Compositionality: categorial variations on a theme

Michael Moortgat, Utrecht University

Tutorial, WoLLIC 2023, Halifax

A tutorial

Target audience logic/computer science background

Aim bird's eye view of the architecture of 'categorial' grammars, and the design choices for dealing with natural language form and meaning, and the relation between these two.

More To Explore Some useful general references

- Moot & Retoré, 2012, The Logic of Categorial Grammars. A Deductive Account of Natural Language Syntax and Semantics
- Moot, 2021, Type-logical investigations: proof-theoretic, computational and linguistic aspects of modern type-logical grammars

Background: Lambek's categorial type logics

The original presentation ('Deductive systems as categories') considers statements $A \longrightarrow B$, i.e. derivability is modelled as a relation holding between types.

Pre-order laws

$$A \longrightarrow A \qquad \frac{A \longrightarrow B \quad B \longrightarrow C}{A \longrightarrow C}$$

Residuation laws

$$B \longrightarrow A\backslash C \quad \textit{iff} \quad A \bullet B \longrightarrow C \quad \textit{iff} \quad A \longrightarrow C/B$$

Structural laws

$$A \bullet (B \bullet C) \longleftrightarrow (A \bullet B) \bullet C$$

$$A \bullet B \longrightarrow B \bullet A$$

$$I \bullet A \longleftrightarrow A \longleftrightarrow A \bullet I$$

Pure residuation logic: **NL** [L61]; **L**=**NL** plus associativity [L58,88];
LP=**L**+commutativity, Lambek-Van Benthem calculus, a.k.a. MILL

Models: residuated monoids/groupoids

(N)L intended models for the syntactic calculi are the multiplicative systems freely generated by the words of the language under concatenation.

Types as sets of expressions, i.e. subsets of a groupoid/semigroup/monoid $\langle M, \cdot \rangle$ with

$$\begin{array}{lcl} A \bullet B & = & \{a \cdot b \in M \mid a \in A \wedge b \in B\} \\ C/B & = & \{a \in M \mid \forall_{b \in B}\ a \cdot b \in C\} \\ A \backslash C & = & \{b \in M \mid \forall_{a \in A}\ a \cdot b \in C\} \\ I & = & \{1\} \end{array}$$

- groupoid [L61], types assigned to phrases, bracketed strings
- semigroup [L58], types assigned to strings, associative multiplication
- monoid [L88], multiplicative unit, empty string

LP Calculus of semantic types, abstracting from word order/constituent structure.

Parsing as deduction

Natural Deduction format left of turnstile: words instead of their types

$$
\dfrac{\dfrac{\text{paper}}{n} \qquad \dfrac{\dfrac{\text{that}}{(n\backslash n)/(s/np)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\text{Bob}}{np} \qquad \dfrac{\dfrac{\text{rejected}}{(np\backslash s)/np} \qquad np \vdash np}{\text{rejected} \cdot np \vdash np\backslash s}\ /E}{\text{Bob} \cdot (\text{rejected} \cdot np) \vdash s}\ \backslash E}{(\text{Bob} \cdot \text{rejected}) \cdot np \vdash s}\ \mathsf{A}^r}{\text{Bob} \cdot \text{rejected} \vdash s/np}\ /I}{\text{that} \cdot (\text{Bob} \cdot \text{rejected}) \vdash n\backslash n}\ /E}{\text{paper} \cdot (\text{that} \cdot (\text{Bob} \cdot \text{rejected})) \vdash n}\ \backslash E
$$

- axiom leaves: lexical type assignments; A^r semi-associativity
- $/, \backslash E$: slash Elim $\simeq$ modus ponens; $/, \backslash I$: slash Intro $\simeq$ hypothetical reasoning

Alternative formats sequent calculus, display logic, proof nets, . . .

Natural Deduction

Structures, sequents Sequents $\Gamma \vdash A$ with A a type, Γ a structure.

Structures: trees with type formulas at the leaves:

$$\Gamma, \Delta \quad ::= \quad A \mid \Gamma \cdot \Delta$$

where 2-place $\cdot$ is the structural counterpart of $\bullet$.

