Unifying formulaic, geometric, and algebraic theories of semantics

András Kornai
SZTAKI Computer Science Research Institute

ESSLLI, August 4 2021


ตายอด
คลำช้าง

Outline

Intro

- In Lecture 1 we outlined five classes of models, and in Lecture 2 we discussed one of the five in detail, word vectors.
- Today we discuss (hyper)graph-based theories, what Kornai and Kracht, 2015 called "Algebraic Conceptual Representations" but we can simply call "the ear". Emphasis on current theories (NSM, AMR, UD, 4lang) at the expense of classic "GOFAI" systems (Quillian, Schank, Woods, . . .)
- Why the ear? Thoughts and concepts are things in the head. With proper instrumentation (a biological exception) they can be **heard**.

Conjoined twins Krista and Tatiana Hogan really do have a different outlook on life – they see through **each other's** eyes. The four-year-olds have a **conjoined** section of the brain, allowing them to **hear each other's thoughts** and see through **each other's** eyes. Nov 21, 2010

- Remember, our goal is to model concepts (things in the head)

Motivation

- The bulk of the information in a sentence is carried by the lexicon not by the logical structure: numerically about 85-95%.
- Cannot perform textual inference without taking word meaning into account. *Reading as reasoning* (Thorndike 1917) *NL understanding* (McCarthy 1976)
- Need to formalize arguments based on the meaning of the predicates. *God can't create a mountain without creating a valley*.
- Excellent semi-formal material (Jackendoff, Fauconnier, Lakoff, Langacker, Talmy, Wierzbicka, etc) cries out for formalization.

Are the nodes words?

- Our whipping boy will be Fodor, 1998, who is very lucid about points we diametrically oppose, e.g. that words are atomic
- Fodor is situated in philosophy and psychology. We only want to do linguistic semantics here, an undertaking that Fodor declares impossible "English has no semantics".
- We will have subword units, and a fair amount of semantics that goes beyond "well, it's noncompositional"
- We will also have definitions of the kind used in lexicography, except more formal
- We can't do cogsci, but the ACR approach is quite consonant with modern fMRI work like Fedorenko et al., 2020 that goes against "modularity"

Why algebra?

The generative concept of the lexicon implies a set of primitives, the algebraic concept developed here does not. Consider multiplication in Z_3:

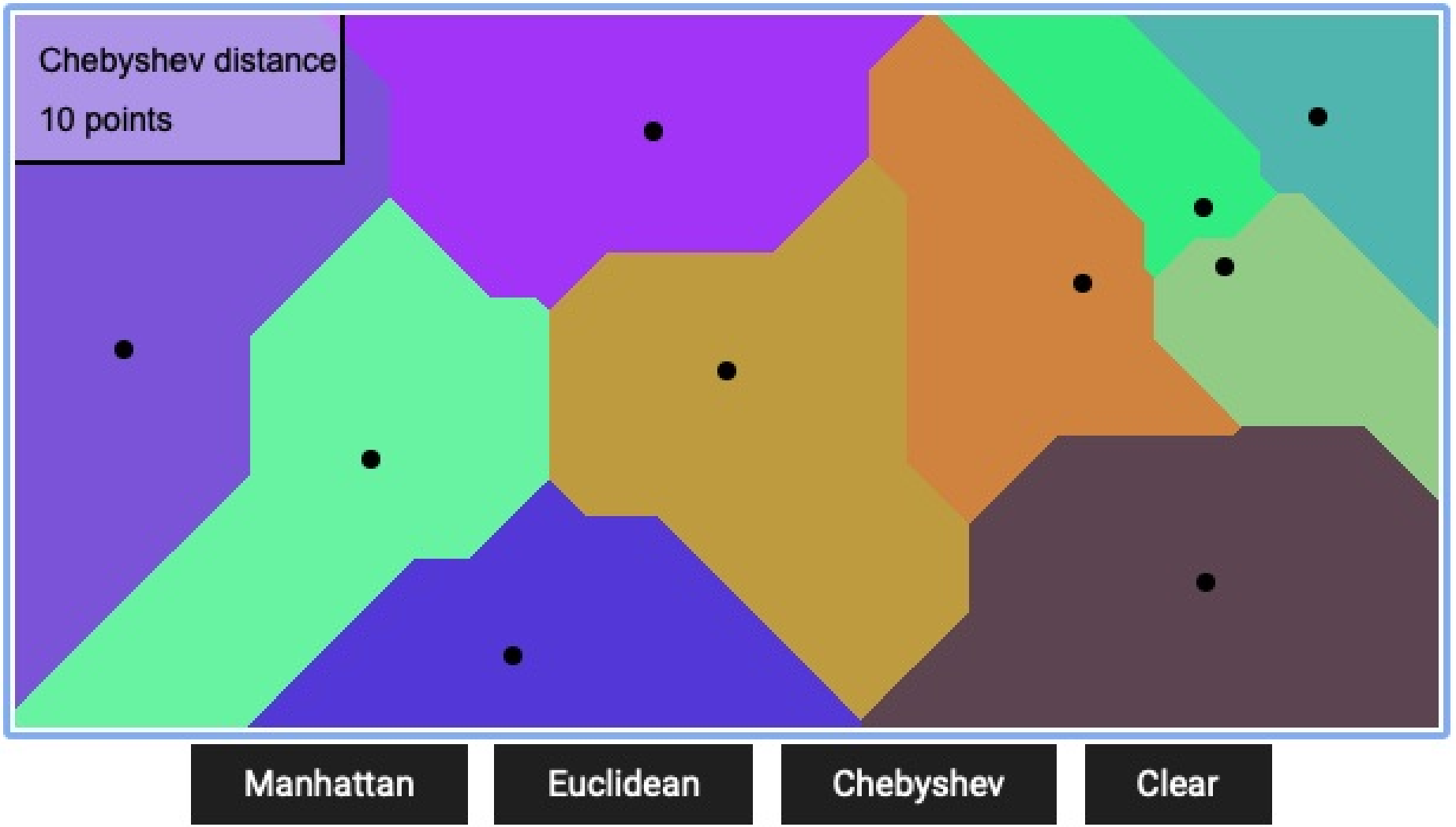
	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

