

Sample Problems for Quiz 1

with worked answers

CS 135: Computational Semantics – From Logic to Geometry

Brandeis University · Fall 2026

What this is. These are not the quiz questions. They are the same *kinds* of task on different data, drawn from the same material – so working them is practice rather than memorization. The quiz has four sections in this order, and so does this sheet.

How to use it. Work each problem with a pen before looking at the answer. Every number here is small enough to do by hand, and you may leave an answer as $\log_2$ of a fraction, or as an unsimplified fraction, wherever that is the exact form of it. **No calculator is needed on the quiz and none is permitted**, so do not use one here either.

If you can do every problem on this sheet, you are prepared. If one of them stops you, that is the thing to ask about before the quiz.

Given. These are printed on the quiz as well. You are not being tested on whether you memorized them.

$$\text{PMI}(w, c) = \log_2 \frac{p(w, c)}{p(w)p(c)}$$

$$\text{PPMI}(w, c) = \max(0, \text{PMI}(w, c))$$

$$\cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}$$

$$p_\alpha(c) \propto \text{count}(c)^\alpha$$

Useful for hand computation: the association ratio can be written entirely in counts, with no probabilities, as $\frac{p(w, c)}{p(w)p(c)} = \frac{N \cdot n(w, c)}{n(w)n(c)}$.

1. Sets, relations, and types

Throughout this section the domain is fixed:

$$\mathcal{D}_e = \{ \text{Mia}, \text{Noah}, \text{Omar}, \text{Jaco} \}$$

with coordinates in that order, so $\mathbf{e}_1 = \text{Mia}$ and $\mathbf{e}_4 = \text{Jaco}$ (a dog). Two predicates:

$$\text{teacher} = (1, 0, 1, 0), \quad \text{travels} = (1, 1, 1, 0)$$

and one relation:

Q (<i>trusts</i>)	Mia	Noah	Omar	Jaco
Mia	0	1	0	1
Noah	1	0	0	0
Omar	0	1	0	1
Jaco	0	0	0	0

$Q_{ij} = 1$ when individual i trusts individual j . Rows are the one who trusts.

A1. Compute $\text{teacher} \odot \text{travels}$ and say in English which set it denotes. Then compute $\text{teacher} \cdot \text{travels}$ and say what the number counts.

Answer. $\text{teacher} \odot \text{travels} = (1, 0, 1, 0)$, the set $\{\text{Mia}, \text{Omar}\}$ – the teachers who travel. The Hadamard product is intersection on indicator vectors.

$\text{teacher} \cdot \text{travels} = 1 + 0 + 1 + 0 = 2$. The inner product counts the members of the intersection, $|A \cap B|$. Note that these are two readings of the same four multiplications: the Hadamard product keeps them separate, the inner product adds them up.

A2. Evaluate *Every teacher travels* and *Every traveler teaches*, using only cardinalities. State the test you applied, and explain why the two sentences do not get the same answer.

Answer. The test for EVERY is $|A \cap B| = |A|$, where A is the restrictor – the noun – and B is the scope.
Every teacher travels: $A = \text{teacher}$, so $|A| = 2$ and $|A \cap B| = 2$. They are equal, so the sentence is **true**.
Every traveler teaches: now $A = \text{travels}$, so $|A| = 3$ while $|A \cap B|$ is still 2. They differ, so the sentence is **false**. Noah is the witness: he travels and does not teach.
 They differ because EVERY is not symmetric in its two arguments. The intersection count is shared, but the comparison is against the restrictor, and swapping the sentence around swaps which set that is.

A3. Evaluate *Some teacher travels* and *No teacher travels*. Give the test in each case.

Answer. SOME tests $|A \cap B| > 0$: here $2 > 0$, so **true**.
 NO tests $|A \cap B| = 0$: here $2 \neq 0$, so **false**.
 Unlike EVERY, both of these are symmetric – *Some traveler teaches* and *No traveler teaches* get the same answers, because the test looks only at the intersection and never at the restrictor.

A4. Compute $Q \mathbf{e}_{\text{Noah}}$ and $Q^\top \mathbf{e}_{\text{Noah}}$. Name each set in English, and say in one sentence why they are not the same set.

