

FOUNDATIONS OF NON-COMPOSITIONAL SEMANTICS

András Kornai
SZTAKI Computer Science Research Institute

HLT

Which part of *is* don't you understand?
Zoom, May 18 2020

PLAN

- Brief history of 4lang
- The informal lexicon (dictionary) as something of a total mess
- The main classes of semantic representations: formulaic, algebraic, geometric
- Voronoids
- Smolensky's (1990) program
- *be*, subj, obj
- Coming attractions

HISTORY OF 4lang

- Word meaning long studied in traditional semantics
- Intriguing, but informal, results
- Mainstream formal semantics had nothing to say
- Closely tied to morphology
- (Kornai, 2010) builds on computational lexicographic tradition (COBUILD, LDOCE, ...), used Eilenberg machines
- Contributors since include Ács, Borbély, Gyenis, Kovács, Kracht, Makrai, Nemeskey, Vásárhelyi, Zséder
- Recently shifted from machines to word embeddings

WEBSTER'S 3RD be/1

of
ket
alls
jah
ms
ge
ks)
nas
k>
the
res
na-
of
by
ter

bdy abbr boundary

be \ (ˈ) bē, -bi \ vb, past 1st & 3d sing **was** \ (ˈ) wəz, (ˈ) wəz also (ˈ) wōz \ or dial **were** \ see below \ or **war** \ (ˈ) wār \; 2d sing **were** \ wə(r); 1wər, + V 1wər, 1wē, + V " or 1wər, also 1wēr; archaic or Brit (ˈ) w|a(a) (ə) r or |e(ə) r or |a(a) ə or |eə \ or dial & archaic **was** or dial **war** (with you) or archaic **wast** (with thou) \ wəst, (ˈ) wāst also (ˈ) wōst \ or **wert** (with thou) \ wə(r) |t, 1wər |t, (ˈ) wē |t, (ˈ) wē |t, usu |d. + V \ pl **were** or substand & archaic **was** or dial **war**; past subjunctive **were** or substand & archaic **was** or archaic 2d sing **wert** (with thou); past part **been** \ (ˈ) bin, bən; ben (in standard speech more often unstressed or with secondary stress than with primary stress); Brit usu & US sometimes (ˈ) bēn \ or dial **ben** \ (ˈ) ben \ pres part **be-ing** \ 1bēɪŋ, 1bēēŋ, rapid (ˈ) bēŋ \ or dial Brit & archaic **been** \ (ˈ) bēn, 1bēən \ pres 1st sing **am** \ əm, (ˈ) am also (ˈ) aa(ə) m; after "I" often m \ or dial **is** or **be**; 2d sing **are** (with you) \ ə(r), (ˈ) ār, (ˈ) ā(r); after a vowel-final pronoun often r \ or archaic **art** (with thou) \ ə(r) |t, (ˈ) ār |t, (ˈ) ā |t, + V usu |d. \ or dial **be** or **is** (with you) or dial Brit **beest** (with thou) \ (ˈ) bēst, 1bēəst \ or **bist** (with thou) \ (ˈ) bist \ 3d sing **is** \ (ˈ) iz, əz; after vowels & after voiced consonants other than z, zh, & j: often z; after voiceless consonants other than s, sh, & ch: often s \ or dial Brit **be**, pl **are** or substand **is** or dial & archaic **be** or archaic **been** \ the last like past part **BEEN** \ pres subjunctive