- *unique* does not mean *irreducible*
- one generator (a) is just as good as the other (b) – there is no unique/distinguished primitive
- Similar examples abound: for example in a linear space any basis is just as good as any other to define all vectors in the space.

Examples from the lexicon

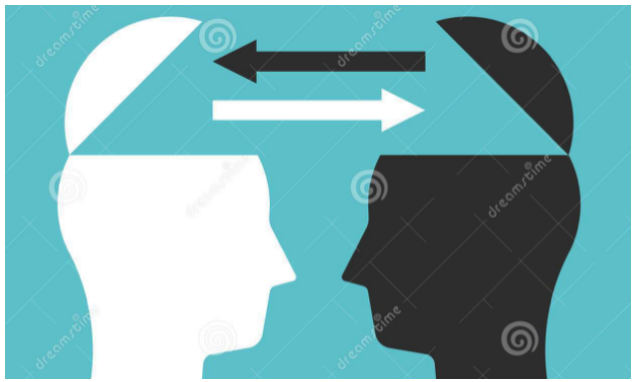
- Consider the Hungarian verbal stem *toj* and the derived *tojó* 'hen', *tojás* 'egg', and *tojni* 'to lay egg'. It is evident that eggs are what hens lay, hens are what lay eggs, and laying of eggs is what hens do.
- Now consider *prison* 'place where inmates are kept by guards', *guard* 'person who keeps inmates in prison', and *inmate* 'person who is kept in prison by guards'.
- Finally consider the imaginary language where prison guards are called *keepers*, inmates *keepees*, and the prison itself a *keep*.

Voronoids


Chebyshev distance
10 points
Manhattan
Euclidean
Chebyshev
Clear

A very rough theory of communication

- Discrete system of signs (nodes) embedded in regions
- Explains "telementation" (person A has a thought, says something, now person B has reasonably similar thought)



- Remember word vectors just a tiny subspace of thought vectors
- What we care about is the projection of Ψ on L
- Full vector may encode sensory stuff, e.g. prototypical images Rosch, 1975 Lakoff, 1987, Gärdenfors, 2000

Communication by voronoids

If word vectors are the distinguished points of voronoids, we have a simple theory of communication.

