

- Brief intro to the five main theory types This lecture
- Relatively painless intro to vector semantics Second lecture
- Foundations of non-compositional semantics Third lecture
- What do bound morphemes mean? Third lecture
- Spatiotemporal semantics, 'projection mapping', indexicals Fourth lecture
- Negation, probability, implicature, modality Fifth lecture
- How to work on the elephant by synchronous rewriting Time permitting

Work of many people

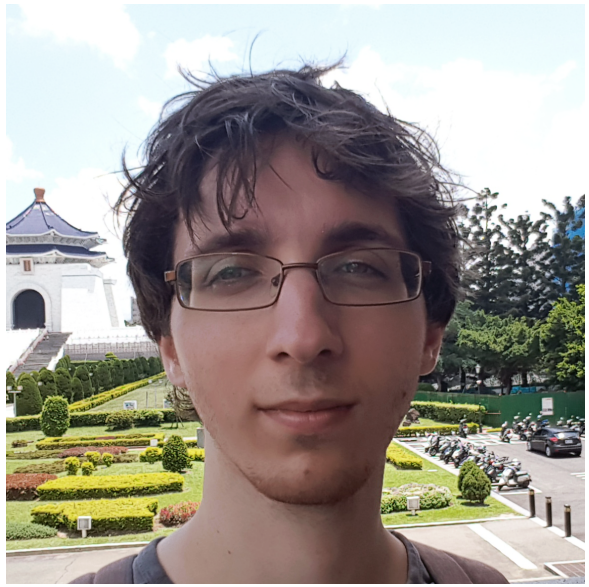
Judit Ács

Márton Makrai

Dávid Nemeskey

Ádám Kovács

Gábor Recski



Dániel Lévai

Format

- Lectures are kept reasonably modular
- Each lecture about two 40 minutes halves
- Followed by question/answer session
- Readings/slides made available in advance, video afterwards
- Course website at kornai.com/2021/ESSLLI
- People interested can do project or term paper

Theories of semantics classified by mathematical apparatus

- 1. Logic-based: the Frege–Russell–Tarski–Montague mainstream, henceforth MG (including lineal descendants like Discourse Representation Theory, Dynamic Predicate Logic, Inquisitive Semantics, etc)
- 2. Based on (hyper)graphs: Traditional AI/KR, AMR, 4lang
- 3. Based on linear algebra: distributional semantics (CVS)
- 4. Based on automata theory: Finite State models (operational semantics, see e.g. Fernando, 2018)
- 5. Based on rejection of formal apparatus: cognitive semantics \ Jackendoff

The tusk: logical theories

The mainstream: MG

Attendees are likely to know this, and if not, plenty of great textbooks and advanced courses are available, I recommend Kracht, 2011 and Jacobson, 2014.

Ignoring theory-internal problems, such as hyperintensionals, there are three main issues:

- Logic is too powerful (which makes it unlearnable)
- Meaning postulates are brittle, word meaning remains a mystery
- Creates problems where there are none

The tusk is too sharp

- Problem: logic is not the right tool, it's too sharp. Natural language is incapable of arithmetic
- Everyday language is "a rough and ready instrument incapable of expressing Truth with a capital T" Russell, 1940
- In natural language "it seems to be impossible to define the notion of truth or even to use this notion in a consistent manner and in agreement with the laws of logic" Tarski and Blaustein, 1936 (English tr. 1956)
- "people who put knowledge into computers need mathematical logic, including quantifiers, as much as engineers need calculus" McCarthy, 2005)
- The key takeaway: either you consider *The atomic weight of mercury is 200.592(3)* a natural language sentence, and want a theory that can deal with it, or you are content to consider it a sentence of technical language, outside of scope for natural language semantics, but you can't have both.

The tusk is too weakly attached

The word entropy of natural language is about 12–16 bits/word M08:7.1. Capitalization and punctuation (our best proxies for intonation and related factors) contribute less than 7% (0.12 bits of 1.75 bits per character Brown et al. (1992).