WEBSTER'S 3RD *be/2*

archaic been (the last like past part *BEEN* \ pres subjunctive *be* or obs 2d sing *beest* (with *thou*) [ME *been*, fr. OE *bēon*; akin to OHG *bim* am, ON *būa* to live, dwell, Goth *bauan*, L *fui* I have been, *futurus* about to be, *fieri* to become, Gk *phyein* to bring forth, *phynai* to be born, be, Skt *bhavati* he is] vi **1 a** : to equal in meaning : have the same connotation (sense 3) as <God is love> <January is the first month> <let $x \sim 10$ > : represent symbolically <the seven lampstands are the seven churches — Rev 1:20 (RSV)> **b** : to constitute the same idea or object as : have individual identity with <the first person I met *was* my brother> <the pianist himself *was* the composer of the piece> <\$50 *was* all I had> **c** : to constitute the same class as <these three books *are* the authoritative works on the president's life> **d** : to have a meaning that includes or implies the meaning of <fish *are* vertebrates> <red *is* a color> : have a (specified) qualification or characterization <the leaves *are* green> <this book *is* heavy> **e** : to belong as an individual to the class of <the fish you caught *was* a trout> **f** : to belong as a class to the larger class of <some animals with horns and divided hoofs *are* graminivorous animals> — used regularly in senses 1a through 1f as the copula of simple predication **g** : SIGNIFY : amount to <her death *was* nothing to him> **h** : to show oneself as an out-

WEBSTER'S 3RD *be*/3

was a verb
<some animals with horns and divided hoofs *are* graminivorous animals> — used regularly in senses 1a through 1f as the copula of simple predication **g** : SIGNIFY : amount to <her death *was* nothing to him> **h** : to show oneself as an outstanding example of — used with main stress in spoken sentences <the doctor pleased the parents by commenting, “That *is* a baby”> **i** : to constitute genuinely : actualize well the type of <one of the few great elegies which *are* elegies — Douglas Bush> **j** : to seem to consist of : show oneself gripped or dominated by (a feeling) <she *was* all scorn at the proposition> : become completely covered with <road *was* all mud> **2 a** : to exist either absolutely or in relations or under conditions specified : have an objective existence : have reality or actuality : LIVE <Thee, which *wert* and *art* and evermore shalt *be* — Reginald Heber> <I think, therefore I *am*> — often used with *there* <once upon a time *there was* a knight> <*there is* a wreck ahead> **b** : to have, maintain, or occupy a place, situation, or position : show a certain characteristic — often used with a prepositional phrase <the book *is* on the table> <he *was* at ease> **c** : to remain unmolested, unbothered, or uninterrupted — used only in infinitive form <let him ~ ; stop pestering him> **d** : HAPPEN, OCCUR : take place <the concert *was* last night> <where will the meeting ~>

WEBSTER'S 3RD *be*/4

⟨let him ~ ; stop pestering him⟩ **a** : HAPPEN, OCCUR : take place ⟨the concert *was* last night⟩ ⟨where will the meeting ~⟩
e *archaic* : BELONG, PERTAIN ⟨to thine and Albany's issue ~ this perpetual —Shak.⟩ **f** (1) : to come or go : JOURNEY ⟨we will ~ on our way shortly⟩ ⟨have you *been* home since Christmas⟩ (2) : to make a stay : show oneself or be present ⟨they will ~ in town all week⟩ ⟨*was* your sister at the party last night⟩ — not used in the present; use of the past tense followed by *to* ⟨I *was* to town yesterday⟩ often considered nonstandard **g** : to come around in due course often in following a schedule or appointed round — used only in perfect forms ⟨has the postman *been* this morning⟩ **h** *substand* : ACT — used only in the perfect; used as an intensive ⟨see what you have *been* and done⟩ **3** *now dial Brit* : to stand good for expense (as in a treat) ⟨offering to ~ his friend's dinner⟩ ~ *verbal auxiliary* **1** : to undergo an action — used with the past participle of transitive verbs as a passive-voice auxiliary ⟨the money *was* found⟩ ⟨German *is* spoken here⟩ ⟨the house *is being* built⟩ **2** : to perform a continuous action : be supposed to perform a future action — used as the auxiliary of the present participle in the so-called progressive tenses, usu. expressing continuous action ⟨he *is* reading⟩ ⟨I have *been* cleaning⟩

