

# The Structure of Mathematical Expressions

*An ARXIV Case Study*

Deyan Ginev and Bruce R. Miller

National Institute of Standards and Technology

March 21, 2012



---

# Contents

---

<b>Contents</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Motivation . . . . .	3
1.2 Related Resources . . . . .	4
1.3 Experimental Setup . . . . .	4
<b>2 A Study of Mathematical Syntax</b>	<b>7</b>
2.1 Basics . . . . .	7
2.2 Discrete math . . . . .	7
2.3 Continuous math . . . . .	9
2.4 Other fields . . . . .	9
<b>3 Discussion</b>	<b>11</b>
<b>4 Conclusion</b>	<b>13</b>

## *Chapter 1*

---

# **Introduction**

---

In this study, we survey the notational diversity of present-day mathematical expressions, in order to uncover their linguistic phenomena. A practical motivation for this study is to provide a foundation for determining the boundary between syntactic and semantic phenomena in said expressions, from the perspective of language modeling. The ultimate goal of this project is to construct a grammar of mathematical expressions, which captures all relevant syntactic properties established in this study, and allows for the semantic analysis necessary to model and observe the semantic relationships.

## **1.1 Motivation**

We want to enable machine-reading of formulas, in order to provide a variety of user-assistance services, such as semantic search, text-to-speech synthesis, semantic interactions (definition lookup), as well as computer algebra support (“evaluate subexpressions on demand”) and ultimately computer verification (“does that proof step really hold?”).<sup>1</sup>

EdN:1

---

<sup>1</sup>EdNOTE: expand

## 1.2 Related Resources

Notation census, beginnings of study are in Deyan's thesis, Naproche and FMathL have examples, but no real systematic study.<sup>2</sup>

## 1.3 Experimental Setup

The primary corpus on which we base this investigation is the Cornell pre-print archive "ARXIV"<sup>3</sup>, consisting of over 700,000 articles in 37 scientific subfields.

### arXiv Sandbox

<sup>4</sup>

As a secondary resource, we we will also consult entry-level literature on highschool mathematics, in order to exhibit basic phenomena, as well as to demonstrate phenomena apriori known to the authors.<sup>5</sup>

---

<sup>2</sup>EDNOTE: expand

<sup>3</sup>EDNOTE: cite here

<sup>4</sup>EDNOTE: Say that, on the ARXIV front, we first start with the train sandbox from Deyan's thesis

<sup>5</sup>EDNOTE: Wikipedia? PEMDAS?

Train1	Differential Geometry <a href="http://arxmliv.kwarc.info/files/9609/dg-ga.9609012">http://arxmliv.kwarc.info/files/9609/dg-ga.9609012</a>
Train2	Quantum Physics <a href="http://arxmliv.kwarc.info/files/0910/0910.5733/">http://arxmliv.kwarc.info/files/0910/0910.5733/</a>
Train3	High Energy Physics - Theory <a href="http://arxmliv.kwarc.info/files/9407/hep-th.9407125/">http://arxmliv.kwarc.info/files/9407/hep-th.9407125/</a>
Train4	Commutative Algebra <a href="http://arxmliv.kwarc.info/files/0809/0809.4873/">http://arxmliv.kwarc.info/files/0809/0809.4873/</a>
Train5	Statistics Theory <a href="http://arxmliv.kwarc.info/files/0905/0905.1486/">http://arxmliv.kwarc.info/files/0905/0905.1486/</a>
Train6	General Relativity and Quantum Cosmology <a href="http://arxmliv.kwarc.info/files/0807/0807.2507/">http://arxmliv.kwarc.info/files/0807/0807.2507/</a>
Train7	Cosmology and Extragalactic Astrophysics <a href="http://arxmliv.kwarc.info/files/0908/0908.2548">http://arxmliv.kwarc.info/files/0908/0908.2548</a>
Train8	Exactly Solvable and Integrable Systems <a href="http://arxmliv.kwarc.info/files/0905/0905.2033">http://arxmliv.kwarc.info/files/0905/0905.2033</a>
Train9	Geometric Topology <a href="http://arxmliv.kwarc.info/files/0809/0809.4477">http://arxmliv.kwarc.info/files/0809/0809.4477</a>
Train10	Algebraic Geometry <a href="http://arxmliv.kwarc.info/files/0704/0704.0537">http://arxmliv.kwarc.info/files/0704/0704.0537</a>

