Cover Page

Name	Tikhon Pshenitsyn
Email	tpshenitsyn@mi-ras.ru
Phone Number	+79199682119
Title of the Work	Hypergraph Lambek Calculus
Award Category	Outstanding Master Thesis Award
Degree Awarded	30.06.2023
University	Lomonosov Moscow State University

Summary of the Thesis "Hypergraph Lambek Calculus"

Tikhon Pshenitsyn

June 1, 2024

The thesis *Hypergraph Lambek Calculus* is devoted to developing an extension of the Lambek calculus and Lambek categorial grammars to hypergraphs. The research is inspired by the formal grammar theory. There exist several major paradigms for describing formal languages, which include automata, grammars, definability by formuli etc. One of them is the categorial grammar approach. This concept arises in the works by Ajdukiewicz, Bar-Hillel and Lambek. In the article "The Mathematics of Sentence Structure" [2], Lambek introduced a logical calculus which formulas are used as syntactic categories in linguistics. The Lambek calculus L is a substructural logic, more precisely, it can be characterised as the non-commutative multiplicative intuitionistic linear logic. Its formulas are built using the product operation $\cdot$ and two divisions $\backslash, /$. On the basis of the Lambek calculus, one can define Lambek categorial grammars as follows. A Lambek categorial grammar consists of an alphabet Σ, a finite binary relation $\triangleright$ between symbols of Σ and formulas of the Lambek calculus, and a distinguished formula S. The grammar generates the language of all strings $a_1 \dots a_n$ for which there exist formulas $T_1, \dots, T_n$ such that $a_i \triangleright T_i$, and the sequent $T_1 \cdot \ldots \cdot T_n \to S$ is derivable in L.

The Lambek calculus L can be modified in several ways in order to model certain linguistic phenomena the basic system cannot capture. First, one can change the basic rules for the product $A \cdot B$. In the Lambek calculus, this operation is associative but not commutative. By adding commutativity one obtains the multiplicative intuitionistic linear logic MILL, which is also denoted by LP in the literature. On the other hand, by rejecting associativity of $\cdot$ in L, one obtains the system NL, which also has applications in linguistics. Secondly, one can enrich the Lambek calculus with additional operations. For example, in [3], the non-associative Lambek calculus with modalities is presented; in [4], the displacement calculus is introduced, which includes discontinuous operations. These are many different logical systems, which, however, are quite similar in their properties. All of them can be called *multiplicative* since the operations used in them are subject to the same principles as the multiplicative operations of linear logic. One might ask for a most general calculus that would unify the abovementioned ones.

Our idea of generalising the Lambek calculus is to revise the kind of structures we work with. The Lambek calculus deals with string structures since natural language sentences can be viewed as simply strings. However, if we allow commutativity, then it would be more natural to work with multisets instead of strings; on contrary, if we disallow associativity, then we have to work with trees. Hence, many modifications of the Lambek calculus are simply modifications of underlying structures. To make the theory sufficiently general, one should assume that the structures behind the Lambek calculus are hypergraphs. Namely, while the product operation $A \cdot B$ of L can be considered as a combination of the resources A and B in a linear, string-like way, there should be many other ways of resource combinations. This is made possible in the *hypergraph Lambek calculus* HL introduced in [8, 5, 6]. This is a substructural Gentzen-style logic such that its formulas are built using hypergraphs. We use the definition of a hypergraph from [1] where hyperedge replacement grammars are introduced. Let us present the definition of a formula of HL from [8] (with some simplifications).

Definition 1. The set Fm of formulas is defined inductively as follows:

1. Let Pr be the set of propositional variables (called *primitive formulas*). Then $Pr \subseteq \mathrm{Fm}$.

2. Let M be a hypergraph such that its hyperedges are labeled by formulas from Fm. Then $T = \times(M)$ belongs to Fm.

3. Let \$ stand for a special label ("placeholder"). Let N be a formula ($N \in \mathrm{Fm}$) and let D be a hypergraph such that exactly one of its hyperedges (call it e_0) is labeled by \$, while other edges are labeled by formulas from Fm. Then $T = N \div (D)$ also belongs to Fm.

