

The Coproduct Behind the Product

Bilinear forms, Hopf algebras, and the geometric product

James Pustejovsky

Brandeis University
Department of Computer Science
CS 135: Computational Semantics

October 16, 2026

Roadmap

I	Two halves of a pair	what a form reads; what one number cannot keep
II	Splitting	coproducts of strings, bags, oriented sets; the axioms
III	Pairing one leg	contraction from the coproduct; queries; two slots
IV	Split, pair, join	the geometric product from any bilinear form
V	The other half of the form	same algebra, different split; attention; RoPE
VI	Other coproducts, and close	copying; Merge; summary; exercises; reading

Sessions 11 and 13 built the product of two vectors from a bilinear form. Today's question is how a form on vectors becomes a product of everything – planes, volumes, event records. The answer runs through one map, the coproduct, which lists the ways to cut an object in two.

Where We Are

The running example of Sessions 11 and 13: an attention head with form $G = W_Q^\top W_K$, and two tokens $\mathbf{x}_1 = (3, 1)$ and $\mathbf{x}_2 = (1, 3)$.

$$G = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}}_{S, \text{ symmetric}} + \underbrace{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}}_{A, \text{ antisymmetric}}$$

	1 $\rightarrow$ 2	2 $\rightarrow$ 1	
the score $\mathbf{x}^\top G \mathbf{y}$	24	8	what the head computes
$\mathbf{x}^\top S \mathbf{y}$	16	16	how alike
$\mathbf{x}^\top A \mathbf{y}$	+8	-8	which way

Session 13 declared only the quadratic form $Q(\mathbf{x}) = \mathbf{x}^\top G \mathbf{x} = (x_1 + x_2)^2$, which sees only S , and asked for a product with $\mathbf{v}\mathbf{v} = Q(\mathbf{v})$. Polarization fixed the symmetric half, and the wedge carried the rest:

$$\mathbf{x}_1 \mathbf{x}_2 = 16 + 8 \mathbf{e}_1 \wedge \mathbf{e}_2, \quad \mathbf{x}_2 \mathbf{x}_1 = 16 - 8 \mathbf{e}_1 \wedge \mathbf{e}_2,$$

and in a basis every product reduces by two moves: swap neighbors for a sign, cancel equal neighbors for a number.

Three Questions for Today

- ① **Is the coproduct the symmetric/antisymmetric split in disguise?** A Hopf algebra comes with a map Δ , the coproduct. On two vectors it gives

$$\Delta(\mathbf{u} \wedge \mathbf{v}) = (\mathbf{u} \wedge \mathbf{v}) \otimes 1 + \mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u} + 1 \otimes (\mathbf{u} \wedge \mathbf{v}),$$

and the middle term looks like the antisymmetric half of a matrix. Is that what it is?

- ② **How do you multiply a plane by a plane?** For two vectors, polarization fixed the symmetric half of the product and the wedge kept the rest; longer products reduce by two moves in a basis. Is there one rule that starts from a bilinear form and multiplies any two multivectors, with no basis?
- ③ **What would the antisymmetric half do if we kept it?** The head learned all of G ; the algebra was told only Q . What changes if the product is built from G itself?

One construction answers all three: *split, pair, join*.

Part I

Two Halves of a Pair

A Bilinear Form Reads the Dyad

Two vectors give a **dyad**, the matrix of all their products, $\mathbf{xy}^\top$; as a tensor it is $\mathbf{x} \otimes \mathbf{y}$. A bilinear form is a weighted sum of the dyad's entries:

$$\mathbf{x}^\top G \mathbf{y} = \sum_{i,j} G_{ij} x_i y_j = \langle G, \mathbf{xy}^\top \rangle, \quad \langle M, N \rangle = \sum_{ij} M_{ij} N_{ij}$$

Split the dyad of the two tokens the same way as the form:

$$\mathbf{x}_1 \mathbf{x}_2^\top = \begin{pmatrix} 3 & 9 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 5 \\ 5 & 3 \end{pmatrix} + \begin{pmatrix} 0 & 4 \\ -4 & 0 \end{pmatrix}$$

Symmetric and antisymmetric matrices are orthogonal under $\langle \cdot, \cdot \rangle$, so **each half of the form reads only the matching half of the dyad**:

$\langle \textcolor{green}{S}, \text{ symmetric half of the dyad} \rangle$	$3 + 5 + 5 + 3$	$= 16$
$\langle \textcolor{red}{A}, \text{ antisymmetric half of the dyad} \rangle$	$4 + 4$	$= 8$
$\langle G, \mathbf{x}_1 \mathbf{x}_2^\top \rangle$	$3 + 18 + 0 + 3$	$= 24$

The Antisymmetric Half Reads a Plane

Pair the (i, j) and (j, i) terms of $\mathbf{x}^\top A \mathbf{y}$, using $A_{ji} = -A_{ij}$:

$$\mathbf{x}^\top A \mathbf{y} = \sum_{i < j} A_{ij} (x_i y_j - x_j y_i) = \sum_{i < j} A_{ij} [\mathbf{x} \wedge \mathbf{y}]_{ij}$$

The numbers $x_i y_j - x_j y_i$ are the coordinates of the bivector $\mathbf{x} \wedge \mathbf{y}$, one per coordinate plane. So **the antisymmetric half of a form is a linear reading of the plane its two arguments span.**

- **In two dimensions** there is one plane, and $\mathbf{x}^\top A \mathbf{y} = A_{12} [\mathbf{x} \wedge \mathbf{y}]_{12}$. The head has $A_{12} = 1$, so its 8 is the signed area $[\mathbf{x}_1 \wedge \mathbf{x}_2]_{12} = 3 \cdot 3 - 1 \cdot 1$. That is why Session 13's Clifford product reproduced it.
- **In d dimensions** A weighs $\binom{d}{2}$ planes separately: 294,528 of them for $d = 768$.
- **Rotary position embeddings** (Session 14), plane by plane: a query and a key whose positions differ by δ score $\cos(\delta\theta) (\mathbf{q} \cdot \mathbf{k}) - \sin(\delta\theta) [\mathbf{q} \wedge \mathbf{k}]_{12}$. Distance moves weight between the dot and the plane.

What One Number Cannot Keep

A bilinear form adds its two readings: $24 = 16 + 8$. From 24 alone no one can say how much was likeness and how much was direction; the reverse pair gives $8 = 16 - 8$.

object	what it holds	for $(\mathbf{x}_1, \mathbf{x}_2)$
the dyad $\mathbf{x} \otimes \mathbf{y}$	all n^2 products, in order	$\begin{pmatrix} 3 & 9 \\ 1 & 3 \end{pmatrix}$
the geometric product $\mathbf{x}\mathbf{y}$	one number and the whole plane	$16 + 8\mathbf{e}_1 \wedge \mathbf{e}_2$
the score $\mathbf{x}^\top G \mathbf{y}$	one number	24

The geometric product keeps the two readings in different places, grade 0 and grade 2. From the symmetric half of the dyad it keeps one number, the declared form applied to it; the antisymmetric half it keeps whole, as a plane.

For two vectors that was enough. To multiply planes and volumes, a form has to be handed the right pieces to read, and something has to list the pieces. That is the job of the coproduct.

