# Towards Structure-Aware Neural Representations of Agda Programs

Konstantinos Kogkalidis, Orestis Melkonian, Jean-Philippe Bernardy

Theorem Proving and Machine Learning in the age of LLMs April 2025, Edinburgh

Publication : Learning Structure-Aware Representations of Dependent Types @ NeurIPS 2024

ACK : Funds from **EuroProofNet CA2011** (2 STSMs in 2023)

ı

# Mandatory (?) Redundant Intro Slide

#### **Automated Theorem Proving**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis.

#### **Automated Theorem Proving in the Times of ML**

Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque.

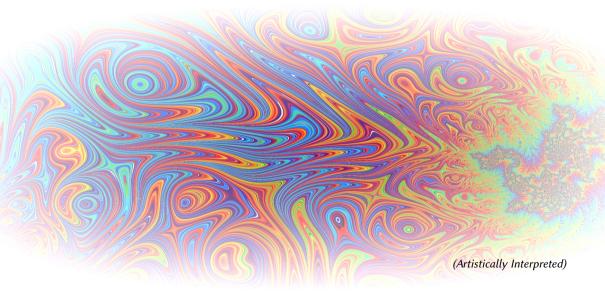
# Proving stuff (in Agda): what you write

open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans) data IN: Set where zero: IN  $suc : \mathbb{N} \to \mathbb{N}$  $+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}$ zero + n = nsuc m + n = suc (m + n)+-comm :  $(m \ n : \mathbb{N}) \to m + n \equiv n + m$ +-comm zero zero = refl +-comm zero (suc n) = cong suc (+-comm zero n) +-comm (suc m) zero = cong suc (+-comm m zero) +-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n)) where +-suc :  $\forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)$ +-suc zero n = refl+-suc (suc m) n = cong suc (+-suc m n)

# ... what Agda shows you

open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans) data N: Set where zero: N  $suc : \mathbb{N} \to \mathbb{N}$  $+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}$ zero + n = nsuc m + n = suc (m + n)+-comm:  $(m n : \mathbb{N}) \to m + n \equiv n + m$ +-comm zero zero = refl +-comm zero (suc n) = cong suc (+-comm zero n)+-comm (suc m) zero = cong suc (+-comm m zero) +-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n)) where +-suc :  $\forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)$ +-suc zero n = refl+-suc (suc m) n = cong suc (+-suc m n)

# ... what Agda really sees



# ... what you show the LLM



Where did all the colors go?



# Doing things "right"





# Doing things right

#### **Contributions:**

- Structured Machine Learning Data for Agda.
- Learning to Represent (the Shapes of) Dependent Types.

```
open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans)
data N: Set where
  zero: N
  suc : \mathbb{N} \to \mathbb{N}
                                                                     1. We go through all definitions.
+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = suc (m + n)
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = refl
+-comm zero (suc n) = cong suc (+-comm zero n)
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
  where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
         +-suc zero n = refl
         +-suc (suc m) n = cong suc (+-suc m n)
```

open import Relation.Binary.PropositionalEquality using (\_≡\_; refl; cong; trans)

1. We go through all definitions.

```
_+_: \mathbb{N} \to \mathbb{N} \to \mathbb{N} 1. We go

zero + n = n

suc m + n = \operatorname{suc}(m + n)

+-comm : (m n : \mathbb{N}) \to m + n \equiv n + m

+-comm zero zero = refl

+-comm (suc m) zero = cong suc (+-comm zero n)

+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))

where +-suc: \mathbb{V} m n \to m + \operatorname{suc} n \equiv \operatorname{suc}(m + n)

+-suc zero n = \operatorname{refl}

+-suc (suc m) n = \operatorname{cong} \operatorname{suc}(+-\operatorname{suc} m n)
```

```
1. We go through all definitions.
 +: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
\operatorname{suc} m + n = \operatorname{suc} (m + n)
  where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
```

```
1. We go through all definitions.
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = refl
+-comm zero (suc n) = cong suc (+-comm zero n)
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
 where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
                       n = refl
         +-suc zero
         +-suc (suc m) n = cong suc (+-suc m n)
```

open import Relation.Binary.PropositionalEquality using ( $\_\equiv\_$ ; refl; cong; trans)

```
zero: \mathbb{N}

suc: \mathbb{N} \to \mathbb{N}

+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
```

```
[-+]: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = \operatorname{suc}(m + n)
```

