

Context

From 7th century India to Today

Andrew Ollett



April 14, 2026

Outline

- 🍁 Two ways of thinking about context
- 🍁 Meaning and context in LLMs
- 🍁 Context in Indian philosophy of language
- 🍁 Comparison and discussion

Two ways of thinking about context

- 1 Something **extra, outside** of “the text itself,” which we consult only when we’re **uncertain** of the meaning of particular expressions.

He went to the bank.

Two ways of thinking about context

- 1 Something **extra, outside** of “the text itself,” which we consult only when we’re **uncertain** of the meaning of particular expressions.

He went to the bank to withdraw some cash.

He went to the bank to take a drink of water.

Two ways of thinking about context

1 “Extra” in what sense?

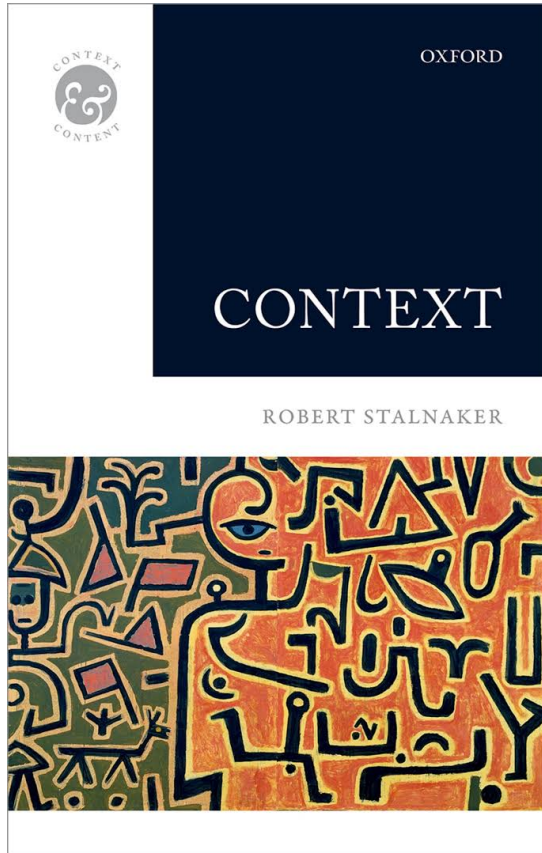
- ✿ Statements that occur outside of the sentence (or “text”) under interpretation (**discursive context**)
 - ✿ basically preceding or following statements
- ✿ Information that we can understand from the circumstances in which the sentence is spoken (**situational context**)
- ✿ Information that we can take for granted (**real-world context**)

Two ways of thinking about context

The main **theoretical** problem with this approach is that meaning has to be determined in **at least** two stages:

1. First we determine the “basic,” “literal,” or “proper” meanings according to some procedure (e.g., looking them up in a dictionary).
2. Then we “refine” these meanings by consulting context.

Two ways of thinking about context



Robert Stalnaker, *Context* (Oxford 2016)

An extremely good overview from one of the pioneers in this field, which focuses on the notion of context as **background information** (and hence closer to view #1).

Two ways of thinking about context

But what if **contextual meaning** is all there is?

That takes us to the second way:

2 Context is **constitutive of meaning itself**.

What might this actually mean?

Two ways of thinking about context

2 Context is **constitutive of meaning itself**.

This is a challenging and counterintuitive suggestion put forward in the 7th c. CE by the philosopher **Prabhākara**, which people have mostly either ignored or ridiculed...

Two ways of thinking about context

2 Context is **constitutive of meaning itself**.

This is a challenging and counterintuitive suggestion put forward in the 7th c. CE by the philosopher **Prabhākara**, which people have mostly either ignored or ridiculed...

But it is also **operationalized in modern large-language models**, which have transformed the world, so maybe it is worth trying to figure out what Prabhākara was saying.

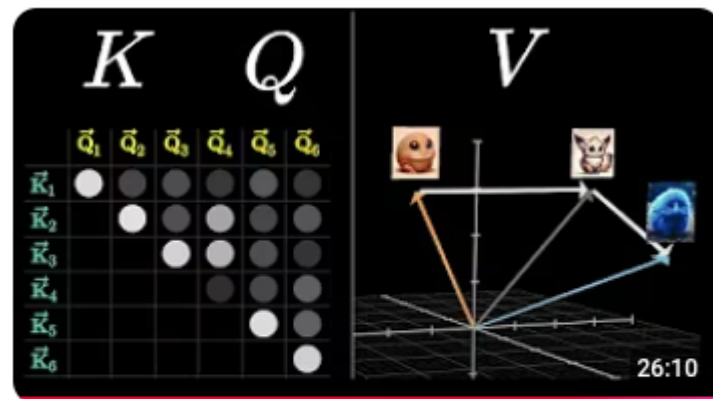
Two ways of thinking about context

We'll look at context-sensitivity in LLMs, then return to Prabhākara, and I'll try to explain **why** Prabhākara came up with the idea that “context is everything.”

Meaning and Context in LLMs

- 🌿 **Transformers** are the most common kind of language model in use today.
- 🌿 They implement context-sensitivity with a mechanism known as **attention**.

For a much better presentation than I can give you see this video by Grant Sanderson:

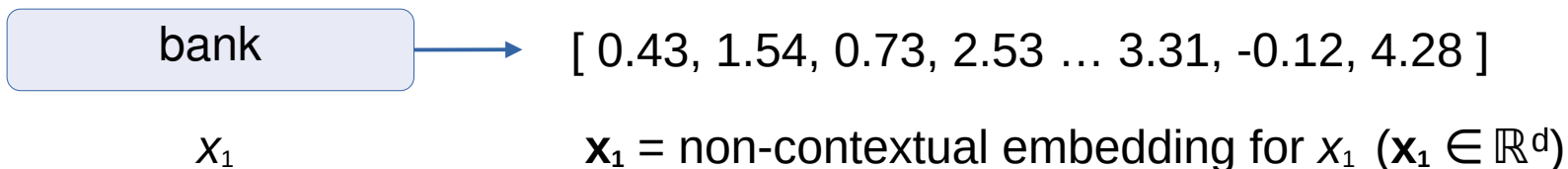


Attention in transformers, step-by-step | Deep Learning Chapter 6

3.9M views • 2 years ago

Word embeddings

- ✿ The first step in a transformer is to convert the word (or rather a subword token) into a mathematical representation, i.e., a **vector**, called the **embedding** of the token.
- ✿ The initial embedding of a token is **static** or **non-contextual**. The vectors for the token “bank” are the same, regardless of its context (and hence regardless of its different meanings).

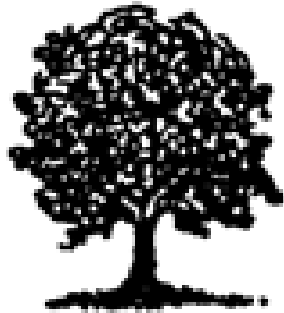


Word embeddings

- ✿ What do the values in the non-contextual embedding represent?
- ✿ We won't talk much about it here (the point is to get to a **contextual** embedding) but the idea is that an **embedding** represents a word's meaning.

Word embeddings

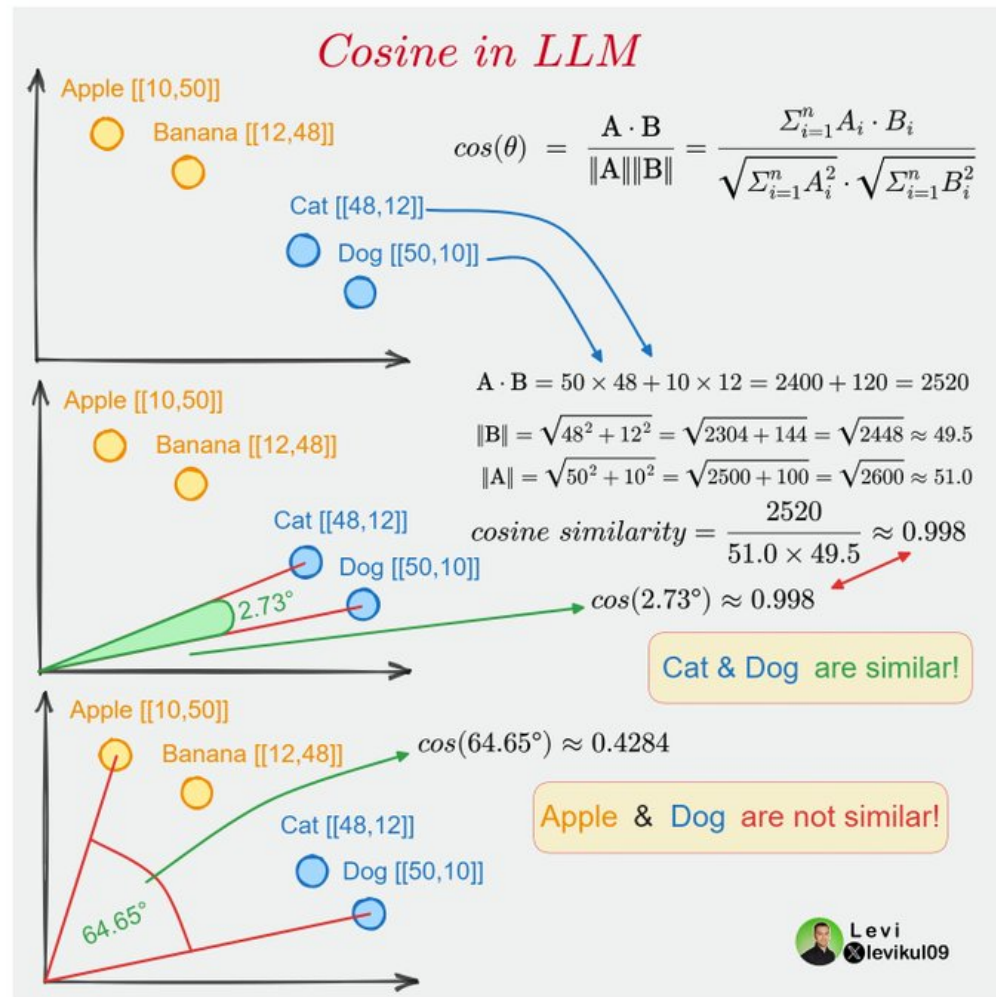
- Obviously “meaning” cannot mean a relationship to something in **the real world**. The operative concept here is not reference or a signifier-signified relation:



