Language Models and Relational Models of the Multiplicative-Additive Lambek Calculus

Max Kanovich, Stepan L. Kuznetsov, and Andre Scedrov

LAP 2021

Introduction: Algebra of Formal Languages

- Let Σ be a finite alphabet.
- By Σ^+ we denote the set of all non-empty words over Σ.
- $\mathcal{P}(\Sigma^+)$ is the set of all formal languages over Σ without the empty word.
- We introduce the following algebraic operations on $\mathcal{P}(\Sigma^+)$:

$$A \cdot B = \{uv \mid u \in A, v \in B\}$$
$$A \setminus B = \{u \in \Sigma^+ \mid (\forall v \in A)\ vu \in B\}$$
$$B \,/\, A = \{u \in \Sigma^+ \mid (\forall v \in A)\ uv \in B\}$$
$$A \vee B = A \cup B; \quad A \wedge B = A \cap B$$

- The most interesting operations are two divisions, $\setminus$ and $/$. They are connected to product in the following way:

$$B \subseteq A \setminus C \iff A \cdot B \subseteq C \iff A \subseteq C \,/\, B$$

Relational Algebras

- Another class of algebraic structures we are going to keep in mind is formed by the *algebras of binary relations.*
- Let W be a non-empty set. Fix a transitive binary relation $U \subseteq W \times W$, which we shall call the "universal" one.
- We take $\mathcal{P}(U)$, the set of all subrelations of U, and introduce algebraic operations in the same signature as on $\mathcal{P}(\Sigma^+)$:

$$
\begin{aligned}
&R \cdot S = R \circ S \\
&R \setminus S = \{\langle y, z\rangle \in U \mid (\forall \langle x, y\rangle \in R)\ \langle x, z\rangle \in S\} \\
&S \,/\, R = \{\langle x, y\rangle \in U \mid (\forall \langle y, z\rangle \in R)\ \langle x, z\rangle \in S\} \\
&R \vee S = R \cup S; \quad R \wedge S = R \cap S
\end{aligned}
$$

- Again,

$$S \subseteq R \setminus T \iff R \cdot S \subseteq T \iff R \subseteq T \,/\, S$$

Residuated Lattices

- Both algebras of languages and relational algebras are special kinds of a more general class of algebraic structures, *residuated lattices*.
- A residuated lattice is a tuple $\mathfrak{A} = (\mathcal{A}, \preceq, \cdot, \backslash, /, \vee, \wedge)$, where:
  - $\preceq$ is a partial order which forms a lattice, $\vee$ and $\wedge$ are lattice join and meet;
  - $(\mathcal{A}, \cdot)$ is a semigroup;
  - $b \preceq a \,\backslash\, c \iff a \cdot b \preceq c \iff a \preceq c \,/\, b$, for any $a, b, c \in \mathcal{A}$.
- Residuated lattices give algebraic semantics to substructural logics, like, for example, Heyting algebras do for intuitionism.

  N. Galatos, P. Jipsen, T. Kowalski, H. Ono. Residuated Lattices: An Algebraic Glimpse at Substructural Logics. Springer, 2007.
- The logic of residuated lattices is the multiplicative-additive Lambek calculus.

Multiplicative-Additive Lambek Calculus (MALC)

$$\frac{}{A \vdash A}\ \mathrm{Id}$$

$$\frac{\Pi \vdash A \quad \Gamma, B, \Delta \vdash C}{\Gamma, \Pi, A \backslash B, \Delta \vdash C}\ \mathrm{L}\backslash \qquad \frac{A, \Pi \vdash B}{\Pi \vdash A \backslash B}\ \mathrm{R}\backslash \quad (\Pi \text{ is not empty})$$

$$\frac{\Pi \vdash A \quad \Gamma, B, \Delta \vdash C}{\Gamma, B / A, \Pi, \Delta \vdash C}\ \mathrm{L}/ \qquad \frac{\Pi, A \vdash B}{\Pi \vdash B / A}\ \mathrm{R}/ \quad (\Pi \text{ is not empty})$$

$$\frac{\Gamma, A, B, \Delta \vdash C}{\Gamma, A \cdot B, \Delta \vdash C}\ \mathrm{L}\cdot \qquad \frac{\Pi_1 \vdash A \quad \Pi_2 \vdash B}{\Pi_1, \Pi_2 \vdash A \cdot B}\ \mathrm{R}\cdot$$

$$\frac{\Gamma, A, \Delta \vdash C \quad \Gamma, B, \Delta \vdash C}{\Gamma, A \vee B, \Delta \vdash C}\ \mathrm{L}\vee \qquad \frac{\Pi \vdash A}{\Pi \vdash A \vee B} \quad \frac{\Pi \vdash B}{\Pi \vdash A \vee B}\ \mathrm{R}\vee$$

$$\frac{\Gamma, A, \Delta \vdash C}{\Gamma, A \wedge B, \Delta \vdash C} \quad \frac{\Gamma, B, \Delta \vdash C}{\Gamma, A \wedge B, \Delta \vdash C}\ \mathrm{L}\wedge \qquad \frac{\Pi \vdash A \quad \Pi \vdash B}{\Pi \vdash A \wedge B}\ \mathrm{R}\wedge$$