Syntax is an information source of its own. There are many formalisms, we just conider binary trees over n words. These contribute *at most* $\log_2 C_n$ bits. C_n is hard to compute exactly, but asymptotically $C_n \sim 4^n/\sqrt{\pi} n^{1.5}$, so encoding a parse tree requires less than 2 bits/word. (The masoretes used 2 bits for parsing the Bible, Aronoff (1985))

The key takeaway: Information is carried by the words. Logical structure accounts for no more than 12–16% of the information conveyed by a sentence, a number that actually goes down with increased sentence length, and emotive content for even less, perhaps 5–7%.

The tusk is a weapon

Weapons are necessary for certain purposes, but their overuse, actually their very presence, can create problems. When you have a tusk, everything looks like a tree to be debarked.

- Sharp or fuzzy boundaries: are you *fat* when your weight, expressed in kilograms, divided by your height (expressed in meters) squared, is over 30?
- Superfluous readings, "metaphoric usage", and "metonymy" – ordinary language use gets demoted to special status.
- "The first step is to measure whatever can be easily measured. This is OK as far as it goes. The second step is to disregard that which can't be easily measured or to give it an arbitrary quantitative value. This is artificial and misleading. The third step is to presume that what can't be measured easily really isn't important. This is blindness. The fourth step is to say that what can't be easily measured really doesn't exist. This is suicide." (WP on the McNamara fallacy)

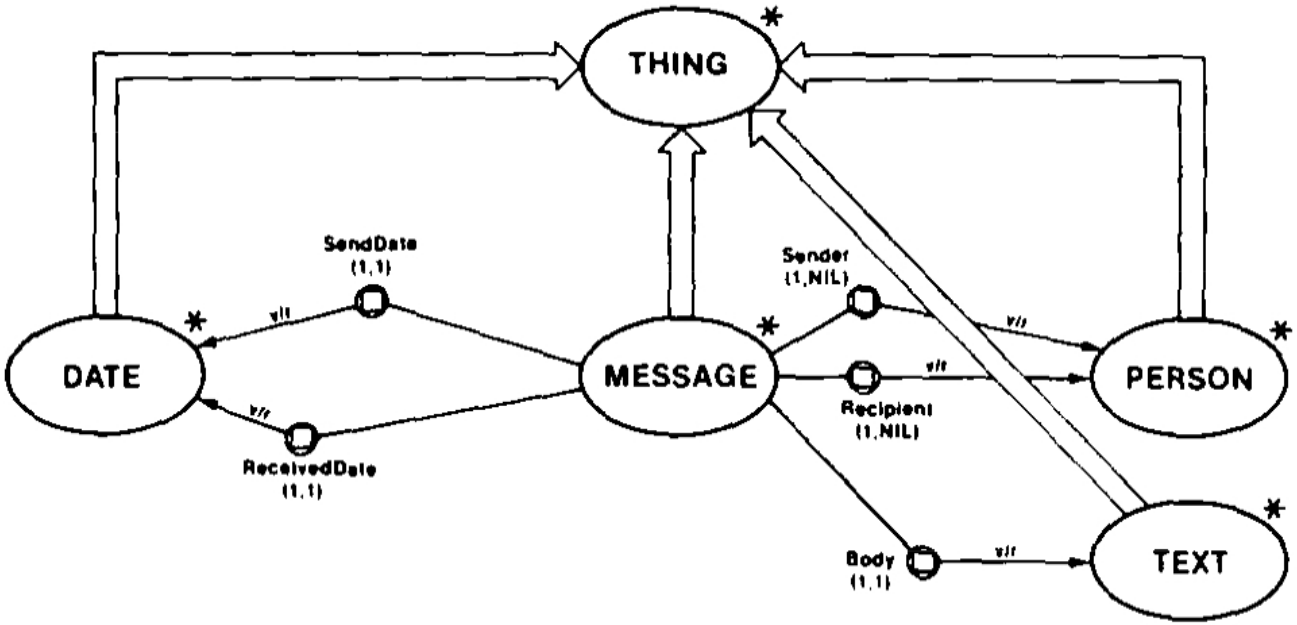
Advantages of truth-theoretical semantics

- Nice clean fun with λ-calculus
- Good (albeit imperfect) stories about intension, modality, temporal reasoning
- Beats the naive "pictures in the mind" theory every possible way
- Good (actually too good) account of quantifiers
- Somewhat good account of pronouns
- Led to the discovery of exciting phenomena (Bach-Peters sentences, paycheck pronouns, non-constituent coordination)
- Fits well with type-theoretical work, programming lg semantics
- Model theory fits well with *reist/concretist* philosophy

The ear: graph representations

- Mainstream approach in AI, its popularity moves in tandem with the AI hype cycle
- Linguists always had their own graphs (constituency, dependency, trees/DAGs, LFG diagrams, . . .)
- Modern, linguistically inspired versions: AMR (Banarescu et al., 2013); `4lang` (Kornai, 2010)
- Now terascale, primary tool in XAI
- Does not require reist underpinnings: there can be 'real things' not made of atoms, e.g. feelings, attitudes, circles, . . .