ARBOR

Word embeddings

- 🦉 The operative concept is rather one of **similarity** between words, which has a precise mathematical definition (cosine similarity) once words are converted into vectors:



Word embeddings

- ✿ The initial, **non-contextual** embedding represents a “placement” of a word within a multidimensional space **exclusively based on its co-occurrence with other words** throughout the corpus.
- ✿ Already we see that the concept of meaning operationalized here is one in which **context is everything** (although there is no difference between **global** and **local** context yet!).
- ✿ Words with multiple meanings (like **bank**) are assigned a single representation that combines those meanings.

Attention

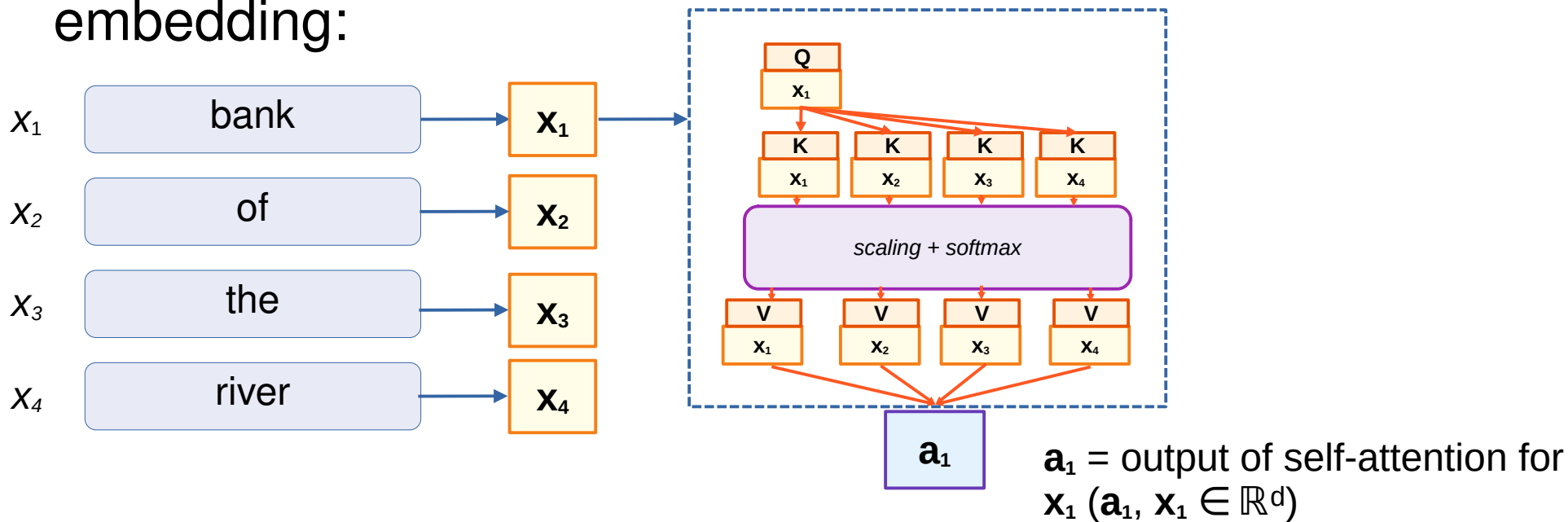
- Now we are going to **update** the embedding: for a given word **in context**, we make changes to the values of the vector that will change its direction vis-à-vis the embeddings of all of the other words.
- This allows us to capture the differences between, e.g., the two instances of the “same” word **bank** in the sentences “he went to the **bank** to get cash” and “he want to the **bank** to drink water.”

Attention

- ✿ The purpose of the transformer is to **update** these non-contextual embeddings by taking local context into account.
- ✿ The mechanism for doing so, **attention**, involves:
 - ✿ computing a measure of **similarity** between the current token and the preceding (and in bidirectional transformers, following) tokens
 - ✿ updating the embedding with new values based on this measure of similarity
 - ✿ allowing the parameters for this process to be learned in training

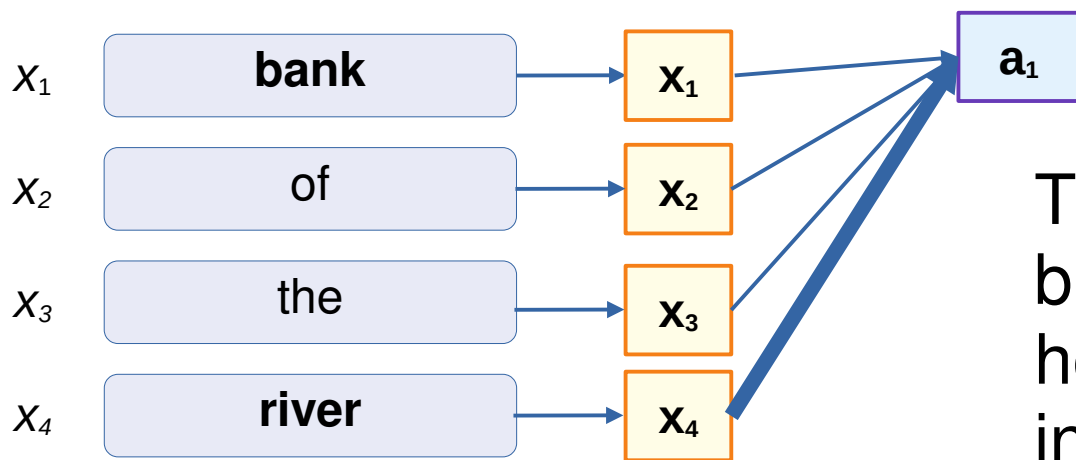
Attention

- These computations are packed into a layer of the model, the **self-attention** layer, which we can think of as outputting a **contextually enriched** version of the input embedding:



Attention

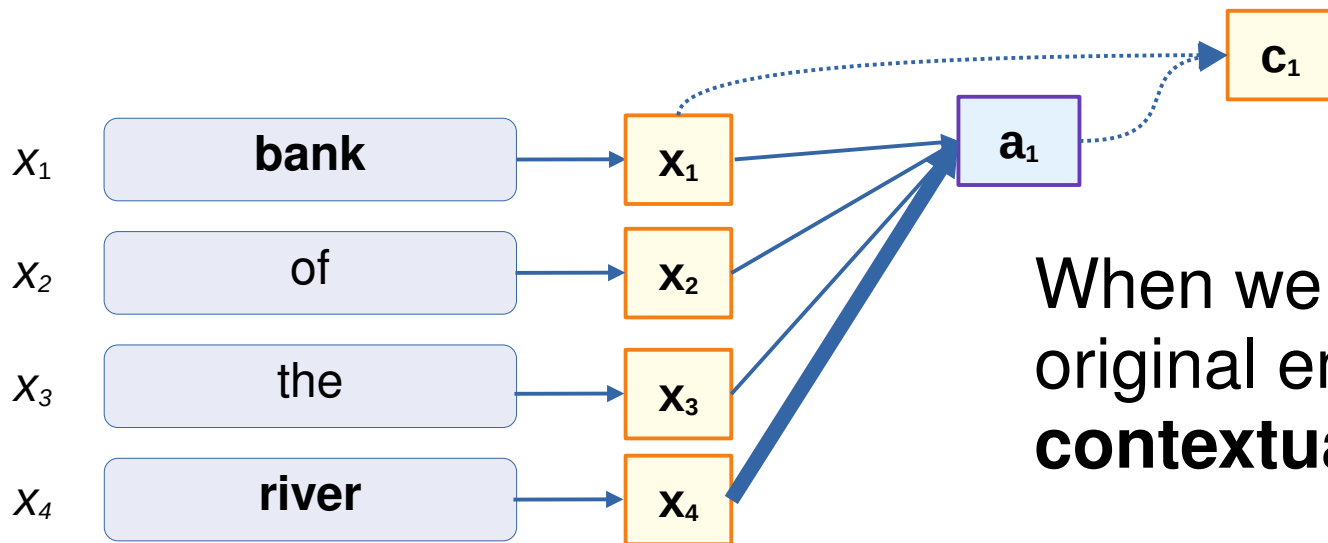
- ✿ The learned parameters will take the embedding of “bank” in the direction indicated by the word “river.”
(This is a bidirectional example, with no masking.)