Answer. $Q \mathbf{e}_{\text{Noah}}$ picks out the second *column*, $(1, 0, 1, 0)$: the people who trust Noah, namely $\{\text{Mia}, \text{Omar}\}$.
 $Q^\top \mathbf{e}_{\text{Noah}}$ picks out the second *row*, $(1, 0, 0, 0)$: the people Noah trusts, namely $\{\text{Mia}\}$.
 The two indices of Q are different argument positions. Multiplying by Q fixes the second argument and asks who can fill the first; multiplying by $Q^\top$ does the reverse.

A5. Is *Omar trusts Jaco* true? Write the sentence as a bilinear form before you evaluate it. Then do the same for *Jaco trusts Omar*.

Answer. $\mathbf{e}_{\text{Omar}}^\top Q \mathbf{e}_{\text{Jaco}} = Q_{34} = 1$, so **true**.
 $\mathbf{e}_{\text{Jaco}}^\top Q \mathbf{e}_{\text{Omar}} = Q_{43} = 0$, so **false**.
 The point of writing the bilinear form is that the subject vector and the object vector enter on different sides of Q , which is the whole of what makes the two sentences different.

A6. Evaluate *Every teacher trusts Jaco*. You will need to build one of the two sets first.

Hint. A quantifier needs two sets of individuals. The scope here is not given to you directly – extract it from Q .

Answer. The scope set is “the individuals who trust Jaco”, which is the fourth column of Q : $Q \mathbf{e}_{\text{Jaco}} = (1, 0, 1, 0)$, namely $\{\text{Mia}, \text{Omar}\}$.
 Now apply the EVERY test with $A = \text{teacher} = (1, 0, 1, 0)$: $|A| = 2$ and $|A \cap B| = (1, 0, 1, 0) \cdot (1, 0, 1, 0) = 2$. Equal, so the sentence is **true**.
 This is the composite move worth practicing: a relation plus one argument gives a predicate, and a predicate is the kind of thing a quantifier can take.

A7. Give the semantic type of *trusts* and the tensor order of the object representing it. What order results from applying it to one entity, and what type is that result?

Answer. Type $\langle e, \langle e, t \rangle \rangle$, tensor order 2 – a matrix.
 Applying it to one entity contracts one index. The result has order 1, a vector, of type $\langle e, t \rangle$ – a predicate, exactly the kind of object *teacher* and *travels* already are.

A8. A sentence of the form *Noah trusts Mia* evaluates to a single number. In terms of tensor order, account for how two order-1 objects and one order-2 object produce an order-0 object.

Answer. Each application contracts one index. Starting from order 2 and applying two entities, each of order 1, removes both of the matrix’s indices: $2 \rightarrow 1 \rightarrow 0$. An order-0 tensor is a scalar, which is what a truth value is in this setting – a member of $\mathcal{V}(t) = \{0, 1\}$.
 Equivalently, in the bilinear form $\mathbf{u}^\top Q \mathbf{v}$ both of Q ’s indices are summed over, and nothing is left free.

2. The distributional pipeline

A term–context co-occurrence table over three words and three contexts. Margins and total are given, so you never have to add a row up yourself.

	c_1	c_2	c_3	row
w_1	150	25	25	200
w_2	25	50	25	100
w_3	25	25	50	100
col	200	100	100	$N = 400$

For reference while grading: every association ratio in this table is an exact fraction, and every PMI value except one is an exact integer.

$$\frac{N n(w, c)}{n(w) n(c)} = \begin{pmatrix} 3/2 & 1/2 & 1/2 \\ 1/2 & 2 & 1 \\ 1/2 & 1 & 2 \end{pmatrix} \implies \text{PMI} = \begin{pmatrix} \log_2 \frac{3}{2} & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \approx \begin{pmatrix} 0.585 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

B1. Compute $\text{PMI}(w_1, c_1)$ and $\text{PMI}(w_2, c_1)$. Leave each as $\log_2$ of a fraction, and give a decimal only where it is exact.

Answer. Using $\frac{N n(w, c)}{n(w) n(c)}$ throughout:

$$\frac{400 \cdot 150}{200 \cdot 200} = \frac{60000}{40000} = \frac{3}{2}, \text{ so } \text{PMI}(w_1, c_1) = \log_2 \frac{3}{2} \approx 0.585. \text{ Not exact; leave it as } \log_2 \frac{3}{2}.$$

$$\frac{200 \cdot 200}{400 \cdot 25} = \frac{40000}{10000} = \frac{4}{1} = 4, \text{ so } \text{PMI}(w_2, c_1) = \log_2 \frac{1}{4} = -2, \text{ exactly.}$$

B2. Compute the whole PMI row for w_2 . One of the three entries is exactly zero – say what that zero means about w_2 and that context.