Multiplicative-Additive Lambek Calculus (MALC)

- The cut rule of the following form:

$$\frac{\Pi \vdash A \quad \Gamma, A, \Delta \vdash C}{\Gamma, \Pi, \Delta \vdash C}\ \mathrm{Cut}$$

  is admissible in MALC.
- As said above, algebraic models of MALC are *residuated lattices:* variables and formulae are interpreted as elements of $\mathcal{A}$, and a sequent $A_1, \ldots, A_n \vdash B$ is interpreted as $A_1 \cdot \ldots \cdot A_n \preceq B$.
- Models on algebras of formal languages and models on relational algebras are called *L-models* and *R-models* respectively.
- MALC can be also viewed as a non-commutative intuitionistic version of linear logic (J.-Y. Girard, 1987). This was noticed by V. M. Abrusci (1990).

Lambek Categorial Grammars

- The original motivation for the Lambek calculus is its usage for describing natural language syntax (J. Lambek, 1958).
- This usage is connected to L-models.
- For each letter $a \in \Sigma$ the grammar associates one or more syntactic types, which are formulae of the Lambek calculus: $a \triangleright A$.
- A word $a_1 \dots a_n$ is considered grammatically correct, if the corresponding sequent $A_1, \dots, A_n \vdash s$ is derivable.
- The standard example is "John loves Mary," with the corresponding sequent $np, (np \setminus s) / np, np \vdash s$.

Part I: Distributivity

- Both L-models and R-models are distributive (as lattices): $(A \wedge B) \vee C \equiv (A \vee C) \wedge (B \vee C)$.
- In general, however, residuated lattices can be non-distributive.
- Thus, $(A \vee C) \wedge (B \vee C) \vdash (A \wedge B) \vee C$ is not derivable MALC, which prevents the latter from being L-complete or R-complete.
- Indeed, if this sequent were derivable, then it would be true in **all** residuated lattices, which would make them all distributive (which is not the case).
- There exists a natural, non-distributive modification of L-models which avoids this problem and gains completeness (C. Wurm 2017).

Partial Completeness Results

- $\mathrm{L}\wedge$, i.e., MALC without $\vee$, is R-complete (H. Andréka & Sz. Mikulás 1994)
- The Lambek calculus without $\vee$ and $\wedge$, is L-complete (M. Pentus 1995)
- $\mathrm{L}(\backslash, /, \wedge)$, that is, MALC with only three connectives: $\backslash, /, \wedge$, is L-complete (W. Buszkowski 1982)
- Open question: L-completeness of $\mathrm{L}\wedge$ (i.e., MALC without $\vee$).
- It is also unknown whether adding distributivity as an extra axiom yields completeness.
- *We show that the situation with* $\mathrm{L}\vee$ (i.e., MALC without $\wedge$) *is different from the one with* $\mathrm{L}\wedge$.

Distributivity without $\wedge$

- **Theorem.** The sequent

$$((x \,/\, y) \vee x) \,/((x \,/\, y) \vee (x \,/\, z) \vee x), (x \,/\, y) \vee x, \\ ((x \,/\, y) \vee x) \backslash((x \,/\, z) \vee x) \vdash (x \,/(y \vee z)) \vee x$$

  is not derivable in $\mathrm{L}\vee$, but can be derived using the distributivity axiom (and cut).
- Thus, $\mathrm{L}\vee$ is neither L-complete nor R-complete (because L-models and R-models are distributive).

How to Guess the Sequent?