The output of the attention block gives us a measure of how much we change the initial representation of “bank” based on its local context.

Attention

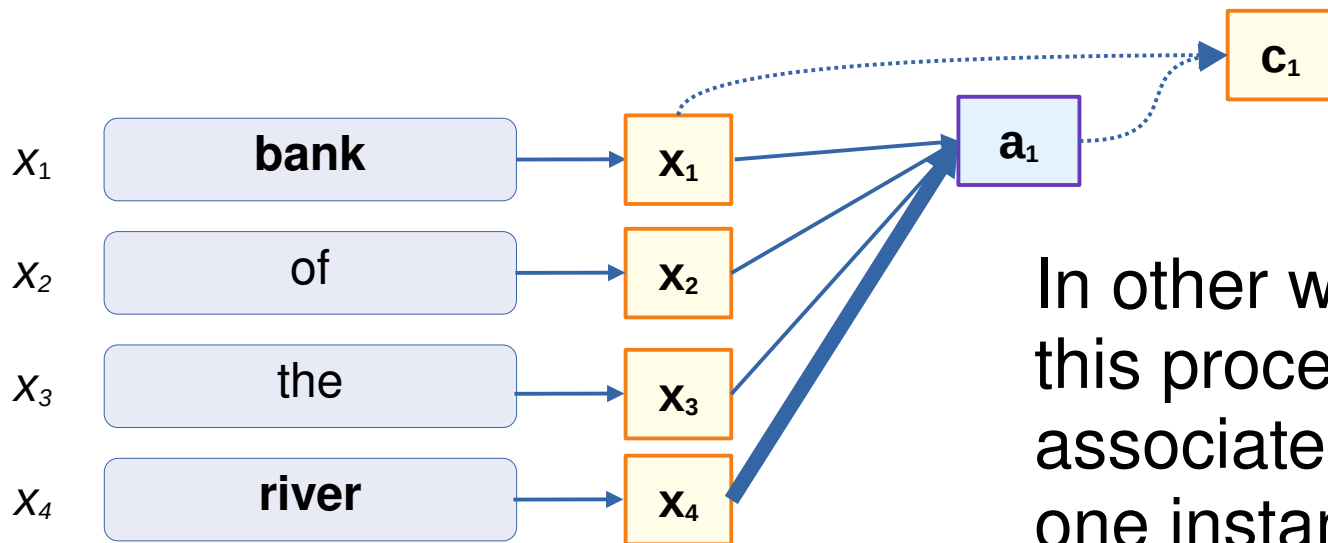
- ✿ The learned parameters will take the embedding of “bank” in the direction indicated by the word “river.”
(This is a bidirectional example, with no masking.)



When we add attention to the original embeddings, we get **contextual embeddings**.

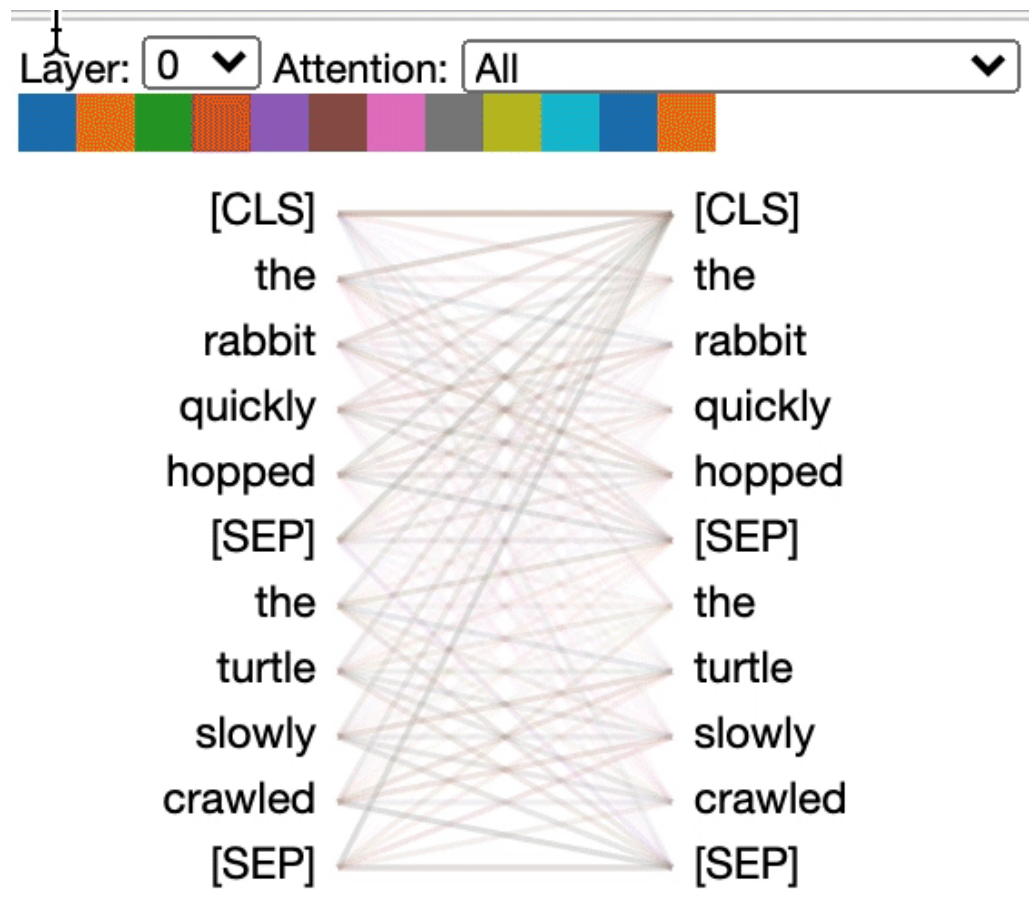
Attention

- ✿ The learned parameters will take the embedding of “bank” in the direction indicated by the word “river.”
(This is a bidirectional example, with no masking.)



In other words, by the end of this process, the vector associated with “bank” (in this one instance!) will represent a **relational meaning**.

Attention



<https://github.com/jessevig/bertviz>

Attention

- ✿ This is a massive simplification, not just because I have left out all of the matrix multiplication.
- ✿ Attention is usually performed multiple times in parallel (**multi-head attention**).
- ✿ This (in principle) allows the model to capture different ways in which context can influence a word's relational meaning.

