Semicompositionality

András Kornai SZTAKI Computer Science Research Institute and Dept of Algebra, Budapest Institute of Technology

NASSLLI, June 26 2025









SIMILE AND EXTENDED METAPHOR

Feature	Simile	Metaphor
Example	He fights like a lion	He is a lion
Form	Explicit comparison	Implicit identification
Connective	Uses <i>like</i> , <i>as</i>	absent
Syntax	Binary comparison	Predicative/nominal

- **Simile**: an explicit comparison using words like *like*, *as*, or *similar to*
- We account for these by the semantics of *like*, *as*, *similar to*
- **Metaphor**: an implicit identification where one thing is said to *be* another
- We account for these by the semantics of be
- Very frequent!
- Extended metaphor: (Lakoff and Johnson, 1980)
 - Metaphor maps entire conceptual domains (e.g., ARGUMENT IS WAR).
 - Simile suggests partial mapping or surface analogy

SIMILAR

=agt has quality, =pat has quality, "to _" mark_ =pat

- 41ang lexical entry lists only English syntax, which demands x similar to y. In Hungarian we would have x-NOM hasonló y-ALL (allative case, in many situations translated by the preposition to, but not always)
- quality is underdetermined. Joe is like a lion can be 'has big mane', 'eats meat all the time', 'brave', ... – any lexical property is fair game
- Encyclopedic properties not so great: Joe is like a lion ???⇒ Joe lives in Africa
- =agt, =pat are the only variables in the formal language of 41ang, but the Quine-like ontological claim 'only those things exist that a variable can take on as a value' is false

WHICH PART OF *is* DON'T YOU UNDERSTAND?

- What makes similes similar to metaphor is the underdetermination of the aspect/quality that is relevant at the moment. When making an appointment for him in a barbershop *He is (like) a lion* can in fact mean 'has a huge mane'
- But *similar/like/as* have this referential ambiguity built in whereas *is* does not
- Typical usage of *is*
- Well what is it? Serious challenge not just to monosemic, but to any and all approaches to lexical semantics
- Embarrassingly many subcases, embarrasingly vague definitions of each

of bdy abbr boundary

Ibe \(!)be, bi\ vb, past 1st & 3d sing was \(!)waz, (!)waz also ket ()woz or dial were \see below or war (()war : 2d sing ills were \ wa(r); 'war, + V 'war., 'wa, + V " or 'war. also 'war: iah archaic or Brit (;)wa(a)()r or e()r or a(a) or eo or dial & ms archaic was or dial war (with you) or archaic wast (with thou) ge west, (;)wast also (;)wost or wert (with thou) we(r) t. KS) wərlt, (')wəlt, (')wəilt, usu |d.+V\ pl were or substand & las archaic was or dial war; past subjunctive were or substand & :k> archaic was or archaic 2d sing wert (with thou); past part the been (;) bin, bon; ben (in standard speech more often res unstressed or with secondary stress than with primary stress): na-Brit usu & US sometimes (;)ben or dial ben (!)ben of pres part being \'bein, 'been, rapid (')ben or dial Brit & by archaic been \() ben, 'bean pres 1st sing am \am, () am also (:)aa(a)m; after "I" often m or dial is or be; 2d sing are (with ter you) (-2(r), (-)ar, (-)a(r; after a vowel-final pronoun often r)or archaic art (with thou) $\geq (r)|t$, (;) ar|t, (;) a|t, + V usu |d. or dial be or is (with you) or dial Brit beest (with thou) $\langle (:) best$, $b\bar{c}$ best or **bist** (with thou) (!) bist 3d sing is (!) iz, \bar{z} ; 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, plare or substand is or dial & archaic be or archaic been \the last like past part BEEN \ pres subjunctive

Semicompositionality

NASSLLI, June 26 2025

tast like past part BEEN pres subjunctive be or obs 2d sing beest (with thou) [ME been, fr. OE beon; 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 $\langle God is love \rangle \langle January is the first month \rangle$ $(\text{let } x \sim 10)$: 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) **1**: 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-