- **Lemma.** If $A \vdash D$ and $B \vdash D$ are derivable (join), then for $C = (A \,/\, D) \cdot A \cdot (A \,\backslash\, B)$ we have $C \vdash A$ and $C \vdash B$ (meet). (see Lambek 1958, Pentus 1994)
- In particular, $C = (A \,/ (A \vee B)) \cdot A \cdot (A \,\backslash\, B)$ is a meet for A and B.
- Take $A = (x \,/\, y) \vee x$ and $B = (x \,/\, z) \vee x$.
- By distributivity,

$$((x \,/\, y) \vee x) \wedge ((x \,/\, z) \vee x) \vdash ((x \,/\, y) \wedge (x \,/\, z)) \vee x$$

- The succedent is equivalently replaced by $(x \,/ (y \vee z)) \vee x$.
- The antecedent is replaced by a stronger meet $C = (A \,/ (A \vee B)) \cdot A \cdot (A \,\backslash\, B)$ (it is stronger, since $C \vdash A$, $C \vdash B$, thus $C \vdash A \wedge B$).
- This yields, using cut, derivability of our sequent in the presence of distributivity.

Proving Non-Derivability in L$\vee$

- Non-derivability of our sequent in L$\vee$ does *not* come automatically from non-derivability of the distributivity law, since our new meet C is stronger than $A \wedge B$.
- However, the derivability problem is decidable, so we can just use derivability-checking software (developed by P. Jipsen, available online), which gives the answer in several seconds.
- In our WoLLIC 2019 paper, we also do manual proof search.
- One can also construct an algebraic countermodel (shorter, but requires some creativity).

Commutative and Affine Generalizations

- Adding the permutation rule of the following form

$$\frac{\Gamma, A, B, \Delta \vdash C}{\Gamma, B, A, \Delta \vdash C}\ \mathrm{P}$$

  to MALC (that is, making things commutative) gives the multiplicative-additive fragment of intuitionistic linear logic (ILL).

- If one additionally adds weakening

$$\frac{\Gamma, \Delta \vdash C}{\Gamma, A, \Delta \vdash C}\ \mathrm{W}$$

  this will give the multiplicative-additive fragment of intuitionistic affine logic (IAL).

Commutative and Affine Generalizations

- The sequent

$$((x\,/\,y) \vee x)\,/((x\,/\,y) \vee (x\,/\,z) \vee x), (x\,/\,y) \vee x,$$
$$((x\,/\,y) \vee x)\backslash((x\,/\,z) \vee x) \vdash (x\,/(y \vee z)) \vee x$$

  is still not derivable if we add commutativity (permutation rule), that is, in ILL.
- For the affine case (IAL, with weakening rule), the sequent should be slightly modified

$$((x\,/\,y) \vee w)\,/((x\,/\,y) \vee (x\,/\,z) \vee w), (x\,/\,y) \vee w,$$
$$((x\,/\,y) \vee w)\backslash((x\,/\,z) \vee w) \vdash (x\,/(y \vee z)) \vee w$$

Part II: Systems with the Unit

- In intuitionistic linear logic, the unit constant (multiplicative truth) is axiomatized as follows:

$$\frac{\Gamma, \Delta \vdash C}{\Gamma, 1, \Delta \vdash C}\ \mathrm{L1} \qquad \frac{}{\vdash 1}\ \mathrm{R1}$$

- Thus, adding 1 requires abolishing antecedent non-emptiness restriction.
- In residuated lattices, this corresponds to moving from arbitrary semigroups (recall that, in any residuated lattice, $(\mathcal{A}, \cdot)$ is a semigroup) to monoids: $(\mathcal{A}, \cdot, 1)$.
- In particular, we modify the definition of L-models by allowing the empty word in languages.

Undecidability with the Unit

- The multiplicative unit constant, 1, is necessarily interpreted in L-models as $\{\varepsilon\}$ (due to $A \cdot 1 \vdash A$).
- Axiomatising the unit as multiplicative truth in linear logic yields incomplete systems: for example, $(1 \wedge G) \cdot F \equiv F \cdot (1 \wedge G)$ is true in L-models, but not derivable in non-commutative linear logic.
- We present a minimal system $\mathsf{L}^{+\varepsilon}(\backslash, \wedge, 1)$, which captures the following L-correct principles: $A \cdot \{\varepsilon\} = \{\varepsilon\} \cdot A$ ("commuting") and $\{\varepsilon\} \cdot \{\varepsilon\} = \{\varepsilon\}$ ("doubling").
- Notice that it is in the language of $\backslash, \wedge, 1$ only.