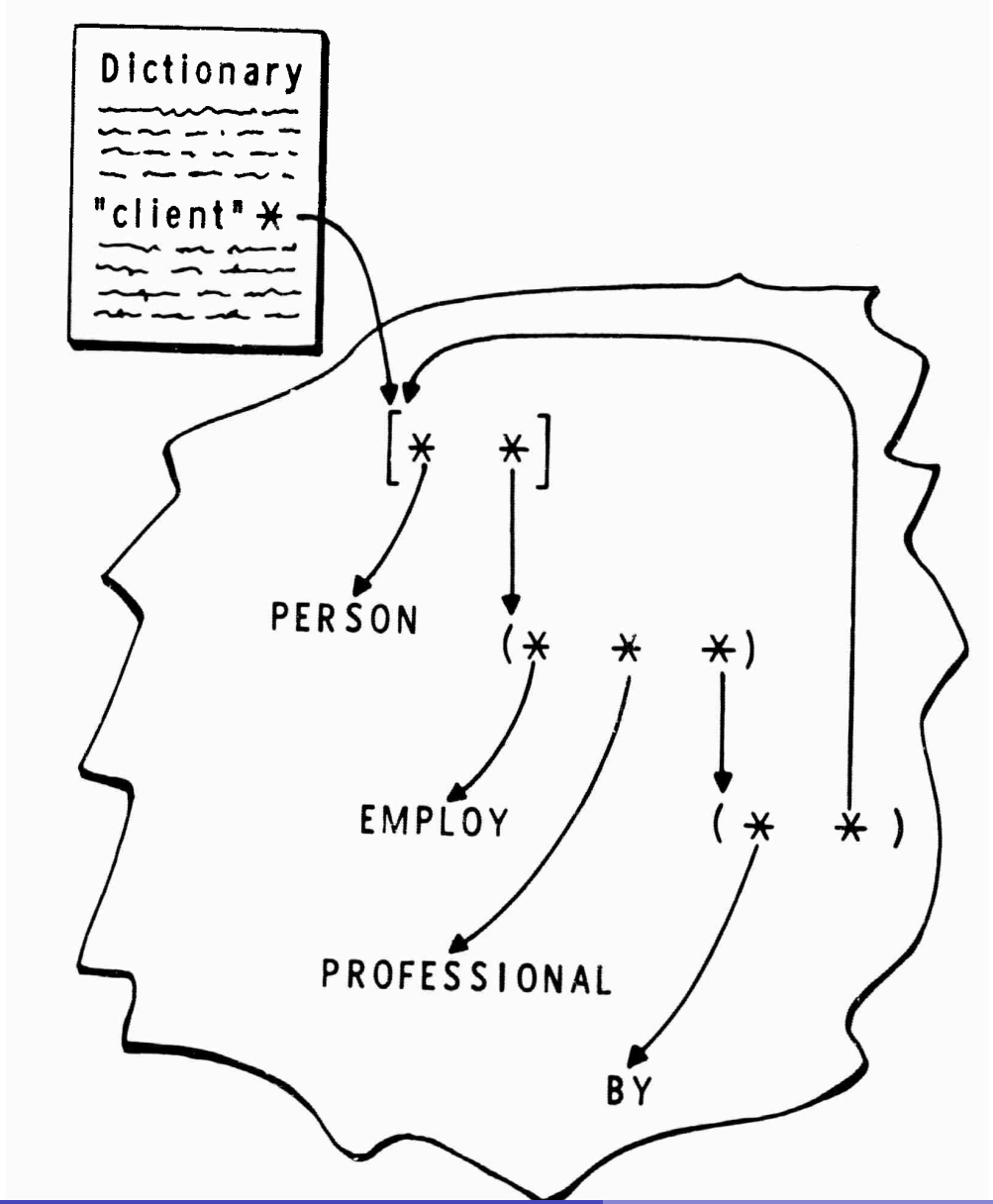
Classic KR



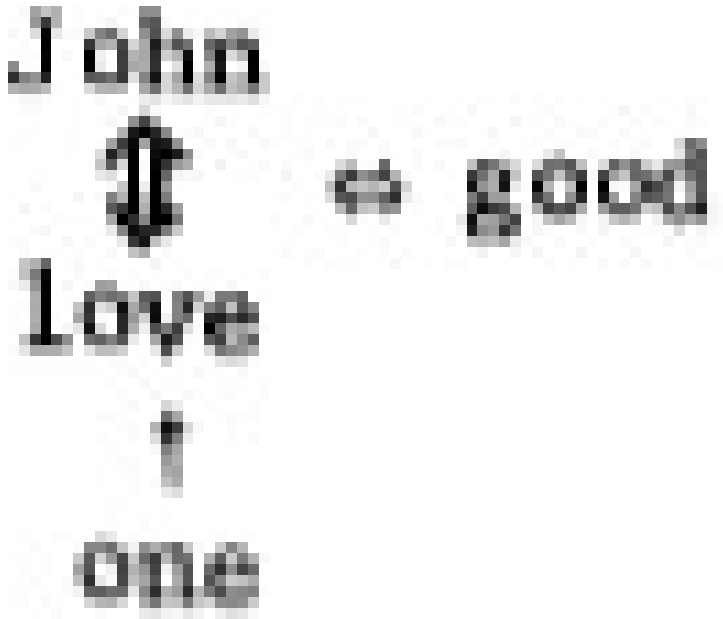
"A MESSAGE is, among other things, a THING with at least one Sender, all of which are PERSONs, at lease one Recipient, all of which are PERSONs, a Body, which is a TEXT, a SendDate, which is a DATE, and a ReceivedDate, which is a DATE."

