

Merging two Hierarchies of External Contextual Grammars with Subregular Selection

Bianca Truthe

Justus-Liebig-Universität Giessen

DCFS, Potsdam, July 4–6, 2023

Introduction

- In the area of formal languages and automata theory, regular languages and finite automata are widely studied.

Introduction

- In the area of formal languages and automata theory, regular languages and finite automata are widely studied.
- Several classes of specific finite automata and their accepted languages have been investigated separately.

Introduction

- In the area of formal languages and automata theory, regular languages and finite automata are widely studied.
- Several classes of specific finite automata and their accepted languages have been investigated separately.
- Many papers have been published by different authors on subregular families of languages.

Introduction

- In the area of formal languages and automata theory, regular languages and finite automata are widely studied.
- Several classes of specific finite automata and their accepted languages have been investigated separately.
- Many papers have been published by different authors on subregular families of languages.
- Focus is often on the decrease of descriptive or computational complexity when going from arbitrary regular languages to special ones.

Introduction

- In the area of formal languages and automata theory, regular languages and finite automata are widely studied.
- Several classes of specific finite automata and their accepted languages have been investigated separately.
- Many papers have been published by different authors on subregular families of languages.
- Focus is often on the decrease of descriptive or computational complexity when going from arbitrary regular languages to special ones.
- Here, the generative capacity of contextual grammars with special selection languages is considered.

Contextual Grammars

Contextual grammars

- introduced by Solomon Marcus in 1969,
- formal model for generating languages,
- starting with an initial finite set of words,
- wrapping a context around a (sub)word

Solomon Marcus: Contextual grammars.
Revue Roum. Math. Pures Appl. 14 (1969), 1525–1534.

Contextual Grammars

Contextual grammars

- introduced by Solomon Marcus in 1969,
- formal model for generating languages,
- starting with an initial finite set of words,
- wrapping a context around a (sub)word

External: a context (u, v) wrapped around a word x yields the word uxv

Solomon Marcus: Contextual grammars.
Revue Roum. Math. Pures Appl. 14 (1969), 1525–1534.

Contextual Grammars

Contextual grammars

- introduced by Solomon Marcus in 1969,
- formal model for generating languages,
- starting with an initial finite set of words,
- wrapping a context around a (sub)word

External: a context (u, v) wrapped around a word x yields the word uxv

Internal: (u, v) added to x gives all words $x_1ux_2vx_3$ where $x_1x_2x_3 = x$

Solomon Marcus: Contextual grammars.
Revue Roum. Math. Pures Appl. 14 (1969), 1525–1534.

Contextual Grammars

Contextual grammars

- introduced by Solomon Marcus in 1969,
- formal model for generating languages,
- starting with an initial finite set of words,
- wrapping a context around a (sub)word

External: a context (u, v) wrapped around a word x yields the word uxv

Internal: (u, v) added to x gives all words $x_1ux_2vx_3$ where $x_1x_2x_3 = x$

Contexts equipped with conditions where they can be applied

Solomon Marcus: Contextual grammars.
Revue Roum. Math. Pures Appl. 14 (1969), 1525–1534.

Definitions

Contextual grammar with selection in \mathcal{F} is a construct

$$G = (V, \{(S_1, C_1), (S_2, C_2), \dots, (S_n, C_n)\}, A)$$

where

- V is an alphabet,
- for $1 \leq i \leq n$, S_i is a language over some $U \subseteq V$ in \mathcal{F} (selection language),
- for $1 \leq i \leq n$, C_i is a finite set of pairs (u, v) (contexts) with $u, v \in V^*$,
- A is a finite subset of V^* (the set of axioms).

Definitions

Contextual grammar with selection in \mathcal{F} is a construct

$$G = (V, \{(S_1, C_1), (S_2, C_2), \dots, (S_n, C_n)\}, A)$$

where

- V is an alphabet,
- for $1 \leq i \leq n$, S_i is a language over some $U \subseteq V$ in \mathcal{F} (selection language),
- for $1 \leq i \leq n$, C_i is a finite set of pairs (u, v) (contexts) with $u, v \in V^*$,
- A is a finite subset of V^* (the set of axioms).