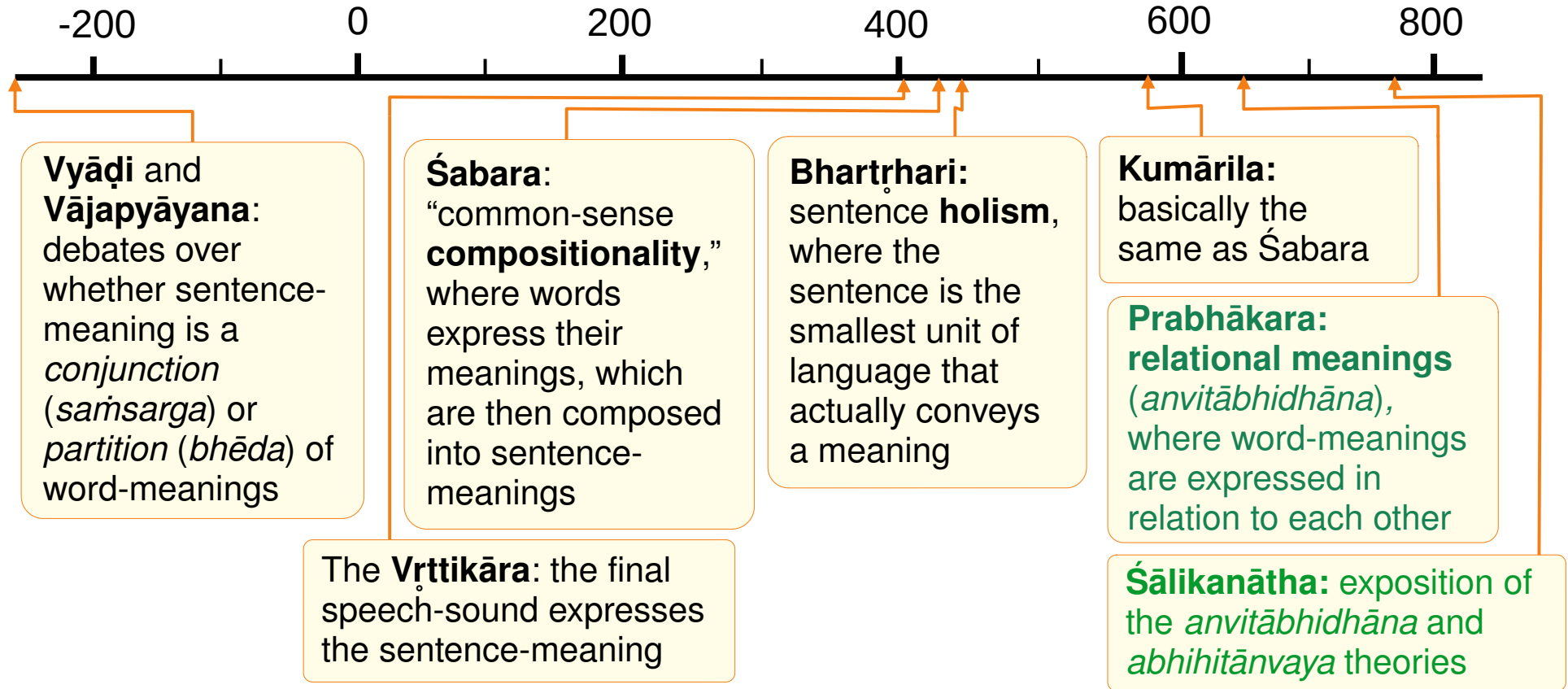
Attention

- ✿ The **representation** of every word is based on its **co-occurrence** with other words, both across the corpus (the contribution of a static embedding) and in a local context window (the contribution of self-attention).
 - ✿ Increasing the size of these context windows has been a major priority in AI development.
- ✿ Hence the concept of meaning in modern LLMs: **context is everything**. You **never** start by hard-coding meanings from a dictionary; you **always** look to the other words with which a given word co-occurs.

... back to 7th-century India

- ✿ I find it most helpful to make sense of Indian philosophy in terms of the **debates** that thinkers conducted with other thinkers.

Timeline



... back to 7th-century India

- ✿ For Prabhākara, the idea of **relational meanings** was a solution to what appeared to be a conflict between two philosophical commitments:
 - ✿ the **compositionality** principle
 - ✿ the **context** principle

Compositionality

The meaning of the whole depends on:

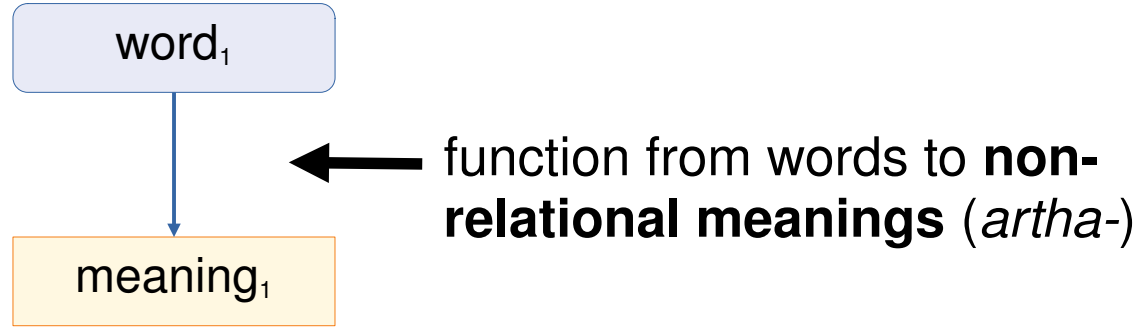
- the meaning of its parts;
- the rules of combination.

✿ This was a fundamental (and common-sense) commitment held by many thinkers, but most explicitly by Prabhākara's predecessors in Mīmāṃsā, **Śabara** and **Kumārila**.

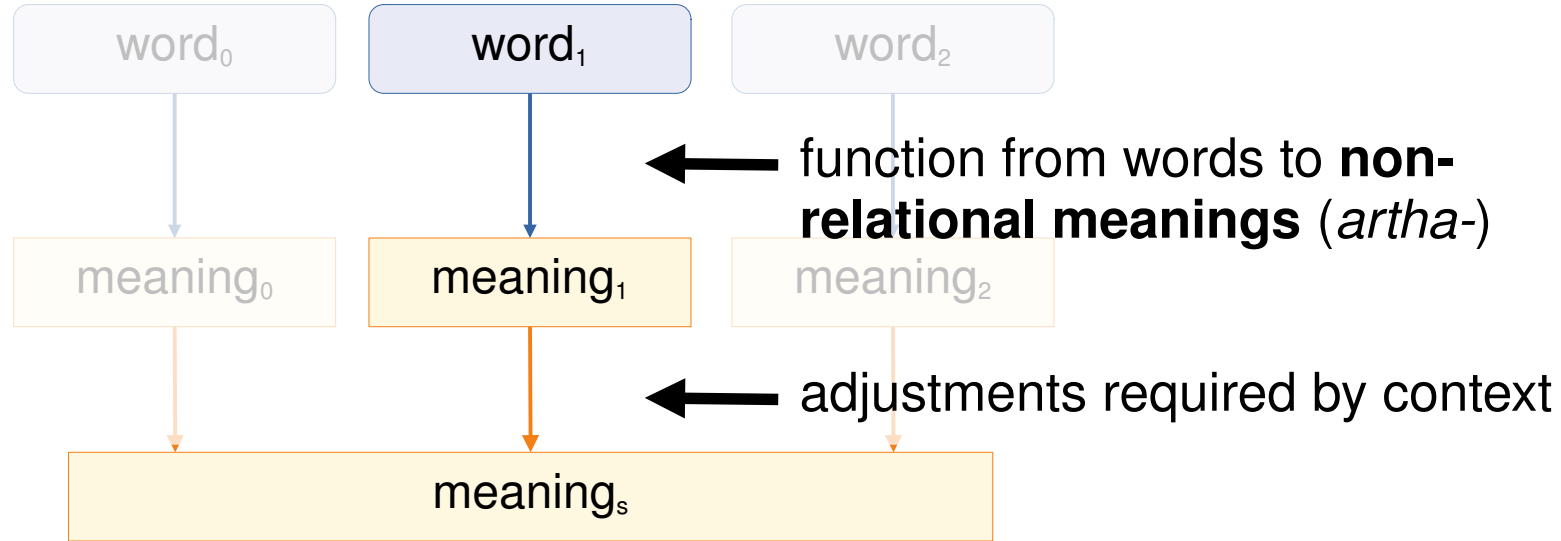
Abhihitānvaya

- ✿ The principle of compositionality is the basis for a theory of sentence-meaning called *abhihitānvaya* (“the relation of expressed meanings”) attributed to Kumāṛila.
 - ✿ Please bear in mind, however, that Kumāṛila (who lived a little bit before Prabhākara) **never** used the term himself, and in fact never used *anvaya* in the sense of a “relation” between word-meanings.
 - ✿ Śālikanātha (probably) came up with the term as a description of a theory he thought Kumāṛila might have been committed to.

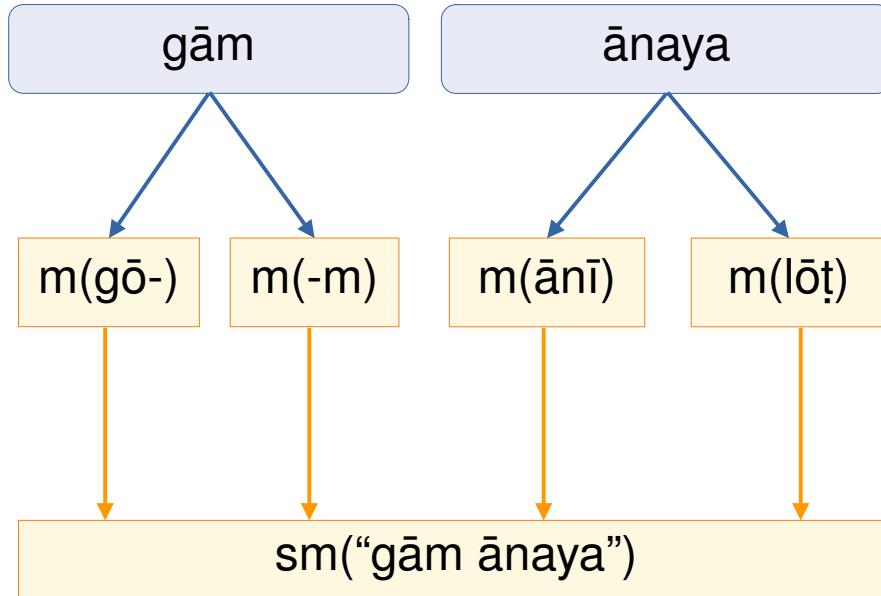
Abhihitānvaya



Abhihitānvaya

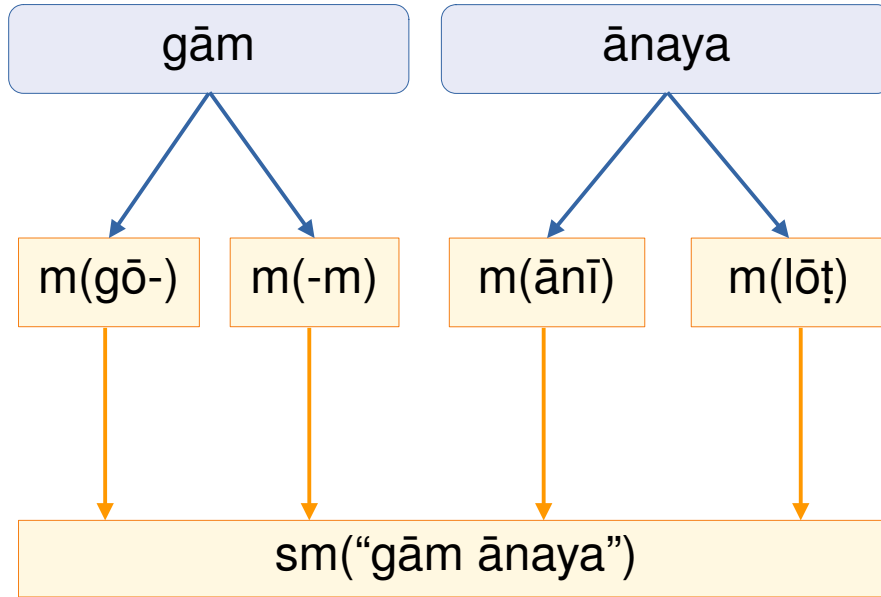


Abhihitānvaya



“Bring [the] cow”