Part II

Splitting

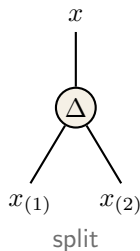
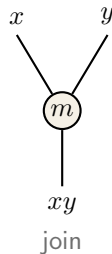
Products Join; Coproducts Split

A **product** takes two things to one, $m : A \otimes A \rightarrow A$. A **coproduct** goes the other way. It takes one thing to the list of all the ways of cutting it in two, written as a sum of pairs:

$$\Delta(x) = \sum x_{(1)} \otimes x_{(2)}$$

This is **Sweedler's notation**: $x_{(1)}$ is a left piece, $x_{(2)}$ the matching right piece, and the sum runs over the cuts. Joni and Rota (1979) used this idea to organize large parts of combinatorics, where a coproduct records how an object breaks into parts.

The $\otimes$ keeps the two pieces apart and in order. It does not multiply them.



Cutting a Sentence

A string of words can be cut at any of its gaps, including the two ends. Write 1 for the empty string:

the	dog	barked
0	1	2

$$\Delta(\text{the dog barked}) = 1 \otimes \text{the dog barked} + \text{the} \otimes \text{dog barked} \\ + \text{the dog} \otimes \text{barked} + \text{the dog barked} \otimes 1$$

This is the **deconcatenation** coproduct, one term per cut, and chart parsing already relies on it. The inside algorithm for a grammar in Chomsky normal form sums a span's score over exactly these cuts, the ones that leave two nonempty pieces:

$$\beta_A(w) = \sum_{w=w_{(1)}w_{(2)}} \sum_{A \rightarrow BC} P(A \rightarrow BC) \beta_B(w_{(1)}) \beta_C(w_{(2)})$$

Split, score each piece, combine. A sum of the form $\sum f(w_{(1)}) g(w_{(2)})$ is called a **convolution** over the coproduct; semiring parsing (Goodman 1999) runs the same recursion with other ways of combining scores.

The Rules a Coproduct Obeys

rule	formula	in words
coassociative	$(\Delta \otimes \text{id}) \Delta = (\text{id} \otimes \Delta) \Delta$	cutting into three pieces does not depend on which cut comes first; the dog barked has 10 three-piece cuttings either way
counit	$(\varepsilon \otimes \text{id}) \Delta = \text{id} = (\text{id} \otimes \varepsilon) \Delta$	splitting off nothing changes nothing; ε keeps only the empty piece
compatible	$\Delta(xy) = \Delta(x) \Delta(y)$	to cut a product, cut each factor and multiply the pieces

An algebra with a coproduct obeying all three is a **bialgebra**. Add one more map, the antipode, and it is a **Hopf algebra** (Sweedler 1969). The next two slides apply the third rule to two familiar algebras, where it does all the work.

For strings, contiguous cuts are compatible with *shuffling* as the product; concatenation pairs with cutting out subsequences. Both give Hopf algebras.

Bags of Words: The Symmetric Algebra

A **bag of words** forgets order and counts repeats. Bags are the monomials of the **symmetric algebra** $\text{Sym}(V)$: polynomials in word vectors, with $uv = vu$. Two rules fix the coproduct:

- 1 a single word goes left or right: $\Delta(w) = w \otimes 1 + 1 \otimes w$;
- 2 compatibility: $\Delta(xy) = \Delta(x) \Delta(y)$.

$$\begin{aligned}\Delta(\text{dog barked}) &= (\text{dog} \otimes 1 + 1 \otimes \text{dog})(\text{barked} \otimes 1 + 1 \otimes \text{barked}) \\ &= \text{dog barked} \otimes 1 + \text{dog} \otimes \text{barked} + \text{barked} \otimes \text{dog} + 1 \otimes \text{dog barked}\end{aligned}$$

Every way to divide the bag into two bags. Repeats are counted:

$$\Delta(\text{dog}^2) = \text{dog}^2 \otimes 1 + 2 \text{dog} \otimes \text{dog} + 1 \otimes \text{dog}^2, \quad \Delta(w^k) = \sum_j \binom{k}{j} w^j \otimes w^{k-j}$$

The 2 counts which copy goes left. The middle of $\Delta(uv)$ is $u \otimes v + v \otimes u$: symmetric, because in a bag a swap costs nothing.

A document in a bag-of-words model is one such monomial: order gone, counts kept.

Oriented Sets: The Exterior Algebra

In the **exterior algebra** $\Lambda(V)$, $\mathbf{u} \wedge \mathbf{v} = -\mathbf{v} \wedge \mathbf{u}$ and $\mathbf{u} \wedge \mathbf{u} = 0$. Use the same two rules, with one change: moving a vector past a vector costs a sign, also between tensor factors,

$$(1 \otimes \mathbf{u})(\mathbf{v} \otimes 1) = -\mathbf{v} \otimes \mathbf{u}.$$

$$\begin{aligned}\Delta(\mathbf{u} \wedge \mathbf{v}) &= (\mathbf{u} \otimes 1 + 1 \otimes \mathbf{u})(\mathbf{v} \otimes 1 + 1 \otimes \mathbf{v}) \\ &= (\mathbf{u} \wedge \mathbf{v}) \otimes 1 + \mathbf{u} \otimes \mathbf{v} + (1 \otimes \mathbf{u})(\mathbf{v} \otimes 1) + 1 \otimes (\mathbf{u} \wedge \mathbf{v}) \\ &= (\mathbf{u} \wedge \mathbf{v}) \otimes 1 + \mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u} + 1 \otimes (\mathbf{u} \wedge \mathbf{v})\end{aligned}$$

The middle term is antisymmetric because the swap costs a sign. In the bag the swap cost nothing, and the middle term was $u \otimes v + v \otimes u$. Same rules, one sign.

Check: $\Delta(\mathbf{u} \wedge \mathbf{u}) = 0 + \mathbf{u} \otimes \mathbf{u} - \mathbf{u} \otimes \mathbf{u} + 0 = 0$, as it must be.

With a sign on every swap, $\Lambda(V)$ is a Hopf algebra in the graded sense, often called a Hopf superalgebra.

Strings, Bags, Oriented Sets

	strings , $T(V)$	bags , $\text{Sym}(V)$	oriented sets , $\Lambda(V)$
order of words	kept	forgotten	kept up to sign
a repeated word	kept	counted	gives 0
size at length k , from n words	n^k	$\binom{n+k-1}{k}$	$\binom{n}{k}$
$n = 4, k = 2$	16	10	6
middle of Δ , two words	$u \otimes v$ (one cut)	$u \otimes v + v \otimes u$	$u \otimes v - v \otimes u$
in NLP	sentences, n -grams, TPR	bag-of-words models	role-filler planes, extension blades

One idea, three times: a coproduct lists the ways to take an object apart. What changes from column to column is what counts as the same object, and so what a swap costs.

The last column is the one geometric algebra is built on: Session 13's Clifford algebra has the same sizes $\binom{n}{k}$, grade by grade.