Label function

The inverse of the embedding $\vec{v}$ is a function $l : \mathbb{R}^n \to D^*$ that is lifted to the entire polytope surrounding a word vector

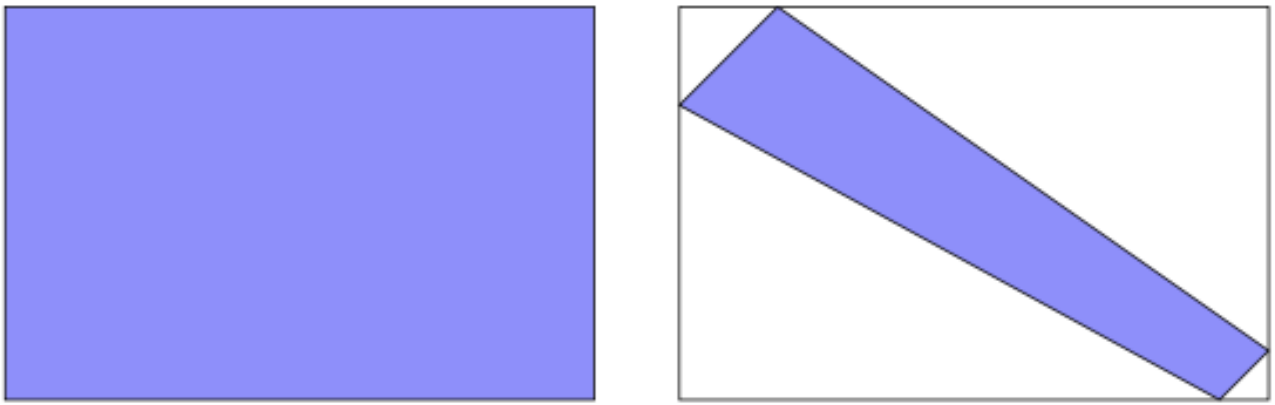
If speaker and hearer have similar voronoids, simple ideas or sensations can be communicated by uttering the label of the polytope where it falls: I see a candle, and say *candle*. This is sufficient for the hearer to know which polytope was meant, and thereby gain some rough understanding of my mental activity. This is not because our P_{candle} polytopes have identical boundaries, but rather because the boundaries cover so much of the π(candle) probability mass that the symmetric difference between the polytopes of speaker and hearer is negligible.

Characteristics of algebraic description

- Defining equations are quite often redexes – there is no guarantee of normal form. *kill* $\rightarrow$ `cause to die` $\rightarrow$ `cause not to live` $\rightarrow$ `cause not to have life functions` $\rightarrow$ `cause not to metabolize, grow, multiply` $\rightarrow$ ...
- Typical decompositions are subdirect – no *eidopoios diaphora.* Example: *ropeladder* 'ladder made of rope'; *manslaughter* 'slaughter undergone by man'; *testtube* is 'tube used for test', so the overall semantics can only specify that $N_1 N_2$ is 'N_2 that is V-ed by N_1'
- Definitions are typically monosemic (no 'metaphoric usage'), incredibly simple, politically incorrect/conservative, neutral across lexical categories *divorce, secure, ...*

Nodes are morphemes

- Recall that mouse:mice::ox:oxen works for word vectors, so there is a vector we can reasonably call the vector for the plural morpheme.
- But if *ropeladder* means 'ladder made of rope', why doesn't *paperback* mean 'back made of paper'?
- In algebraic theories we can use subdirect decomposition:



(a) Direct product

(b) Subdirect product

Edges require arity reduction

kill means `cause to die`: *x kill y* means CAUSE(x, `die`(y)). Fodor's (1971) critique relied on naive string substitution, this doesn't survive the advent of binding theory. His main problem, that *die* itself can be further decomposed as `become not alive` is a non-issue for the algebraic approach.