Abhihitānvaya

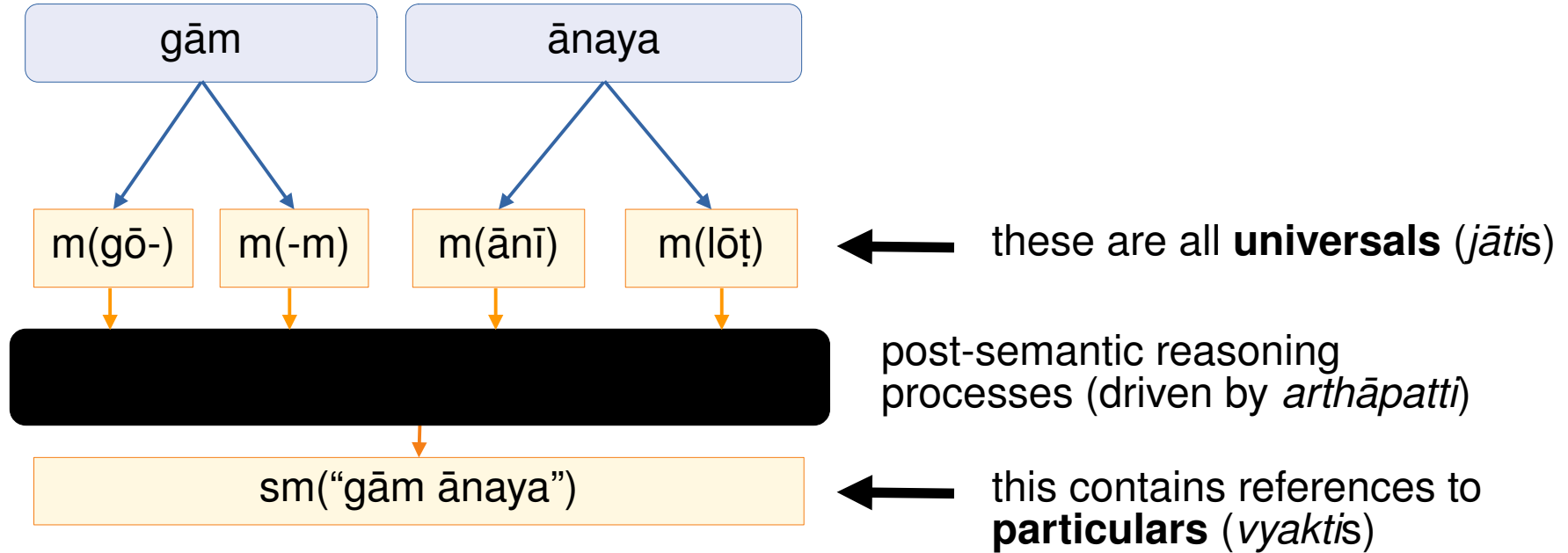


“Bring [the] cow”

← these are all **universals** (*jātis*)

← this contains references to **particulars** (*vyaktis*)

Abhihitānvaya



Abhihitānvaya

- ✿ In this view, the **meaning** of a word is invariant (“static”) across contexts. Larger units of meaning are **compositional** as much as possible, which means that we build larger units of meaning by putting smaller units of meaning together and making whatever minimal adjustments are necessary.
- ✿ But it turns out that we **always** have to update the non-contextual word-meanings with information from the context.

Context

- ✿ The **context** principle, as it has been formulated in Western philosophy of language following Frege, is:

Parts only have meaning in the context of a whole.

The best-known advocate of this view (to the extent that we can understand what he was actually arguing) was the grammarian **Bhartrhari**.

Context

- ✿ In Bhartṛhari's view, we always **start from** the sentence-meaning (which he mysteriously calls *pratibhā*, “insight”). Words (and word-meanings) do not really exist; they are isolated for analytic purposes only from the sentence (and sentence-meaning).
- ✿ Hence both “contextual” and “non-contextual” meanings are just more or less useful fictions, abstracted from larger contexts of use.

Context

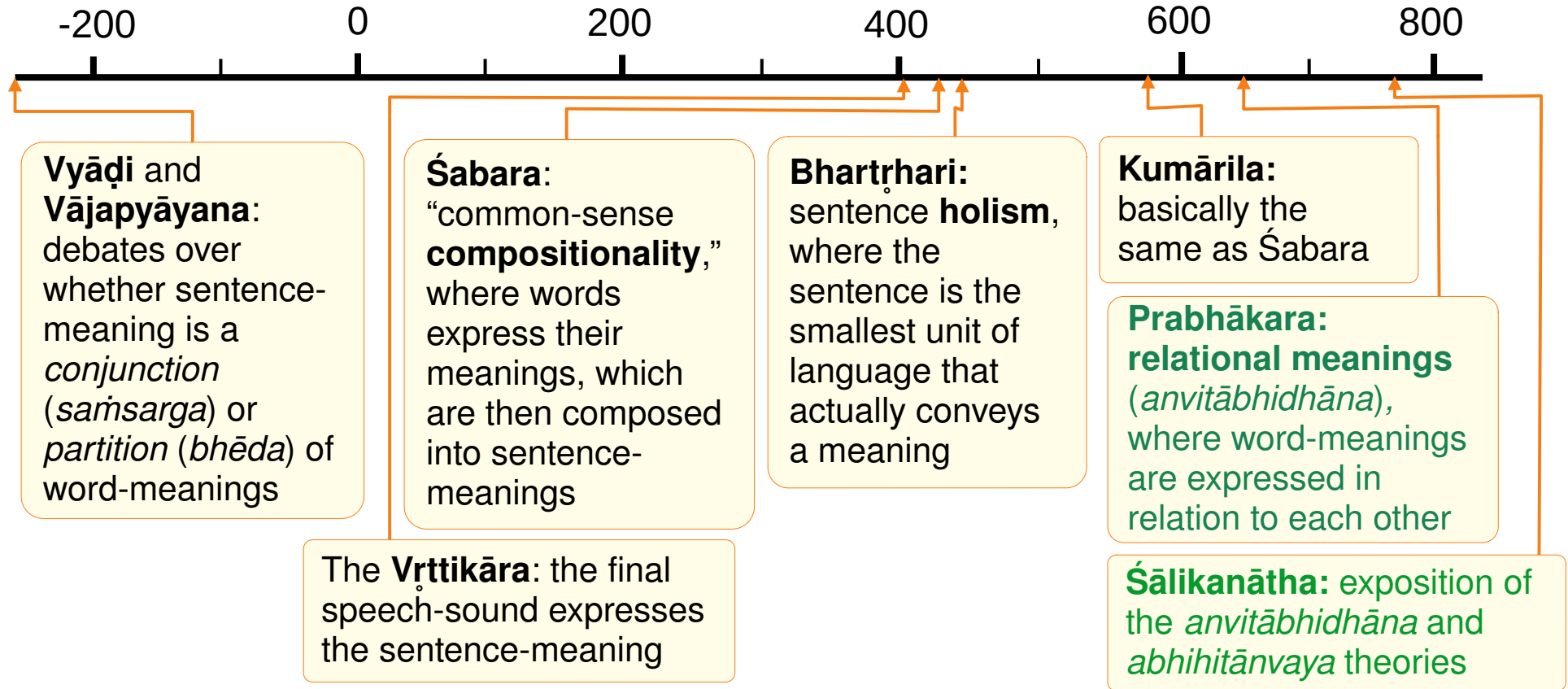
- ✿ Prabhākara thought of this position as mystical gibberish, impossible to even represent in a diagram, but he retained one critical insight: **linguistic expressions are only meaningful in the context of the whole of which they form a part.**

Putting it all together

- ✿ For Prabhākara, theories of meaning (especially meaning) needed to be able to account for several aspects of language use:
 - ✿ we always communicate in **sentences**, not words
 - ✿ we **learn language** by observation of complex states of affairs
 - ✿ we are able to **reliably understand** sentences we've never heard before

These require a combination of context and compositionality.