WEBSTER'S 3RD *be*/5

gressive tenses, usu. expressing continuous action <he *is* reading> <I have *been* sleeping> <the house *is* being built> but sometimes in present-tense form expressing future non-continuous action <he *is* leaving tomorrow> **3** : to have changed place or condition as a result of completing an action — used with the past participle of certain intransitive verbs as an auxiliary forming archaic perfect tenses <Christ *is* risen from the dead — 1 Cor 15:20 (DV)> <the minstrel boy to the war *is* gone — Thomas Moore> **4** : become supposed : become destined — used with the infinitive with *to* to express futurity, arrangement in advance, or obligation <I *am* to interview him today> <he *was* to become one of the most famous men of his century> <you *are* to repay the loan in monthly installments> — usu. not used in the form of an infinitive or participle **5** : to undergo a continuous action : be in the process of — used in a passive sense with the present participle or with the gerund preceded by the prefix *a-* or the preposition *a* <while the ark *was* building — 1 Pet 3:20 (NCE)> <when the ark *was* a building — 1 Pet 3:20 (DV)>; now usu. replaced by the passive construction with *being* followed by the past participle, as in 1 and 2

CAMBRIDGE ENGLISH DICT *be/1*

- 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*

CAMBRIDGE ENGLISH DICT *be/2*

- 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).*

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
- ② *not all that is can be understood*
- ③ to pay a visit, go
- ④ *have you been to Spain?*
- ⑤ to take place, occur
- ⑥ *my birthday was last Thursday*
- ⑦ 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

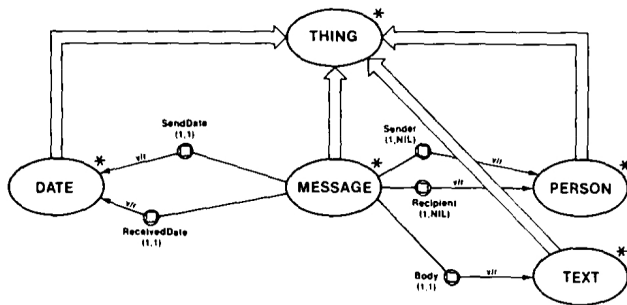
SEMANTIC REPRESENTATION

- 1 The mainstream standard in linguistic semantics, at least for the compositional case, is formulaic (logic-based)
- 2 Typically, formulas are from higher order intensional logic.
4lang has its own (weaker) formulas
- 3 The mainstream standard in AI/KR is the use of (hyper)graphs (algebra-based)
- 4 Typically, the graphs have nodes for individuals and classes (aka 'concepts' and edges for relations (hyperedges for higher arity relations)
- 5 The mainstream standard in computational linguistics is mapping words onto vectors of a low-dimensional Euclidean space \mathbb{R}^n (typically $n = 300$)
- 6 4lang doesn't always combine vectors by addition
- 7 We use IRTG (Koller, 2015) for synchronous rewriting of the formula, graph, and vector
- 8 If customers demand it, we may resurrect the Eilenberg branch

FORMULAS

- ❶ Resemble dictionary definitions, designed for human readability
- ❷ Conjunctive top-level, avg 2.7 conjuncts (max 9 found so far)
camera kamera cinematographica_machinula kamera
1221 N machine, make photograph, HAS lens
- ❸ make =AGT CAUSE [=PAT[exist]]
- ❹ Syntactic sugar: $x \text{ IS_A } y$ can be written as $x[y]$ or as $y(x)$
- ❺ Defaults are in $\langle \rangle$, grouping by $\{ \}$
- ❻ Only one quantifier: gen, treated as a noun (no scope). Two thematic roles (VBTOs), no ternary relations (Kornai, 2012)
- ❼ < 20 binaries, < 600 core (uroboros) vocabulary elements

CLASSIC KR



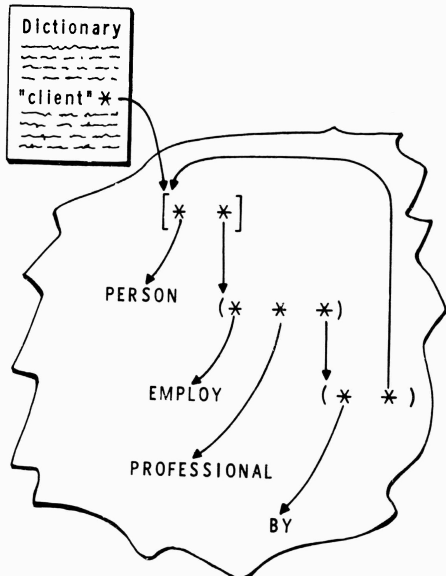
"A MESSAGE is, among other things, a THING with at least one Sender, all of which are PERSONs, at least one Recipient, all of which are PERSONs, a Body, which is a TEXT, a SendDate, which is a DATE, and a ReceivedDate, which is a DATE."

