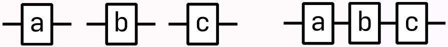
Context-Free Languages of String Diagrams

Matt Earnshaw
Tallinn University of Technology

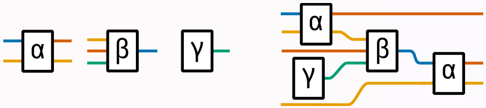
Mario Román
Oxford University

Categories for Automata and Language Theory
Dagstuhl, April 2025

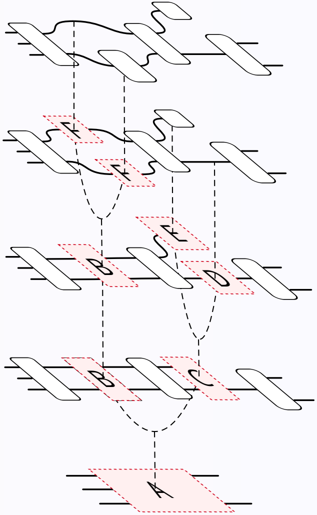
Language of *words* live in free monoids:



This talk is about languages that live in free strict *monoidal categories*:

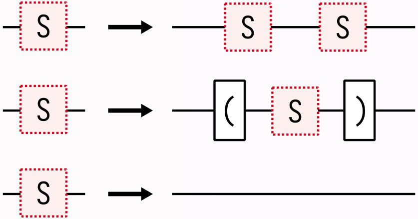


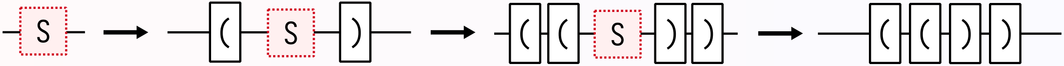
I shall focus on the case of *context-free* languages of string diagrams, which capture context-free languages of words, trees, hypergraphs, etc.



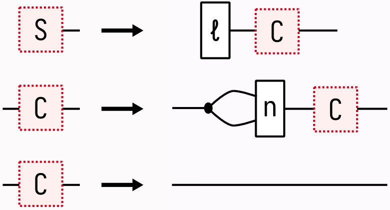
E. and Román, 2025 – *Context-Free Languages of String Diagrams*, FoSSaCS
E., 2025 – *Languages of String Diagrams*

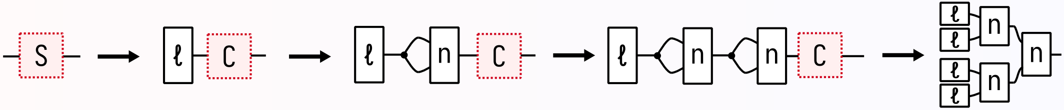
Example: context-free grammars





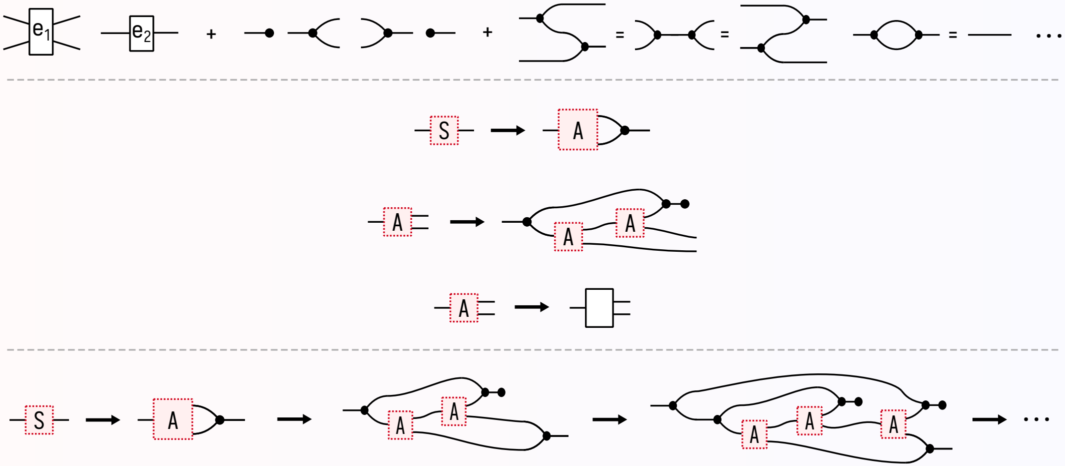
Example: context-free tree grammars





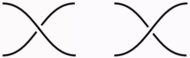
Rounds, 1969

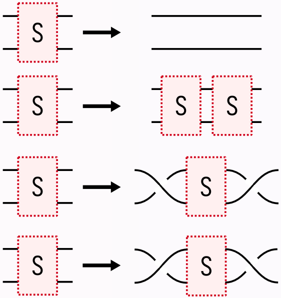
Example: context-free hypergraph (HR) grammars

