Remarks on the history of categorial grammar

J. Lambek, McGill University, Montreal[0)]

A number of mathematicians have taken a professional interest in linguistics: Eratosthenes, Wallis and Grassmann, to name just a few. But a case can be made for claiming that grammar[1)] is a branch of mathematics.

The school of Pythagoras divided mathematics into four disciplines: arithmetic, geometry, astronomy and music. About a thousand years later, Boethius[2)] declared that this so-called "quadrivium" was to constitute the core of undergraduate university studies. It remained so for another thousand years or so, although it was supplemented by a more elementary "trivium", consisting of rhetoric, logic and grammar. Together, these formed the seven "liberal arts".

Two of the more "trivial" subjects were admitted into mathematics in more recent times: logic in the nineteenth century and grammar, reluctantly, in the twentieth. Grammar entered computer science via formal language theory and led to such questions as whether natural languages may be context-free [Savitch 1987]. My own research has been concerned with a different mathematical approach to linguistics, the study of what is known as "categorial grammar".

The word "category" here means "type" and was given this meaning by Aristotle, who was influenced by a word denoting the public prosecutor, derived from *cata* + *agora*, meaning "against" and "forum". Categories as types must be distinguished from the categories in mathematics, introduced by Eilenberg and Mac Lane.[3)]

Categorial grammarians claim that much of the grammar of a natural language can be squeezed into the dictionary, by assigning appropriate types to the words listed there. These types are elements of an algebraic system, or terms of a logical one, and it is claimed that the grammaticality of a string of words can be decided by a calculation on the corresponding string of types. In particular, the types should trigger the correct word order of a sentence.

Categorial grammar was initiated by the Polish logician Ajdukiewicz [1935]. He had been influenced by ideas of Lesniewski and Husserl, and was to influence further developments by Bar-Hillel [1953]. I arrived at similar ideas [1958] while writing a book on ring theory [1966] and debating a suitable notation for sets of homomorphisms between bimodules with my friend and colleague George Findlay. Given bimodules ${}_RA_S$, ${}_SB_T$ and ${}_RC_T$, we wrote

$$A \otimes B = A \otimes_S B,\ C/B = \mathrm{Hom}_R(B, C),\ A\backslash C = \mathrm{Hom}_R(A, C)$$

and noticed that

$$\mathrm{Hom}_{R,S}(A, C/B) \simeq\ \mathrm{Hom}_{R,T}(A \otimes B, C) \simeq\ \mathrm{Hom}_{S,T}(B, A\backslash C)$$

and that a similar notation was suitable for calculating with two-sided ideals in a ring:

$$A \subseteq C/B \Leftrightarrow AB \subseteq C \Leftrightarrow B \subseteq A\backslash C.$$

Having come across the type theory of Alonzo Church [1940], it occurred to me that a similar idea might be applicable to natural languages. Given two basic types n (for names) and s (for sentences), one ought to be able to analyze some sample sentences as follows:

$$\begin{array}{l} \textit{John sleeps} \\ n\ (n\backslash s) \qquad\qquad \to s \end{array}$$

$$\begin{array}{l} \textit{he} \quad \textit{sleeps} \\ (s/(n\backslash s))\ (n\backslash s) \qquad \to s \end{array}$$

$$\begin{array}{l} \textit{John} \quad \textit{likes} \quad \textit{Jane} \\ n\ ((n\backslash s)/n)\ n \to n(n\backslash s) \quad \to s \end{array}$$

$$\begin{array}{l} \textit{John} \quad \textit{likes} \quad \textit{her} \\ n\ (n\backslash(s/n))\ ((s/n)\backslash s) \quad \to (s/n)((s/n)\backslash s) \ \to s \end{array}$$

I proposed a *syntactic calculus*: a logical system with operations $\backslash$, $\cdot$, $/$ satisfying

$$(a \cdot b) \cdot c \leftrightarrow a \cdot (b \cdot c)$$

$$a \to c/b \quad \text{iff} \quad a \cdot b \to c \quad \text{iff} \quad b \to a\backslash c.$$