Quillian, Schank

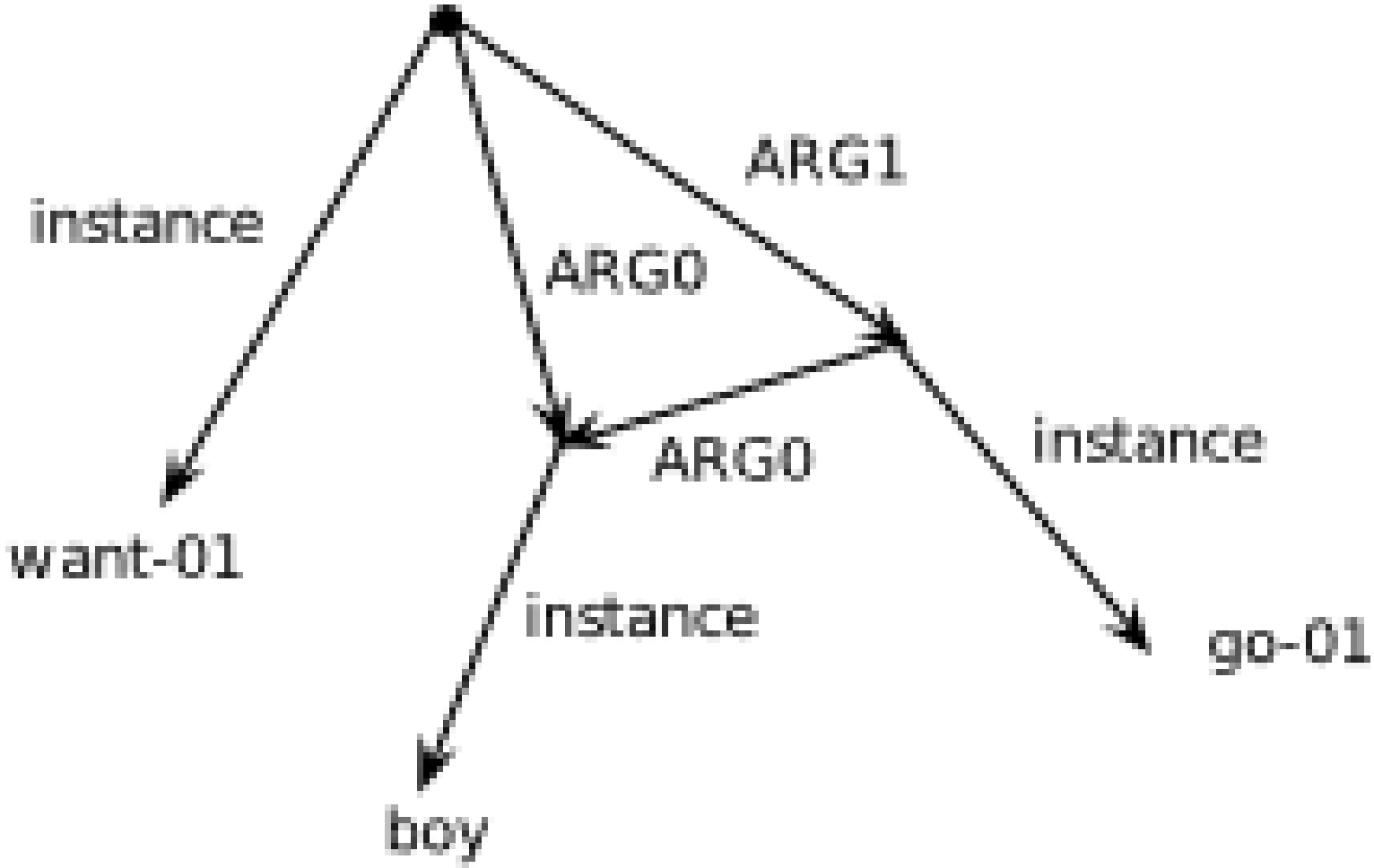
Semantic Memory



Conceptual Dependency



AMR graphs

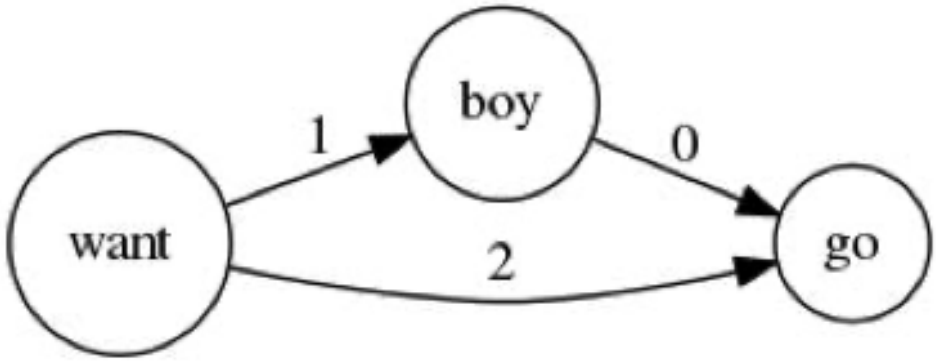


The boy wants to go

AMR graphs *cont'd*

- Rooted, directed, edge- and leaf-labeled graphs
- $\sim$ 100 relations: :accompanier, :age, :beneficiary, :cause, :compared-to, :concession, :condition, :consist-of, :degree, :destination, :direction, :domain, :duration, :employed-by, :example, :extent, :frequency, :instrument, :li, :location, :manner, :medium, :mod, :mode, :name, :part, :path, :polarity, :poss, :purpose, :source, :subevent, :subset, :time, :topic, :value, :quant, :unit, :scale, :day, :month, :year, :weekday, :time, :timezone, :quarter, :dayperiod, :season, :year2, :decade, :century, :calendar, :era
- Neo-Davidsonian graph nodes for entities, events, properties, and states.
- Standardized AMR-parsed corpora (SemBanks) exist for English (60k sentences) and Chinese (5k)

4lang graphs



The boy wants to go

- Have three kinds of links: 0 (is/is_a); 1 (subject); 2 (object)
- In contrast, Cyc has over 45.000 link types, and contemporary efforts like DBpedia or YAGO have $10^5 - 10^6$. The vast majority of these are like *isSpouseOf*, obviously compositional
- `4lang` graphs can be built on RDF-like "triple stores", explicitly addressing known difficulties with these such as **negation, quantifier scope, nested modals** and relations of seemingly **higher arity** *LA is between San Diego and SF along US101*
- Effort to provide semantics for the entire vocabulary

