The 2017 S.-Y. Kuroda Prize

The Association for Mathematics of Language (SIGMOL) awards the third S.-Y. Kuroda Prize for Lasting Contributions to the Mathematics of Language to

Professor James Rogers

of Earlham College Computer Science Department. The prize recognizes Jim for his groundbreaking work on the use of monadic second-order logic to formalize theories of natural language syntax as well as his role in establishing model-theoretic syntax firmly as a central tool in the mathematics of language.

Starting around 1980, so-called constraint-based or principle-based approaches to natural language syntax were gaining support among linguists and computational linguists, where the well-formed linguistic structures are all those that satisfy the constraints spelled out by a linguistic theory, rather than those generated by some sort of rewriting mechanism. By the late 1980s, many of the formal treatments of systems of constraints were couched in formal logic, and by the early 1990s, a number of people advocated the use of more or less standard logical languages to express constraints, abandoning extra-logical mechanisms found in earlier works. In these approaches, linguistic structures to be described are more or less ordinary mathematical structures, on which the vocabulary for expressing linguistic constraints is interpreted in the usual manner of model theory. Among the people who pushed for such approaches, Jim was the first to promote the use of monadic second-order logic (MSO) to describe tree structures commonly used in syntax.

Quantification over arbitrary sets of nodes available in MSO makes it possible to express various principles of syntactic theories directly and faithfully, while the equivalence between the MSO-definable sets of finite trees (of bounded branching factor) and the recognizable tree languages (Thatcher and Wright, 1968; Doner, 1970) serves as a yardstick for locating the generative capacity of a given syntactic theory relative to the context-free/non-context-free divide. Moreover, the question of satisfiability of any finite set of MSO sentences over the class of finite trees is decidable by reduction to the decidability of the MSO theory of the infinite, complete binary-branching tree (S2S) (Rabin, 1969). While earlier work in computer science used MSO languages with multiple successor functions (denoting ith child), Jim, in his 1998 book *A Descriptive Approach to Language-Theoretic Complexity* (based on his 1994 dissertation), used $L^2_{K,P}$, a monadic second-order language equipped with the linear precedence relation separately from the child-of relation, which is particularly well-suited for describing properties of trees with no fixed branching factor. He proved that the above formal properties of the MSO languages with multiple successor functions also hold of $L^2_{K,P}$.

In his book, Jim went on to examine how much of Chomsky's Government and Binding Theory of English syntax can be formalized in $L^2_{K,P}$. His findings were two-fold. First, he showed that the mechanism of free indexation, which is used widely in the GB framework, is not definable in $L^2_{K,P}$, and in fact makes it possible to express a class of constraints for which satisfiability is undecidable. Second, he showed that almost all of GB principles about English syntax can be formulated in such a way as to avoid the use of free indexation and are in fact definable in $L^2_{K,P}$, provided that a constant bound is placed on the number of overlapping chains. Consequently, this portion of the GB theory of English defines a recognizable tree language, and thus only generates a context-free string language.

In a closely related paper published in 1997, '"Grammarless" phrase structure grammar', Jim reformalized GPSG (Gazdar et al., 1985) in $L^2_{K,P}$, demonstrating once again that a model-theoretic reinterpretation can bring new insights to an existing linguistic theory. In his paper published in 1999, 'The descriptive complexity of generalized local sets', Jim proved that an unranked tree language (set of trees with no finite bound on the branching factor) is MSO-definable if and only if it is recognizable in a suitably extended sense. (Along with Neven (1999), this seems to be one of the first published proofs of this equivalence, which Libkin (2004) called "part of folklore".)

Beyond these specific results about MSO and its applications to the mathematics of language, Jim has been extremely influential in the dissemination of descriptive complexity viewpoints in the mathematics of language. Nowadays model-theoretic methods in general and MSO in particular inform not only research in the realm of syntax, but also in phonology. Very recent examples include Graf (2017) (in syntax) and Jardine (2017) (in phonology). It is safe to say that the research program Jim initiated by his 1998 book continues to inspire a new generation of mathematical linguists, and is alive and well today.

July 13, 2017
S.-Y. Kuroda Prize Committee

References

John Doner. 1970. Tree acceptors and some of their applications. *Journal of Computer and System Sciences* **4**: 406–451.

Gerald Gazdar, Ewan Klein, Geoffrey Pullum, and Ivan Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

Thomas Graf. 2017. A computational guide to the dichotomy of features and constraints. *Glossa: A Journal of General Linguistics* **2**(1): article 18.

Adam Jardine. 2017. On the logical complexity of autosegmental representations. In Makoto Kanazawa, Philippe de Groote, and Mehrnoosh Sadrzadeh, editors, *Proceedings of the 15th Meeting on the Mathematics of Language*, pages 22–35. Association for Computational Linguistics.

Leonid Libkin. 2004. *Elements of Finite Model Theory*. Berlin: Springer.

Frank Neven. 1999. *Design and Analysis of Query Languages for Structured Documents: A Formal and Logical Approach*. Doctoral disseration, Limburgs Universitair Centrum.

Michael O. Rabin. 1969. Decidability of second-order theories and automata on infinite trees. *Transactions of the Americal Mathematical Society* **141**: 1–35.

James Rogers. 1994. *Studies in the Logic of Trees with Applications to Grammar Formalisms*. Doctoral dissertation, Department of Computer and Information Sciences, University of Delaware.

James Rogers. 1997. "Grammarless" phrase structure grammar. *Linguistics and Philosophy* **20**: 721–746.

James Rogers. 1998. *A Descriptive Approach to Language-Theoretic Complexity*. Stanford, CA: CSLI Publications.

James Rogers. 1999. The descriptive complexity of generalized local sets. In Hans-Peter Kolb and Uwe Mönnich, editors, *The Mathematics of Syntactic Structure: Trees and their Logics*, pages 21–40. Berlin: Mouton de Gruyter.

J. W. Thatcher and J. B. Wright. 1968. Generalized finite automata theory with an application to a decision problem of second-order logic. *Mathematical Systems Theory* **2**: 57–81.