Answer. The three ratios are $\frac{1}{2}$, 2 and 1, so the row is $(-1, +1, 0)$.

The zero sits at c_3 . It means w_2 and c_3 co-occur exactly as often as independence would predict: $p(c_3 | w_2) = p(c_3) = 1/4$. Knowing you are looking at w_2 tells you nothing at all about whether the context is c_3 . Note what the zero is *not*: it is not a statement that the pair is rare. The cell has a count of 25, the same count as several others. Zero PMI is about independence, not about frequency.

B3. Write out the PPMI matrix for the whole table. Then describe its shape in one sentence, and say what PPMI has destroyed in this particular table.

Answer.

$$\text{PPMI} = \begin{pmatrix} \log_2 \frac{3}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Every negative entry was clipped, and what survives is a single positive value per row – the matrix has become diagonal.

What was destroyed: the row for w_1 had *two* negative entries, -1 at c_2 and -1 at c_3 , each saying that w_1 occurs with that context only half as often as chance predicts. Both are now 0, and so is $\text{PMI}(w_2, c_3)$, which was a genuine zero. So “actively avoids” and “is indifferent to” have become indistinguishable. The usual defense of clipping is that negative PMI is estimated unreliably from sparse counts, and that is a claim about data quality rather than about meaning.

B4. Compute $\cos(w_2, w_3)$ on the PPMI rows. Then compute it on the *raw count* rows. The second one is an exact fraction; find it.

Hint. For the raw rows, notice that w_2 and w_3 have the same norm.

Answer. On PPMI. The rows are $(0, 1, 0)$ and $(0, 0, 1)$. Their dot product is 0, so $\cos(w_2, w_3) = 0$ – orthogonal.

On raw counts. The rows are $(25, 50, 25)$ and $(25, 25, 50)$.

$$w_2 \cdot w_3 = 625 + 1250 + 1250 = 3125, \quad \|w_2\|^2 = \|w_3\|^2 = 625 + 2500 + 625 = 3750$$

Since the norms are equal, the denominator is just 3750, and

$$\cos(w_2, w_3) = \frac{3125}{3750} = \frac{5}{6} \approx 0.833$$

B5. You have just computed $5/6$ and 0 for the same pair of words. Explain why the two numbers disagree so sharply, in terms of what each one is measuring.

Answer. The raw rows agree because both words are moderately frequent and both occur with all three contexts. Most of what the cosine is picking up is *shared frequency mass* – the two vectors point in similar directions largely because both have substantial entries everywhere, not because the words behave alike. PMI divides out exactly that expected-by-frequency component and keeps only the departure from chance. Once it is removed, w_2 and w_3 have nothing in common: w_2 's only above-chance context is c_2 , w_3 's is c_3 , and each is at chance or below everywhere else. The cosine is then 0.

So $5/6$ was largely a statement that both words are common, and 0 is a statement that their selectional preferences do not overlap. Those are different questions, and the pipeline is built to ask the second.

B6. Under PPMI, all three pairwise cosines in this table come out to 0. Is that a bug in the data, a bug in PPMI, or neither? What does it suggest about the next step in the pipeline?

Answer. Neither, exactly – it is PPMI working as designed on a table that is too small. Each row has one surviving context, and no two rows share one, so every pair is orthogonal and the similarity structure is empty.

This is the sparsity problem in miniature. On real data the same thing happens in a weaker form: PPMI matrices are very sparse, most pairs of words share no contexts at all, and a great many cosines come out at or near zero even for words that are related.

What it motivates is **dimensionality reduction** – factoring the PPMI matrix so that words which never co-occur with the same context can still land near each other because their contexts are themselves similar. Accept any answer naming SVD, a low-rank factorization, or dense embeddings generally.

B7. The context counts are $(200, 100, 100)$. Applying context distribution smoothing with $\alpha = 0.75$ gives $p_\alpha \approx (0.457, 0.272, 0.272)$, where the unsmoothed p was $(0.50, 0.25, 0.25)$. The PMI row for w_2 changes from $(-1, +1, 0)$ to approximately $(-0.87, +0.88, -0.12)$. Account for both the direction of the change and the reason anyone would want it.