Machine learning on one slide

- Strict separation (typically 80-10-10) of *train*, *dev* and *test* data
- Train is used for building the model, dev for finetuning, test typically hidden from the model builder
- A model optimizes some figure of merit (e.g. word error rate in speech recognition)
- Strong culture of shared tasks (teams working on the same data)
- Generally requires large datasets (gigaword is now typical)
- Supervised methods rule – unsupervised learning still in its infancy
- `aclweb.org/aclwiki/POS_Tagging_(StateOfTheArt)` (see also Manning, 2011)
- For a recent summary see Bengio, LeCun, and Hinton, 2021

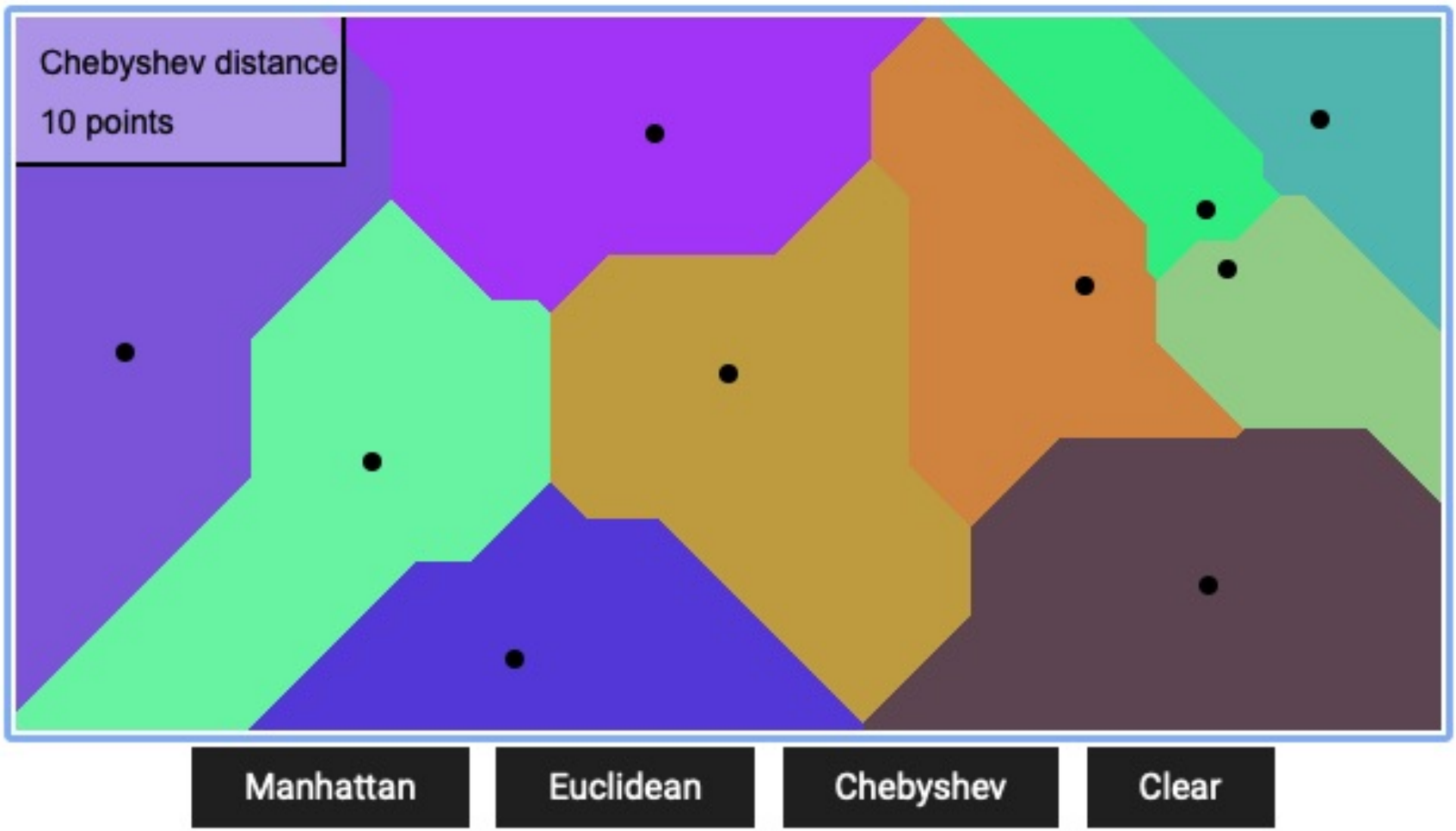
The body: continuous vector space (CVS) semantics