Axiom, logical rules For the base logic **NL**, we have the axiom $A \vdash A$ and as logical inference rules, for each connective an elimination rule and an introduction rule.

$$\frac{\Gamma \vdash A \quad \Delta \vdash A\backslash B}{\Gamma \cdot \Delta \vdash B}\ \backslash E \qquad \frac{\Gamma \vdash B/A \quad \Delta \vdash A}{\Gamma \cdot \Delta \vdash B}\ /E$$

$$\frac{A \cdot \Gamma \vdash B}{\Gamma \vdash A\backslash B}\ \backslash I \qquad \frac{\Gamma \cdot A \vdash B}{\Gamma \vdash B/A}\ /I$$

$$\frac{\Gamma \vdash A \quad \Delta \vdash B}{\Gamma \cdot \Delta \vdash A \bullet B}\ \bullet I \qquad \frac{\Delta \vdash A \bullet B \quad \Gamma[A \cdot B] \vdash C}{\Gamma[\Delta] \vdash C}\ \bullet E$$

Notation: $\Gamma[\Delta]$ for a structure Γ containing a substructure Δ

N.D.: explicit structural rules

Postulate extensions of the base logic take the form of structural rules.

Formula variables $\rightsquigarrow$ structure variables (in context).

Associativity Compare right vs left rotation:

$$(A \bullet B) \bullet C \longrightarrow A \bullet (B \bullet C) \quad \rightsquigarrow \quad \frac{\Gamma[\Delta \cdot (\Delta' \cdot \Delta'')] \vdash D}{\Gamma[(\Delta \cdot \Delta') \cdot \Delta''] \vdash D}\ \mathsf{A}^r$$

$$A \bullet (B \bullet C) \longrightarrow (A \bullet B) \bullet C \quad \rightsquigarrow \quad \frac{\Gamma[(\Delta \cdot \Delta') \cdot \Delta''] \vdash D}{\Gamma[\Delta \cdot (\Delta' \cdot \Delta'')] \vdash D}\ \mathsf{A}^l$$

Semi-associativity **NL** $+ \{\mathsf{A}^l, \mathsf{A}^r\} =$ fully associative **L**, but see

Zeilberger, LMCS 2019: Tamari order on well-bracketed strings/binary trees in terms of a semi-associative calculus.

Grammars

A categorial grammar consists of a universal and a language-specific component.

- universal: a type calculus, **(N)L**
- language specific: a lexicon assigning each word a finite number of types

Language Given a categorial grammar G and a type B we write $L(G, B)$ for the strings of type B recognized by G. $w_1 \cdots w_n \in L(G, B)$ if the following hold:

- $(w_i, A_i) \in$ Lex for $1 \leq i \leq n$;
- $\Gamma_{[A_1,\ldots,A_n]} \vdash B$, for Γ an antecedent structure with yield $A_1, \ldots, A_n$

Idealization?

Proofs and terms: syntactic calculi (N)$L_{/,\backslash}$

Types, terms p atomic

$$A, B ::= p \mid A\backslash B \mid B/A \qquad M, N ::= x \mid \lambda^r x.M \mid \lambda^l x.M \mid (M \ltimes N) \mid (N \rtimes M)$$

Wansing, 1990, Formulas-as-types for a Hierarchy of Sublogics of Int Prop Logic

Typing rules Axiom $x : A \vdash x : A$ var Γ, Δ all distinct

$$\frac{\Gamma \cdot x : A \vdash M : B}{\Gamma \vdash \lambda^r x.M : B/A}\ I/ \qquad \frac{x : A \cdot \Gamma \vdash M : B}{\Gamma \vdash \lambda^l x.M : A\backslash B}\ I\backslash$$

$$\frac{\Gamma \vdash M : B/A \quad \Delta \vdash N : A}{\Gamma \cdot \Delta \vdash (M \ltimes N) : B}\ E/ \qquad \frac{\Gamma \vdash N : A \quad \Delta \vdash M : A\backslash B}{\Gamma \cdot \Delta \vdash (N \rtimes M) : B}\ E\backslash$$

Compare: $LP_{\multimap}$ **L** extended with product commutativity, a.k.a. MILL, Multiplicative Intuitionistic Linear Logic. In MILL, the slashes $/, \backslash$ collapse to linear implication $\multimap$.

$$\frac{\Gamma, x : A \vdash M : B}{\Gamma \vdash \lambda x.M : A \multimap B}\ (\multimap I) \qquad \frac{\Gamma \vdash M : A \multimap B \quad \Delta \vdash N : A}{\Gamma, \Delta \vdash M\ N : B}\ (\multimap E)$$

Compositionality

Compositional translations

The classical view Homomorphism Montague 1970, Universal Grammar

$$\text{Source} \xrightarrow{h} \text{Target}$$

relating types/proofs of a Source logic to their Target counterparts.

