The Lambek Calculus with Unary Connectives

Stepan Kuznetsov
Steklov Mathematical Institute (Moscow), Russia
(partially based on joint work with M. Kanovich, A. Scedrov, and N. Ryzhkova)

1 The Lambek Calculus

The Lambek calculus **L** was introduced by J. Lambek [7]. Formulae *(types)* of **L** are are built from a countable set of variables *(primitive types)* $\mathrm{Var} = \{p, q, r, \dots\}$ using three binary connectives: $\backslash$ *(left division)*, $/$ *(right division)*, and $\cdot$ *(multiplication)*. The set of all types is denoted by Tp. The Lambek calculus derives *sequents* of the form $\Pi \to A$, where A is a type and Π is a sequence of types. In **L**, Π is required to be non-empty. There exists a variant of the Lambek calculus, $\mathbf{L}^*$, without this restriction.

The axioms and rules of **L** are as follows: $\qquad A \to A$

$$\frac{\Pi, A \to B}{\Pi \to B \,/\, A}\ (\to /) \qquad \frac{A, \Pi \to B}{\Pi \to A \,\backslash\, B}\ (\to \backslash)$$

$$\frac{\Pi \to A \quad \Gamma, B, \Delta \to C}{\Gamma, (B \,/\, A), \Pi, \Delta \to C}\ (/ \to) \qquad \frac{\Pi \to A \quad \Gamma, B, \Delta \to C}{\Gamma, \Pi, (A \,\backslash\, B), \Delta \to C}\ (\backslash \to)$$

$$\frac{\Gamma, A, B, \Delta \to C}{\Gamma, (A \cdot B), \Delta \to C}\ (\cdot \to) \qquad \frac{\Pi_1 \to A \quad \Pi_2 \to B}{\Pi_1, \Pi_2 \to A \cdot B}\ (\to \cdot)$$

Lambek syntactic types can be naturally interpreted as formal languages over an alphabet Σ. For the interpretation w, the following should hold:

$$w(A \cdot B) = w(A) \cdot w(B) = \{uv \mid u \in w(A), v \in w(B)\}$$
$$w(B \,/\, A) = w(B) \,/\, w(A) = \{v \mid (\forall u \in w(A))\, vu \in w(B)\}$$
$$w(A \,\backslash\, B) = w(A) \,\backslash\, w(B) = \{v \mid (\forall u \in w(A))\, uv \in w(B)\}$$

Note that this definition works differently for **L** and $\mathbf{L}^*$ (for **L**, the empty word is not included in the languages). A sequent $A_1, \dots, A_n \to B$ is true under interpretation w, if $w(A_1) \cdot \ldots \cdot w(A_n) \subseteq w(B)$. Both variants of the Lambek calculus are sound and complete w.r.t. this interpretation:

Theorem 1 (M. Pentus). *A sequent is derivable in* **L** *(resp., in* $\mathbf{L}^*$*) iff it is true under any interpretation* $w\colon \mathrm{Tp} \to \mathcal{P}(\Sigma^+)$ *(resp.,* $w\colon \mathrm{Tp} \to \mathcal{P}(\Sigma^*)$*).* [10]

Lambek *categorial grammars* are finite correspondences between Lambek types and letters of an alphabet. The word $a_1 \dots a_n$ belongs to the language generated by such grammar if there exist types $A_1, \dots, A_n$ in the correspondence with letters $a_1, \dots, a_n$ resp., such that $A_1, \dots, A_n \to H$ is derivable in $\mathbf{L}$ or one of its variants. Here H is a fixed type, usually primitive.