A formula of the form $\times(M)$ is understood as the combination of resources according to the hypergraph M. For example, the formula $p \cdot q$ of L can be represented as the formula $\times\left({\scriptstyle (1)} \bullet \xrightarrow{p} \bullet \xrightarrow{q} \bullet {\scriptstyle (2)} \right)$ of HL, which means that one combines p and q using a string graph[1]. A formula of the form $N \div (D)$ represents a kind of objects such that, if any object of this kind is placed instead of the placeholder hyperedge in D, then the resulting object is of the type N. For example, the formula p/q of L corresponds to the formula $p \div \left({\scriptstyle (1)} \bullet \xrightarrow{\$} \bullet \xrightarrow{q} \bullet {\scriptstyle (2)} \right)$ (such formulas are used in the translation of L into HL).

The main part of the thesis starts with thoroughly motivating the definition of HL (Section 2.1), which are quite large. This is done by carefully comparing the principles of context-free grammars, Lambek categorial grammars and hyperedge replacement grammars. Namely, we aimed to complete the following square in a natural way.

context-free grammar	VS	Lambek categorial grammar *Lambek calculus* L
$\wr$		$\wr$
hyperedge replacement grammar [1]	VS	hypergraph Lambek categorial grammar *hypergraph Lambek calculus* HL

In Section 2.2, we formally define the notion of a formula and a sequent of HL and we present axioms and rules of this calculus[2]. Then, we start investigating properties of the resulting formalism and its connections to the calculi studied in the literature. Our goal was to justify that the presented definitions are indeed correct, natural, even canonical in some sense (i.e. that our way is the rightest way to define a hypergraph extension of the Lambek calculus). First of all, in Section 3.1, we introduce the cut rule (which is formulated in HL by using the hyperedge replacement operation on hypergraphs) and we prove cut elimination. This is a fundamental property one would like a sequent calculus to enjoy. As a corollary, we show that the left rule for $\times$ and the right rule for $\div$ are reversible (Section 3.2).

In Section 4, we justify the role of HL as an umbrella logic which unifies several variants of the Lambek calculus. In Section 4.1, we prove that the Lambek calculus L can be embedded in HL; in Section 4.2, we prove the same for the Lambek calculus with the unit. In Section 4.3, we prove that the Lambek calculus with permutation LP can be embedded in HL. These are just a few examples of embeddings included in the thesis; apart from them, in [7], it is shown how to embed the displacement calculus in HL. This and some other embeddings are not included in the thesis because of their length: we had to include only the most basic and important results.

In Section 5, we turn to the issue of generalising Lambek categorial grammars to hypergraphs. Note that, to our best knowledge, there has been no generalisation of the categorial grammar approach to the area of graph grammars, which is a serious omission. We introduce hypergraph Lambek categorial grammars and prove that any lexicalised hyperedge replacement grammar can be transformed into an equivalent hypergraph Lambek categorial grammar (this result generalises a known relation between context-free grammars and Lambek grammars for strings). We also prove that the converse is not true since hypergraph Lambek categorial grammars are able to generate languages that cannot be generated by hyperedge replacement grammars. Namely, we show that the set of connected graphs can be generated by a hypergraph Lambek categorial grammar.

[1] According to the definition of hypergraph we use (see [1] for details), hypergraphs have labels on hyperedges; the vertices marked by (1) and (2) are external vertices, they are needed for the hyperedge replacement operation.

[2] The logic HL has one axiom, 2 rules for $\times$ and 2 rules for $\div$.

Finally, in Section 6, we investigate algorithmic complexity of HL and prove that the derivability problem for this calculus is NP-complete. This is a nice fact since the Lambek calculus is NP-complete as well, so we can generalise it significantly without increasing its complexity.