Splitting a Plurality

Take three dogs as orthonormal atoms $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$. In the position paper's treatment of quantification (§8.6), a predicate's extension is the blade of its atoms, and **grade is cardinality**: $D = \mathbf{d}_1 \wedge \mathbf{d}_2 \wedge \mathbf{d}_3$ has grade $|\text{dog}| = 3$. Its coproduct lists every way to divide the three dogs into two groups:

$$\begin{aligned}\Delta(D) = D \otimes 1 + & \underbrace{\mathbf{d}_{12} \otimes \mathbf{d}_3 + \mathbf{d}_{23} \otimes \mathbf{d}_1 + \mathbf{d}_{31} \otimes \mathbf{d}_2}_{\text{two dogs, one left over}} \\ & + \underbrace{\mathbf{d}_1 \otimes \mathbf{d}_{23} + \mathbf{d}_2 \otimes \mathbf{d}_{31} + \mathbf{d}_3 \otimes \mathbf{d}_{12}}_{\text{one dog, two left over}} + 1 \otimes D\end{aligned}$$

where $\mathbf{d}_{12} = \mathbf{d}_1 \wedge \mathbf{d}_2$, and so on. With the cyclic order ($\mathbf{d}_{31}$ rather than $\mathbf{d}_{13}$) every sign is $+$.

- $2^3 = 8$ terms; the slice with k dogs on the left has $\binom{3}{k}$ of them.
- The $(2, 1)$ slice lists every two-dog part of the plurality, each with its complement: the parts that *two of the three dogs* ranges over.

Pluralities ordered by *part of* form a lattice (Link 1983), and the coproduct enumerates the parts of one plurality. Whether partitives should be computed this way is an open question.

What Makes It Hopf: The Antipode

A Hopf algebra has one more map, the **antipode** S . It undoes the product in this sense: apply S to the left piece of every cut, multiply back, and everything cancels except the empty object.

$$\sum S(x_{(1)}) x_{(2)} = \varepsilon(x) 1 = \sum x_{(1)} S(x_{(2)})$$

algebra	antipode
oriented sets, $\Lambda(V)$	$S(X) = (-1)^k X$ on grade k : the grade involution , which splits a multivector into its even and odd parts
bags, $\text{Sym}(V)$	$S(w_1 \cdots w_k) = (-1)^k w_1 \cdots w_k$
strings, cut into subsequences	$S(w_1 \cdots w_k) = (-1)^k w_k \cdots w_1$: a signed reversal

Check on $\mathbf{u} \wedge \mathbf{v}$: $S(\mathbf{u} \wedge \mathbf{v}) + S(\mathbf{u}) \wedge \mathbf{v} - S(\mathbf{v}) \wedge \mathbf{u} + \mathbf{u} \wedge \mathbf{v} = \mathbf{u} \wedge \mathbf{v} - \mathbf{u} \wedge \mathbf{v} + \mathbf{v} \wedge \mathbf{u} + \mathbf{u} \wedge \mathbf{v} = 0$. ✓

The rest of the lecture needs only the coproduct. The antipode is what separates a Hopf algebra from a bialgebra; the name honors Heinz Hopf.

A Tempting Shortcut

The shortcut

“ $\Delta(\mathbf{u} \wedge \mathbf{v})$ splits $\mathbf{u} \wedge \mathbf{v}$ into a symmetric part and an antisymmetric part, the way a matrix splits into $S + A$.”

the split of a dyad	the coproduct
acts on $\mathbf{u} \otimes \mathbf{v}$, in $V \otimes V$	acts on $\mathbf{u} \wedge \mathbf{v}$, in $\Lambda^2 V$
two halves, sorted by what a swap does	a list of cuts, of sizes $(2, 0)$, $(1, 1)$, $(0, 2)$
keeps the symmetric half	never sees it: the wedge already dropped it

What is right about it. The $(1, 1)$ slice, $\mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u}$, is the wedge written back as an antisymmetric matrix, $\mathbf{u}\mathbf{v}^\top - \mathbf{v}\mathbf{u}^\top$: Session 11’s dictionary between bivectors and antisymmetric matrices.

What it misses. The number $\mathbf{u} \cdot \mathbf{v}$ is not a piece of $\mathbf{u} \otimes \mathbf{v}$; it is what a bilinear form returns when it pairs two legs. [The coproduct supplies legs, and a form does the pairing.](#)

Part III

Pairing One Leg

Contraction Is Pairing One Leg

To contract a vector $\mathbf{a}$ into X : cut X every way, pair $\mathbf{a}$ with the left piece, keep the right piece.

$$\mathbf{a} \rfloor X = \sum g(\mathbf{a}, X_{(1)}) X_{(2)}, \quad X \rfloor \mathbf{a} = \sum X_{(1)} g(X_{(2)}, \mathbf{a})$$

Here g is the declared form, and $g(\mathbf{a}, \cdot) = 0$ on any piece that is not a vector. In

$\Delta(\mathbf{u} \wedge \mathbf{v}) = (\mathbf{u} \wedge \mathbf{v}) \otimes 1 + \mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u} + 1 \otimes (\mathbf{u} \wedge \mathbf{v})$ only the middle terms have a vector on the left:

$$\mathbf{a} \rfloor (\mathbf{u} \wedge \mathbf{v}) = g(\mathbf{a}, \mathbf{u}) \mathbf{v} - g(\mathbf{a}, \mathbf{v}) \mathbf{u}, \quad (\mathbf{u} \wedge \mathbf{v}) \rfloor \mathbf{a} = \mathbf{u} g(\mathbf{v}, \mathbf{a}) - \mathbf{v} g(\mathbf{u}, \mathbf{a})$$

These are Session 14's formulas, and **the minus sign is the coproduct's sign**.

Into a volume. The $(1, 2)$ slice of $\Delta(\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3)$ is $\mathbf{e}_1 \otimes \mathbf{e}_{23} + \mathbf{e}_2 \otimes \mathbf{e}_{31} + \mathbf{e}_3 \otimes \mathbf{e}_{12}$, so

$$\mathbf{a} \rfloor \mathbf{e}_{123} = a_1 \mathbf{e}_{23} + a_2 \mathbf{e}_{31} + a_3 \mathbf{e}_{12}$$

Each direction goes to the plane perpendicular to it: contraction into the volume is **duality**, as in Session 14, now read off a list of cuts.