Timeline



Anvitābhidhāna in overview

- ✿ The theory that he devised — albeit in a somewhat sketchy way — is called *anvitābhidhāna*:

anvita- “related” [word-meanings]

abhidhāna- “expression”

Hence “the expression of relational word-meanings” by the individual words. The central concept here is that of a **relation** (*anvaya*) between word-meanings that is constitutive of the sentence-meaning.

Anvitābhīdhāna in overview

✿ Essential texts:

- ✿ Prabhākara's *Bṛhatī, tadbhūtādhikaraṇa* (1.1.24–25)
- ✿ Śālikanātha's *Bhāṣyapariśiṣṭa* and *Rjuvimalapañcikā* on the same section
- ✿ Śālikanātha's *Vākyārthamāṭrkā* (especially part 1)

Note that there are few or no translations for these texts, and even specialists have misunderstood the theory. I recommend the work of **Kei Kataoka**, **Shishir Saxena** and **Patrick Cummins** (and myself of course).

Anvitābhīdhāna in overview

- ✿ Note that Prabhākara and Śālikanātha worked in the tradition known as **Mīmāṃsā**, the goal of which is to formulate general rational principles for interpreting the texts of the Vēdas (the sacred texts of Hinduism). They were therefore primarily concerned with theorizing the meaning of sentences such as:

darśapūrṇamāsābhyām svargakāmō yajēta

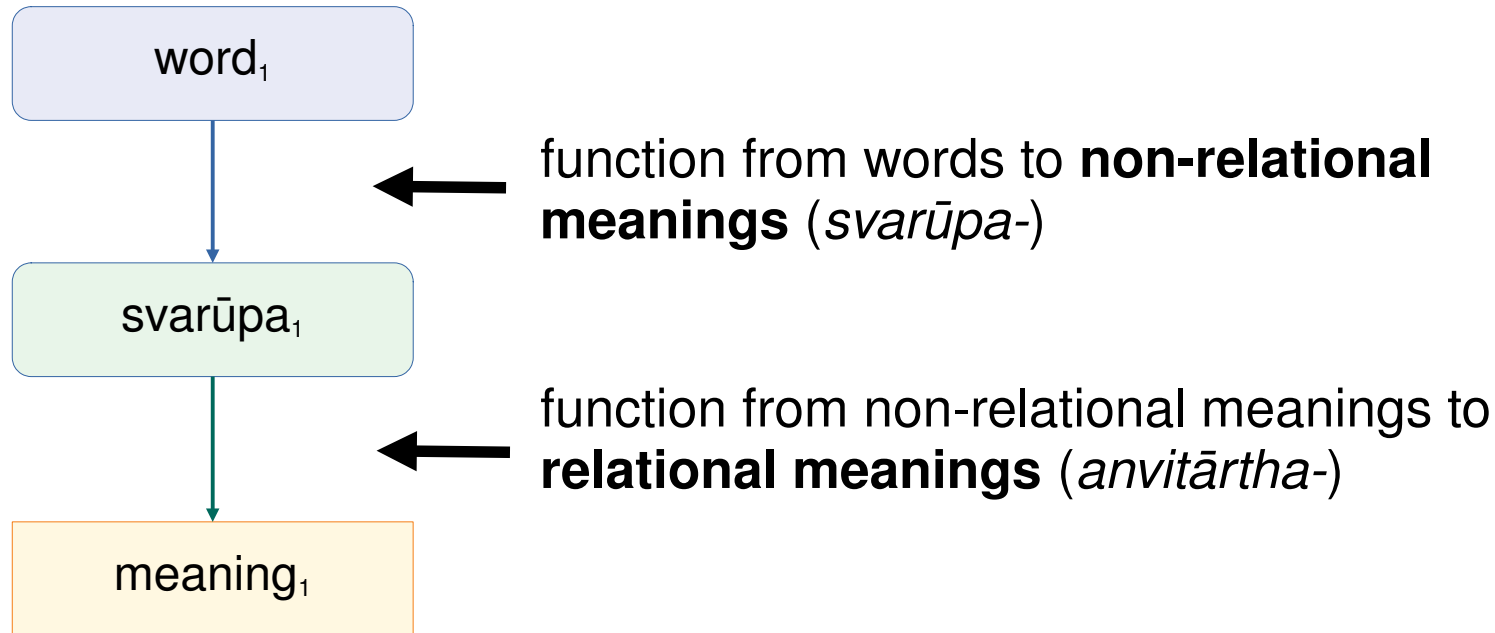
One who desires heaven should sacrifice with the new- and full-moon sacrifices.

Anvitābhīdhāna in overview

- ✿ Please also note that **Prabhākara** provided a “big picture” but did not work out this theory in all of the details. Most of that was left to his follower, **Śālikanātha**, but he too is less detailed than we would like. Nobody ever gave this a formal interpretation (although **Gaṅgēśa** in the 13th century came close), and obviously nobody put it in a form that could be operationalized by computers.

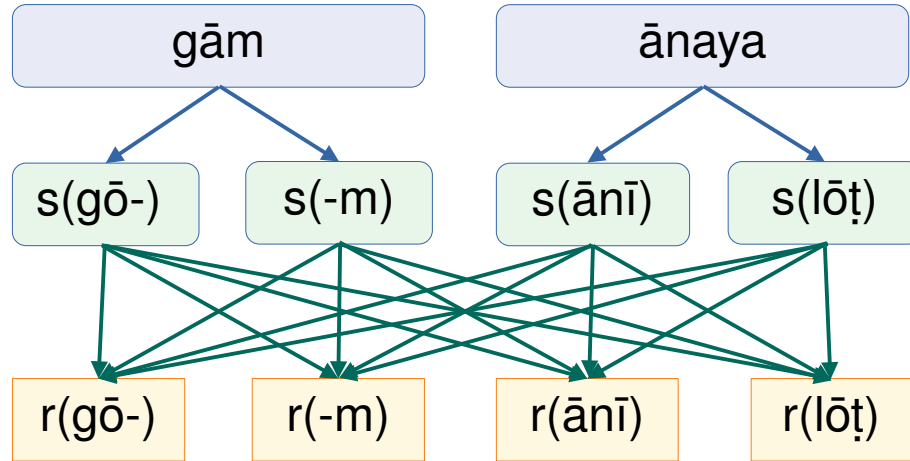
Anvitābhīdhāna in overview

- ✿ In the reconstruction of Śālikanātha (8th/9th c.) it works as follows:



Anvitābhidhāna in overview

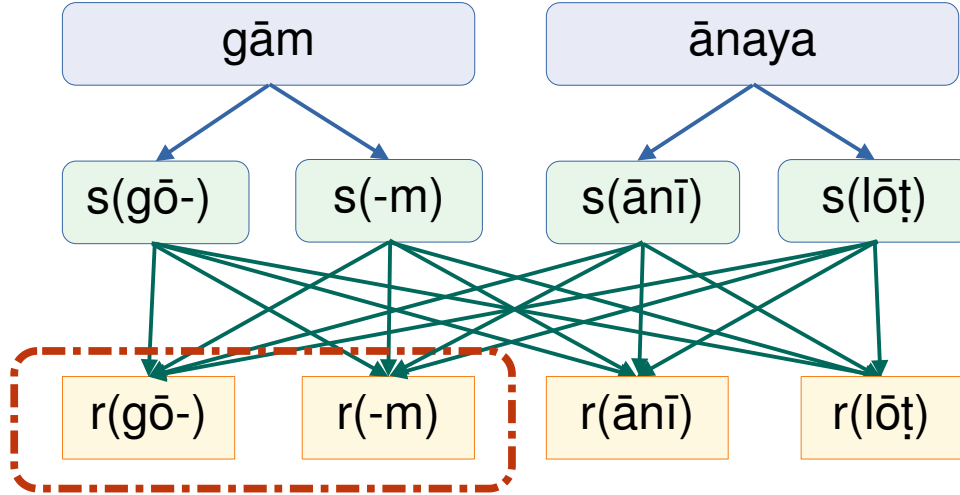
- ✿ In the reconstruction of Śālikanātha (8th/9th c.) it works as follows:



“Bring [the] cow”

Anvitābhīdhāna in overview

- ✿ In the reconstruction of Śālikanātha (8th/9th c.) it works as follows:



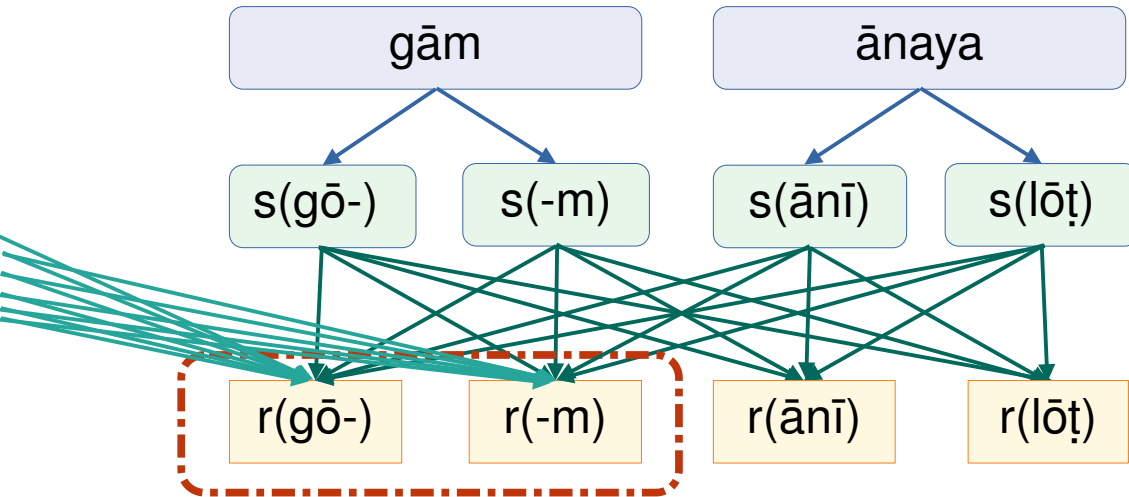
“Bring [the] cow”

In Sanskrit *gō-* famously has multiple meanings (including the sky, water, the eye, etc.).