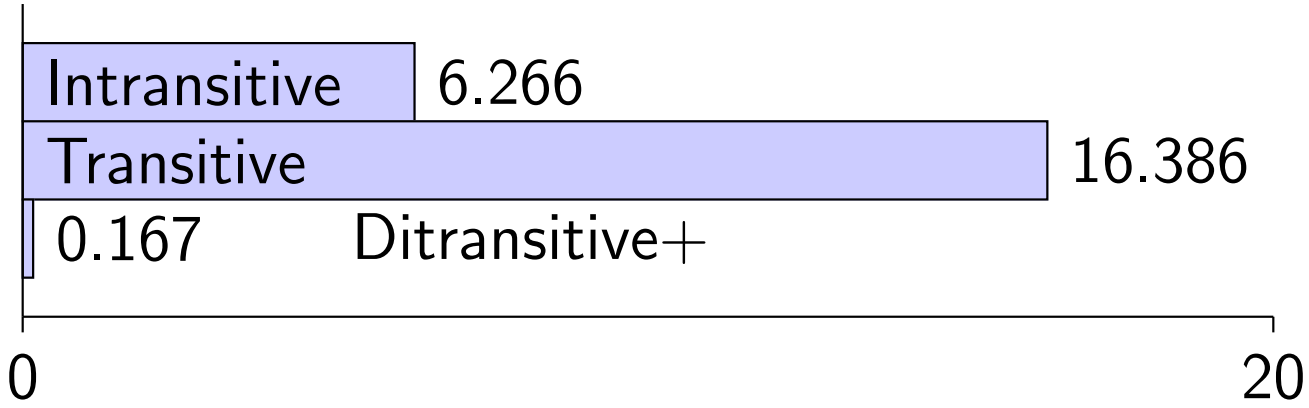
Notation: PRED(x, y) is treated as a pair of edges from x to `pred` and from `pred` to y, labeled '1' (subject) and '2' (object) respectively. In graphs we also permit '0' (`is_a` links), but these are abbreviatory devices, ideally computable form the definition of words.

We eliminate ditransitives (indirect objects, instruments, etc.) entirely. For example, *give(x,y,z)* is analyzed as CAUSE(x,HAS(z,y)) 'cause to have'.

The idea gave him the shivers the shivers were not originally in the idea's possession

Mary gave him typhoid Mary didn't cease to have typhoid just by giving it to him.

Is arity reduction a big problem?



If one looks at the verbs deemed ditransitive or even higher arity, one sees all the classic decomposition cases going back to the 1970s (Generative Semantics) and many that don't even require that much effort

John hardboiled an egg for Mary `John hardboil egg, egg for Mary`

Key takeaway: no **edge** labeled by *hardboil* or *for*. These are **nodes** with their own subjects and (prepositional) objects.

Lexicography: desiderata

- *Lexemes* reconstructing the intensional meaning (as bundles of defeasible properties) `dog kutya canis pies 1465 e N animal, has leg(four), bark, bite, faithful`
- Homogeneous formalization of *entire* lexical entry (can't be based on examples)
- Reconstruction of variable binding without variables or VBTOs – for graphs we use unification, for vectors we use equalizers
- Reconstruction of `is_a` links without explicit provision for them
- Reconstruction of analytic/synthetic division
  - Synthetic: *calcium carbonate is used for cleaning white gloves* True, but uninteresting (belongs in encyclopedia not in lexicon)
  - Analytic: *friendly dogs don't bite* Can be derived from meaning of *friendly* and *bite*

Lexical entry composed of form and meaning

vertigo 'a peculiar dizzy sensation caused by great heights' (1950s) 'often caused by damage or disease in the inner ear' (American Heritage, current)

Definition in `4lang`: `dizzy, height cause, disease cause`

For a computational dictionary, we have three major desiderata: **Universality, Reducability**, and **No encyclopedic knowledge** (see VS 1.3)

Ignore encyclopedic knowledge (see next slide), ignore subtle distinctions among forms of dizziness *unless proven relevant for naive reasoning*, factor inheritable knowledge (e.g. that it's a `feeling` or `sensation` – this will be stored at the entry of `dizzy`. Also ignore *abnormal* because it follows from `disease cause`


Patient.co.uk
Trusted medical information and support
Home
Diseases & Conditions
Medicines & Drugs
Professional Reference
Discussion Fo
Like
Tweet 0
+1 0
Home > Professional Reference > Vertigo
Vertigo
This PatientPlus article is written for healthcare professionals so the language may be more technical than the condition leaflets. You may find the abbreviations list helpful.
On this page
Causes
Epidemiology
Presentation
Differential diagnosis
Investigations
Management
Complications
Prognosis
Document references
Internet and further reading
Vertigo is a type of dizziness and involves a false sensation that oneself or the surroundings are moving or spinning, usually accompanied by nausea and loss of balance. Causes of vertigo are often differentiated into:
Central (cerebral cortex, cerebellum, brainstem), e.g. cerebrovascular disease, migraine, multiple sclerosis, acoustic neuroma, diplopia, alcohol intoxication.
Peripheral (vestibular labyrinth, semicircular canals or vestibular nerve), e.g. viral labyrinthitis, vestibular neuritis, benign paroxysmal positional vertigo (BPPV), Ménière's disease, motion sickness, ototoxicity (e.g. gentamicin), Herpes zoster (Ramsay Hunt syndrome)
Causes
See also
Benign Paroxysmal Positional Vertigo
Dizziness
See 6 more »
In more detail
Dizziness, Giddiness and Feeling Faint
Labyrinthitis

Which part of *is* don't you understand?