A Query Meets Every Leg

Session 13's record of *John loves Mary*: roles $\mathbf{r}_{\text{ag}} = \mathbf{e}_1$ and $\mathbf{r}_{\text{th}} = \mathbf{e}_2$; fillers $\text{john} = \mathbf{e}_3$ and $\text{mary} = 0.6\mathbf{e}_3 + 0.8\mathbf{e}_4$, with cosine 0.6.

$$\mathbf{E} = \mathbf{r}_{\text{ag}} \wedge \text{john} + \mathbf{r}_{\text{th}} \wedge \text{mary}$$

$$\text{middle of } \Delta(\mathbf{E}) = \mathbf{r}_{\text{ag}} \otimes \text{john} - \text{john} \otimes \mathbf{r}_{\text{ag}} + \mathbf{r}_{\text{th}} \otimes \text{mary} - \text{mary} \otimes \mathbf{r}_{\text{th}}$$

query	the legs it meets	answer
$\mathbf{r}_{\text{ag}} \downharpoonright \mathbf{E}$: who is the agent?	left legs; only $\mathbf{r}_{\text{ag}} \cdot \mathbf{r}_{\text{ag}} = 1$ is nonzero	john, exactly
$\mathbf{E} \downharpoonright \text{mary}$: what role does Mary fill?	right legs; $\text{john} \cdot \text{mary} = 0.6$ and $\text{mary} \cdot \text{mary} = 1$	$0.6\mathbf{r}_{\text{ag}} + \mathbf{r}_{\text{th}}$, mixed
$\mathbf{E} \downharpoonright \text{mary}^*$, with $\text{mary}^* = 1.25\mathbf{e}_4$	right legs; $\text{mary}^* \cdot \text{john} = 0$ and $\text{mary}^* \cdot \text{mary} = 1$	$\mathbf{r}_{\text{th}}$, exactly

The coproduct lists every leg a query will meet. Orthonormal roles keep the role legs clean, so filler queries are exact. Fillers are not orthonormal, so a role query needs their **dual basis**, here mary^* : the position paper's propositions on querying (§5.6).

Why a Relation Needs Two Slots, Again

In a single entity space, *John loves himself* would be $\text{john} \wedge \text{john} = 0$. There is nothing for the coproduct to cut, and no leg for a query to meet.

FGA gives each argument its own isometric copy of the entity space, a subject slot V_E^s and an object slot V_E^o , entered through ι_s and ι_o . Both legs now survive, and their subspaces say which is which:

$$\text{middle of } \Delta(\iota_s(\text{john}) \wedge \iota_o(\text{john})) = \iota_s(\text{john}) \otimes \iota_o(\text{john}) - \iota_o(\text{john}) \otimes \iota_s(\text{john})$$

The head's whole form, stored across the slots as $L = \sum_{i,k} G_{ik} \mathbf{e}_i^s \wedge \mathbf{e}_k^o$, is read by two queries, one on each leg:

$$\iota_s(\mathbf{x}) \rfloor (L \llcorner \iota_o(\mathbf{y})) = \mathbf{x}^\top G \mathbf{y}$$

store	$(\mathbf{x}_1, \mathbf{x}_2)$	$(\mathbf{x}_2, \mathbf{x}_1)$	$(\mathbf{x}_1, \mathbf{x}_1)$
the symmetric form S , all that Q can see	16	16	16
a bivector in one space, holding A	8	-8	0
across two slots, holding G	24	8	16

In-Class Exercise 1

- 1 Write $\Delta(\text{the old dog})$ with contiguous cuts. How many terms are there, and how many three-piece cuttings?
- 2 In $\text{Sym}(V)$, compute $\Delta(\text{salt}^2 \text{ pepper})$, the bag salt, pepper, salt. Explain each coefficient as a choice of copies.
- 3 From the coproduct of $e_1 \wedge e_2 \wedge e_3$, compute $e_2 \rfloor e_{123}$ and $e_{123} \rfloor e_3$.
- 4 ★ In the record $\mathbf{E}$ of *John loves Mary*, compute $\text{john} \rfloor \mathbf{E}$. Where does each term come from, and what would you query with instead?

Six minutes. Item 4 is the starred one.

In-Class Exercise 1: Answers

1. $1 \otimes \text{the old dog} + \text{the} \otimes \text{old dog} + \text{the old} \otimes \text{dog} + \text{the old dog} \otimes 1$: four terms, one per gap. Three pieces need two cuts, possibly at the same gap: $\binom{4}{2} + 4 = 10$.

2. Write s for salt and p for pepper. Then $\Delta(s^2p) = (s \otimes 1 + 1 \otimes s)^2 (p \otimes 1 + 1 \otimes p)$, which is

$$s^2p \otimes 1 + s^2 \otimes p + 2sp \otimes s + 2s \otimes sp + p \otimes s^2 + 1 \otimes s^2p.$$

The 2s count which of the two salts goes left. The coefficients add to $2^3 = 8$, one for each way of sending the three tokens left or right.

3. From $\mathbf{e}_2 \otimes \mathbf{e}_{31}$: $\mathbf{e}_2 \rfloor \mathbf{e}_{123} = \mathbf{e}_{31}$. From $\mathbf{e}_{12} \otimes \mathbf{e}_3$: $\mathbf{e}_{123} \rfloor \mathbf{e}_3 = \mathbf{e}_{12}$. Every other cut pairs to 0.

4. ★ john meets the left legs. The term $-\text{john} \otimes \mathbf{r}_{\text{ag}}$ gives $-\mathbf{r}_{\text{ag}}$, and $-\text{mary} \otimes \mathbf{r}_{\text{th}}$ gives $-0.6 \mathbf{r}_{\text{th}}$, so $\text{john} \rfloor \mathbf{E} = -\mathbf{r}_{\text{ag}} - 0.6 \mathbf{r}_{\text{th}}$. The sign comes from asking on the wrong side, the 0.6 from the fillers' overlap. Query on the right with the dual vector instead: $\mathbf{E} \rfloor \text{john}^* = \mathbf{r}_{\text{ag}}$, with $\text{john}^* = \mathbf{e}_3 - 0.75 \mathbf{e}_4$.

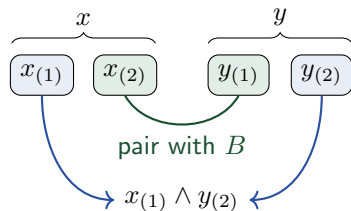
Part IV

Split, Pair, Join

The Recipe

$$x \circ_B y = \sum B(x_{(2)}, y_{(1)}) x_{(1)} \wedge y_{(2)}$$

- 1 **Split** x and y every way: the coproduct.
- 2 **Pair** the inner pieces, $x_{(2)}$ with $y_{(1)}$, using the bilinear form B . Pieces of different grade pair to 0.
- 3 **Join** the outer pieces with the wedge, $x_{(1)} \wedge y_{(2)}$.
- 4 **Add** over all the splits.



Rota and Stein (1994) introduced the Clifford product this way; the construction is now called **cliffordization**. It is associative for every bilinear form B , symmetric or not (Fauser 2002).

Nothing is declared but B . The coproduct does the bookkeeping that the two reduction moves did by hand.

Two Vectors

$\Delta \mathbf{u} = \mathbf{u} \otimes 1 + 1 \otimes \mathbf{u}$ and $\Delta \mathbf{v} = \mathbf{v} \otimes 1 + 1 \otimes \mathbf{v}$: two cuts each, four combinations.

$x_{(1)}$	$x_{(2)}$	$y_{(1)}$	$y_{(2)}$	pairing	term
$\mathbf{u}$	1	1	$\mathbf{v}$	$B(1, 1) = 1$	$\mathbf{u} \wedge \mathbf{v}$
$\mathbf{u}$	1	$\mathbf{v}$	1	$B(1, \mathbf{v}) = 0$	0
1	$\mathbf{u}$	1	$\mathbf{v}$	$B(\mathbf{u}, 1) = 0$	0
1	$\mathbf{u}$	$\mathbf{v}$	1	$B(\mathbf{u}, \mathbf{v})$	$B(\mathbf{u}, \mathbf{v})$

$$\mathbf{u} \circ_B \mathbf{v} = B(\mathbf{u}, \mathbf{v}) + \mathbf{u} \wedge \mathbf{v}$$

With $B = S$, the declared form: $\mathbf{x}_1 \circ_S \mathbf{x}_2 = 16 + 8 \mathbf{e}_1 \wedge \mathbf{e}_2$. That is Session 13's product, recovered without polarization.