Embedding (static)

Given a dictionary D, a static embeddig is a function $\vec{v}$ that assigns for each word $w \in D$ a vector $\vec{v}(w) \in \mathbb{R}^n$

- First computational treatment by Schütze, 1993 (but goes back to Firth, 1957)
- First implementation that really worked (Bengio et al., 2003)
- NLP "almost from scratch" POS, CHUNK, NER, role labeling (Collobert et al., 2011)
- Has linear structure (king–queen=man–woman) (Mikolov, Yih, and Zweig, 2013)
- Why? (Pennington, Socher, and Manning, 2014; Arora et al., 2015; Gittens, Achlioptas, and Mahoney, 2017)

Voronoi diagram



http://yunzhishi.github.io/voronoi.html

Voronoids

Definition

A *voronoid* $V = \langle \mathcal{P}, P \rangle$ is a pairwise disjoint set of polytopes $\mathcal{P} = \{P_i\}$ in $\mathbb{R}^n$ together with exactly one point p_i in the inside of each P_i.

1. Voronoi diagrams are used in psychological classification (Gärdenfors, 2000). Voronoids are more general, no requirement that
2. the p_i to be at the center of the P_i
3. the facets of the polytopes to lie equidistant from to labelled points
4. the union of the P_i to cover the space almost everywhere – there can be entire regions missing (not containing a distinguished point)

PAC learning + sparsity objective

Linearity

A *linear voronoid* is a voronoid defined by hyperplanes h_j such that every facet of every polytope lies in one of these.

PAC learning

Each concept c corresponds to a probability distribution π_c over $\mathbb{R}^n$

(A concept like *candle* is associated to to other verbal descriptors 'cylindrical, has a wick at the axis, is made of wax, used on festive occasions' and to nonverbal ones, such as a picture of 'the candle' or even the characteristic smell of burning candles.)
We have two objectives: first, to enclose the bulk of each concept set c in some P_i so that $\pi_c(P_i)$ is sufficiently close to 1, and second, to reduce the cardinality of the hyperplane set.
More on word vectors next time

The feet: automata theory

- Operational semantics a la Plotkin/Hennessy "small step" will not be discussed
- This has more to do with the limitations of my understanding than with unworthiness of the approach
- FSTs may get a mention as they are excellent for morphophonological computation
- Eilenberg machines will not be discussed (but see S19:5.8,6.6)
- Will discuss operational aspects for (hyper)graphs and word vectors as we go along
- These are vaguely analogous to "big step" or "natural" semantics a la Kahn, but the analogy will not be exploited

The trunk: cognitive semantics

- Clear linguistic appeal
- Intriguing, but informal, results
- Mainstream formal semantics has nothing to say
- Often insightful, rarely verifiable
- Langacker at the anti-formal extreme, Jackendoff at the formal end, Talmy in between.

Captatio benevolentiae

Textual motivation:

There is in Sullivan's makeup [] an Oxford debater's ready access to the rhetoric of condescending scorn *Jonathan Raban, NYRB 4/12/07*

- What is the extension E (or intension I) of *Sullivan's makeup*?
- Who is *an Oxford debater*?
- Can the *rhetoric of condescending scorn* be analyzed as the genitive of material (just like *a bar of gold*)?
- These concerns are anything but new, see `McCarthy (1976)`

We may not have a full understanding of the relation *x has ready access to y*, but we do know that having ready access to something means that the possessor can deploy it swiftly and with little effort. What the sentence means is simply that Raban finds Sullivan capable of doing so, in fact as capable as those highly skilled in the style of debate practiced at the Oxford Union where condescension and scorn are approved, even appreciated, rhetorical tools.

Long term goals

- characteristica universalis – progress is being made
- calculus ratiocinator – not quite there, especially painful gap in formalizing natural language arguments the way we can formalize mathematical arguments
- Central takeaway from first lecture: Word meaning matters

Thank you!