A chained view Interpretation as a two-step process $h'' \circ h'$

$$\text{Source} \xrightarrow{h'} \text{Target}_{der} \xrightarrow{h''} \text{Target}_{lex}$$

- h' derivational semantics, source constants (words) as black boxes
- h'' lexical semantics, unpacking word-internal semantics

Toy example: (N)L to LP/MILL

Source atoms: s, np, n; target atoms e (entities), t (truth values).

$$\textbf{(N)L}^{s,np,n}_{/,\backslash} \quad \xrightarrow{\lceil\cdot\rceil} \quad \textbf{LP}/\text{MILL}^{e,t}_{\multimap}$$

Types $\lceil s \rceil = t$, $\lceil np \rceil = e$, $\lceil n \rceil = e \multimap t$, $\lceil A\backslash B \rceil = \lceil B/A \rceil = \lceil A \rceil \multimap \lceil B \rceil$.

Proofs $\lceil x \rceil = \widetilde{x}$ translates Axioms; for Intro/Elim rules:

$$\lceil \lambda^l x.M \rceil = \lceil \lambda^r x.M \rceil = \lambda \widetilde{x}.\lceil M \rceil \qquad \lceil N \rtimes M \rceil = \lceil M \ltimes N \rceil = \lceil M \rceil \ \lceil N \rceil$$

Example

$$\begin{array}{rcl} M & = & \text{paper} \rtimes (\text{that} \ltimes \lambda^r x.(\text{Bob} \rtimes (\text{rejected} \ltimes x))) : n \\ \lceil M \rceil & = & ((\lceil\text{that}\rceil\ \lambda x.((\lceil\text{rejected}\rceil\ x)\ \lceil\text{Bob}\rceil))\ \lceil\text{paper}\rceil) : e \multimap t \end{array}$$

Remark $\lceil\cdot\rceil$ sends source atoms to target types, not necessarily atomic.

Beyond linearity, $\mathrm{MILL}^{!}_{\multimap}$

To express word-internal meaning recipes, we need expressivity beyond simple MILL:

- IL? too much, free copying (Contraction), deletion (Weakening)
- MILL+exponential for controlled copying/deletion, ILL ! too strong, Soft LL?

Target signature $\textsc{Bob}^{e}$, $\textsc{rejected}^{e \to e \to t}$, $\textsc{paper}^{e \to t}$, $\wedge^{t \to t \to t}$, $\quad A \to B = \,!\,A \multimap B$

$$\mathsf{that}^{(n\backslash n)/(s/np)} \xrightarrow{\lceil\cdot\rceil} \lambda x \lambda y \lambda z.((y\ z) \wedge (x\ z))^{(e\to t)\to(e\to t)\to(e\to t)}$$

Substituting the lexical translations in $\lceil M \rceil$ and simplifying:

$$\begin{aligned} \lceil M \rceil &= (\lceil \mathsf{that} \rceil\ \lambda x.((\lceil \mathsf{rejected} \rceil\ x)\ \lceil \mathsf{Bob} \rceil))\ \lceil \mathsf{paper} \rceil \\ &= ([\lambda x \lambda y \lambda z.((y\ z) \wedge (x\ z))]\ \lambda x.((\textsc{rejected}\ x)\ \textsc{Bob}))\ \textsc{paper} \\ &= \lambda x.((\textsc{paper}\ x) \wedge ((\textsc{rejected}\ x)\ \textsc{Bob})) : e \to t \end{aligned}$$

de Groote & Retoré, 1996, On the Semantic Readings of Proof Nets

Remark $\lceil\cdot\rceil$ sends source constants to target terms, not necessarily atomic.