The number came from pairing the right piece of $\mathbf{u}$ with the left piece of $\mathbf{v}$; the plane came from pairing nothing.

A Vector and a Plane

$\Delta(\mathbf{v} \wedge \mathbf{w}) = (\mathbf{v} \wedge \mathbf{w}) \otimes 1 + \mathbf{v} \otimes \mathbf{w} - \mathbf{w} \otimes \mathbf{v} + 1 \otimes (\mathbf{v} \wedge \mathbf{w})$. The right piece of $\mathbf{u}$ is 1 or $\mathbf{u}$; pair it with each left piece:

$$\mathbf{u} \circ_B (\mathbf{v} \wedge \mathbf{w}) = \underbrace{\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w}}_{\text{pair nothing}} + \underbrace{B(\mathbf{u}, \mathbf{v}) \mathbf{w} - B(\mathbf{u}, \mathbf{w}) \mathbf{v}}_{\text{pair one leg}} = \mathbf{u} \wedge (\mathbf{v} \wedge \mathbf{w}) + \mathbf{u} \rfloor (\mathbf{v} \wedge \mathbf{w})$$

$$(\mathbf{v} \wedge \mathbf{w}) \circ_B \mathbf{u} = \mathbf{v} \wedge \mathbf{w} \wedge \mathbf{u} + B(\mathbf{w}, \mathbf{u}) \mathbf{v} - B(\mathbf{v}, \mathbf{u}) \mathbf{w} = (\mathbf{v} \wedge \mathbf{w}) \wedge \mathbf{u} + (\mathbf{v} \wedge \mathbf{w}) \lrcorner \mathbf{u}$$

The contraction of Part III is the “pair one leg” part of the product: a vector times anything is its wedge plus its contraction. Chevalley (1954) built the Clifford product on the exterior algebra from exactly this rule.

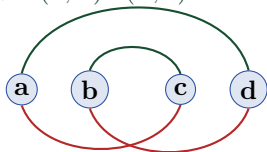
Check, with the dot product: $\mathbf{e}_1 \circ (\mathbf{e}_1 \wedge \mathbf{e}_2) = 0 + 1 \cdot \mathbf{e}_2 - 0 = \mathbf{e}_2$, matching $\mathbf{e}_1 \mathbf{e}_1 \mathbf{e}_2 = \mathbf{e}_2$; and $(\mathbf{e}_1 \wedge \mathbf{e}_2) \circ \mathbf{e}_1 = 0 + 0 - 1 \cdot \mathbf{e}_2 = -\mathbf{e}_2$, matching $\mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_1 = -\mathbf{e}_2$.

Pairing a Plane With a Plane

When both inner pieces are planes, B is extended by a determinant:

$$B(\mathbf{a} \wedge \mathbf{b}, \mathbf{c} \wedge \mathbf{d}) = B(\mathbf{a}, \mathbf{d}) B(\mathbf{b}, \mathbf{c}) - B(\mathbf{a}, \mathbf{c}) B(\mathbf{b}, \mathbf{d})$$

$+ B(\mathbf{a}, \mathbf{d}) B(\mathbf{b}, \mathbf{c})$: nested



$- B(\mathbf{a}, \mathbf{c}) B(\mathbf{b}, \mathbf{d})$: crossed

Pair the inner legs first and then the outer ones, as with nested brackets, and subtract the crossed pairing.

Dot product:

$B(\mathbf{e}_1 \wedge \mathbf{e}_2, \mathbf{e}_1 \wedge \mathbf{e}_2) = 0 \cdot 0 - 1 \cdot 1 = -1$, the scalar part of $(\mathbf{e}_1 \mathbf{e}_2)^2 = -1$.

Lorentz form $\text{diag}(1, -1)$: $0 - (1)(-1) = +1$. A plane that squares to $+1$ generates Session 13's boosts.

Higher grades use larger determinants, a Laplace expansion. The determinant is not a convenience: extended this way, the recipe is associative (Rota and Stein 1994).

Two Planes: Every Way to Match Legs

Run the recipe on $\Delta(\mathbf{a} \wedge \mathbf{b})$ and $\Delta(\mathbf{c} \wedge \mathbf{d})$, four cuts each. The product $(\mathbf{a} \wedge \mathbf{b}) \circ_B (\mathbf{c} \wedge \mathbf{d})$ is the sum of three parts:

grade 0, both legs paired: $B(\mathbf{a}, \mathbf{d})B(\mathbf{b}, \mathbf{c}) - B(\mathbf{a}, \mathbf{c})B(\mathbf{b}, \mathbf{d})$

grade 2, one leg paired: $B(\mathbf{b}, \mathbf{c}) \mathbf{a} \wedge \mathbf{d} - B(\mathbf{b}, \mathbf{d}) \mathbf{a} \wedge \mathbf{c} - B(\mathbf{a}, \mathbf{c}) \mathbf{b} \wedge \mathbf{d} + B(\mathbf{a}, \mathbf{d}) \mathbf{b} \wedge \mathbf{c}$

grade 4, none paired: $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c} \wedge \mathbf{d}$

dot product	value	shared directions	as in the table
$\mathbf{e}_{12} \circ \mathbf{e}_{12}$	-1	both	$\mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_1 \mathbf{e}_2 = -1$
$\mathbf{e}_{12} \circ \mathbf{e}_{23}$	$\mathbf{e}_{13}$	one	$\mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_2 \mathbf{e}_3 = \mathbf{e}_1 \mathbf{e}_3$
$\mathbf{e}_{12} \circ \mathbf{e}_{34}$	$\mathbf{e}_{1234}$	none; needs four dimensions	$\mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3 \mathbf{e}_4$

The product of two planes sorts them by what they share: shared legs are paired away, and the rest survive, grade by grade.

Comparing Two Relations by Their Product

Relations across two slots: *John loves Mary* is $R = \mathbf{s}_j \wedge \mathbf{o}_m$, with $\mathbf{s}_x = \iota_s(x)$, $\mathbf{o}_y = \iota_o(y)$, and the two slots orthogonal. Take $\cos(\text{john}, \text{kim}) = \cos(\text{mary}, \text{sue}) = 0.6$.