The thesis is written in the Russian language; however, it is based on the articles [8, 5, 6] published in English. The article [8] is published in the Journal of Logical and Algebraic Methods in Programming, and it is a journal version of the paper presented at ICGT 2021 (featured as the best theoretical paper by EATCS); the article [5] is published in the Post-proceedings of the International Workshop on Graph Computation Models; the article [6] is an arXiv preprint. I attach all the three papers to the submission (they are in the same pdf file as the thesis). Below, I present the table of contents of the thesis translated into English and indicate how the chapters of the thesis correspond to articles' content.

1. Preliminaries [8, Section 2]
   (a) Basic notions and notation
   (b) Context-free grammar
   (c) Lambek calculus and Lambek grammar
   (d) Modifications of the Lambek calculus
   (e) Hypergraphs, hyperedge replacement
   (f) Hyperedge replacement grammar
2. Definition of the hypergraph Lambek calculus [8, Section 3], [5, Section 3]
   (a) Informal considerations [5, Section 3]
   (b) Formal definition [8, Section 3.1]
   (c) Auxiliary notions
3. Structural properties of HL [8, Section 5]
   (a) Cut elimination [8, Theorem 4], [8, Appendix A]
   (b) Reversibility of the rules $(\times L)$ and $(\div R)$ [8, Proposition 1]
4. Embedding theorems [8, Section 3.3, Section 3.4], [6, Section 4]
   (a) Lambek calculus [8, Section 3.3]
   (b) Lambek calculus with the unit [6, Section 4.4]
   (c) Lambek calculus with the permutation rule [8, Section 3.4]
   (d) On embedding other calculi
5. Hypergraph Lambek grammars and hyperedge replacement grammars [8, Section 6]
6. Algorithmic complexity [8, Section 6]

The thesis is designed as a very detailed and unhurried introduction to the area of the hypergraph Lambek calculus featuring its most fundamental and basic features. The definitions are illustrated by numerous examples. The obtained results show that the problem of finding an appropriate generalisation of L and its variants to the hypergraph setting as well as the problem of combining the areas of categorial grammars and graph grammars are successfully solved. The logic HL allows one to reason about resources organised in an arbitrary way, in particular, about commutative resources, nonassociative resources etc.

Some of the results concerning HL were not included in the thesis due to size limitations. E.g., in [8], we prove that hypergraph Lambek categorial grammars generate all finite intersections of languages generated by hyperedge replacement grammars, which shows that the former are significantly more powerful than the latter.

The work as well as the research area itself is novel, I did this work solely. The work falls in the area of computational logic, namely, to non-classical logics, hence it is relevant to areas of interest of the VCLA Award.

References

[1] Frank Drewes, Hans-Jörg Kreowski, and Annegret Habel. Hyperedge replacement graph grammars. In Grzegorz Rozenberg, editor, *Handbook of Graph Grammars and Computing by Graph Transformations, Volume 1: Foundations*, pages 95–162. World Scientific, 1997.

[2] Joachim Lambek. The mathematics of sentence structure. *The American Mathematical Monthly*, 65(3):154–170, 1958.

[3] Michael Moortgat. Multimodal linguistic inference. *J. Log. Lang. Inf.*, 5(3/4):349–385, 1996.

[4] Glyn Morrill, Oriol Valentín, and Mario Fadda. The displacement calculus. *J. Log. Lang. Inf.*, 20(1):1–48, 2011.

[5] Tikhon Pshenitsyn. Grammars based on a logic of hypergraph languages. In Berthold Hoffmann and Mark Minas, editors, Proceedings Twelfth International Workshop on *Graph Computational Models,* Online, 22nd June 2021, volume 350 of *Electronic Proceedings in Theoretical Computer Science*, pages 1–18. Open Publishing Association, 2021.

[6] Tikhon Pshenitsyn. Introduction to a hypergraph logic unifying different variants of the lambek calculus, 2021.

[7] Tikhon Pshenitsyn. Displacement calculus and hypergraph Lambek calculus. In *Proceedings International Junior Scientific Forum "Lomonosov-2022" (in Russian)*. MAKS Press, 2022.

[8] Tikhon Pshenitsyn. Hypergraph Lambek grammars. *Journal of Logical and Algebraic Methods in Programming*, 129:100798, 2022.