Kornai

Semicompositionality

NASSLLI, June 26 2025

(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) i : 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 \sim ; stop pestering him) d: HAPPEN, OCCUR : take

Kornai

Semicompositionality

NASSLLI, June 26 2025

(let nim \sim ; stop pestering nim) a : HAPPEN, OCCUR : take place (the concert was last night) (where will the meeting \sim) e archaic : BELONG, PERTAIN (to thine and Albany's issue \sim this perpetual —Shak.) f (1) : to come or go : JOURNEY (we will \sim on our way shortly) (have you been home since Christmas) (2) : to make a stay : show oneself or be present (they will \sim 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 \sim 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

Kornai

Semicompositionality

NASSLLI, June 26 2025

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 noncontinuous 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 minstre 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

- 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!
- 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.
- 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

Kornai

CAMBRIDGE ENGLISH DICT be/2

- 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?
- 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.
- 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.
- used to say what can happen The exhibition of modern prints is currently to be seen at the City Gallery.
- 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

- used with past participles to form the
- used in sentences about an imagined situation
- used in sentences to introduce an aim when you are saying what must be done in order to achieve it
- used to say that someone or something is the same as the subject of the sentence
- used to say where something or someone is
- used to say when something happens
- used to describe someone or something, or say what group or type they belong to
- to behave in a particular way
- used to say how old someone is
- used to say who something belongs to
- used to talk about the price of something
- to be equal to a particular number or amount
- 3 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*

- to have presence in the realm of perceived reality, exist, live
- not all that is can be understood
- to pay a visit, go
- have you been to Spain?
- to take place, occur
- my birthday was last Thursday
- 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

CAMBRIDGE ENGLISH DICT be/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!
- 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.
- 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

Kornai

- 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!
- 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?
- 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.
- 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.
- used to say what can happen The exhibition of modern prints is currently to be seen at the City Gallery.
- 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).

THE KEY IDEA FOR be

- 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 l(A) is_a l(B) translated by set-theoretic containment.
- It is not true by definition that John sleeps. The entire thought vector whose persistent (linguistic) part is \vec{v} (John) will be wandering around, sometimes falling in the \vec{v} (sleep) polytope sometimes not, $\langle \vec{v}$ (John), \vec{v} (sleep) $\rangle = 0$. But when we say John sleeps the listener fixes the John vector inside the sleep polytope.
- But if $\vec{v}(John)$ and $\vec{v}(sleep)$ are fixed, how can we do this? We change the scalar product!

THOUGHT MATRICES

- Instead of *thought vectors*, the mental state is defined by matrices
- Intransitive predication corresponds to *is*, not ISA. In the static case we have eternal truths: boats are small open ships (by definition of boat, but when we say *the boat is old* this is not true by definition, just for the situation described by the sentence.
- P(t+1) = P(t) + s|=agt⟩ ⟨=pat| (Here the agent is the boat, and the patient is 'being old'.)
- For transitive verbs V we have $P(t+1) = P(t) + s |= agt \rangle \langle V + = pat |$ so for John eats fish we obtain an update on the thought matrix by a matrix that is the outer product of the John (one-hot) row vector and the eat+fish column vector

CONSEQUENCES OF MONOSEMY

- Typically, there is **no** 'metaphoric usage'
- *My new tv is a lemon* relies on a definitional property of *lemon*, being *sour* 'less enjoyable, pleasant, or satisfactory' (LDOCE)
- Lexical entries are conjunctive, but in a metaphor typically only one conjunct is taken (my new tv is not a fruit, is not yellow, doesn't have characteristic citrus smell, ...)
- Polysemy can't be fully eliminated, cf. *bitter* 'having a sharp strong taste like black coffee without sugar', 'feeling angry, jealous, and upset because you think you have been treated unfairly'
- 'a type of dark beer that is popular in Britain, or a glass of this'
- Are the taste, the feeling, and the beer 1, 2, or 3 subentries?