- +-comm:  $(m n : \mathbb{N})$  →  $m + n \equiv n + m$ +-comm zero zero = refl +-comm zero (suc n) = cong suc (+-comm zero n) +-comm (suc m) zero = cong suc (+-comm m zero) +-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) where +-suc:  $\forall m n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)$ +-suc zero n = refl
- We go through all definitions.
   For each definition, we record:
  - its name

open import Relation.Binary.PropositionalEquality using (\_≡\_; refl; cong; trans)

```
data \mathbb{N}: Set where zero : \mathbb{N} suc : \mathbb{N} \to \mathbb{N}
```

```
-+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = \text{suc}(m + n)
```

- +-comm :  $(m n : \mathbb{N}) \to m + n \equiv n + m$
- +-comm zero zero = rei
- +-comm zero (suc n) = cong suc (+-comm zero n)
- +-comm (suc m) zero = cong suc (+-comm m zero
- +-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n)
  - where +-suc :  $\forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)$ 
    - +-suc zero n = ref
    - +-suc (suc m) n = cong suc (+-suc m n)

- 1. We go through **all definitions**. For each definition, we record:
  - its name
  - its type

open import Relation.Binary.PropositionalEquality using (\_≡\_; refl; cong; trans)

```
zero : \mathbb{N} suc : \mathbb{N} \to \mathbb{N}
```

```
\begin{bmatrix} -+ : \mathbb{N} \to \mathbb{N} \to \mathbb{N} \\ \text{zero} + n = n \\ \text{suc } m + n = \text{suc } (m + n) \end{bmatrix}
```

- +-comm :  $(m \ n : \mathbb{N}) \rightarrow m + n \equiv n + m$
- +-comm zero zero = ref
- +-comm zero (suc n) = cong suc (+-comm zero n)
- +-comm (suc m) zero = cong suc (+-comm m zero
- +-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n)
  - where +-suc :  $\forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)$

```
+-suc zero n = ref
```

+-suc (suc m) n = cong suc (+-suc m n)

- 1. We go through **all definitions**. For each definition, we record:
  - its name
  - its type
  - its term (/proof)

```
open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans)
data N: Set where
  zero: N
  suc : \mathbb{N} \to \mathbb{N}
                                                                    2. We go through all subterms.
+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = suc (m + n)
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = refl
+-comm zero (suc n) = cong suc (+-comm zero n)
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
  where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
         +-suc zero n = refl
         +-suc (suc m) n = cong suc (+-suc m n)
```

```
open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans)
data N: Set where
  zero: N
  suc : \mathbb{N} \to \mathbb{N}
                                                                    2. We go through all subterms.
+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = suc (m + n)
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = ?
+-comm zero (suc n) = cong suc (+-comm zero n)
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
  where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
         +-suc zero n = refl
         +-suc (suc m) n = cong suc (+-suc m n)
```

```
open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans)
data N: Set where
  zero: N
  suc : \mathbb{N} \to \mathbb{N}
                                                                    2. We go through all subterms.
+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = suc (m + n)
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = refl
+-comm zero (suc n) = cong suc (+-comm zero ?)
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
  where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
         +-suc zero n = refl
         +-suc (suc m) n = cong suc (+-suc m n)
```

```
open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans)
data N: Set where
  zero: N
  suc : \mathbb{N} \to \mathbb{N}
                                                                    2. We go through all subterms.
+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = suc (m + n)
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = refl
+-comm zero (suc n) = cong suc (+-comm ? n)
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
  where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
         +-suc zero n = refl
         +-suc (suc m) n = cong suc (+-suc m n)
```

```
open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans)
data N: Set where
  zero: N
  suc : \mathbb{N} \to \mathbb{N}
                                                                     2. We go through all subterms.
+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
zero + n = n
suc m + n = suc (m + n)
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = refl
+-comm zero (suc n) = cong suc ?
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
  where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
         +-suc zero n = refl
         +-suc (suc m) n = cong suc (+-suc m n)
```

```
open import Relation.Binary.PropositionalEquality using ( = ; refl; cong; trans)
data N: Set where
 zero: N
 suc : \mathbb{N} \to \mathbb{N}
                                                                   2. We go through all subterms.
+: \mathbb{N} \to \mathbb{N} \to \mathbb{N}
                                                                       For each subterm, we record:
zero + n = n
                                                                        • its type (/the goal)
suc m + n = suc (m + n)
+-comm: (m n : \mathbb{N}) \to m + n \equiv n + m
+-comm zero zero = refl
+-comm zero (suc n) = cong suc ?
+-comm (suc m) zero = cong suc (+-comm m zero)
+-comm (suc m) (suc n) = cong suc (trans (+-suc m n) (+-comm (suc m) n))
 where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
         +-suc zero n = refl
         +-suc (suc m) n = cong suc (+-suc m n)
goal: n \equiv (n + zero)
```