Theorem 2 (M. Pentus)**.** *Grammars based on* $\mathbf{L}$ *(resp., on* $\mathbf{L}^*$*) generate precisely the class of context-free languages without the empty word (resp., the class of all context-free languages).* [9]

2 The Reversal Operation

The unary *reversal* operation is defined as follows: $M^{\mathrm{R}} = \{a_n \dots a_1 \mid a_1 \dots a_n \in M\}$ for any formal language M. The extension of $\mathbf{L}$ with the unary $(\cdot)^{\mathrm{R}}$ connective, $\mathbf{L}^{\mathrm{R}}$, is obtained from $\mathbf{L}$ by adding the following rules ($\Gamma^{\mathrm{R}} = A_n^{\mathrm{R}}, \dots, A_1^{\mathrm{R}}$ for $\Gamma = A_1, \dots, A_n$):

$$\frac{\Gamma \to C}{\Gamma^{\mathrm{R}} \to C^{\mathrm{R}}}\ ({}^{\mathrm{R}} \to {}^{\mathrm{R}}) \qquad \frac{\Gamma, A^{\mathrm{RR}}, \Delta \to C}{\Gamma, A, \Delta \to C}\ ({}^{\mathrm{RR}} \to)_{\mathrm{E}} \qquad \frac{\Gamma \to C^{\mathrm{RR}}}{\Gamma \to C}\ (\to {}^{\mathrm{RR}})_{\mathrm{E}}$$

The good properties of the Lambek calculus keep valid for its extension with the reversal operation ([4] for $\mathbf{L}$, [5] for $\mathbf{L}^*$):

Theorem 3. *The calculi* $\mathbf{L}^{\mathrm{R}}$ *and* $\mathbf{L}^{*\mathrm{R}}$ *are sound and complete w.r.t. interpretations of types as formal languages.*

Theorem 4. $\mathbf{L}^{\mathrm{R}}$*-grammars (resp.,* $\mathbf{L}^{*\mathrm{R}}$*-grammars) generate precisely the class of context-free languages without the empty word (resp., the class of all context-free languages).*

3 The (Sub)exponential

The *exponential* modality, ! (called "bang"), is inherited from linear logic. It is governed by the following rules. Here we consider only the $\mathbf{L}^*$ case, since the $\mathbf{L}$ one is much more subtle (see [1]).

$$\frac{\Gamma, A, \Delta \to B}{\Gamma, !A, \Delta \to B}\ (! \to) \qquad \frac{!A_1, \dots, !A_n \to B}{!A_1, \dots, !A_n \to !B}\ (\to !) \qquad \frac{\Gamma, \Delta \to B}{\Gamma, !A, \Delta \to B}\ (\mathrm{weak})$$

$$\frac{\Gamma, !A, !A, \Delta \to B}{\Gamma, !A, \Delta \to B}\ (\mathrm{contr}) \qquad \frac{\Gamma, !A, \Delta, \Phi \to B}{\Gamma, \Delta, !A, \Phi \to B}\ (\mathrm{perm}_1) \qquad \frac{\Gamma, \Delta, !A, \Phi \to B}{\Gamma, !A, \Delta, \Phi \to B}\ (\mathrm{perm}_2)$$

We also consider a less powerful modality, for which we impose contraction and permutation, but not weakening. We also denote it by ! and call a *subexponential.* This modality is motivated from the linguistic side [8].

Theorem 5. *Grammars based on the extension of* $\mathbf{L}^*$ *with* ! *(both with and without* (weak)*) generate all recursively enumerable languages.* [2]

This is obtained by encoding finite *theories* over $\mathbf{L}^*$ inside sequents using the (sub)exponential modalities.

Corollary 6. *The derivability problem for* $\mathbf{L}^*$ *extended with* ! *is undecidable.*

However, in linguistical practice ! is usually applied only to variables. For this case, the derivability problem is decidable and belongs to NP [2].