Table 1.1: Sandbox of Ten Random ARXIV Papers from Diverse Scientific Subfields



## Chapter 2

---

# A Study of Mathematical Syntax

---

## 2.1 Basics

### Foundations

6 7 8

EdN:6  
EdN:7  
EdN:8

### High School

9 10

EdN:9  
EdN:10

## 2.2 Discrete math

### Set Theoretic Notations

11 12

EdN:11  
EdN:12

---

<sup>6</sup>EdNOTE: arithmetic, grouping fences and equality

<sup>7</sup>EdNOTE: basic relations and orderings

<sup>8</sup>EdNOTE: arithmetic and algebraic sequences?

<sup>9</sup>EdNOTE: geometry here, otherwise a separate geometry subsection

<sup>10</sup>EdNOTE: trigonometry, complex and rational numbers

<sup>11</sup>EdNOTE: elementhood, inclusions, set constructors, overloaded arith ops

<sup>12</sup>EdNOTE: also maps : domains -j codomains, xRy notations

## Logical Operators

EdN:13 13

## Combinatorics

EdN:14 14 15  
EdN:15

## Number Theory

EdN:16 16 17 18 19  
EdN:17  
EdN:18  
EdN:19

## Graph Theory

EdN:20 20 21 22  
EdN:21  
EdN:22

## Algebra

EdN:23 23 24 25 26  
EdN:24  
EdN:25  
EdN:26

## Functions Theory

EdN:27 27

---

<sup>13</sup>EDNOTE: classic logic, HOL, type theories

<sup>14</sup>EDNOTE: Infinite sums

<sup>15</sup>EDNOTE: binomials, combinations, permutations,

<sup>16</sup>EDNOTE: modulo modifiers

<sup>17</sup>EDNOTE: tuples

<sup>18</sup>EDNOTE: divisibility notations  $a \mid b$  and  $b/a$

<sup>19</sup>EDNOTE: DLMF sneaky notations

<sup>20</sup>EDNOTE: edge and vertex notations

<sup>21</sup>EDNOTE: incidence and adjacency notations

<sup>22</sup>EDNOTE: Wiki is very nice: [http://en.wikipedia.org/wiki/Glossary\\_of\\_graph\\_theory](http://en.wikipedia.org/wiki/Glossary_of_graph_theory)

<sup>23</sup>EDNOTE: vectors

<sup>24</sup>EDNOTE: maps and complements

<sup>25</sup>EDNOTE: groups

<sup>26</sup>EDNOTE: lattices

<sup>27</sup>EDNOTE: talk about associativity of application and composition, “;” and “o” as notation variants, discuss complex examples



## 2.3 Continuous math

### Calculus

28

EdN:28

### Probability

29 30

EdN:29

EdN:30

### Interval Notation and Arithmetic

31

EdN:31

### Topology

32

EdN:32

### Differential Geometry

Some intro text?

## 2.4 Other fields

### Quantum Physics

33 34 :

EdN:33

EdN:34

---

<sup>28</sup>EDNOTE: differentials, integrals, limits, remember brownian motion integral notations!

<sup>29</sup>EDNOTE: Bayes formula with multiple denotations of P

<sup>30</sup>EDNOTE: Various conditional and joint probability notations

<sup>31</sup>EDNOTE: introduce interval notations, then move to interval arithmetic

<sup>32</sup>EDNOTE: manifold constructors and notations

<sup>33</sup>EDNOTE: Bra-ket notation

<sup>34</sup>EDNOTE: computer science, biology, chemistry...

	Expression	Meaning	Syntax
1.	$(\mathcal{V}/\mathcal{Z}, k\omega)$ <b>Discussion:</b>	symplectic torus ?	circumfix constructor
2.	$\mathcal{Z}$ <b>Discussion:</b>	self-dual lattice ?	atom abbreviation
3.	$(\mathcal{V}, \omega)$ <b>Discussion:</b>	symplectic vector space ?	circumfix constructor
4.	$Lag(\mathcal{V})$ <b>Discussion:</b>	Lagrangian Grassmannian ?	circumfix constructor
5.	$Lag_4(\mathcal{V})$ <b>Discussion:</b>	4-fold covering space ?	complex object, applicative constructor

Table 2.1: Differential Geometry Notations

## *Chapter 3*

---

# **Discussion**

---



## *Chapter 4*

---

# **Conclusion**

---