Answer. What the exponent did. Raising each count to 0.75 and renormalizing *flattens* the context distribution, moving probability mass from common contexts toward rare ones. Here c_1 falls from 0.500 to 0.457 while c_2 and c_3 each rise from 0.250 to 0.272.

Why the PMI values move as they do. $p_\alpha(c)$ sits in the denominator, so a context whose probability went up now faces a larger denominator and its PMI is pushed *down*; a context whose probability went down has its PMI pushed *up*. That is exactly the pattern observed: c_1 rises from -1 to -0.87 , while c_2 falls from $+1$ to $+0.88$ and c_3 from 0 to -0.12 .

Why anyone wants it. PMI is biased in favor of rare contexts: a context seen only a few times can post a very high PMI simply because its marginal is tiny, and that score is not reliable. Smoothing inflates the marginal of rare contexts and so damps their scores. The cost is that genuinely informative rare contexts are damped too. The 0.75 exponent is an empirical compromise, not a derived constant.

3. Metrics and bilinear maps

C1. Let $\mathbf{u} = (2, 3)$ and $\mathbf{v} = (4, 1)$.

(a) Compute $\mathbf{u}^\top I \mathbf{v}$.

(b) Compute $\mathbf{u}^\top G \mathbf{v}$ for $G = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$, writing out all four terms of the sum before you add them.

Answer. (a) $2 \cdot 4 + 3 \cdot 1 = 11$.

(b) The bilinear form expands as $\mathbf{u}^\top G \mathbf{v} = u_1 G_{11} v_1 + u_1 G_{12} v_2 + u_2 G_{21} v_1 + u_2 G_{22} v_2$, giving

$$(2)(1)(4) + (2)(0)(1) + (3)(0)(4) + (3)(5)(1) = 8 + 0 + 0 + 15 = 23$$

Write the two zero terms out. The point of the exercise is to see where each entry of G lands in the sum, and the off-diagonal entries are the ones people lose track of.

C2. A query $\mathbf{q} = (3, 1)$ is compared against two candidates, $\mathbf{a} = (4, 1)$ and $\mathbf{b} = (1, 4)$.

(a) Under the ordinary dot product, which is returned?

(b) Find the range of g for which $G = \text{diag}(1, g)$ reverses that ranking, and verify one value of g by computing both scores.

(c) What happens at the boundary value of g ?

Answer. (a) $\mathbf{q} \cdot \mathbf{a} = 12 + 1 = 13$ and $\mathbf{q} \cdot \mathbf{b} = 3 + 4 = 7$. $\mathbf{a}$ is returned.

(b) With $G = \text{diag}(1, g)$ we get $\mathbf{q}^\top G \mathbf{a} = 12 + g$ and $\mathbf{q}^\top G \mathbf{b} = 3 + 4g$. Then $\mathbf{b}$ wins when

$$3 + 4g > 12 + g \iff 3g > 9 \iff g > 3$$

Verifying at $g = 4$: $\mathbf{q}^\top G \mathbf{a} = 12 + 4 = 16$ and $\mathbf{q}^\top G \mathbf{b} = 3 + 16 = 19$. Reversed.

(c) At exactly $g = 3$ both scores are 15 – the two candidates tie, and the ranking is undefined. Every ranking reversal produced by a continuous change in G passes through such a tie.

The lesson is that “nearest” is a statement about G and not about the vectors alone. Changing the metric changes the answer to a retrieval query without changing a single stored vector.

C3. Let $\mathbf{u} = (2, 1, 0)$ and $\mathbf{v} = (1, 0, 3)$. Write out $\mathbf{u}\mathbf{v}^\top$ as a 3×3 matrix, state its rank, and say why the rank had to come out that way.

Answer.

$$\mathbf{u}\mathbf{v}^\top = \begin{pmatrix} 2 & 0 & 6 \\ 1 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{rank} = 1$$

Every row is a scalar multiple of $\mathbf{v}^\top = (1, 0, 3)$ – the multipliers are 2, 1 and 0, which are the entries of $\mathbf{u}$. So the row space is spanned by a single vector. Equivalently, every column is a multiple of $\mathbf{u}$.

An outer product of two nonzero vectors always has rank exactly 1. That is what it means for a matrix to factor into a property of the first argument times a property of the second, with no interaction between them.