Lecture and supporting materials available at
http://kornai.com/2021/ESSLLI

Second lecture: vectors
Third lecture: (hyper)graphs, lexicon, non-compositionality, morphology with vectors
Fourth lecture: spatial and temporal semantics, coercion, indexicals
Fifth lecture: negation, modality, probability, implicature
Possible reading: Kornai *Vector Semantics* book draft
https://kornai.com/Drafts/advsem.pdf

Aronoff, Mark (1985). "Orthography and Linguistic Theory: The Syntactic Basis of Masoretic Hebrew Punctuation". In: *Language* 61.1, pp. 28–72.

Arora, Sanjeev et al. (2015). "Random Walks on Context Spaces: Towards an Explanation of the Mysteries of Semantic Word Embeddings". In: *arXiv:1502.03520v1* 4, pp. 385–399.

Banarescu, Laura et al. (2013). "Abstract Meaning Representation for Sembanking". In: *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 178–186. URL: https://www.aclweb.org/anthology/W13-2322.

Bengio, Yoshua, Yann LeCun, and Geoffrey Hinton (2021). "Deep Learning for AI". In: 64.7, pp. 58–65. DOI: 10.1145/3448250.

Bengio, Yoshua et al. (2003). "A Neural Probabilistic Language Model". In: *Journal of Machine Learning Research* 3, pp. 1137–1155. URL: http://www.jmlr.org/papers/v3/bengio03a.html.

Brown, P. et al. (1992). "An Estimate of an Upper Bound for the Entropy of English". In: *Computational Linguistics* **18**/1, pp. 31–40.

Collobert, R. et al. (2011). "Natural Language Processing (Almost) from Scratch". In: *Journal of Machine Learning Research (JMLR)*.

Fernando, Tim (2018). "Finite-state Methods Featuring Semantics". In: *Tokens of Meaning: Papers in Honor of Lauri Karttunen*. Ed. by Cleo Condoravdi and Tracy Holloway King, pp. 525–542. URL: http://kornai.com/Drafts/dare.pdf.

Firth, John R. (1957). "A synopsis of linguistic theory". In: *Studies in linguistic analysis*. Blackwell, pp. 1–32.

Gärdenfors, Peter (2000). *Conceptual Spaces: The Geometry of Thought*. MIT Press.

Gittens, Alex, Dimitris Achlioptas, and Michael W. Mahoney (2017). "Skip-Gram – Zipf + Uniform = Vector Additivity". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, pp. 69–76.

DOI: 10.18653/v1/P17-1007. URL: http://aclweb.org/anthology/P17-1007.

Jacobson, Pauline (2014). *Compositional Semantics*. Oxford University Press.

Johnstone, Peter T. (2002). *Sketches of an Elephant: A Topos Theory Compendium*. Oxford University Press.

Kornai, András (2010). "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.

Kracht, Marcus (2011). *Interpreted Languages and Compositionality*. Vol. 89. Studies in Linguistics and Philosophy. Berlin: Springer.

Manning, Christopher D. (2011). "Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?" In: *Computational Linguistics and Intelligent Text Processing*. Ed. by Alexander F. Gelbukh. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 171–189. ISBN: 978-3-642-19400-9. DOI: 10.1007/978-3-642-19400-9_14.

McCarthy, John J. (2005). *Human-level AI is harder than it seemed in 1955*. URL: http://www-formal.stanford.edu/%20jmc/slides/wrong/wrong-sli/wrong-sli.html.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig (2013). "Linguistic Regularities in Continuous Space Word Representations". In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*. Atlanta, Georgia: Association for Computational Linguistics, pp. 746–751.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (2014). "Glove: Global Vectors for Word Representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543. DOI: 10.3115/v1/D14-1162. URL: http://www.aclweb.org/anthology/D14-1162.

Russell, Bertrand (1940). *An Inquiry into Meaning and Truth*. Allen and Unwin.

Schütze, Hinrich (1993). "Word Space". In: *Advances in Neural Information Processing Systems 5*. Ed. by SJ Hanson, JD Cowan, and CL Giles. Morgan Kaufmann, pp. 895–902.

Tarski, Alfred and Leopold Blaustein (1936). "Der Wahrheitsbegriff in den formalisierten Sprachen". In: *Studia philosophica*, pp. 261–405.