Chameleon words: lexical polymorphism

Coordination represents another case of ostensible copying:

a (Alice sings)$_s$ and (Bob dances)$_s$
b Alice (sings and dances)$_{np\backslash s}$
c Bob (criticized and rejected)$_{(np\backslash s)/np}$ the paper
d (Alice praised)$_{s/np}$ but (Bob criticized)$_{s/np}$ the paper

Syntactically deriving (b–d) types from initial $(s\backslash s)/s$ goes beyond linearity:

$$\dfrac{\dfrac{\dfrac{\vdots}{(\boxed{np} \cdot np\backslash s) \cdot ((s\backslash s)/s \cdot (\boxed{np} \cdot np\backslash s) \vdash s}}{\boxed{np} \cdot (np\backslash s \cdot ((s\backslash s)/s \cdot np\backslash s) \vdash s}\ \text{Copy!}}{(s\backslash s)/s \vdash ((np\backslash s)\backslash(np\backslash s))/(\boxed{np}\backslash s)}\ /, \backslash\text{Intro}$$

Generalized coordination

An alternative to copying in the syntax: type-restricted form of polymorphism

Partee & Rooth 1982, Generalized Conjunction and Type Ambiguity

Conjoinable types

- $s \in$ CType;
- $A\backslash B, B/A \in$ CType if $B \in$ CType, $A \in$ Type

Generalized interpretation scheme $\sqcap^X$ (infix): coordinator of type $X \to X \to X$

- $P \sqcap^t Q := P \wedge Q$ coordination in type t amounts to boolean conjunction
- $P \sqcap^{A \to B} Q := \lambda x^A.(P\ x) \sqcap^B (Q\ x)$ distributing the x^A parameter over the conjuncts

Remark Emms 1994, undecidability of general polymorphic **L**

Variations

Variations: source

Hybrid Typelogical Grammar (Kubota & Levine 2012,...) Layered architecture, mixing Lambek slashes and linear implication $\multimap$

$$A, B ::= p \mid A\backslash B \mid A/B \qquad \mathcal{A}, \mathcal{B} ::= A \mid \mathcal{A} \multimap \mathcal{B}$$

with **L** types: strings/concatenation; $\mathrm{MILL}_{\multimap}$ types: functions string $\to$ string

see also Abrusci & Ruet, 1999

Abstract Categorial Grammar (De Groote 2001, ...)

- **LP**/$\mathrm{MILL}_{\multimap}$ for abstract syntax, Curry's 'tectogrammatical' structure
- surface form, meaning composition, ... derived from abstract source
- refinement of Chomsky hierarchy via type homomorphisms of growing complexity

purely applicative AS source

ACG: from abstract syntax to surface form

- Target signature: string as function type $* \multimap *$ (abbrev σ). Concat: composition; empty string: id function. Constants: word forms $\mathtt{w} :: \sigma$

$$+ := \lambda sri.s(r(i)) \qquad \epsilon := \lambda i.i$$

- type homomorphism: $\lceil A \rceil = \sigma$, for source atoms A.
- translating the abstract source constants:

CONSTANT	SOURCE TYPE	$\lceil\cdot\rceil$	TARGET TYPE
PAPER	n	$\mathtt{paper}$	σ
THAT	$(np \multimap s) \multimap n \multimap n$	$\lambda Zs.s + \mathtt{that} + (Z\ \epsilon)$	$(\sigma\sigma)\sigma\sigma$
BOB	np	$\mathtt{Bob}$	σ
REJECTED	$np \multimap np \multimap s$	$\lambda sr.r + \mathtt{rejected} + s$	$\sigma\sigma\sigma$

$$\begin{array}{rcl} M & = & ((\lceil\text{that}\rceil\ \lambda x.((\lceil\text{rejected}\rceil\ x)\ \lceil\text{Bob}\rceil))\ \lceil\text{paper}\rceil) : s \\ \lceil M \rceil & = & \mathtt{paper} + \mathtt{that} + \mathtt{Bob} + \mathtt{rejected} : \sigma \end{array}$$

Variations: target

DisCoCat Compositionality, vector-based (Coecke et al 2010, ...)

CCC A compact closed category (CCC) is monoidal, i.e. it has associative $\otimes$ with unit I; and for every object there is a left and a right adjoint satisfying

$$A^l \otimes A \xrightarrow{\epsilon^l} I \xrightarrow{\eta^l} A \otimes A^l \qquad A \otimes A^r \xrightarrow{\epsilon^r} I \xrightarrow{\eta^r} A^r \otimes A$$

In a symmetric CCC, the tensor moreover is commutative, and we can write A^* for the collapsed left and right adjoints.