open import Relation. Binary. Propositional Equality using ( = : refl; cong; trans) data N . Set where 2. We go through all subterms. +  $: \mathbb{N} \to \mathbb{N} \to \mathbb{N}$ For each subterm, we record: • its type (/the goal) • its scope +-comm:  $(m n : \mathbb{N}) \to m + n \equiv n + m$ +-comm zero (suc n) = cong suc ?where +-suc :  $\forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)$ 

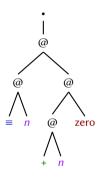
```
2. We go through all subterms.
                                                                       For each subterm, we record:
                                                                         • its type (/the goal)
                                                                         • its scope

    its context

+-comm zero (suc n) = cong suc ?
 where +-suc : \forall m \ n \rightarrow m + \text{suc } n \equiv \text{suc } (m + n)
```

3. Each term & type is recorded as both:a pretty string

goal:  $n \equiv (n + zero)$ 



- 3. Each term & type is recorded as both:
  - a pretty string
  - ullet the underlying AST

goal:  $n \equiv (n + zero)$ 

### Data: TL;DR

#### **Niceties:**

- · among first ML datasets for Agda
- ullet subterm iteration  $\Longrightarrow$  type-checked data augmentation for free
- extraction explicitly preserving type-structure

### Data: TL;DR

#### **Niceties:**

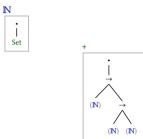
- · among first ML datasets for Agda
- subterm iteration  $\implies$  type-checked data augmentation for free
- · extraction explicitly preserving type-structure

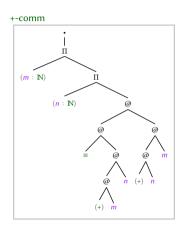
#### Numerical<sup>1</sup>:

- 800 modules
- 11.751 definitions
- 67.255 "holes" read: data points

1: passing extracts from agda-stdlib 1.7.2

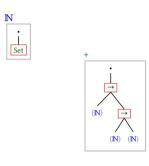
What's to represent?

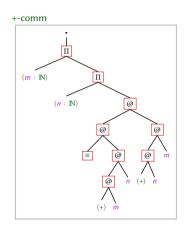




A sequence of ASTs

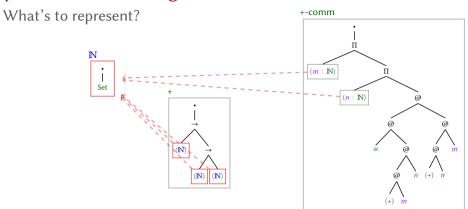
What's to represent?





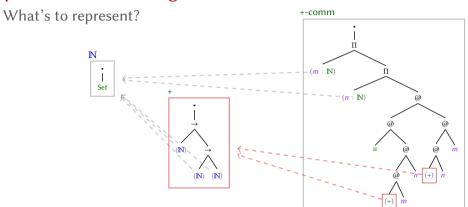
A sequence of ASTs, where nodes are:

• TT primitives



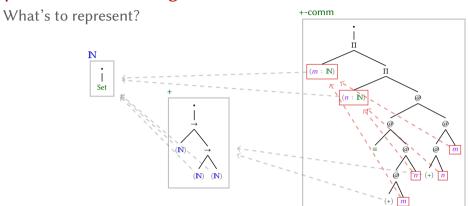
A sequence of ASTs, where nodes are:

- TT primitives
- references to (other) lemmas (inter-AST)



A sequence of ASTs, where nodes are:

- TT primitives
- references to (other) lemmas (inter-AST)



A sequence of ASTs, where nodes are:

- TT primitives
- references to (other) lemmas (inter-AST)
- references to bound variables (intra-AST)

How to represent it?