1. used to say something about a person, thing, or state, to show a permanent or temporary quality, state, job, etc. *He is rich. It's cold today. I'm Andy. That's all for now. What do you want to be (= what job do you want to do) when you grow up? These books are (= cost) $3 each. Being afraid of the dark, she always slept with the light on. Never having been sick himself, he wasn't a sympathetic listener. Be quiet! The problem is deciding what to do. The hardest part will be to find a replacement. The general feeling is that she should be asked to leave. It's not that I don't like her - it's just that we rarely agree on anything!*
2. used to show the position of a person or thing in space or time *The food was already on the table. Is anyone there? The meeting is now (= will happen) next Tuesday. There's a hair in my soup.*
3. used to show what something is made of *Is this plate pure gold? Don't be so cheeky! Our lawyers have advised that the costs could be enormous. You have to go to college for a lot of years if you want to be a doctor. Come along - we don't want to be late! Oranges, lemons, limes and grapefruit are types of citrus fruit.*
4. used to say that someone should or must do something *You're to sit in the corner and keep quiet. Their mother said they were not to (= not allowed to) play near the river. There's no money left - what are we to do?*
5. used to show that something will happen in the future *We are to (= we are going to) visit Australia in the spring. She was never to see (= she never saw) her brother again.*
6. used in conditional sentences to say what might happen *If I were to refuse they'd be very annoyed. (formal) Were I to refuse they'd be very annoyed.*
7. used to say what can happen *The exhibition of modern prints is currently to be seen at the City Gallery.*
8. to exist or live *(formal) Such terrible suffering should never be. (old use or literary) By the time the letter reached them their sister had ceased to be (= had died).*

https://dictionary.cambridge.org/dictionary/english/be

LDOCE *be*

1. used with past participles to form the
2. used in sentences about an imagined situation
3. used in sentences to introduce an aim when you are saying what must be done in order to achieve it
4. used to say that someone or something is the same as the subject of the sentence
5. used to say where something or someone is
6. used to say when something happens
7. used to describe someone or something, or say what group or type they belong to
8. to behave in a particular way
9. used to say how old someone is
10. used to say who something belongs to
11. used to talk about the price of something
12. to be equal to a particular number or amount
13. to exist

LDOCE *be+*

This is the fun part: 1,429 expressions from *be taken aback (by sth)* to *be yours for the taking/asking*. Most of these appear to be perfectly compositional *be able to do sth* 'to have the skill, strength, knowledge etc needed to do something'; *be wrong (about sb/sth)* 'to not be right in what you think or believe about someone or something'

The difficulties predate Webster's 3rd: *Webster's New World* (going back to 1951) uses even more vague terms in the definition, such as 'used to express futurity, possibility, obligation, intention, etc'; *The Concise Oxford* (1911) has, distributed among several senses, 'exist, occur, live, remain, continue, occupy such a position, experience such a condition, have gone to such a place, busy oneself so, hold such a view, be bound for such a place, belong under such a description, coincide in identity with, amount to, cost, signify'.

Collins *be*

1. to have presence in the realm of perceived reality, exist, live *not all that is can be understood*
2. to pay a visit, go *have you been to Spain?*
3. to take place, occur *my birthday was last Thursday*
4. used as a linking verb between the subject of a sentence and its noun or adjective complement or complementing phrase. Has no intrinsic meaning of its own but rather expresses the relationship of either essential or incidental equivalence or identity or to specify an essential or incidental attribute. It is also used with an adverbial complement to indicate a relationship of location in space or time