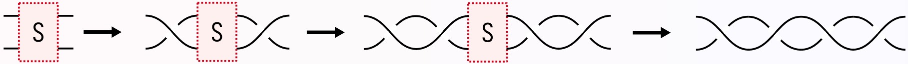


Bauderon and Courcelle, 1987   Habel, 1992

Example: context-free grammar of unbraids





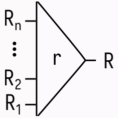


Context-free grammars

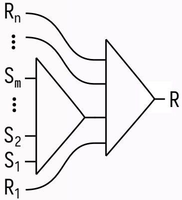
Idea (Melliès and Zeilberger, 2023; Walters, 1989)
A context-free grammar is a *morphism of multigraphs*.

A *multigraph* or *species* comprises sets of sorts S, generators G and functions,

$$S^* \xleftarrow{s} G \xrightarrow{t} S.$$



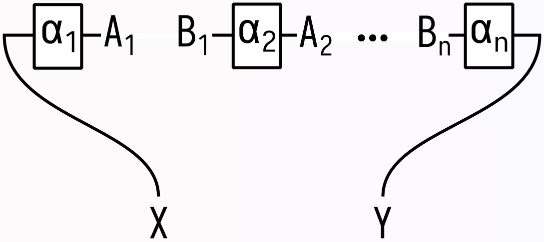
Multigraphs are to *multicategories* (Lambek, 1969) as graphs are to categories.



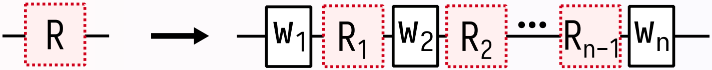
Multicategory of spliced morphisms

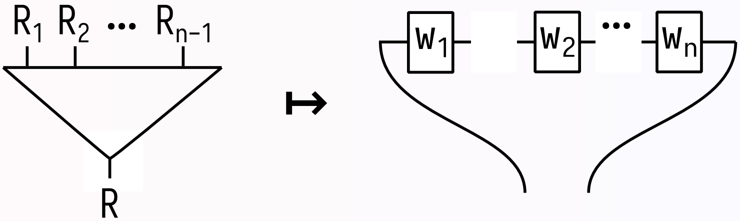
For $\mathbb{C}$ a category, there is a multicategory $\mathcal{W}\mathbb{C}$ with:

- Objects: pairs of objects of $\mathbb{C}$, denoted ${}^{A}_{B}$
- $\mathcal{W}\mathbb{C}(; {}^{X}_{Y}) := \mathbb{C}(X; Y)$,
- $\mathcal{W}\mathbb{C}({}^{A_1}_{B_1}, \dots, {}^{A_n}_{B_n}; {}^{X}_{Y}) := \mathbb{C}(X; A_1) \times \prod_{i=1}^{n-1} \mathbb{C}(B_i; A_{i+1}) \times \mathbb{C}(B_n; Y)$,
- composition, splicing into holes using composition in $\mathbb{C}$



Context-free languages á la Melliès and Zeilberger, 2023; Walters, 1989





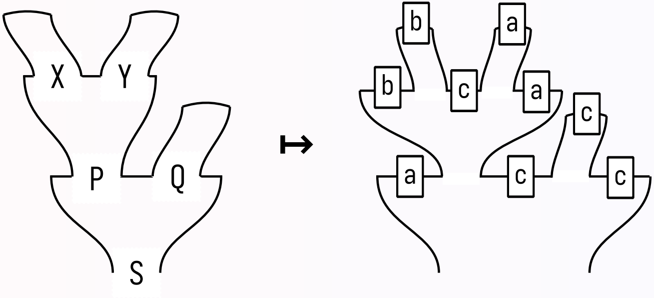
$$M \longrightarrow |W\mathbb{C}|$$

Context-free languages á la Melliès and Zeilberger, 2023; Walters, 1989

$$M \xrightarrow{\Phi} |\mathcal{W}\mathbb{C}|$$

$$FM \xrightarrow{\Phi^*} \mathcal{W}\mathbb{C}$$

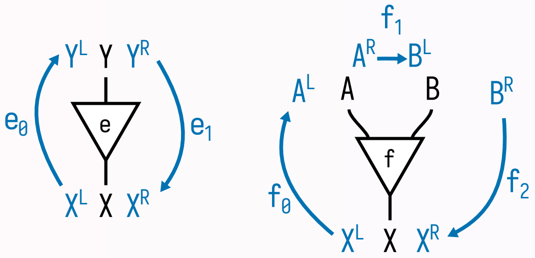
Language of the grammar: $\Phi^*[FM(;S)] \subseteq \mathcal{W}\mathbb{C}(;\tfrac{I}{F}) := \mathbb{C}(I;F)$