Sanskrit Derivatives of ankh (Whitney 1885)

- aṅkha () Noun
 - Root vankh + nominal stem
 - "narrowness", "tightness", "distress"
- ankhana () Noun
 - √ankh + suffix -ana (act noun)
 - "act of pressing together"
- ankhin () Adjective
 - √ankh + suffix -in (possessive)
 - "compressing", "constricting"
- ankhata () Noun
 - √ankh + suffix -ața (instrumental noun)
 - "instrument of compression"
- ankhațin () Adjective
 - $\sqrt{ankh} + suffix -tin$
 - "compressive", "pertaining to compression"

Derivatives of $\sqrt{\text{ANKH}}$ (Monier-Williams)

V

V

Ν

Ν

А

А

Ν

Ν

Ν

А

form ankhayati paryankhayate paryanka paryankikā paryankikrta paryankabaddha paryankabandha paryankabandhana paryankabhogin paryankastha

segments ,∕aṅkh+ya+ti $pary+_{ankh+ya+te}$ pary+_/ankh+a paryanka+ikā paryanka+krta paryanka+baddha paryanka+bandha paryanka+bandhana paryanka+bhogin paryanka+stha

gloss stir up, mix encircle, clasp couch, meditation posture palanguin, bed made into a bed seated cross-legged cloth-binding posture posture with tied cloth kind of serpent seated on a sofa

TYPICAL CASES

- Person/work *She is reading Dostoevsky*
- Location/activity Hollywood experimented with the idea
- Container/content *She drank two glasses. The suits made the decision.*
- Pars pro toto Nice wheels! Paris sent troops.
- (Man showing a copy of the NYT) I work for this newspaper!
- The more you try to subtype the harder it becomes: Is *Paris* pars pro toto for the entire country? Is it the location where government activity is taking place? Who sent the troops, the country or the government?
- The office takes up the entire 7th floor. The office rejected the payment.
- We have *one* monosemous vector, and we choose which subspace to project it to

Kornai

PROJECTION MAPPING

- Name comes from Fauconnier, 1985
- Very similar to the 'extended metaphor' of Lakoff and Johnson, 1980
- The phenomenon is everywhere. Today (as I'm writing the slides, 5/28/25) the WSJ has *Trump's Emcee Act Fails in Ukraine ... the long-promoted Ukraine spectacular is flopping ... In six months or a year, a new episode may be in the offing*
- Today (6/26/25) *More Homeowners Find Themselves Underwater* Some who bought around the market peak in pandemic boomtowns owe more on their mortgages than their homes are worth (INVESTMENTS ARE VEHICLES)

SCHEMAS

- Schemas, sometimes called scripts, are often used in GOFAI Minsky, 1975; Schank and Abelson, 1977 and go back to 1930s psychology
- We have seen the commercial_transaction schema
- This has four participants: *a buyer, a seller, some goods* and *money* or some other 'thing of value'
- Before the transaction, seller has the goods and buyer has the money, afterwards buyer has the goods and seller has the money
- The same schema is applicable both for the word *buy* and for *sell*

THE EXCHANGE SCHEMA AS A VORONOID



FIGURE: exchange_

(Temporal aspects will be added shortly)

Gen 25:29-34

- 29 And Jacob sod pottage: and Esau came from the field, and he was faint:
- 30 And Esau said to Jacob, Feed me, I pray thee, with that same red pottage; for I am faint: therefore was his name called Edom.
- 31 And Jacob said, Sell me this day thy birthright.
- 32 And Esau said, Behold, I am at the point to die: and what profit shall this birthright do to me?
- 33 And Jacob said, Swear to me this day; and he sware unto him: and he sold his birthright unto Jacob.
- 34 Then Jacob gave Esau bread and pottage of lentiles; and he did eat and drink, and rose up, and went his way: thus Esau despised his birthright.