FVect, linear maps concrete instance of sCCC

- unit I: the field $\mathbb{R}$; bases: fixed, so $V^* \cong V$ one can ignore $\cdot^*$
- ϵ map: inner products; η map: identity tensor (with $\lambda = 1$) or multiples

$$\epsilon_V : V \otimes V \mapsto \mathbb{R} \quad \text{given by} \quad \sum_{ij} v_{ij}(\vec{e}_i \otimes \vec{e}_j) \mapsto \sum_i v_{ii}$$

$$\eta_V : \mathbb{R} \mapsto V \otimes V \quad \text{given by} \quad \lambda \mapsto \sum_i \lambda(\vec{e}_i \otimes \vec{e}_i)$$

From (N)L to sCCC

Types assign a vector space to the source atoms: $\lceil np \rceil = \lceil n \rceil =$ N, $\lceil s \rceil =$ S;

$$\lceil A \bullet B \rceil = \lceil A \rceil \otimes \lceil B \rceil \qquad \lceil A/B \rceil = \lceil A \rceil \otimes \lceil B \rceil^* \qquad \lceil A \backslash B \rceil = \lceil A \rceil^* \otimes \lceil B \rceil$$

Proofs Syntactic derivations $f : A \longrightarrow B$ in (N)L are interpreted as linear maps.

We give the translation for a categorical presentation equivalent to (display) sequent calculus:

- axioms $A \vdash A$ with A atomic
- residuation rules
- monotonicity rules

Wijnholds & MM 2017, ...

Interpretation: proofs

Identity, composition $\lceil 1_A \rceil = 1_{\lceil A \rceil}, \quad \lceil g \circ f \rceil = \lceil g \rceil \circ \lceil f \rceil$

Residuation

$$\left\lceil \frac{f : A \bullet B \longrightarrow C}{\triangleright f : A \longrightarrow C/B} \right\rceil$$

$$\lceil \triangleright f \rceil = \lceil A \rceil \xrightarrow{1_{\lceil A \rceil} \otimes \eta_{\lceil B \rceil}} \lceil A \rceil \otimes \lceil B \rceil \otimes \lceil B \rceil^* \xrightarrow{\lceil f \rceil \otimes 1_{\lceil B \rceil^*}} \lceil C \rceil \otimes \lceil B \rceil^*$$

$$\left\lceil \frac{g : A \longrightarrow C/B}{\triangleright^{-1} g : A \bullet B \longrightarrow C} \right\rceil$$

$$\lceil \triangleright^{-1} g \rceil = \lceil A \rceil \otimes \lceil B \rceil \xrightarrow{\lceil g \rceil \otimes 1_{\lceil B \rceil}} \lceil C \rceil \otimes \lceil B \rceil^* \otimes \lceil B \rceil \xrightarrow{1_{\lceil C \rceil} \otimes \epsilon_{\lceil B \rceil}} \lceil C \rceil$$

similarly for $\triangleleft, \triangleleft^{-1}$

Interpreting proofs

Monotonicity The case of parallel composition is immediate: $\lceil f \bullet g \rceil = \lceil f \rceil \otimes \lceil g \rceil$.

For / we have
$$\left\lceil \frac{f : A \longrightarrow B \quad g : C \longrightarrow D}{f/g : A/D \longrightarrow B/C} \right\rceil$$

where $\lceil f/g \rceil \quad =$
$$\begin{array}{c}
\lceil A \rceil \otimes \lceil D \rceil^* \\
\Big\downarrow \lceil f \rceil \otimes \eta_{\lceil C \rceil} \otimes 1_{\lceil D \rceil^*} \\
\lceil B \rceil \otimes \lceil C \rceil^* \otimes \lceil C \rceil \otimes \lceil D \rceil^* \\
\Big\downarrow 1_{\lceil B \rceil \otimes \lceil C \rceil^*} \otimes \lceil g \rceil \otimes 1_{\lceil D \rceil^*} \\
\lceil B \rceil \otimes \lceil C \rceil^* \otimes \lceil D \rceil \otimes \lceil D \rceil^* \\
\Big\downarrow 1_{\lceil B \rceil \otimes \lceil C \rceil^*} \otimes \epsilon_{\lceil D \rceil} \\
\lceil B \rceil \otimes \lceil C \rceil^*
\end{array}$$