Contour ⊣ Splice

Theorem (Melliès and Zeilberger, 2023)

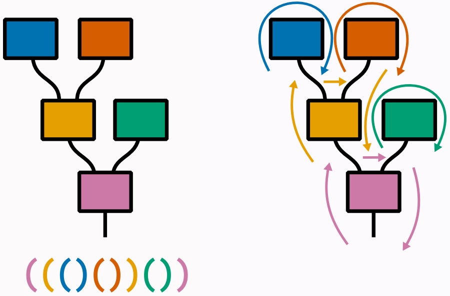
$\mathcal{W}$: Cat → MultiCat *has a left adjoint* $\mathcal{C}$: MultiCat → Cat *given by contours.*



Contours provide a structural notion of *linearization of derivation trees*.

Chomsky-Schützenberger-Melliès-Zeilberger representation theorem

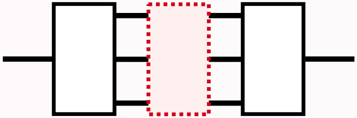
Theorem (Chomsky and Schützenberger, 1963) Every context-free language arises as the image of the intersection of a regular language and a Dyck language.

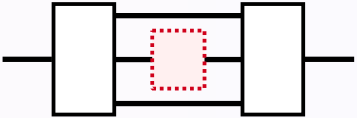


Theorem (Melliès and Zeilberger, 2023) Every context-free language of morphisms in a category arises as the functorial image of the intersection of a regular language in a category and a ($\mathcal{C}$-chromatic) tree contour language

Towards context-free monoidal grammars

What is an appropriate notion of multicategory of contexts in a monoidal category?





Symmetric multicategory of diagram contexts, $\boxed{\mathbb{C}}$

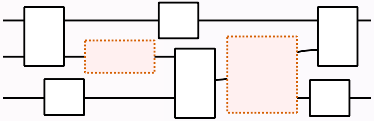
For a monoidal category $\mathbb{C}$ we freely add holes of each type

$$\frac{}{\vdash \mathrm{id} : \genfrac{}{}{0pt}{}{X}{X}} \qquad \frac{}{\vdash f : \genfrac{}{}{0pt}{}{X_1,\dots,X_n}{Y_1,\dots,Y_m}} \qquad \frac{}{\boxed{X} : \genfrac{}{}{0pt}{}{A}{B} \vdash \boxed{X} : \genfrac{}{}{0pt}{}{A}{B}}$$

$$\frac{\Gamma \vdash t_1 : \genfrac{}{}{0pt}{}{A}{B} \qquad \Delta \vdash t_2 : \genfrac{}{}{0pt}{}{B}{C} \qquad \Psi \in \mathrm{Shuf}(\Gamma;\Delta)}{\Psi \vdash t_1 ; t_2 : \genfrac{}{}{0pt}{}{A}{C}}$$

$$\frac{\Gamma \vdash t_1 : \genfrac{}{}{0pt}{}{A_1}{B_1} \qquad \Delta \vdash t_2 : \genfrac{}{}{0pt}{}{A_2}{B_2} \qquad \Psi \in \mathrm{Shuf}(\Gamma;\Delta)}{\Psi \vdash t_1 \otimes t_2 : \genfrac{}{}{0pt}{}{A_1 {+}{+} A_2}{B_1 {+}{+} B_2}}$$

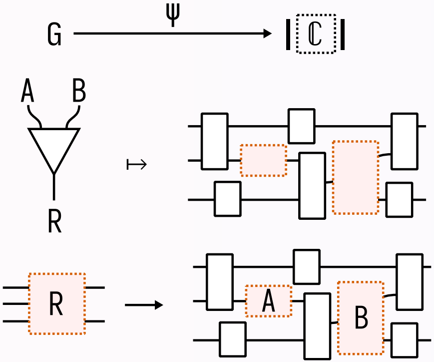
Derivable judgements $\Gamma \vdash f : \genfrac{}{}{0pt}{}{A}{B}$ are *diagram contexts* $f : \Gamma \to \genfrac{}{}{0pt}{}{A}{B}$



$$f : \genfrac{}{}{0pt}{}{1}{1}, \genfrac{}{}{0pt}{}{1}{2} \to \genfrac{}{}{0pt}{}{3}{2}$$

Context-free monoidal grammars

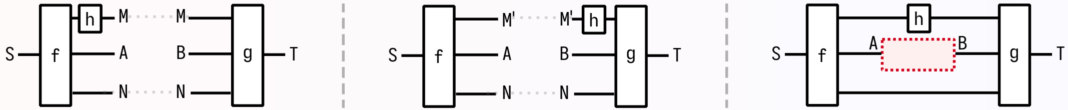
A *context-free monoidal grammar* over a strict monoidal category $(\mathbb{C}, \otimes, I)$ is a morphism of (symmetric) multigraphs Ψ as below, where $\mathcal{G}$ is finite, and a start sort $S \in \mathcal{G}$.