$R \circ R'$, with $R' =$	product	shared
<i>John loves Mary</i>	-1	both
<i>John loves Sue</i>	$-(\mathbf{m} \cdot \mathbf{s}) - \mathbf{o}_m \wedge \mathbf{o}_s = -0.6 - 0.8 \mathbf{e}_{12}^o$	the subject
<i>Kim loves Mary</i>	$-(\mathbf{j} \cdot \mathbf{k}) - \mathbf{s}_j \wedge \mathbf{s}_k = -0.6 - 0.8 \mathbf{e}_{12}^s$	the object
<i>Kim loves Sue</i>	$-(\mathbf{j} \cdot \mathbf{k})(\mathbf{m} \cdot \mathbf{s}) - (\mathbf{m} \cdot \mathbf{s}) \mathbf{s}_j \wedge \mathbf{s}_k - (\mathbf{j} \cdot \mathbf{k}) \mathbf{o}_m \wedge \mathbf{o}_s + \mathbf{s}_j \wedge \mathbf{o}_m \wedge \mathbf{s}_k \wedge \mathbf{o}_s$	neither exactly

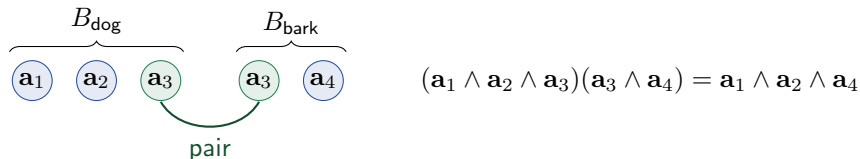
A shared participant is paired away. The participants that differ survive as a plane in their slot, and the number is minus their similarity. *What two relations share goes into the number; what differs is kept, sorted by grade.*

A toy with unit vectors: $\mathbf{e}_{12}^s$ is the plane of the subject slot, $\mathbf{e}_{12}^o$ that of the object slot.

Why Extension Blades Multiply to the Symmetric Difference

The position paper's quantification (§8.6): one orthonormal atom per individual, and a predicate's extension is the blade of its atoms, $B_A = \bigwedge_{a \in A} \mathbf{a}$, of grade $|A|$. Then $B_A B_C = \pm B_{A \Delta C}$, so $|A \cap C| = \frac{1}{2}(|A| + |C| - \text{grade}(B_A B_C))$. The recipe shows why only one term survives:

- Orthonormal atoms pair only with themselves, so the two inner pieces must be the same set of atoms T , inside $A \cap C$.
- A shared atom left unpaired would appear in both outer pieces, and the wedge would be 0. So T is all of $A \cap C$.
- One split survives: pair the shared atoms, join the rest.



Three dogs and two barkers, the paper's example: the product has grade 3, so $|\text{dog} \cap \text{bark}| = \frac{1}{2}(3 + 2 - 3) = 1$. The relations of the last slide behave the same way, with similarity in place of identity.

In-Class Exercise 2

- 1 With the dot product, use the recipe to compute $(\mathbf{e}_1 \wedge \mathbf{e}_2) \circ (\mathbf{e}_2 \wedge \mathbf{e}_3)$ and $(\mathbf{e}_2 \wedge \mathbf{e}_3) \circ (\mathbf{e}_1 \wedge \mathbf{e}_2)$. Are they equal? Why not?
- 2 Declare $B = \text{diag}(1, -1, 1)$. Compute $(\mathbf{e}_1 \wedge \mathbf{e}_2) \circ_B (\mathbf{e}_1 \wedge \mathbf{e}_2)$ and $(\mathbf{e}_1 \wedge \mathbf{e}_3) \circ_B (\mathbf{e}_1 \wedge \mathbf{e}_3)$. Which plane generates boosts, and which rotations?
- 3 Extension blades for $A = \{a_1, a_2\}$ and $C = \{a_2, a_3\}$: find the one split that survives, the product, and $|A \cap C|$.
- 4 ★ From the recipe, show that $\mathbf{u} \circ_B \mathbf{u} = B(\mathbf{u}, \mathbf{u})$ and $\mathbf{u} \circ_B \mathbf{v} + \mathbf{v} \circ_B \mathbf{u} = B(\mathbf{u}, \mathbf{v}) + B(\mathbf{v}, \mathbf{u})$, for any B . What happened to the wedges, and to the antisymmetric half of B ?

Eight minutes. Item 4 sets up Part V.

In-Class Exercise 2: Answers

1. In $\mathbf{e}_{12} \circ \mathbf{e}_{23}$ the inner legs $\mathbf{e}_2, \mathbf{e}_2$ pair to 1 and the outer legs join: $\mathbf{e}_1 \wedge \mathbf{e}_3$. In $\mathbf{e}_{23} \circ \mathbf{e}_{12}$ only the term $B(\mathbf{a}, \mathbf{d}) \mathbf{b} \wedge \mathbf{c}$ survives, with $\mathbf{a} = \mathbf{d} = \mathbf{e}_2$: the result is $\mathbf{e}_3 \wedge \mathbf{e}_1 = -\mathbf{e}_{13}$. Both pair the shared $\mathbf{e}_2$, but the outer legs come out in opposite orders.
2. $(\mathbf{e}_1 \wedge \mathbf{e}_2)^2 = 0 - (1)(-1) = +1$: boosts. $(\mathbf{e}_1 \wedge \mathbf{e}_3)^2 = 0 - (1)(1) = -1$: rotations.
3. $T = \{a_2\}$: the cuts $(\mathbf{a}_1 | \mathbf{a}_2)$ and $(\mathbf{a}_2 | \mathbf{a}_3)$. The product is $\mathbf{a}_1 \wedge \mathbf{a}_3$, of grade 2, so $|A \cap C| = \frac{1}{2}(2 + 2 - 2) = 1$.
4. ★ $\mathbf{u} \circ_B \mathbf{u} = B(\mathbf{u}, \mathbf{u}) + \mathbf{u} \wedge \mathbf{u} = B(\mathbf{u}, \mathbf{u})$. Adding the two orders, $\mathbf{u} \wedge \mathbf{v} + \mathbf{v} \wedge \mathbf{u} = 0$, and for the same reason the antisymmetric half of B cancels: $A(\mathbf{u}, \mathbf{v}) + A(\mathbf{v}, \mathbf{u}) = 0$. What is left is $2S(\mathbf{u}, \mathbf{v})$.
Neither rule can see A .

Part V

The Other Half of the Form

Use the Whole Head

Session 13 declared only Q , so its product saw only S . The recipe accepts any bilinear form. Give it the head's whole form, $G = S + A$:

$$\mathbf{x}_1 \circ_G \mathbf{x}_2 = G(\mathbf{x}_1, \mathbf{x}_2) + \mathbf{x}_1 \wedge \mathbf{x}_2 = 24 + 8 \mathbf{e}_1 \wedge \mathbf{e}_2, \quad \mathbf{x}_2 \circ_G \mathbf{x}_1 = 8 - 8 \mathbf{e}_1 \wedge \mathbf{e}_2$$

The number is now the head's whole score, likeness and direction together. Yet the two rules that defined the Clifford algebra still hold (Exercise 2.4):

$$\mathbf{x} \circ_G \mathbf{x} = Q(\mathbf{x}), \quad \mathbf{u} \circ_G \mathbf{v} + \mathbf{v} \circ_G \mathbf{u} = 2 S(\mathbf{u}, \mathbf{v})$$

Here $\mathbf{x}_1 \circ_G \mathbf{x}_1 = 16$, and $\mathbf{x}_1 \circ_G \mathbf{x}_2 + \mathbf{x}_2 \circ_G \mathbf{x}_1 = 32 = 2 \cdot 16$. A drops out of both rules, because it vanishes on the diagonal and changes sign under the swap.