The **relational meaning** of the word *gām* (*gō-* in the accusative singular) should encode **a specific meaning** of the word, namely **cow**, based on its cooccurrence with other words in a sentence.

Anvitābhidhāna in overview

- ✿ In the reconstruction of Śālikanātha (8th/9th c.) it works as follows:

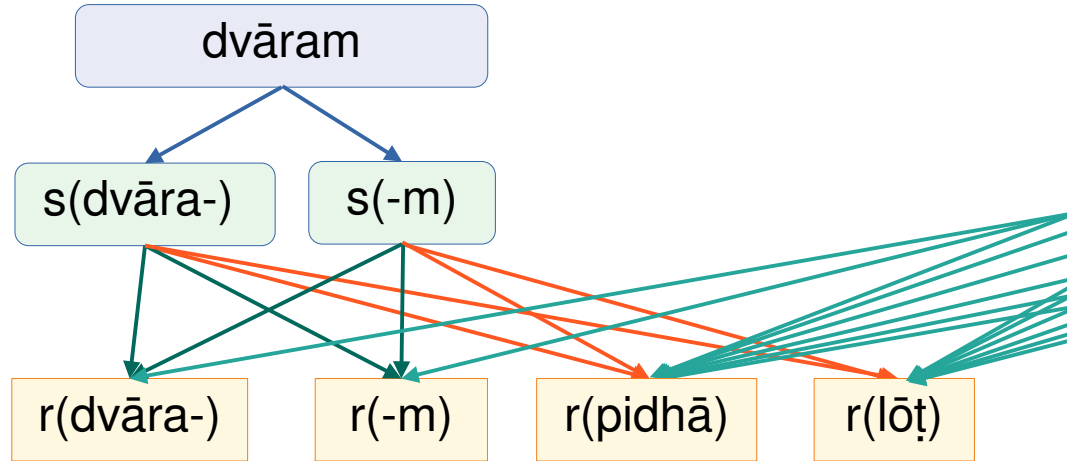


“Bring [the] cow”

In fact there is no reason why we cannot take into consideration **other words** from the context, which would further determine the **relational meaning** of the word *gām*, e.g., a particular milk-cow named Bessie that is known to the listener.

Anvitābhidhāna in overview

- ✿ In the reconstruction of Śālikanātha (8th/9th c.) it works as follows:



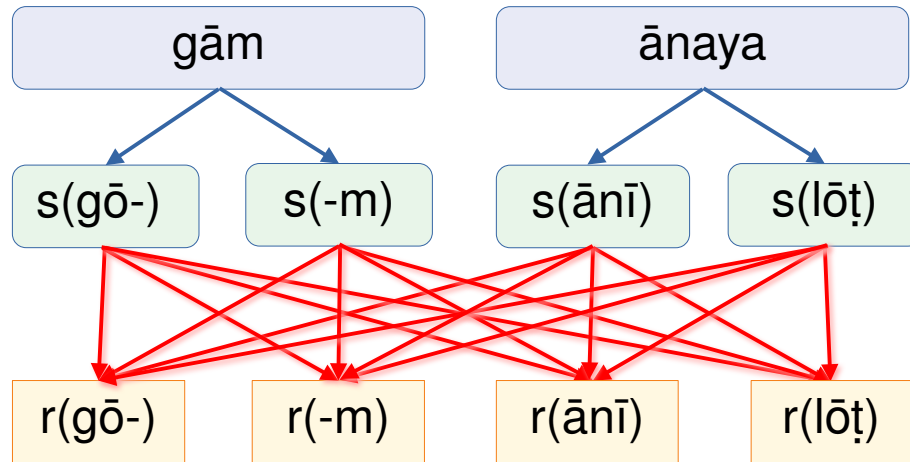
“Door!”

Śālikanātha uses the traditional example of filling in words from context (*adhyāhāra*) to explain how we arrive at an appropriate meaning from single-word sentences like “door!”

We understand a **related meaning**, namely “open” (if you are on a bus in Chicago), and the word “door” conveys its meaning in relation to this (unexpressed) verb.

Anvitābhidhāna in overview

- ✿ In the reconstruction of Śālikanātha (8th/9th c.) it works as follows:

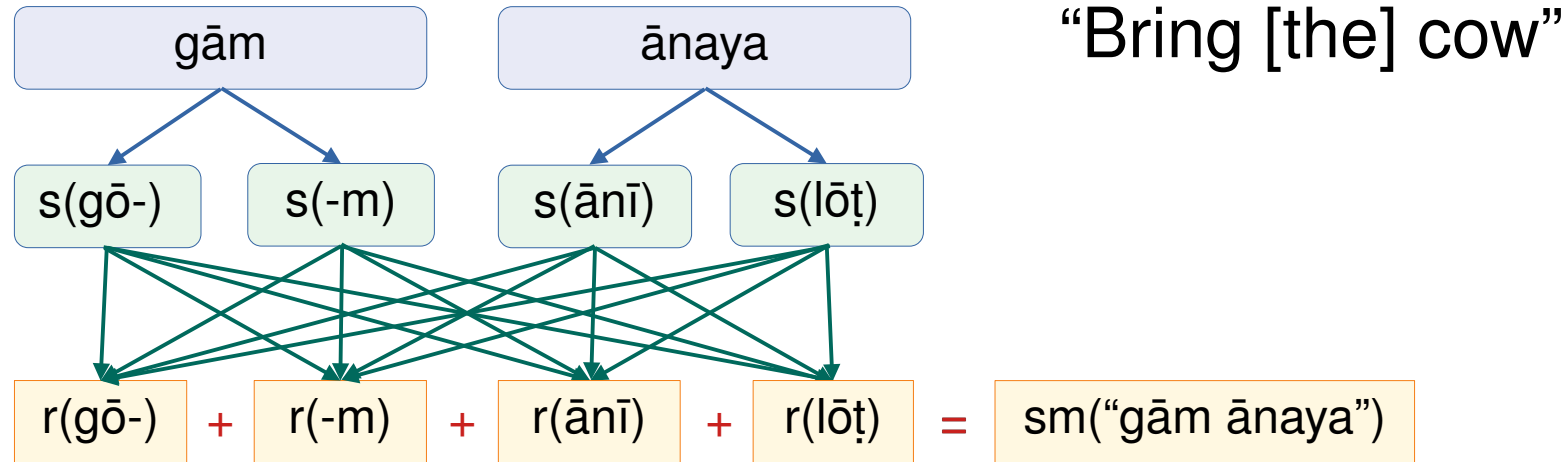


“Bring [the] cow”

1. the sentence-meaning is the **mutual relation** of the word-meanings

Anvitābhidhāna in overview

- ✿ In the reconstruction of Śālikanātha (8th/9th c.) it works as follows:



2. the sentence-meaning is the **relational word-meanings themselves**

Anvitābhīdhāna in overview

- ✿ In both interpretations, we end up with a striking suggestion: **context** is **constitutive** of meaning; the meaning of a linguistic expression is **defined by** its relation, in some context, to other such meanings.

Anvitābhīdhāna and Language Models

- 🦊 **Rohan Pandey** (OpenAI) has anticipated most of the interesting comparisons between *anvitābhīdhāna* and modern LMs, including especially **contextual embeddings**.

Review: Mīmāṃsā & Neural NLP

- Contextualized embeddings appear to follow Anvitābhīdhāna
 - Words in a sentence capture contextual meaning
- Abhihitānvaya may instead describe how words compose into a single sentence representation
 - Future work exploring syntax in CLS token
- Other psycholinguistic evidence clearly prefers Anvitābhīdhāna
 - Garden path sentences show how context influences syntax parse and word meanings
 - Ex: “the old man the boat”



Anvitābhīdhāna and Language Models

- ✿ In both attention and *anvitābhīdhāna*, we aim to produce a representation of a word's meaning (using both “word” and “meaning” loosely) that **incorporates** the effects that other words in the context might have on it.
- ✿ Both involve a **relation** (*anvaya* or attention) between each element (a word-meaning or embedding) and every other element.
- ✿ Functionally, both involve transforming a static representation (*svarūpa*) into a contextual representation (*anvitārtha*).