To this was later added a unity element 1 satisfying

$$a \cdot 1 \leftrightarrow a \leftrightarrow 1 \cdot a.$$

Algebraically, this was of course just a *residuated* semigroup or monoid; but, logically, it was a kind of propositional calculus lacking Gentzen's structural rules of *weakening, contraction* and *interchange*. Algebraically, these take the form

$$a \to 1, \quad a \to a \cdot a, \quad a \cdot b \to b \cdot a.$$

When these rules are in force, we obtain the familiar positive intuitionistic propositional calculus, with

$$a \wedge b = a \cdot b,\ T = 1,\ a \Rightarrow b = a\backslash b = b/a.$$

Having realized the possible application to linguistics, I rushed to the library, which contained just two linguistics journals: Language and Word. In the most recent issue of Language, there was an article by Bar-Hillel [1953], who had developed essentially the same arithmetical notation, although he adopted the symbol \ only in a later publication [1964].

While Bar-Hillel's system admitted the "contractions"

$$(a/b)b \to a,\ b(b\backslash a) \to c,$$

the syntactic calculus also allowed "expansions"

$$a \to (ab)/b,\ a \to b\backslash(ba)$$

and what linguists were to call "type raising"

$$b \to (a/b)\backslash a,\ b \to a/(b\backslash a).$$

I felt thus encouraged to submit a paper to the American Mathematical Monthly. Not surprisingly, Bar-Hillel was the referee. He had two objections:

(1) My typist was blamed for writing "categorical" in place of "categorial".

(2) My system had no obvious decision procedure.

I did not confess that I was to blame for (1), not the typist, but I was able to answer (2) by invoking a Gentzen style sequent calculus, though without his structural rules. As it turned out, his cut-elimination theorem was even easier to prove in the absence of the latter.

A few years later, I was invited to a symposium of the American Mathematical Society [1961] on applications of mathematics to linguistics. Hard pressed to find anything new to say, I discussed a non-associative version of the syntactic calculus. I also pointed out that Bourbaki's introduction of the tensor product, which was intended to replace bilinear mappings by linear ones, resembles one of Gentzen's introduction rules.

These ideas did not catch on with the linguistic community and even met with some antagonism by certain graduate students, although two linguists, Noam Chomsky and Robert Lees, tried to encourage me to continue. Still, I could see the new wave of Chomsky's generative-transformational grammar sweeping everything else aside and turned my attention to other things.

Many years later, there was a revival of interest in the syntactic calculus. Theoretical questions were answered by Buszkowski, Andrejka, Pentus and others. The connection with Montague semantics was pointed out by van Benthem [1991] and exploited in textbooks by Morrill [1994] and Carpenter [1998]. Moortgat and Oehrle introduced modalities into the non-associative syntactic calculus to license associativity and commutativity when needed (see Moortgat [1997]).

The development starting with Ajdukiewicz dealt largely with syntax, but could be extended to morphology, mainly by enlarging the set $\{s, n\}$ of basic types to a larger partially ordered set which made allowance for grammatical features such as person, number, case and tense. A parallel development for semantics was initiated by Haskell Curry [1961] who retained Gentzen's structural rules and allowed terms of the positive intuitionistic propositional calculus to stand for semantic types. At the same time, it was realized that deductions in this intuitionistic calculus could be represented by terms of the lambda calculus or, as Curry preferred, combinators that avoided bound variables.

Combinators continued to be exploited by linguists, e.g. Keenan and Steedman, while lambda-terms played a basic rôle in Montague semantics, e.g. in the texts by Dowty [1979] and Janssen [1983]. In my personal view, combinators or lambda-terms are just ways of describing the morphisms of a cartesian closed category, an elegant abstraction introduced by Bill Lawvere. In fact, I viewed semantics as a functor from a residuated monoidal category to a cartesian closed one, an idea that had also been in the mind of computer scientists such as Benson [1970] and Hotz [1966].

In all the grammars discussed here so far, a basic rôle is played by one or two binary operations, called *division* by algebraists or *implication* by logicians.