THE LINGUISTIC TAKEWAY

- This whole story makes no sense whatsoever unless we interpret it in the exchange_ frame
- It takes a great deal of work to link up the language with the schema: only in 31, *sell me thy birthright* is the schema triggered
- This already requires sophisticated anaphor resolution: we need to know that Esau will be *seller*, Jacob will be *buyer*. Morphology for nomen agentis -*er/3627* stem_-er is_a =agt, "_ -er" mark_ stem_
- *sell* =agt cause_ buyer has =pat, buyer cause_ =agt has money_, dative_ mark_ buyer
- We need to figure out the other two slots, that the goods are the birthright, and the thing of value is the bowl of lentils.
- The mechanism is coercive, it is not that Esau IsA seller, Esau is *the* seller, Jacob buy birthright is a valid inference.

PATTERN RECOGNITION

The four vectors { $\vec{v}(buyer)$, $\vec{v}(seller)$, $\vec{v}(goods)$, $\vec{v}(money)$ } are the defining elements of the exchange schema as a set (the curly braces emphasize that their order is immaterial). Together, they define a polytope, the intersection of the positive half-spaces. The other 4 vectors in the example, $\vec{v}(Jacob)$, $\vec{v}(Esau)$, $\vec{v}(birthright)$ and $\vec{v}(food)$ are just points (or small polytopes) in *L*. What we are looking for is an equaliser *Q* such that after applying *Q* to the representation space *R* that reflects the state of affairs before Jacob making the offer, we obtain *R'*, where not just

- (1) $\vec{v}(buyer) = \vec{v}(Jacob)$ and
- (2) $\vec{v}(seller) = \vec{v}(Esau)$ but also
- (3) $\vec{v}(goods) = \vec{v}(birthright)$ and
- (4) $\vec{v}(thing_of_value) = \vec{v}(bowl_of_lentils)$ hold

PATTERN RECOGNITION CONT'D

These equations are created by several different mechanisms. Equations (1-2) come from resolving pronouns corresponding to speaker and hearer: the sentence Sell me (this day) your birthright is addressed to Esau, making him the seller, and spoken by Jacob, making him the buyer. Since the birthright appears in the patient slot of this sentence, we obtain the third equation by a general syntactic mechanism (whatever follows the verb is the object). The last equation is supported by the mechanism of pragmatic inference discussed in S19:5.7: we know from earlier sentences that the *food* and the *bowl of lentils* are the same, we know Esau is faint, and he himself acknowledges that at this point the food is more important to him than his birthright: Behold, I am at the point to die: and what profit shall this birthright do to me? This establishes, from the seller's perspective, that the thing of value to be received for the goods is indeed the food.

EQUALITY, NOT JUST is_a

- We could use the scalar product mechanism we used for *be* to establish that Esau is in the seller halfspace, but we want more, we want to say that Esau is *the* seller of this instance of the schema
- This is why use equalizers (linear transformations that have e.g. $\vec{v}(seller) \vec{v}(Esau)$ in the kernel) to the entire representation space L^2 .
- Equalizers are used instead of variable binding both for pronouns and for other indexicals
- We will talk about *coercion* for what Fauconnier, 1985 calls 'projection mapping' to emphasize the Procrustean aspect of the process
- Heider-Simmel test see discussion in Gordon and Hobbs, 2017

- Fauconnier, Gilles (1985). *Mental Spaces*. MIT Press.
 - Gordon, Andrew and Jerry Hobbs (2017). A Formal Theory of Commonsense Psychology: How People Think People Think. Cambridge University Press.
- Lakoff, George and Mark Johnson (1980). *Metaphors we live by*. University of Chicago Press. ISBN: 978-0226468013.
- Minsky, Marvin (1975). "A framework for representing knowledge". In: *The Psychology of Computer Vision*. Ed. by P.H. Winston. McGraw-Hill, pp. 211–277.
- Schank, Roger C. and Robert P. Abelson (1977). Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures. Hillsdale, NJ: Lawrence Erlbaum.
- Whitney, William Dwight (1885). "The roots of the Sanskrit language". In: Transactions of the American Philological Association (1869–1896) 16, pp. 5–29. DOI: 10.2307/2935779.