Choosing appropriate monoidal categories, we get the examples shown previously.

Raw contexts, Raw($\mathbb{C}$)

Do we have a left adjoint to forming diagram contexts? Not quite...



$$\mathrm{Raw}(\mathbb{C}) \to \boxed{\mathbb{C}}$$

Proposition

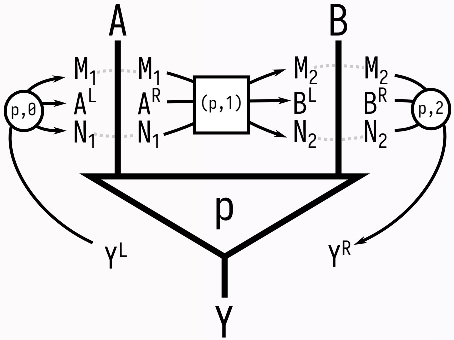
Taking raw contexts in a monoidal category extends to a functor Raw : MonCat $\to$ MultiCat.

Optical contour

Theorem

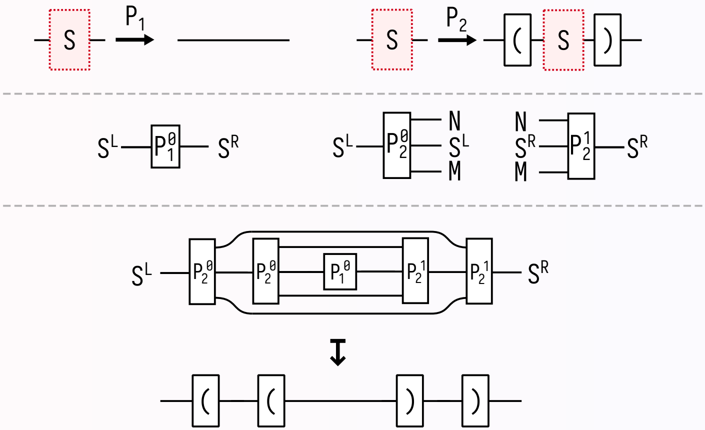
Raw : MonCat $\to$ MultiCat *has a left adjoint, given by the optical contour of a multicategory,*

$$\text{OptC} : \text{MultiCat} \to \text{MonCat}$$



A representation theorem for context-free languages of string diagrams

M. E. and Sobociński, 2022 introduced *regular languages of string diagrams*.



Every context-free language of string diagrams arises as the image of a regular language of string diagrams under a monoidal functor.

References

- Bauderon, Michel and Bruno Courcelle (1987). "Graph expressions and graph rewritings". In: *Mathematical systems theory* 20.1, pp. 83–127. ISSN: 1433-0490. DOI: 10.1007/BF01692060.
- Chomsky, Noam and M.P. Schützenberger (1963). "The Algebraic Theory of Context-Free Languages". In: *Computer Programming and Formal Systems*. Ed. by P. Braffort and D. Hirschberg. Vol. 35. Studies in Logic and the Foundations of Mathematics. Elsevier, pp. 118–161. DOI: 10.1016/S0049-237X(08)72023-8.
- E. (2025). "Languages of String Diagrams". PhD thesis. DOI: 10.23658/TALTECH.2/2025.
- E., M. and Pawel Sobociński (2022). "Regular Monoidal Languages". In: *MFCS 2022*. Vol. 241. LIPIcs, 44:1–44:14. DOI: 10.4230/LIPIcs.MFCS.2022.44.
- E. and Mario Román (2025). *Context-Free Languages of String Diagrams*. arXiv: 2404.10653 [cs.FL].
- Habel, Annegret (1992). *Hyperedge Replacement: Grammars and Languages*. Vol. 643. Lecture Notes in Computer Science. Springer. ISBN: 3-540-56005-X. DOI: 10.1007/BFB0013875.
- Lambek, Joachim (1969). "Deductive systems and categories II. Standard constructions and closed categories". In: *Category Theory, Homology Theory and their Applications I*. Ed. by Peter J. Hilton. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 76–122. ISBN: 978-3-540-36095-7.
- Melliès, Paul-André and Noam Zeilberger (2023). *The categorical contours of the Chomsky-Schützenberger representation theorem*. URL: https://hal.science/hal-04399404.
- Rounds, William C. (1969). "Context-Free Grammars on Trees". In: *ACM Symposium on Theory of Computing*. STOC '69. New York, NY, USA: Association for Computing Machinery, pp. 143–148. DOI: 10.1145/800169.805428.
- Walters, R.F.C. (1989). "A note on context-free languages". In: *Journal of Pure and Applied Algebra* 62.2, pp. 199–203. ISSN: 0022-4049. DOI: 10.1016/0022-4049(89)90151-5.