A recent revival of interest in the syntactic calculus was set in motion by Girard's linear logic. This differed from the former in two respects: it was classical rather than intuitionistic and it permitted commutativity, that is, Gentzen's interchange rule. Both Abrusci and I were led to investigate non-commutative linear logic, combining features of the syntactic calculus and linear logic.[4)]

Although I was investigating *bilinear logic*, as I preferred to call it, I did not realize that it had interesting applications to linguistics, until I heard a lecture by Claudia Casadio, who proposed such an application. During her lecture, it struck me that the De Morgan dual of the tensor product[5)] was without linguistic significance and that one could identify the tensor product with its dual, thus obtaining what might be called *compact bilinear logic*.[6)]

As an algebraic system, classical bilinear logic may be described as a residuated monoid with a *dualizing* element 0 such that, for all elements a,

$$a^{r\ell} = a = a^{\ell r},$$

where

$$a^r = a\backslash 0, \quad a^\ell = 0/a.$$

The De Morgan dual of the tensor product, here expressed by juxtaposition, can then be defined as

$$a + b = (b^r a^r)^\ell = (b^\ell a^\ell)^r,$$

the last equality being provable. Now, if we postulate

$$a + b = ab, \;\; 0 = 1$$

for the compact case, a simpler description is available: a partially ordered monoid in which every element a has both a *left adjoint* a^ℓ and a *right adjoint* a^r such that

$$a^\ell a \to 1 \to a a^\ell, \;\; a a^r \to 1 \to a^r a.$$

I decided to call such a system a *pregroup*.

Pregroups constitute a simple generalization of partially *ordered groups*, that is, partially ordered monoids which happen to be groups. This generalization seems to have escaped the attention of mathematicians, perhaps because of the paucity of examples of pregroups which are not already partially ordered groups. My favorite such example is the po-monoid of unbounded order-preserving numerical functions $\mathbf{Z} \to \mathbf{Z}$ under composition. For instance, when $f(z) = 2z$ is the doubling function,

$$f^r(z) = [z/2], \;\; f^\ell(z) = [(z+1)/2],$$

$[\rho]$ being the greatest integer less than or equal to ρ. For linguistic applications, one works with the free pregroup generated by the poset of basic types. Here is a simple example:

$$\begin{array}{l} \textit{he kissed her} \\ \pi_3 \; (\pi^r s_2 o^\ell) \; o \;\; \to \;\; [\pi_3 \pi^r] s_2 [o^\ell o] \to s_2 \end{array}$$

where

π_3 =third person singular subject

s_2 = declarative sentence in past tense

o = object

π = subject when person and number are irrelevant

$\pi_3 \to \pi$ [the arrow denoting the partial order].

As I learned from Avarind Joshi, the simple kind of grammar illustrated in the above example had already been proposed by Zellig Harris [1966].[7)] As was pointed out to me by Claudia Casadio, it even had antecedents in the ideas of Charles Sanders Peirce [1897], who would have called π^r and o^ℓ *unsaturated bonds* or *valencies* of the verb *kissed*, in a chemical analogy.

The algebraic system proposed now allows, in addition to single adjoints a^ℓ and a^r of a basic type a, also double adjoints $a^{\ell\ell}$ and a^{rr}. These turn out to be useful in modern languages. To present only one example, first consider the question

$$\begin{array}{l} \textit{would he kiss her ?} \\ (q_2 i^\ell \pi^\ell)\ \pi_3\ (i o^\ell)\ o\ \to\ q_2[i^\ell[\pi^\ell \pi_3] i][o^\ell o] \to q_2 \end{array}$$

where

q_2 = question in the past tense

i = infinitive of intransitive verb,

hence

io^ℓ = infinitive of transitive verb, with a valency o^ℓ to signal that an object is required on the right.

But now we can also analyze

$$\begin{array}{ll} \textit{whom would he kiss ?} & \\ (q o^{\ell\ell} q^\ell)\ (q_2 i^\ell \pi^\ell) \pi_3\ (i o^\ell) & \to\ q[o^{\ell\ell}[q^\ell q_2][i^\ell[\pi^\ell \pi_3] i] o^\ell] \\ & \to q \end{array}$$

where

q = question if the tense is irrelevant

and

$q_2 \to q$.

The double adjoint reflects what Chomskyan linguists used to call a *trace*, here indicated by a dash. Double adjoints are also helpful in analyzing clitic pronouns in Romance languages. I have not yet seen any appearance of triple adjoints, and even double adjoints have so far only been found in modern European languages. Preliminary work on Latin, Arabic and Turkish have only uncovered single adjoints. I am tempted to conjecture that double adjoints were only introduced in the early Renaissance.