C4. Still with $\mathbf{u} = (2, 1, 0)$ and $\mathbf{v} = (1, 0, 3)$: compute $\mathbf{v}\mathbf{u}^\top$, and state its relationship to $\mathbf{u}\mathbf{v}^\top$. Why does this matter for representing a relation?

Answer.

$$\mathbf{v}\mathbf{u}^\top = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 0 & 0 \\ 6 & 3 & 0 \end{pmatrix} = (\mathbf{u}\mathbf{v}^\top)^\top$$

The two orders give transposes of one another, and they are different matrices whenever $\mathbf{u}$ and $\mathbf{v}$ are not parallel.

Why it matters: the order of the two arguments is recoverable from the result. If binding two things produced the same object either way, there would be no way to say which was the subject and which the object. The transpose relationship is a record of which came first, and it is the cleanest such record available.

C5. Take $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (0, 1)$.

(a) For $G = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, compute $\mathbf{u}^\top G \mathbf{v}$ and $\mathbf{v}^\top G \mathbf{u}$.

(b) Do the same for $G' = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$.

(c) State the general fact your two answers illustrate.

Answer. (a) $\mathbf{u}^\top G \mathbf{v} = G_{12} = 2$ and $\mathbf{v}^\top G \mathbf{u} = G_{21} = 2$. **Equal.**

(b) $\mathbf{u}^\top G' \mathbf{v} = G'_{12} = 2$ and $\mathbf{v}^\top G' \mathbf{u} = G'_{21} = 0$. **Not equal.**

(c) In general $\mathbf{u}^\top G \mathbf{v} = \mathbf{v}^\top G^\top \mathbf{u}$, because the quantity is a scalar and so equals its own transpose. When G is symmetric, $G^\top = G$ and the two orders agree: **a symmetric form cannot distinguish $(\mathbf{u}, \mathbf{v})$ from $(\mathbf{v}, \mathbf{u})$** . Choosing these basis vectors makes the point visible, since each form reads off a single entry of G . This is the reason an asymmetric matrix is needed to represent a relation with two distinct argument positions.

C6. Let $\mathbf{u} = (1, 2)$ and $G = \text{diag}(4, 1)$. Compute $\|\mathbf{u}\|^2$ and $\|\mathbf{u}\|_G^2 = \mathbf{u}^\top G \mathbf{u}$, and say in one sentence what a change of G does besides reranking neighbors.

Answer. $\|\mathbf{u}\|^2 = 1 + 4 = 5$ and $\|\mathbf{u}\|_G^2 = (1)(4)(1) + (2)(1)(2) = 4 + 4 = 8$.

A change of G changes *lengths*, not only the comparison between two different vectors. G supplies the whole geometry – which directions count as long, which pairs count as orthogonal, and therefore what the angles are – and the identity matrix is one choice among many rather than the absence of a choice.

C7. Return to the matrix Q (*trusts*) from Section 1. Is Q symmetric? Name a specific pair of individuals that settles the question, and then name a different pair for which Q happens to be symmetric. What does the asymmetry represent linguistically?

Answer. Q is **not** symmetric. The pair that settles it: $Q_{14} = 1$ (Mia trusts Jaco) while $Q_{41} = 0$ (Jaco does not trust Mia). Any such pair is fine.

A pair where it happens to agree: $Q_{12} = Q_{21} = 1$ – Mia and Noah trust each other. A matrix can have symmetric entries here and there without being a symmetric matrix; symmetry is a claim about every pair at once.

Linguistically the asymmetry is **argument structure**. The one who trusts and the one trusted are distinct argument positions, and a symmetric matrix cannot tell them apart – it could only ever express mutual trust.

4. Essay practice

On the quiz there is one short essay, chosen from two prompts, worth 10 points and expected to run about a bluebook page. It is marked on whether the argument is correct and specific, not on length.

The three prompts below are on the two themes the quiz essay draws from. The notes under each are a marking guide, not a model answer – a student making the same points in their own

words should receive full credit.

D1. A classmate says: “cosine similarity measures how similar two words are in meaning.” Give the most precise statement you can of what cosine actually measures in the distributional regime, and name two things it does not measure.

Answer. A precise statement: cosine measures the angle between two rows of a weighted co-occurrence matrix – that is, the degree to which two words have *similar distributions over contexts*, after whatever weighting was applied and with magnitude discarded by the normalization.

