Stepan Lvovich Kuznetsov. Research Statement Summary

The Lambek calculus was introduced in 1958 for linguistic purposes, but it enjoys a very natural algebraic interpretation. Consider the free semigroup and its powerset. On this powerset one can naturally define three operations—multiplication, left and right division. Multiplication is defined evidently, and B/A is such a set, that $C \subseteq B/A \iff C \cdot A \subseteq B$ for any C, and the same for the other division. Subsets of the free semigroup are naturally called *formal languages*.

We consider the set of all facts, of the form like $(B/A) \cdot A \subseteq B$, with variables for formal languages, that are true for any assignment of values (formal languages) to the variables. It happens (M. Pentus 1995), that these facts are exactly the formulae derivable in the Lambek calculus. This result is called the *L-completeness theorem*.

The Lambek calculus and its variants also have other interpretations. For example, adding the monotonicity rule gives a system describing the atomic theory of division of semiring ideals (A. Pentus & M. Pentus 2006). Lambek himself introduced interpretations of his calculus in category theory.

It looks interesting whether the Lambek calculus can be extended via new connectives with natural interpretation on formal languages, preserving L-completeness. For the reversal operation (that inverts letter order in all words in the language), and L-complete extension of the Lambek calculus is constructed (S. K. 2012). For Boolean operations (union and intersection) the natural extension is L-incomplete, since it doesn't derive the law of distributivity. Even adding the unit constant leads to incompleteness, giving a fragment equivalent to the trivalent system RM_3.

We're planning to study L-complete extensions of the Lambek calculus with the unit constant, with the iteration operation (Kleene star), and also the commutative variant. In the latter case the interesting question is completeness over the powerset of the infinite cyclic (1-generated) semigroup. For these extensions, we're planning to prove L-completeness, build good Gentzen-style calculi and investigate their algorithmic complexity (generalizing M. Pentus 2003, Yu. Savateev 2008).

The other line of research concerns *categorial grammars.* These grammars use the Lambek calculus and its variants to give finite descriptions of (usually infinite) formal languages. Letters of the alphabet (linguistically: words) are assigned syntactic types (possibly many, but a finite set for each letter), and the word is put into the language iff the corresponding sequent is derivable in the calculus. For the original Lambek calculus, the languages generated by these grammars are exactly context-free languages (M. Pentus 1992). It is also known that for this even the smallest fragment—with only one division and one variable—is sufficient (W. Buszkowski 1985, S. K. 2012). Adding the intersection connective allows to generate all languages generated by context-free grammars enriched with conjunction (S. K. 2013). There are still many open questions: virtually for every non-trivial extension of the Lambek calculus or its commutative variant the exact characterization of the class of languages is not known, and we're planning to answer some of these questions.

Any context-free language can be generated by a Lambek categorial grammar, where every letter is assigned a *single* type (A. Safiullin 2007). The question whether the full Lambek calculus here can be replaced by a weak fragment (for example, only with one division) remains open, and we'll try to give an answer.