So $\circ_G$ and $\circ_S$ define **the same algebra**: $\text{Cl}(1, 0, 1)$, the head's algebra from Session 13 (Fauser 2002).
What, then, did A change?

Same Algebra, Different Split

Both recipes describe the same element $\mathbf{x}_1\mathbf{x}_2$ of $\text{Cl}(1,0,1)$. They disagree only about where its plane ends:

$$\mathbf{x}_1\mathbf{x}_2 = 16 + 8\mathbf{e}_1 \wedge_S \mathbf{e}_2 = 24 + 8\mathbf{e}_1 \wedge_G \mathbf{e}_2, \quad \mathbf{e}_1 \wedge_G \mathbf{e}_2 = \mathbf{e}_1 \wedge_S \mathbf{e}_2 - A(\mathbf{e}_1, \mathbf{e}_2) = \mathbf{e}_1 \wedge_S \mathbf{e}_2 - 1$$

Here $\wedge_S$ is Session 13's wedge, $\frac{1}{2}(\mathbf{uv} - \mathbf{vu})$, and $\wedge_G$ is the wedge that the recipe with G joins with.

	split by S	split by G
the number in $\mathbf{x}_1\mathbf{x}_2$ and in $\mathbf{x}_2\mathbf{x}_1$	16 and 16: how alike	24 and 8: the head's scores
the number, under the swap	unchanged	changes
what counts as the plane of $\mathbf{e}_1, \mathbf{e}_2$	$\mathbf{e}_1 \wedge_S \mathbf{e}_2$	$\mathbf{e}_1 \wedge_S \mathbf{e}_2 - 1$

The symmetric half of a form decides the algebra. The antisymmetric half decides where the number ends and the plane begins. The grading moves; the split into even and odd does not (Fauser 2000).

With three vectors, A also trades between grades 3 and 1: in S 's terms,
 $\mathbf{e}_1 \wedge_G \mathbf{e}_2 \wedge_G \mathbf{e}_3 = \mathbf{e}_{123} - A_{12} \mathbf{e}_3 + A_{13} \mathbf{e}_2 - A_{23} \mathbf{e}_1.$

What the Two Halves Do

	symmetric half S	antisymmetric half A
in a score $\mathbf{x}^\top G \mathbf{y}$	how alike	which way
in the quadratic form Q	all of it	nothing
in the dyad, it reads	the symmetric half	the plane $\mathbf{x} \wedge \mathbf{y}$
in the recipe's algebra	decides it, through the signature	changes nothing
in the recipe's grading	the default split	moves $A(\mathbf{u}, \mathbf{v})$ from the plane into the number

The first question, answered. The coproduct is not the symmetric/antisymmetric split. It lists the legs; a bilinear form pairs them; and the two halves of the form then play different parts in the product they build.

Session 13's puzzle: a declaration that fixes only Q still yields an algebra that keeps order. Its converse: a form with an antisymmetric half gets no new algebra, only a new way of reading it.

Attention Heads, Read This Way

A head scores $\mathbf{q}^\top G \mathbf{k}$ with $G = W_Q^\top W_K$, the head's **QK circuit** (Elhage et al. 2021). Feed G to the recipe:

$$\mathbf{q} \circ_G \mathbf{k} = \underbrace{\mathbf{q}^\top G \mathbf{k}}_{\text{the score}} + \mathbf{q} \wedge \mathbf{k}$$

- S decides an algebra, through the signature of the head's quadratic form. S has rank at most twice the head width, so for a 64-wide head in a 768-dimensional model it is degenerate, and it can be indefinite. The toy head's algebra is $\text{Cl}(1, 0, 1)$.
- A decides the grading: how much of each query–key product is called score, and how much plane. In S 's terms, the directional part of the score, $\mathbf{q}^\top A \mathbf{k}$, belongs to the plane.
- Saponati et al. (ICML 2025) find that bidirectionally trained encoders develop more symmetric query–key matrices, and autoregressive decoders more directional ones. Read this way, the training objective decides how much of the product a head treats as score and how much as plane.

The last point reads their result in this lecture's terms; their paper does not make that claim. The energy fractions $\|S\|^2/\|G\|^2$ and $\|A\|^2/\|G\|^2$ measure the balance.

RoPE Changes the Form With Distance

With rotary embeddings, a token $\mathbf{x}$ at position m and a token $\mathbf{y}$ at position n score $(R_m W \mathbf{x})^\top (R_n W \mathbf{y}) = \mathbf{x}^\top W^\top R_\delta W \mathbf{y}$, with $\delta = n - m$. A tied head ($W_Q = W_K = W$) gets a different form at every distance:

$$G_\delta = W^\top R_\delta W, \quad S_\delta = W^\top \left(\bigoplus_i \cos(\delta \theta_i) I_2 \right) W, \quad A_\delta = W^\top \left(\bigoplus_i \sin(\delta \theta_i) J \right) W$$

δ	$\cos \delta \theta$	$\sin \delta \theta$	signature of S_δ
0	1.00	0.00	(2, 0, 0)
1	0.54	0.84	(2, 0, 0)
2	-0.42	0.91	(0, 2, 0)
3	-0.99	0.14	(0, 2, 0)

one plane, $W = I$, $\theta = 1$ radian

$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. This describes RoPE in the recipe's terms; it is not a result about trained models. Session 14 found the same split, plane by plane.

- At $\delta = 0$ the head is symmetric and positive: all score, no plane.
- With distance, A_δ grows and shrinks, and S_δ can change sign. In the recipe's terms, RoPE changes the grading *and* the algebra with distance, each plane at its own speed.

Part VI

Other Coproducts in Linguistics

Copying Versus Splitting

Compositional distributional semantics already uses coproducts, of another kind: they **copy**.

- **Relative pronouns.** Sadrzadeh, Clark and Coecke (2013) copy a noun's basis vectors, $\Delta(\mathbf{n}_i) = \mathbf{n}_i \otimes \mathbf{n}_i$, so that in *men who like Mary* the noun both heads the phrase and serves as the subject of *like*: $\text{men} \odot (\text{like} \times \text{Mary})$.
- **Quantifiers.** Hedges and Sadrzadeh (2019) build a bialgebra on the subsets of the universe, with product $|A\rangle|B\rangle = |A \cap B\rangle$ and coproduct $|A\rangle \mapsto |A\rangle \otimes |A\rangle$.

on a subset-indexed basis	product of A and C	coproduct of A
Hedges–Sadrzadeh bialgebra	$A \cap C$	$A \otimes A$: copy
exterior algebra	$\pm(A \cup C)$ if disjoint, else 0	$\sum_{T \subseteq A} \pm T \otimes (A \setminus T)$: split
Clifford, orthonormal atoms	$\pm(A \Delta C)$	built by split, pair, join

Copying needs a chosen basis; splitting does not. The position paper notes that FGA's quantification needs no second, powerset-indexed space: the subset-indexed basis is the one the algebra already has (§7, §8.6).