$\mathsf{L}^{+\varepsilon}(\backslash, \wedge, 1)$

$$\frac{}{A \vdash A}\ \mathrm{Id} \qquad \frac{}{A, 1 \vdash A}\ 1$$

$$\frac{\Pi \vdash A \quad \Gamma, B, \Delta \vdash C}{\Gamma, \Pi, A \backslash B, \Delta \vdash C}\ \mathrm{L}\backslash \qquad \frac{A, \Pi \vdash B}{\Pi \vdash A \backslash B}\ \mathrm{R}\backslash$$

$$\frac{\Gamma, A, \Delta \vdash C}{\Gamma, A \wedge B, \Delta \vdash C} \quad \frac{\Gamma, B, \Delta \vdash C}{\Gamma, A \wedge B, \Delta \vdash C}\ \mathrm{L}\wedge \qquad \frac{\Pi \vdash A \quad \Pi \vdash B}{\Pi \vdash A \wedge B}\ \mathrm{R}\wedge$$

$$\frac{\Gamma, A, (1 \wedge G), \Delta \vdash C}{\Gamma, (1 \wedge G), A, \Delta \vdash C}\ \mathrm{L}\varepsilon \qquad \frac{\Gamma, (1 \wedge G), A, \Delta \vdash C}{\Gamma, A, (1 \wedge G), \Delta \vdash C}\ \mathrm{R}\varepsilon$$

$$\frac{\Gamma, (1 \wedge G), (1 \wedge G), \Delta \vdash C}{\Gamma, (1 \wedge G), \Delta \vdash C}\ \mathrm{D}\varepsilon$$

- **Theorem.** Any system which includes $\mathsf{L}^{+\varepsilon}(\backslash, \wedge, 1)$ and is L-sound is undecidable.
- In particular, so is the set of all L-true sequents, but for this set we do not even know whether it is r.e.

Undecidability Proof Sketch

- We encode 2-counter Minsky machines.
- The direction from computations to derivations is established by constructing the corresponding proofs in $\mathsf{L}^{+\varepsilon}(\backslash, \wedge, 1)$.
- The backwards direction is performed via L-models.

Encoding Minsky Machines

- Atoms (propositional variables): e_1, e_2 (start/end markers); p_1, p_2 (the number of p_i's is the value of counter c_i); $\ell_0, \ell_1, \ldots$ (states of the machine); b.
- If the machine is in state L_i, with $c_1 = k_1$ and $c_2 = k_2$, then it is encoded as follows:

$$e_1, \underbrace{p_1, \ldots, p_1}_{k_1 \text{ times}}, \ell_i, \underbrace{p_2, \ldots, p_2}_{k_2 \text{ times}}, e_2$$

Encoding Minsky Machines

- Each instruction I of the machine is encoded by the corresponding formula A_I
  ($F^{bb} = (F \setminus b) \setminus b$ is the pseudo-double-negation):

I	A_I
$L_i : inc(c_1);\ goto\ L_j;$	$\ell_i \setminus (p_1 \cdot \ell_j)^{bb}$
$L_i : inc(c_2);\ goto\ L_j;$	$\ell_i \setminus (\ell_j \cdot p_2)^{bb}$
$L_i : dec(c_1);\ goto\ L_j;$	$(p_1 \cdot \ell_i) \setminus \ell_j^{bb}$
$L_i : dec(c_2);\ goto\ L_j;$	$(\ell_i \cdot p_2) \setminus \ell_j^{bb}$
$L_i : if(c_1 = 0)\ goto\ L_j;$	$(e_1 \cdot \ell_i) \setminus (e_1 \cdot \ell_j)^{bb}$
$L_i : if(c_2 = 0)\ goto\ L_j;$	$(\ell_i \cdot e_2) \setminus (\ell_j \cdot e_2)^{bb}$

Encoding Minsky Machines

- All operations are encoded into a leading $1 \wedge G$, where G is a big conjunction.
- G includes the following formulae:
  - A_I for each instruction I of our Minsky machine. Each A_I is of the form $g_{\alpha,\beta} = \beta \setminus \alpha^{bb}$.
  - $g_{\xi,\xi} = \xi \setminus \xi^{bb}$ for each atom ξ.
  - $(e_1 \cdot \ell_0 \cdot e_2) \setminus b$, for terminating computation. (L_0 is the final state, and the counters are required to be zero.)

Lemma

The sequent $1 \wedge G, \Delta \vdash b$, where Δ encodes the initial configuration of the machine, is derivable in $\mathrm{L}^{+\varepsilon}(\setminus, \wedge, 1)$ if and only if the machine reaches state L_0 with zero counters, starting from this initial configuration.