4 The Kleene Star

Yet another important operation on formal languages is the *Kleene star:* $A^* = \bigcup_{n=0}^{\infty} A^n$. For the Kleene star, we propose the following rules extending $\mathbf{L}^*$:

$$\frac{\Gamma_1 \to A \quad \ldots \quad \Gamma_n \to A}{\Gamma_1, \ldots, \Gamma_n \to A^*} \ (\to {}^*)$$

$$\frac{\Gamma, \Delta \to C \quad \Gamma, A, A^*, \Delta \to C}{\Gamma, A^*, \Delta \to C} \ ({}^* \to)_{\mathrm{L}} \qquad \frac{\Gamma, \Delta \to C \quad \Gamma, A^*, A, \Delta \to C}{\Gamma, A^*, \Delta \to C} \ ({}^* \to)_{\mathrm{R}}$$

In this system we allow infinite branches of proofs.

For the fragment without $\cdot$ and where * is allowed only in subformulae of the form $A^* \setminus B$ or $B \mathbin{/} A^*$, this calculus enjoys completeness w.r.t. interpretations of types as formal languages [6].

There is an open question whether we could take only regular (cyclic) proofs, like in [11]. However, if we take both * and !, the answer is negative: using results from [3] for theories over Kleene algebras and then encoding the theory into the sequent using the construction from [2], one obtains Π_2^0-hardness of the system. Therefore, it is not equivalent to any system with finite proofs.

Acknowledgements

The author thanks all his colleagues who participated in discussions on the topics considered in this abstract, in particular, Lev Beklemishev, Michael Kaminski, Glyn Morrill, Fedor Pakhomov, Mati Pentus, Daniyar Shamkanov, Alexey Sorokin, Stanislav Speranski, and, of course, his co-authors, Max Kanovich, Andre Scedrov, and Nadezhda Ryzhkova.

Stepan Kuznetsov's work was supported by the Russian Foundation for Basic Research (grants 11-01-00281-a, 12-01-00888-a, 14-01-00127-a, and 15-01-09218-a), by the Presidential Council for Support of Leading Scientific Schools (grants NŠ-65648.2010.1, NŠ-5593.2012.1, NŠ-1423.2014.1, and NŠ-9091.2016.1), and by the Scientific and Technological Cooperation Programme Switzerland – Russia (project "Computational Proof Theory", 2010–2012).

References

[1] M. Kanovich, S. Kuznetsov, A. Scedrov. On Lambek's restriction in the presence of exponential modalities. Proc. LFCS 2016 (LNCS vol. 9537), 141–158.

[2] M. Kanovich, S. Kuznetsov, A. Scedrov. Undecidability of the Lambek calculus with a relevant modality. Proc. FG 2015/2016 (LNCS vol. 9804), 2016 (*to appear,* arXiv: 1601.06303).

[3] D. Kozen. On the complexity of reasoning in Kleene algebra. Information and Control **179**, 2002, 152–162.

[4] S. Kuznetsov. L-completeness of the Lambek calculus with the reversal operation. Proc. LACL 2012 (LNCS vol. 7351), 151–160.

[5] S. Kuznetsov. L-completeness of the Lambek calculus with the reversal operation allowing empty antecedents. Lambek 90 Festschrift (LNCS vol. 8222), 2014, 268–278.

[6] S. Kuznetsov, N. Ryzhkova. A fragment of the Lambek calculus with iteration (in Russian). Proc. Mal'tsev Meeting 2015, Novosibirsk.

[7] J. Lambek. The mathematics of sentence structure. Amer. Math. Monthly, 1958, 65(3), 154–170.

[8] G. Morrill, O. Valentín. Computational coverage of TLG: Nonlinearity. Proc. NLCS 2015 (EPiC Series, vol. 32), 51–63.

[9] M. Pentus. Lambek grammars are context free. Proc. 8th IEEE Symposium on Logic in Computer Science, 1993, 429–433.

[10] M. Pentus. Models for the Lambek calculus. APAL **75**, 1995, 179–213.

[11] D. S. Shamkanov. Circular proofs for the Gödel–Löb provability logic. Math. Notes **96**, 2014, 575–585.