similarly for $g \backslash f$

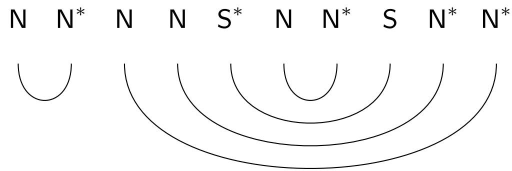
Example

Derivational semantics axiom links of the proof:

$$n_0 \cdot ((n_1 \backslash n_2)/(s_3/np_4) \cdot (np_5 \cdot (np_6 \backslash s_7)/np_8)) \stackrel{\substack{0,2,7,4,5\\1,9,3,8,6}}{\vdash} n_9$$

Tensor contractions corresponding to the axiom links:

$$\lceil M \rceil = \quad \mathbf{paper}_i \otimes \mathbf{that}_{ijkl} \otimes \mathbf{bob}_m \otimes \mathbf{rejected}_{mlk}$$



$$= \quad \mathsf{N}^* \ni \delta^{n,o,p,q,r}_{s,t,u,v,w}\ \mathbf{paper}_n \otimes \mathbf{that}_{souq} \otimes \mathbf{bob}_r \otimes \mathbf{rejected}_{wpv}$$

Lexical semantics

Frobenius operations

Sadrzadeh c.s.

- Duplicate: $V \to V \otimes V$, embed a vector on diagonal of a matrix
- Delete: $V \to \mathbb{R}$, sum the elements of a vector
- Merge: $V \otimes V \to V$, retrieve diagonal of a matrix
- Insert: $\mathbb{R} \to V$, send scalar λ to all-λ vector

Example

$$\mathbf{paper}_i \otimes \mathbf{that}_{ijkl} \otimes \mathbf{bob}_m \otimes \mathbf{rejected}_{mlk}$$

$$\mathsf{N} \quad \mathsf{N}^* \quad \mathsf{N} \quad \mathsf{N} \quad \mathsf{S}^* \quad \mathsf{N} \quad \mathsf{N}^* \quad \mathsf{S} \quad \mathsf{N}^* \quad \mathsf{N}^*$$

Unpacking **that** to obtain 'intersective' $\odot$ interpretation:

- rank reduction of $\mathbf{bob} \otimes \mathbf{rejected} \in \mathsf{S} \otimes \mathsf{N}^*$ to $\mathbf{v} \in \mathsf{N}$
- computation of $\mathbf{paper} \odot \mathbf{v}$ by contraction with tensor $\mathbf{c} \in \mathsf{N} \otimes \mathsf{N} \otimes \mathsf{N}$ with elements $c_{ijk} = 1$ if $i = j = k$ and zero otherwise

Lambdas for Vector Semantics

Muskens & Sadrzadeh 2018

- Target signature. I: index set; R: reals $\mathbb{R}$. Vectors: IR, rank n tensors $I^n R$ (curried). $M := I^2 R$, $C := I^3 R$, etc. Defined operations e.g.

$$
\begin{array}{lll}
\odot & := & \lambda vui.v_i \cdot u_i \qquad VVV \\
\times_1 & := & \lambda mvi.\sum_j m_{ij} \cdot v_j \quad MVV \\
\times_2 & := & \lambda cvij.\sum_k c_{ijk} \cdot v_k \quad CVM
\end{array}
$$

- type homomorphism: $\lceil A \rceil = V$ for source atoms A.
- translating the constants: $\lceil \text{paper}^n \rceil = \texttt{paper}^V$, $\lceil \text{Bob}^{np} \rceil = \texttt{Bob}^V$,

  $\lceil \text{rejected}^{(np\backslash s)/np} \rceil = T^{VVV}$ options:

  - $T = \lambda uv.((\texttt{rejected}^C \times_2 u) \times_1 v)$ tensor contraction
  - $T = \lambda uv.(\texttt{rejected}^V \odot u \odot v)$ multiplicative

Observe low rank target constants are compatible with compositionality.

What's next

Trouble in paradise

- The standard Lambek systems are paragons of mathematical elegance,

  but ill equipped to deal with the harsh facts of language
- Is there a way of extending these systems that increases their linguistic sophistication while maintaining pleasant mathematical and computational properties?

Meeting the challenge Lambek Calculus and its modal extensions