From Computations to Derivations

- Since G includes $g_{\alpha,\beta} = (\beta \setminus \alpha^{bb})$, then derivability of $1 \wedge G, \alpha, \Delta \vdash b$ yields derivability of $1 \wedge G, \Delta, \beta \vdash b$.
  - This enables Minsky commands, but only on the left side of the configuration.
  - This derivation essentially uses "doubling."
- Cyclic transpositions. If G includes $g_{\xi,\xi} = \xi \setminus \xi^{bb}$ for any atom ξ (which is the case), and Δ_1, Δ_2 are all atomic, then derivability of $1 \wedge G, \Delta_1, \Delta_2 \vdash b$ yields derivability of $1 \wedge G, \Delta_2, \Delta_1 \vdash b$.
  - This allows locating $1 \wedge G$ near the necessary place in the configuration.
- Finally, we have $(e_1 \cdot \ell_0 \cdot e_2) \setminus b$ in G.
  - This encodes the finish of computation, $(L_0, 0, 0)$.

From Derivations to Computations

- Let Σ (alphabet) include all atoms.
- Let B_M be the set of "terminating strings," that is, codes of configurations of the Minsky machine M, such that the machine, starting from this configuration, reaches the terminating one $(L_0, 0, 0)$.
- Consider the following L-interpretation:

$$w(a) = \begin{cases} \{a\}, & \text{if } a \neq b \\ \{xy \mid yx \in B_M\}, & \text{for } a = b \end{cases}$$

- **Lemma.** For any instruction I of M, $w(A_I) \ni \varepsilon$. Hence, $w(1 \wedge G) = \{\varepsilon\}$.
- If $1 \wedge G, e_1, \underbrace{p_1, \ldots, p_1}_{k_1 \text{ times}}, \ell_i, \underbrace{p_2, \ldots, p_2}_{k_2 \text{ times}} \vdash b$ is derivable, then interpretation of the antecedent is in $w(b)$, whence the configuration (L_i, k_1, k_2) terminates to $(L_0, 0, 0)$.

Models on Regular Languages

- Recall that the class of regular languages is the minimal class of languages including $\varnothing$, $\{\varepsilon\}$, singletons $\{a\}$ for any $a \in \Sigma$, and closed under language multiplication, union, and iteration (Kleene star): $A^* = \{\varepsilon\} \cup A \cup (A \cdot A) \cup (A \cdot A \cdot A) \cup \dots$
- A specific class of L-models includes only models in which all languages are regular.
- We shall call such models $\mathrm{L}\textsc{reg}$-models.
- This definition is consistent, since the class of regular languages is closed under Lambek operations.
- Without the unit constant 1, the calculus $\mathrm{L}(\backslash, /, \wedge)$ is complete w.r.t. $\mathrm{L}\textsc{reg}$-models (this follows from Buszkowski's and Sorokin's work).

Models on Regular Languages

- The situation changes if we add the unit.
- We still consider theories in the language of MALC with the unit constant.
- As shown by the encoding of Minsky machines above, the theory of **all** L-models in the language of $\backslash, \wedge, 1$ is undecidable; more precisely—Σ_1^0-hard.
- **NB:** we do not claim that it belongs to Σ_1^0, it could be harder!
- On the other hand, the theory of the **subclass** of $\mathrm{L}\textsc{reg}$-models belongs to the Π_1^0 class.
- Indeed, we now have to quantify over regular languages, that is, over regular expressions. This yields an *arithmetical* universal quantifier, thus Π_1^0.

Models on Regular Languages

- Since no Σ_1^0-hard language can belong to Π_1^0, we get the following

Theorem

The theories of L-models and of $\textsc{Lreg}$*-models, in the language of* $\backslash, \wedge, 1$*, are different.*

○ NO regular model property

Theorem 1.1 We can find a sequent of the specific form

$$\mathbb{1} \wedge G, \ \Delta \vdash b$$

so that

(a) We can construct an L-model such that the sequent is not valid in the model,

(b) But the sequent is valid in any LREG-models.

○ The minimalistic propositional systems that are still PSPACE-complete

Main Complexity Results:

Commutative (Linear logic)	Non-commutative (Lambek, circular)
$\mathcal{L}^1(\backslash)$ is NP-complete (Kanovich)	$\mathcal{L}^1(\backslash)$ is polytime (Savateev)
$\mathcal{L}^1(\backslash, \wedge)$ is PSPACE-complete	$\mathcal{L}^1(\backslash, \wedge)$ is PSPACE-complete
$\mathcal{L}^1(\backslash, \vee)$ is PSPACE-complete	$\mathcal{L}^1(\backslash, \vee)$ is PSPACE-complete

One implication, one conjunction or one disjunction. Here $\mathcal{L}^1(\backslash)$, $\mathcal{L}^1(\backslash, \wedge)$, etc., denote fragments with **only one** variable.