4lang GRAPHS

- ① Don't have a huge variety of links: 0 (*is/is_a*); 1 (*subject*); 2 (*object*)
- ② In contrast, Cyc has over 45.000 link types, and more contemporary efforts like DBpedia or YAGO have $10^5 - 10^6$. The vast majority of these are like *isSpouseOf*, obviously compositional
- ③ To get compositional links right, one needs to deal with *is*, *of* as well as with *spouse* 'husband or wife'.
- ④ Note effortless transition from disjunction of common nouns to disjunction of relations *isHusbandOf* or *isWifeOf*! This makes sense iff *is spouse of* is indeed *is (husband or wife) of* i.e. iff the meanings of words (their definitions) can be substituted *salva veritate*
- ⑤ 4lang graphs are built on RDF-like “triple stores”, explicitly addressing known difficulties with these such as **negation**, **quantifier scope**, **nested modals** and relations of seemingly

QUILLIAN, SCHANK

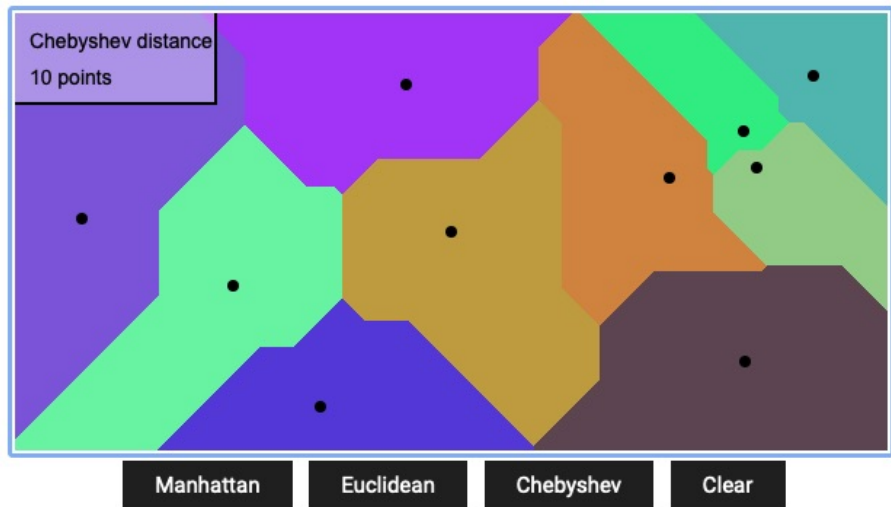
Semantic Memory



Conceptual Dependency

John
‡ ⇒ good
Love
‡
one

VORONOI DIAGRAM



VORONIDS

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 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.

WORD VECTORS – SUCCESS HAS MANY FATHERS

EMBEDDING (STATIC)

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

- Idea first suggested by Schütze, 1993
- 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)

FUNDAMENTAL PROPERTIES OF EMBEDDINGS

1. Frequency

$$\log(p(w)) = \frac{1}{2d} \|\vec{w}\|^2 - \log Z \pm o(1) \quad (1)$$

2. Cooccurrence estimate

$$\log p(w, w') = \frac{1}{2d} \|\vec{w} + \vec{w}'\|^2 - 2 \log Z \pm o(1) \quad (2)$$

3. PMI

$$\langle \vec{w}, \vec{w}' \rangle \sim \frac{\log p(w, w')}{\log p(w) \log p(w')} \quad (3)$$

4. Multi-sense: generally treated as the sum of the vectors belonging to different senses. Today: dynamically disambiguated by context-sensitive embeddings (ELMO, BERT). We use static (but multi-sense) here.