At this point in time, there are about a dozen people who have investigated the application of free pregroups to various natural languages: English, French, German, Polish, Turkish, Arabic, Japanese and Latin. There are plans to look also at isolating and polysynthetic languages, such as Mandarin Chinese and Mohawk respectively.

REFERENCES

V.M. Abrusci, Sequent calculus and phase semantics for pure noncommutative classical propositional calculus, J. Symbolic Logic **56**(1991), 1403-1451.

K. Ajdukiewicz, Die syntactische Konnexität, Studia Philosophica **1**(1935), 1-27. English translation: Syntactic connection, in: S. McCall (ed), Polish Logic, Clarendon Press, Oxford 1967.

Y. Bar-Hillel, A quasi-arithmetic notation for syntactic description, Language **29**(1953), 47-58.

Y. Bar-Hillel, Some linguistic obstacles to machine translation, in: Y. Bar-Hillel, Language and Information, Addison-Wesley, Reading (Mass.), 1964, 75-86.

D.B. Benson, Syntax and semantics, a categorical view, Information and control **17**(1970), 145-160.

B. Carpenter, Type-Logical Semantics, The MIT Press, Cambridge (Mass.), 1998.

C. Casadio, Non-commutative linear logic in linguistics, Grammars **4/3**(2001), 1-19.

C. Casadio and J. Lambek, A tale of four grammars, Studia Logica **71**(2002), 315-329.

A. Church, A formulation of the simple theory of types, J. Symbolic Logic **5**(1940), 56-69.

H.B. Curry, Some logical aspects of grammatical structure, in: R. Jacobson (ed.), Structure of language and its Mathematical Aspects, AMS Proceedings Applied Mathematics 12, Providence 1961, 56-58.

D. Dowty, Word Meaning and Montague Grammar, Reidel, Dordrecht, 1979.

Z.S. Harris, A cyclic cancellation-automaton for sentence well-formedness, International Computation Centre Bulletin **5**(1966), 69-94.

G. Hotz, Eindentigkeit und Mehrdeutigkeit formaler Sprachen. Elektronische Informationsverarbeitung und Kybernetik **2**(1966), 235-247.

T.M.V. Janssen, Foundations and Applications of Montague Grammar, Mathematical Centrum, Amsterdam 1983.

M. Kelly, Many variable functorial calculus I, Springer LNM **281**(1972), 66-105.

J. Lambek, The mathematics of sentence structure, American Mathematical Monthly **65**(1958), 134-170.

J. Lambek, On the calculus of syntactic types, in: R. Jacobson (ed.), Structure of Language and its Mathematical Aspects. AMS Proceedings Applied Mathematics **12**, Providence 1961, 166-178.

J. Lambek,Lectures on Rings and Modules, Blaisdell, Waltham (Mass.) 1966, Chelsea, New York 1976.

M. Moortgat, Categorial type logics, in: J. van Benthem and A. ter Meulen, (eds), Handbook of Logic and Language, Elsevier, Amsterdam 1997, 93-177.

G. Morrill, Type Logical Grammar: Categorial Logic of Signs, Kluwer, Dordrecht 1994.

C.S. Peirce, The logic of relatives, The Monist **7**(1897), 161-217.

W.J. Savitch et al. (eds), The formal complexity of natural language, Reidel, Dordrecht 1987.

J. van Benthem, Language in Action, North Holland, Amsterdam 1991.

ENDNOTES

1) Curiously, our word “glamour” is derived from “grammar”, it being assumed in the middle ages that knowledge of Latin grammar provided magical power.

2) Boethius was a big shot at the court of the Gothic king Theodoric in Ravenna, but sadly ended up being executed as an alleged traitor.

3) In a private communication, the former denied that their choice of the word “category” had been influenced by its then current usage as meaning “type” by Polish logicians.

4) My own paper on the subject had first been submitted to the Proceedings of the Kleene conference in Varna, which never materialized, and was scooped by Abrusci’s [1991] article in the J.S.L.

5) Called “par” by Girard.

6) The word “compact” had been introduced by Max Kelly [1972] in an analogous categorical context.

7) Also known as “Zellig the carpenter”.