Merge and the Coproduct of Trees

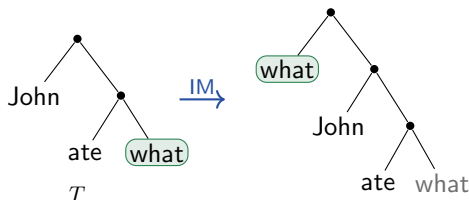
Marcolli, Chomsky and Berwick (2023; book 2025) describe Merge with a Hopf algebra of **workspaces**, forests of binary trees. Its coproduct lists the **accessible terms** T_v of a tree (its subtrees), each with what remains when it is cut out:

$$\Delta(T) = \sum_v T_v \otimes T/T_v + \dots$$

The dots extract several subtrees at once. Merge cuts with the coproduct, selects the terms to be merged, and grafts them under a new root.

Internal Merge merges an accessible term with the rest of its own tree: movement.

The same idea as $\Lambda(V)$, on a different object. The exterior coproduct lists the sub-blades of a blade; this one lists the subtrees of a tree. In both, an operation – contraction there, Merge here – is built by choosing among the cuts.



$\Delta(T)$ contains $\text{what} \otimes T/\text{what}$; Internal Merge regrafts what at the top. The lower copy is not pronounced.

Where This Points

- **Learn the whole form.** The recipe works for any bilinear form. A model that learns a non-symmetric B learns its algebra (through S) and its grading (through A) at once.
- **Probe real heads.** Split query–key forms into S and A , and ask what algebra and what grading each head declares, and whether these track syntactic function.
- **Partitives from the coproduct.** Can *two of the dogs* be computed from the $(2, k - 2)$ slice of $\Delta(B_{\text{dog}})$, the way the position paper computes *every* and *most* from grades?
- **Syntax and relations.** Merge cuts trees; the exterior coproduct cuts relations. Do the cuts of a derivation line up with the contractions that saturate a verb's slots?

Each could seed a final project.

Summary

- 1 A bilinear form reads the dyad: **its symmetric half reads likeness, its antisymmetric half reads the plane $\mathbf{x} \wedge \mathbf{y}$** . One number cannot keep the two apart.
- 2 A **coproduct** lists the ways to cut an object in two. Strings, bags, and oriented sets differ in what a swap costs; for oriented sets, $\Delta(\mathbf{u} \wedge \mathbf{v}) = (\mathbf{u} \wedge \mathbf{v}) \otimes 1 + \mathbf{u} \otimes \mathbf{v} - \mathbf{v} \otimes \mathbf{u} + 1 \otimes (\mathbf{u} \wedge \mathbf{v})$.
- 3 A bialgebra with an antipode is a **Hopf algebra**; on $\Lambda(V)$ the antipode is the grade involution.
- 4 **Contraction pairs one leg**: $\mathbf{a} \rfloor X = \sum g(\mathbf{a}, X_{(1)}) X_{(2)}$. A query meets every leg the coproduct lists.
- 5 **The geometric product is split, pair, join**: $x \circ_B y = \sum B(x_{(2)}, y_{(1)}) x_{(1)} \wedge y_{(2)}$, for any bilinear form B .
- 6 Products sort by sharing: shared legs pair away, and the rest survive by grade. The symmetric difference of extension blades is one case.
- 7 **The symmetric half of B decides the algebra; the antisymmetric half decides the grading**. The coproduct is not the symmetric/antisymmetric split: it hands both halves the legs they read.

Exercises

- 1 Write $\Delta(\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3 \wedge \mathbf{e}_4)$. How many terms fall in each slice $(k, 4 - k)$? Check coassociativity on one three-piece cutting.
- 2 Show that $\Delta(w^k) = \sum_j \binom{k}{j} w^j \otimes w^{k-j}$ in $\text{Sym}(V)$, and that the same rule in $\Lambda(V)$ gives $\Delta(\mathbf{w} \wedge \mathbf{w}) = 0$.
- 3 Use the recipe to compute $(\mathbf{e}_1 \wedge \mathbf{e}_2) \circ (\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3)$ with the dot product, and check it against the table of $\text{Cl}(3, 0, 0)$.
- 4 In two dimensions take $B = \begin{pmatrix} 1 & f \\ -f & 1 \end{pmatrix}$. Show that $\mathbf{e}_1 \circ_B \mathbf{e}_2 = f + \mathbf{e}_1 \wedge \mathbf{e}_2$ and $(\mathbf{e}_1 \circ_B \mathbf{e}_2)^2 = -1$. Which element plays the unit plane, and what is its scalar part?
- 5 For the head, compute $\mathbf{d} \circ_G \mathbf{d}$ for the null difference $\mathbf{d} = \mathbf{x}_1 - \mathbf{x}_2 = (2, -2)$. Why can no choice of A change the answer?
- 6 ★ Write the inside algorithm as a convolution over the coproduct of strings, with the two trivial cuts removed. What plays the role of the product?

Reading: Hopf Algebras and the Clifford Product

- M. E. Sweedler, *Hopf Algebras*, W. A. Benjamin, 1969.
- S. A. Joni and G.-C. Rota, Coalgebras and bialgebras in combinatorics, *Studies in Applied Mathematics* 61, 1979.
- D. Grinberg and V. Reiner, *Hopf Algebras in Combinatorics*, lecture notes, arXiv:1409.8356. A gentle place to start.
- C. Chevalley, *The Algebraic Theory of Spinors*, Columbia University Press, 1954.
- G.-C. Rota and J. A. Stein, Plethystic Hopf algebras, *Proceedings of the National Academy of Sciences* 91, 1994.
- B. Fauser, On the Hopf algebraic origin of Wick normal-ordering, arXiv:hep-th/0007032, 2000.
- B. Fauser, *A Treatise on Quantum Clifford Algebras*, arXiv:math/0202059, 2002.
- R. Abłamowicz and B. Fauser, Mathematics of CLIFFORD: a Maple package for Clifford and Grassmann algebras, arXiv:math-ph/0212031. Computes with any bilinear form.
- L. Dorst, D. Fontijne and S. Mann, *Geometric Algebra for Computer Science*, Morgan Kaufmann, 2007.

Reading: Language and Attention

- M. Sadrzadeh, S. Clark and B. Coecke, The Frobenius anatomy of word meanings I: subject and object relative pronouns, *Journal of Logic and Computation* 23, 2013.
- J. Hedges and M. Sadrzadeh, A generalised quantifier theory of natural language in categorical compositional distributional semantics with bialgebras, *Mathematical Structures in Computer Science* 29(6), 2019.
- M. Marcolli, N. Chomsky and R. C. Berwick, Mathematical structure of syntactic Merge, arXiv:2305.18278, 2023; book, MIT Press, 2025.
- G. Link, The logical analysis of plurals and mass terms: a lattice-theoretical approach, in *Meaning, Use, and Interpretation of Language*, de Gruyter, 1983.
- J. Goodman, Semiring parsing, *Computational Linguistics* 25(4), 1999.
- N. Elhage et al., A mathematical framework for transformer circuits, Transformer Circuits Thread, 2021.
- M. Saponati et al., The underlying structures of self-attention: symmetry, directionality, and emergent dynamics in Transformer training, ICML 2025.
- J. Pustejovsky, *Toward a Functional Geometric Algebra for natural language semantics*, arXiv:2604.25902v2, 2026 – §5.6, §7, §8.6.