Anvitābhīdhāna and Language Models

- ✿ Beyond this, however, the way that meaning is represented in transformers is quite mysterious, and a lot of research has lately gone into why they work as well as they do.
- ✿ What do the millions (or billions) of parameters, stored in multiple matrices across multiple attention heads, actually **mean**? (Current research into **interpretability** of LLMs, especially using sparse autoencoders, addresses this question.)

Anvitābhidhāna and Language Models

- ✿ It may be worth considering other differences between Śālikanātha's *anvitābhidhāna* and the transformer architecture.
- ✿ Some of the features Śālikanātha describes **might** turn out to be represented (albeit in ways that are not currently interpretable) in the dense matrices of parameters in transformers — and might allow us to do “the same things” with less computationally intensive models (along the same lines as sparse autoencoders).

Anvitābhīdhāna and Language Models

- ✿ Conversely, research into LLM interpretability might tell us something about exactly how the “relation” between word-meanings is to be interpreted, which Prabhākara says next to nothing about, and Śālikanātha discusses in a very hand-wavy way.
 - ✿ for Śālikanātha these are grammatical, semantic, and information-structural relations.

Differences: Constraints

- ✿ A major motivation for *anvitābhidhāna* was explaining **learning**, i.e., how we can understand a potentially *infinite* set of contextual meanings based on finite exposure.
- ✿ Transformers also allow for contextual embeddings to vary arbitrarily, but they are calculated with a finite set of parameters (Q , K , V).

Differences: Constraints

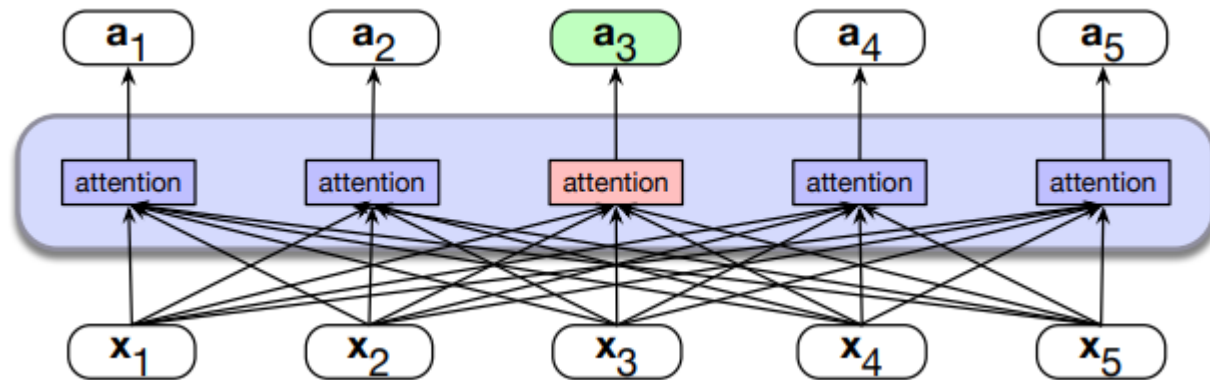
- ✿ Śālikanātha claimed that it is possible (and in fact easy) to “calculate” a contextual meaning given a small number number of **general constraints** on the relational meaning:

✿ **dependency, compatibility, proximity**

It seems **very likely** that such constraints are built into contextual embeddings somehow.

Differences: Hierarchy

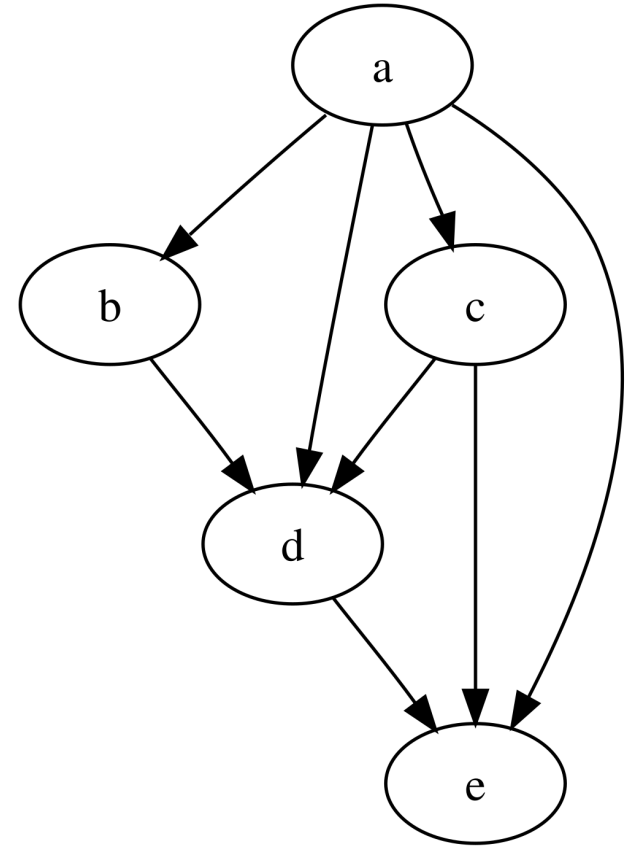
- ✿ The **relation** itself in transformers is flat:



From Daniel Jurafsky and James H. Martin, *Speech and Language Processing* (2024). Draft of January 12, 2025.

Differences: Hierarchy

- For Śālikanātha, the relation is hierarchical (modelled on the inherited Mīmāṃsā concept of *vinīyōga*), and is probably better represented as a directed acyclic graph.



The last token

- ✿ Finally, we haven't been talking about the “upper layers” of the transformer, where the token, having “absorbed” an enormous amount of context from surrounding tokens, is used to perform some task (like predicting the next token).
- ✿ Very briefly, one of the ways that we can think of contextual LMs working is by loading all the information necessary for the task into the last token.

The last token

- ✿ This resembles an important (but generally discarded) view of sentence-meaning associated with someone known only as “the commentator” (*vṛttikāra*) in *Mīmāṃsā*:

It is the final speech-sound, together with the traces produced by previous speech-sounds, that conveys the sentence-meaning.

*pūrvavarṇajanīṭasaṁskārasahitō 'ntyō varṇaḥ
pratyāyakaḥ.* (Śabara's *Mīmāṃsāsūtrabhāṣya* on
1.1.5)

The last token

- ✿ The final “speech-sound” can be compared to the final token (now usually a byte or subword token) — although not meaningful in itself, it “conveys” the meaning associated with the entire sentence, insofar as, by the time we reach the final speech-sound, all of the relational meanings have been computed and taken into account.
- ✿ Śālikanātha does not take this position (he criticizes it in *Vākyārthamāṭrkā*) but it may help in responding to some of the standard critiques of *anvitābhīdhāna* according to which one word suffices to express the entire sentence-meaning.

Thank you!

Br̥hatī p. 350 (on 1.1.24)

in one case a sentence-meaning is understood which is **related** (*anvita-*) to the meaning “white,” and in the next case, one from which the meaning “white” has been subtracted (*apōddhṛta-*) but which is **related** to the meaning “black,” and in the last, one from which the meaning “drive” has been subtracted and which is **related** to the meaning “tie up.”

Here the **relation** is one of co-presence of a particular **sentence-meaning** (s_1, s_2, s_3) and particular **word-meanings** (w_1, w_2, w_3).

Bṛhatī p. 350 (on 1.1.24)

From this it is understood that it is the word-meanings that are understood from the words, and not the sentence-meaning; the sentence-meaning is understood from a group (*samudāya-*) of word-meanings. [...] **Therefore the word-meanings are understood from the words, and the understanding of their mutual relation (*parasparānvayaḥ*) comes from the word-meanings.**

Here the **relation** obtains **between word-meanings themselves** (w_1, w_2, w_3).

Anvaya between word-meanings

Prabhākara will continue to make this equation repeatedly:

Don't word-meanings have a **connection** (*sambandha*) with the sentence-meaning, since the sentence-meaning is just their **relation** (*tadanvayasaiva vākyārthatvāt*)? (p. 375)

For we have already said that **the sentence-meaning is a mutual relation of the word-meanings** (*parasparānvayō hi padārthānām vākyārtha ity uktam*). (p. 381)

The *locus classicus* of *anvitābhidhāna*

SIDDHĀNTIN — It is that [i.e., the mutual relation] that practical activity through linguistic expressions and their meanings serves to convey. And in practical activity, the addition or subtraction [of word-meanings] happens only in accordance with their mutual relation. **Therefore expression is of meanings that are mutually related** (*vyatiṣaktārthābhidhāna*), since one understands the mutual relation only through what is mutually related.

PŪRVAPAKṢIN — What will happen then?

SIDDHĀNTIN — There is no linguistic expression that is uttered without expressing a relational meaning.

(pp. 383–386)