COMMUNICATION BY VORONIDS

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 $l : \mathbb{R}^n \rightarrow D^*$ 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 $\pi(\text{candle})$ probability mass that the symmetric difference between the polytopes of speaker and hearer is negligible.

THOUGHT VECTORS IN ANN MODELS

THOUGHT VECTORS

We will use n recurrent binary neurons, changing state in time. At t they are collectively described by a thought vector

$\Psi(t) = |s_1, \dots, s_n\rangle$. The mental space (Fauconnier, 1985; Talmy, 2000) T will have $\dim n$ (very large, on the order of $10^{10} - 10^{11}$)

Thought vectors move among the 2^n corners of the hypercube according to some 2^n by 2^n transition matrix $P(t)$. By standard methods, the time average $\Gamma(\alpha)$ of the probability of the system being in state α as

$$\Gamma(\alpha) = \frac{\sum_r \lambda_r^M \phi_r^2(\alpha)}{\sum_r \lambda_r^M} \quad (4)$$

where the ϕ_r are the normed eigenvectors of P .

PERSISTENT STATES

If time M is large, This is dominated by the contribution of the eigenvector corresponding to the largest eigenvalue, giving $\Gamma(\alpha) = \phi_1^2(\alpha)$, and we see that and we see that

$$\Gamma(\alpha, \beta) = \phi_1^2(\alpha)\phi_1^2(\beta) = \Gamma(\alpha)\Gamma(\beta) \quad (5)$$

i.e. the long term probability distribution of β is totally uncorrelated to that of α after a large number of steps. If the dominant eigenvalue is degenerate, or just the second one is very close, we have

$$\Gamma(\alpha, \beta) = \frac{\lambda_1^M \phi_1^2(\alpha) + \lambda_2^M \phi_2^2(\alpha)}{\lambda_1^M + \lambda_2^M} \quad (6)$$

Following (Little, 1974), only a small fraction of the eigenvectors will belong to these top eigenvalues, and can this be assumed to be persistent.

THE TAKEAWAY SO FAR

- ① Thought vectors $\Psi(t)$ are very large (dim 10^{10}) and they wander around the hypercube on millisecond timescale
- ② We are only interested in the part Φ 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 $\vec{p}_1, \dots, \vec{p}_d$ (dim $d \approx 200 - 600$) 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
- 2 The lone quantifier gen : $(1/n, \dots, 1/n)$
- 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

1. By definition, carbon IS_A material. We want $\vec{v}(\text{carbon}) \in P_{\text{material}}$ 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 $I(A) \text{ IS_A } I(B)$ 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

$$P_R(t+1) = \lambda I + s|\rangle p_i, p_j\langle| \quad (7)$$

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

$$P_R(t+1) = P_R(t) + s| \rangle = \text{AGT}, = \text{PAT} \langle | \quad (8)$$

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

$$P_R(t+1) = P_R(t) + s| \rangle = \text{AGT}, (= \text{PAT} + \vec{v}) \langle | \quad (9)$$

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.

COMING ATTRACTIONS

- 1 The semantics of morphology
- 2 Time and space
- 3 ER – scalar adjectives without scalars
- 4 LACK – dyadic negation
- 5 Naive probability
- 6 Modality
- 7 Implicature
- 8 Coreference

- 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.
- 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>.
- Collobert, R. et al. (2011). “Natural Language Processing (Almost) from Scratch”. In: *Journal of Machine Learning Research (JMLR)*.
- Fauconnier, Gilles (1985). *Mental Spaces*. MIT Press.
- Gärdenfors, Peter (2000). *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Koller, Alexander (2015). “Semantic construction with graph grammars”. In: *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*. London.
- 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.

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.

Little, W. A. (1974). “The existence of persistent states in the brain”. In: *Mathematical Biosciences* 19, pp. 101–120.

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>.

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.

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

Talmy, Leonard (2000). *Toward a cognitive semantics*. MIT Press.