#### **Candidate Architectures:**

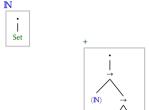
• <del>LLMs</del> -- just no

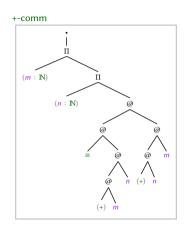
• GNNs -- too generic, oversmoothing

• Tree (R)NNs -- too slow, generally under-performing

• Full Attention [?] -- no structural biases  $--\left(\sum_{t}^{T}n(t)\right)^{2}$  scaling

### Amending self-attention



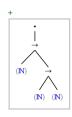


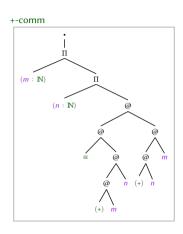
• Treewise Linear (Taylor Series) Attention --  $\#T \times \max_{t}^{T} n(t)$  scaling

(IN) (IN)

### Amending self-attention

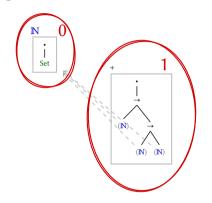


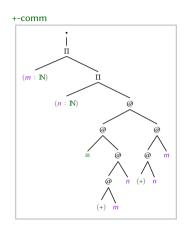




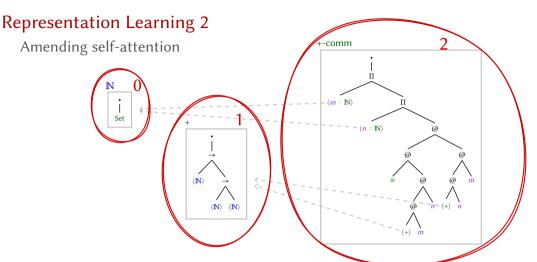
- Treewise Linear (Taylor Series) Attention --  $\#T \times \max_{t=0}^{T} n(t)$  scaling
- Dependency-Level Batching
   -- explicit scope referencing

Amending self-attention





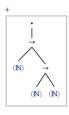
- Treewise Linear (Taylor Series) Attention --  $\#T \times \max_{t=1}^{T} n(t)$  scaling
- Dependency-Level Batching
   -- explicit scope referencing

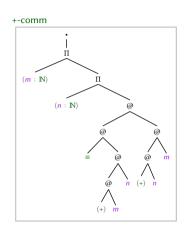


- Treewise Linear (Taylor Series) Attention  $-- \#T \times \max_{t=0}^{T} n(t)$  scaling
- Dependency-Level Batching
   -- explicit scope referencing

Amending self-attention







- Treewise Linear (Taylor Series) Attention --  $\#T \times \max_{t=1}^{T} n(t)$  scaling
- Dependency-Level Batching

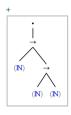
-- explicit scope referencing

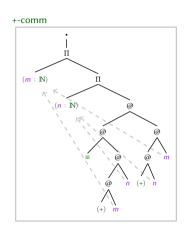
Relative Tree-PE

-- proper inductive biases

### Amending self-attention

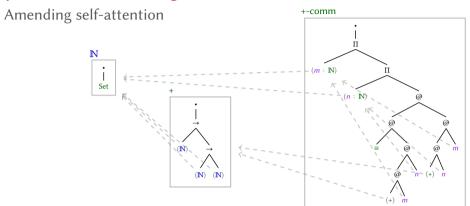






- Treewise Linear (Taylor Series) Attention --  $\#T \times \max_{t}^{T} n(t)$  scaling
- · Dependency-Level Batching
- Relative Tree-PE

- -- explicit scope referencing
- -- proper inductive biases
- -- & dynamic variable indexing



#### Representations informed by type shapes alone:

- invariance to  $\alpha$ -renaming, scope permutations, syntactic distractions, etc.
- ...but a few things get lost in translation

# **Experimental Setup**

#### **Premise Selection**

Contextually rank scope entries by their relevance to the current goal.

### Quill 🦠

Tiny PoC model (6L  $\times$  8H  $\times$  256D; 1 mil. params; 25MB@FP32) trained for ~8h on a V100

#### Data

- train on random holes from 85% of agda-stdlib (ignoring size outliers)
- eval on unseen proofs from:
  - remaining 15% (split between ID and OOD on the basis of size)
  - Unimath & TypeTopology (distant domains)

### ... but does it work?

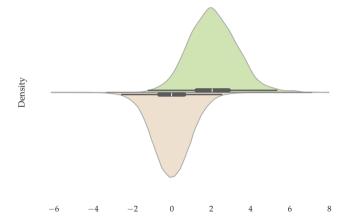
#### Mandatory table with numbers

	Average / R-Precision			
Model	stdlib:ıD	stdlib:ooD	Unimath	ТуреТоро
Quill	<b>50.2</b> / <b>40.3</b>	38.7 / 31.1	<b>27.0</b> / <b>17.4</b>	22.5 / <b>15.4</b>
: Transformer Baseline <sup>1</sup>	10.9 / 3.7	8.5 / 4.5	9.4 / 3.9	5.8 / 0.9

<sup>1:</sup> no Tree-PE, no variable indexing resolution, no Taylor expansion in linear attention

### ... but does it work?

#### Less obscure visualization



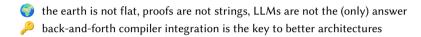
Empirical distribution of selection scores of relevant (green) vs irrelevant (red) lemmas (stdlib:ID).