A more unified treatment seems warranted

Formulas in the 4lang dictionary

1. Resemble dictionary definitions, designed for human readability
2. Conjunctive top-level, avg 2.7 conjuncts (max 9 found so far) `camera kamera cinematographica_machinula kamera 1221 N machine, make photograph, has lens`
3. `make =agt cause [=pat[exist]]`
4. Syntactic sugar: `x is_a y` can be written as `x[y]` or as `y(x)`
5. Defaults are in < >, grouping by { }
6. Only one quantifier: `gen`, treated as a noun (no scope). Two thematic roles, no ternary relations (Kornai, 2012)
7. < 20 primitives, < 800 core (uroboros) vocabulary elements
8. Check out `https://github.com/kornai/4lang/tree/master/Reform` for current state (Release V2 expected together with publication of the Vector Semantics book) Pull requests welcome

Refresher from Tuesday

1. Thought vectors $\Psi(t)$ are very large (dim 10^{11}) and they wander around the hypercube on millisecond timescale
2. We are only interested in the part Ψ that falls in the persistent, linguistic subspace. Here we have a larger (centisecond or beyond) timescale
3. This will contain both word vectors and vectors for representing linguistic knowledge
4. Word vectors will be viewed in a sparse overcomplete basis $\vec{p_1}, \ldots, \vec{p_d}$ (dim $d \approx 300 - 800$) normal to the hyperplanes that bound the polytopes in the voronoid
5. For a vector $\vec{x}$ to fall in the halfspace given by $\vec{p_i}$ means $\langle \vec{x}, \vec{p_i} \rangle > 0$

Smolensky's approach

Smolensky, 1990: whether we do formulaic or algebraic semantics, we are doing symbol manipulation on discrete objects. In the formulaic case we need something like λ-calculus, in the geometric case we need (hyper)graph unification, again something that needs variable binding. In a recurrent net, this can be done by assigning tensors to neurons.

So why is nobody doing this? Because this eats up nodes at a phenomenal rate: for k-tensors we need d^k nodes. With $d = 256, k = 5$ we would have 1.1e12 parameters to train.

We solve this by defending $k = 2$ Word vectors will be the ordinary word vectors, linguistic knowledge will be represented by these and by some d by d matrices. In addition to the word space L given by the word vectors statically, we will use a representation space $R \approx L \times L$ that gets updated based on linguistic (and perhaps sensory) input. We are keeping linear and quadratic terms, making do without cubic and higher.

The easy parts

1. Top (conjunctive) level: intersection of polytopes. Disjunction corresponds to union of polytopes, but negation does *not* correspond to complementation
2. The lone quantifier gen: $(1/n, \ldots, 1/n)$
3. Untyped treatment of A and N: halfspaces, more complex polytopes
4. Smooth transition from lexical to compositional semantics
5. Literal and metaphorical usage
6. Proper nouns get their due
7. Treatment of synonymy

The key idea for *be*

1. By definition, `carbon is_a material`. We want $\vec{v}(\text{carbon}) \in P_{material}$ i.e. the point (vector) for `carbon` to be inside the polytope that defines `material`, or we want the entire `carbon` polytope to lie in the `material` polytope. Generally, we have $I(A)$ is_a $I(B)$ translated by set-theoretic containment.
2. It is not true by definition that *John sleeps*. The entire thought vector whose persistent (linguistic) part is $\vec{v}(\text{John})$ will be wandering around, sometimes falling in the $\vec{v}(\text{sleep})$ polytope sometimes not, $\langle \vec{v}(\text{John}), \vec{v}(\text{sleep}) \rangle = 0$. But when we say *John sleeps* the listener fixes the *John* vector inside the *sleep* polytope.
3. But if $\vec{v}(\text{John})$ and $\vec{v}(\text{sleep})$ are fixed, how can we do this? *We change the scalar product!*

Intransitive predication

In predication, the extensions (vectors, polytopes) of the subject and the predicate remain unchanged. But the component of the matrix P that falls in the linguistic subspace is modified, to encode the fact that is being predicated. Since initially $P = \lambda I$ (remember all persistent vectors belong to the eigenspace defined by the highest eigenvalues), we must update this by

$$P_R(t+1) = \lambda I + s|\rangle p_i, p_j\langle| \tag{1}$$

where s is some positive scaling factor that we use in perturbing the previous matrix, i is the coordinate of *John*, j is the coordinate of *sleep*, and $|\rangle p_i, p_j\langle|$ is the rank 1 matrix that is the outer (Gram) product of the two vectors.
(Note: it is well known that small perturbations leave the eigenvectors invariant, and only change the eigenvalues.)