Things it does not measure, any two of:

- **Relatedness of the right kind.** Antonyms share contexts heavily, so hot and cold score high. Distributional similarity does not distinguish synonymy from antonymy, nor from mere topical association.
- **Frequency, by construction.** Normalizing by the norms throws away magnitude, so a word seen 10 times and a word seen 10,000 times can sit at cosine 1.
- **Any asymmetric relation.** Cosine is symmetric, so it cannot express entailment, hyponymy, or any other directed relation – a point this course returns to.
- **Truth or reference.** There are no individuals in this regime and nothing to be true of.

Full credit needs the phrase “similar distributions over contexts” or a clear equivalent, plus two specific limitations. An answer that treats “similar” as self-explanatory is worth about half.

D2. In Session 6 the inner product $\mathbf{u} \cdot \mathbf{v}$ computed a cardinality, and from it a truth value. In Session 7 the same arithmetic computed a similarity score. Nothing about the formula changed. Account for the difference, and say what was given up in the exchange.

Answer. The expected answer names the change of **regime**, and the way to name it is to say what the coordinates are.

In Session 6 the coordinates are *individuals* in a model, and the vectors are indicator vectors. The inner product counts $|A \cap B|$, and comparing that count against $|A|$ gives a truth value. Everything is exact and everything is interpretable one coordinate at a time.

In Session 7 the coordinates are *contexts*, and the entries are counts or weights estimated from a corpus. The same sum is now a graded score. There is no model behind it, no domain of individuals, and nothing to threshold against.

Given up, any two of: truth conditions, and with them entailment; the interpretability of individual coordinates; reference to a domain; exactness – the second regime is approximate by construction.

Gained: similarity, of which the Session 6 account has no notion whatsoever, and coverage of a vocabulary no model-builder could enumerate by hand.

Full credit needs the regime distinction stated explicitly, in terms of what the coordinates denote, plus at least one specific loss. A fluent answer that never says what the coordinates *are* is worth about half.

D3. You are told only that a proposed composition operation takes two word vectors to a phrase vector and is commutative. Without knowing anything else about it, name a linguistic phenomenon it is guaranteed to get wrong, and explain why this is a matter of algebra rather than a failure of training.

Answer. Guaranteed wrong: any phenomenon where **word order carries meaning**. If $f(\mathbf{a}, \mathbf{b}) = f(\mathbf{b}, \mathbf{a})$ then the predicted vector for AB and the predicted vector for BA are the *same vector*, so their cosine is exactly 1.000.

Good instances: boathouse against houseboat; dog house against house dog; subject and object of a transitive verb; alleged former president against former alleged president.

Why it is algebra. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ and $\mathbf{a} \odot \mathbf{b} = \mathbf{b} \odot \mathbf{a}$ hold for all vectors whatsoever. The conclusion follows from the commutativity alone, with no reference to the corpus, the dimensionality, or how the vectors were obtained. No amount of training moves a number that is 1 by definition. Meanwhile the *observed* vectors for the two phrases are not at cosine 1, which is how we know the prediction is wrong rather than merely unconfirmed.

What would be needed: a **non-commutative** operation. Credit matrix multiplication, the outer product $\mathbf{ab}^\top$ – whose transpose is $\mathbf{ba}^\top$ – or concatenation, with the caveat that concatenation passes the order test only by landing in $\mathbb{R}^{2d}$ rather than $\mathbb{R}^d$, and so changes the problem instead of solving it.

A checklist, if you prefer one. By the day of the quiz you should be able to, without notes:

- turn a set into a vector and back, and say what $\odot$ and $\cdot$ each compute;
- evaluate EVERY, SOME and NO from two cardinalities, and say which of them care about the restrictor;
- get a row or a column out of a relation matrix, and say in English which set you have;
- write a transitive sentence as a bilinear form;
- give the type and tensor order of a predicate, a relation, and a sentence, and say what application does to the order;
- compute a PMI cell from counts alone;
- say what clipping at zero fixes and what it costs;
- compute a cosine by hand, and say when it is undefined;
- say why a raw-count cosine and a PPMI cosine can disagree;
- state what context distribution smoothing does to the denominator, and in which direction;
- expand $\mathbf{u}^\top G \mathbf{v}$ into its individual terms;
- build a G that reverses a given ranking, and verify it;
- write an outer product, give its rank, and say why;
- say what a symmetric G cannot do, and why that matters for relations.

Questions are welcome in class or by email before the quiz.