... but does it work?

#### Findings, TL;DR

- high performance despite limited expressivity & no term exposition
- structure preservation outweights architectural optimizations
- baseline encoder collapses

# Suggested takehome messages



Thank you

- PAPER openreview.net/forum?id=e397soEZh8
   Published manuscript & reviews.
- AGDA2TRAIN github.com/omelkonian/agda2train
   Data extraction as an Agda compilation backend (in Haskell).
- AGDA-QUILL github.com/konstantinoskokos/quill
   ML model; ML-facing Python interface for dataset reading & processing.

open import Data.List.Relation.Binary.Permutation.Propositional.Properties
open PermutationReasoning

```
private variable

--ℓ: Level
A B: Set ℓ
x y: A
xs ys zs ws: List A
xss yss: List (List A)

--concat* = 100

--concat* = 100

--concat* (prep xs p) = ++*1 xs (--concat* p)

--concat* (prep xs p) = ++*1 xs (--concat* p)

--concat* (prep xs p) = ++*1 xs (--concat* p)

--concat* (rone xs p) = ++*1 xs (--concat* p)

--concat* (xss = _ :: _ :: xss|_1 :: _ :: yss|_1 sws px ys p) = begin
xs ++y s ++ concat xss -( *+** ys (++** xs (--concat* p)) )
ys ++x s ++ concat xss -( *+** ys (++** xs (--concat* p)) )
ys ++x s ++ concat xss -( *+** ys (++** xs (--concat* p)) )
--concat* (trans xs-ys ys-zs) = trans (--concat* xs-ys) (--concat* ys-zs)
```

```
4 Data.List.Relation.Binary.Permutation.Propositional.Properties.++*1
  5 Data.List.Relation.Binary.Permutation.Propositional.....trans
  6 Data.List.Relation.Binary.Permutation.Propositional. -- sym
  7 Data.List.Relation.Binary.Permutation.Propositional.Properties.zoom
  8 Data.List.Relation.Binary.Permutation.Propositional.Properties.↔-sym-invo
  lutive
  9 Data.List.Relation.Binary.Permutation.Propositional.Properties.shift
 10 Data List Relation Binary Permutation Propositional . ..
 11 Data List Relation Binary Permutation Propositional Properties ++--*++
 12 Data.List.Base._++_
 13 Data List Base *++
 14 Data List Base reverseAcc
 15 Data.List.Relation.Binary.Permutation.Propositional.Properties.drop-mid
 16 Data.List.Relation.Binary.Permutation.Propositional.Properties.drop-::
 17 Data List Relation Binary Permutation Propositional Properties drop-mid-
 18 Data.List.Base.intercalate
 19 Data.List.Relation.Binary.Permutation.Propositional.Properties.++*
 20 Data.List.Relation.Binary.Permutation.Propositional.Properties.inject
 21 Data List Base man
 22 Agda. Builtin. List. List
 23 Data.List.Base.concatMap
 24 Data List Relation Binary Permutation Propositional PermutationReasoning.
    sten-prep
 25 Data.List.Relation.Binary.Permutation.Propositional._⊸_.prep
 26 Data.List.Relation.Binary.Permutation.Propositional.↔-reflexive
 27 Data List Relation Binary Permutation Propositional Properties --- singleto
    n-inv
 28 Data.List.Base.tails
 29 Data List Base inits
 30 Data List Base concat
 31 Agda.Builtin.List.List._::_
 32 Data List Base an
 33 Data.List.Base.reverse
 34 Data.List.Relation.Binary.Permutation.Propositional.Properties.⇔-reverse
 35 Agda.Builtin.List.List.
 36 Data.List.Relation.Binary.Permutation.Propositional.PermutationReasoning.

    step-swap

 37 Data List Relation Binary Permutation Propositional ... swap
 38 Data List Relation Binary Permutation Propositional Properties --- empty-ine
```

2 Data.List.Relation.Binary.Permutation.Propositional.Properties.shifts
3 Data.List.Relation.Binary.Permutation.Propositional.Properties.+++\*

1 REPL. -- concat\*