be, TRANSITIVE PREDICATION

The meaning of *be* is therefore

$$P_R(t+1) = P_R(t) + s|\rangle{=}\texttt{AGT}, {=}\texttt{PAT}\langle| \tag{2}$$

In a critical sense, *be* doesn't contribute much (time to go back and check all the meanings we listed at the beginning) but other transitive verbs do: for these we have

$$P_R(t+1) = P_R(t) + s|\rangle{=}\texttt{AGT}, ({=}\texttt{PAT} + \vec{v})\langle| \tag{3}$$

where $\vec{v}$ is the word vector corresponding to the transitive verb in question. Note that intransitives smoothly extend to transitives: in the translation of *John eats fish* the meaning of `fish-eat` is subset of the meaning of `eat`.

Joint work with Marcus Kracht

Compositionality in vector semantics

Several combination operations have been proposed for vectors, including vector addition Mitchell and Lapata, 2008, coordinatewise (weighted) multiplication Dinu and Lapata, 2010, function application Coecke, Sadrzadeh, and Clark, 2010 and substitution into recurrent neural nets Socher et al., 2013a. We use $\otimes$ to denote any composition operation (most factor through tensor products).
A key point is that $\otimes$ itself may be parametrized, more similar to the 'type-driven' versions of MG Klein and Sag, 1985. Context Vector Grammars (see Socher et al., 2013b) use what looks like a single binary function $\otimes$, however it is parametrized by part of speech. These grammars work on ordered pairs $(\vec{v}, X)$ where $\vec{v}$ contributes the semantics, and X is some part of speech category (including nonterminals such as NP). In our notation $(\vec{v}, X)$ combines with $(\vec{w}, Y)$ by two square matrices L_{XY}, R_{XY} and a bias $\vec{b}_{XY}$ that depend on X and Y (but not on $\vec{v}$ or $\vec{w}$) to yield $\vec{v} \otimes \vec{w} = \tanh(L\vec{v} + R\vec{w} + \vec{b})$

Simplifying $\otimes$

Since tanh is strictly monotonic, we have $x = y$ iff $\tanh(x) = \tanh(y)$, so the last step of squishing can be ignored in the kind of equational deduction that we will deal with. As an example, consider the `gram3-comparative` task. The semantics should support equations such as $\vec{big} \otimes \vec{er} - \vec{nice} \otimes \vec{er} = \vec{big} - \vec{nice}$ (being an analogy) or, equivalently,
$\tanh(L\vec{big} + R\vec{er} + \vec{b}) - \tanh(L\vec{nice} + R\vec{er} + \vec{b}) = \vec{big} - \vec{nice}$

In reality both the matrix and the vector coefficients are small enough for $\tanh(x) = x$ to be a reasonable approximation, so we have

$$L\vec{big} - L\vec{nice} = \vec{big} - \vec{nice} \tag{4}$$

or, what is the same, $(L - I)(\vec{big} - \vec{nice}) = 0$ not just for *big* and *nice* but for every pair of adjective vectors $\vec{u}, \vec{v}$. This is possible only if $\langle A \rangle$, the subspace generated by the adjectives, is contained in $\mathrm{Ker}(L - I)$.

Simplification of $\otimes$ cont'd

Since L does not even need to be defined outside $\langle A \rangle$, and must coincide with I within $\langle A \rangle$, the simplest assumption is $L = I$ everywhere. Now, R and b are fixed for the comparative task, so $R\vec{er} + \vec{b}$ is some constant vector $\vec{c}$ on $\langle A \rangle$, so that we finally get

$$\forall \vec{x} \in \langle A \rangle : \vec{x} \otimes \vec{er} = \vec{x} + \vec{c} \tag{5}$$

The same argument can be made (with different constant $\vec{c}$) for every derivational and inflexional suffix such as the *-ly* of the `gram1-adjective-to-adverb` or the *-ing* of the `gram5-present-participle` Google task. Further, the same must hold for every case where a fixed formative is used to derive a higher constituent, such as PP[from] from a base NP and a prefix *from*, or NP from a base N and the prefix *the*. Remarkably, just as PP[from] can differ from PP[by] only by a fixed offset, the difference between the constant for *from* and that for *by*, NP[every] and NP[some] can also differ only in a fixed offset irrespective of what the base N was.

Reasoning can be extended to binaries

Consider *the mayor's hat* and *the hat of the mayor* which should get the same vector assigned compositionally through two different routes. If $\vec{m}$ and $\vec{h}$ are the vectors for *mayor* and *hat*, we have some $\vec{m} + \vec{c_1}$ for *the mayor* and $\vec{h} + \vec{c_1}$ for *the hat*. If the *'s* possessive construction is defined by matrices L_1, R_1 and bias $\vec{b_1}$, and the *of*-possessive by $L_2, R_2, \vec{b_2}$, the fact that these mean the same will be expressed, again ignoring the squishing, by
$L_1(\vec{m} + \vec{c_1}) + R_1\vec{h} + \vec{b_1} = L_2(\vec{h} + \vec{c_1}) + R_2(\vec{m} + \vec{c_1}) + \vec{b_2}$
By collecting like terms together, this means
$(L_1 - R_2)\vec{m} + (R_1 - L_2)\vec{h} + \vec{c_4} = \vec{0}$ for some constant $\vec{c_4}$ and for all noun vectors $\vec{m}, \vec{h}$. This of course requires $L_1 = R_2, L_2 = R_1$ and $\vec{c_4} = 0$, meaning that the two constructions differ only in the order they take the possessor and possessed arguments.

Information content of $\otimes$

Looking at the 882 L and R matrices (25 by 25 dimensions) in the CVG instance avaliable as part of the Stanford Dependency Parser, we note that over half (55% for L, 53% for R) of the variance in this set is explained by the first 25 eigenmatrices, so the structure is likely considerably simpler than the full CVG model allows for.

k	no I	I first
1	81.02	
5	82.85	84.59
25	86.88	89.32
50	88.50	90.07
100	89.47	90.24
200	90.08	90.32
882	90.36	90.36

Main takeaways

- Compositionality (direct composition) is the special case of noncompositionality (subdirect composition) when the *differentia specifica* can be neglected. When it can't be reasonably neglected, as in English noun-coun compounding, *you need to add in the extra terms*
- We use hypergraphs, but these only have hypernodes (entire subgraphs considered a single node) not hyperedges (edges touching more than two nodes). We work hard to eliminate hyperedges (higher arity predicates)
- Morphological marking only requires the addition of a constant vector. This makes it commutative, explaining the cross-linguistic variability in the order of affixation.

Thank you!

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

Tomorrow: Spatial and temporal semantics, coercion, indexicals
Friday: negation, modality, probability
Possible reading: Kornai *Vector Semantics* book draft https://kornai.com/Drafts/advsem.pdf Chapters 1 and 2, Kornai and Kracht (2015)

Coecke, Bob, Mehrnoosh Sadrzadeh, and Stephen Clark (2010). "Mathematical Foundations for a Compositional Distributional Model of Meaning". In: *arXiv:1003.4394v1*.

Dinu, Georgiana and Mirella Lapata (2010). "Measuring Distributional Similarity in Context". In: pp. 1162–1172.

Fedorenko, Evelina et al. (2020). "Lack of selectivity for syntax relative to word meanings throughout the language network". In: *Cognition* 203. DOI: https://doi.org/10.1016/j.cognition.2020.104348.

Fodor, Jerry A. (1998). *Concepts*. Clarendon Press.

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

Klein, Ewan and Ivan Sag (1985). "Type-driven translation". In: *Linguistics and Philosophy* 8, pp. 163–201.

Kornai, András (2012). "Eliminating ditransitives". In: *Revised and Selected Papers from the 15th and 16th Formal Grammar Conferences*. Ed. by Ph. de Groote and M-J Nederhof. LNCS 7395. Springer, pp. 243–261.

Kornai, András and Marcus Kracht (July 2015). "Lexical Semantics and Model Theory: Together at Last?" In: *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 14)*. Chicago, IL: Association for Computational Linguistics, pp. 51–61.

Lakoff, George (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press. ISBN: 978-0-226-46803-7.

Mitchell, Jeff and Mirella Lapata (2008). "Vector-based Models of Semantic Composition". In: *Proceedings of ACL-08: HLT*. Columbus, Ohio: Association for Computational Linguistics, pp. 236–244.

Rosch, Eleanor (1975). "Cognitive Representations of Semantic Categories". In: *Journal of Experimental Psychology* 104.3, pp. 192–233.

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

Socher, Richard et al. (2013a). "Parsing with Compositional Vector Grammars". In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 455–465.

Socher, Richard et al. (2013b). "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, pp. 1631–1642.