External derivation step: $x \xRightarrow{\text{ex}} y$ if $x \in S_i$ for some $i \in \{1, \dots, n\}$ and $y = uxv$ for some $(u, v) \in C_i$.

Definitions

Contextual grammar with selection in \mathcal{F} is a construct

$$G = (V, \{(S_1, C_1), (S_2, C_2), \dots, (S_n, C_n)\}, A)$$

where

- V is an alphabet,
- for $1 \leq i \leq n$, S_i is a language over some $U \subseteq V$ in \mathcal{F} (selection language),
- for $1 \leq i \leq n$, C_i is a finite set of pairs (u, v) (contexts) with $u, v \in V^*$,
- A is a finite subset of V^* (the set of axioms).

External derivation step: $x \xRightarrow{\text{ex}} y$ if $x \in S_i$ for some $i \in \{1, \dots, n\}$ and $y = uxv$ for some $(u, v) \in C_i$.

Internal derivation step: $x \xRightarrow{\text{in}} y$ if $x = x_1x_2x_3$ with $x_2 \in S_i$ for some $i \in \{1, \dots, n\}$ and $y = x_1ux_2vx_3$ for some $(u, v) \in C_i$.

Definitions

Languages generated:

$$L_{\text{ex}}(G) = \{z \mid x \xrightarrow[\text{ex}]{*} z \text{ for some } x \in A\},$$

$$L_{\text{in}}(G) = \{z \mid x \xrightarrow[\text{in}]{*} z \text{ for some } x \in A\}$$

Definitions

Languages generated:

$$L_{\text{ex}}(G) = \{z \mid x \xrightarrow[\text{ex}]{*} z \text{ for some } x \in A\},$$

$$L_{\text{in}}(G) = \{z \mid x \xrightarrow[\text{in}]{*} z \text{ for some } x \in A\}$$

$\mathcal{EC}(\mathcal{F})$ and $\mathcal{IC}(\mathcal{F})$: family of all languages generated externally or internally by contextual grammars with selection in \mathcal{F} .

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *monoidal* if and only if $L = V^*$,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *monoidal* if and only if $L = V^*$,
- *nilpotent* if and only if it is finite or its complement $V^* \setminus L$ is finite,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *monoidal* if and only if $L = V^*$,
- *nilpotent* if and only if it is finite or its complement $V^* \setminus L$ is finite,
- *combinational* if and only if it has the form $L = V^*A$ for some subset $A \subseteq V$,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *monoidal* if and only if $L = V^*$,
- *nilpotent* if and only if it is finite or its complement $V^* \setminus L$ is finite,
- *combinational* if and only if it has the form $L = V^*A$ for some subset $A \subseteq V$,
- *definite* if and only if it can be represented in the form $L = A \cup V^*B$ where A and B are finite subsets of V^* ,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *monoidal* if and only if $L = V^*$,
- *nilpotent* if and only if it is finite or its complement $V^* \setminus L$ is finite,
- *combinational* if and only if it has the form $L = V^*A$ for some subset $A \subseteq V$,
- *definite* if and only if it can be represented in the form $L = A \cup V^*B$ where A and B are finite subsets of V^* ,
- *suffix-closed* (or *fully initial* or *multiple-entry* language) if and only if, for any words $x \in V^*$ and $y \in V^*$, the relation $xy \in L$ implies $y \in L$,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *ordered* if and only if the language is accepted by some deterministic finite automaton $\mathcal{A} = (V, Z, z_0, F, \delta)$ where (Z, \preceq) is a totally ordered set and, for any $a \in V$, the relation $z \preceq z'$ implies $\delta(z, a) \preceq \delta(z', a)$,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *ordered* if and only if the language is accepted by some deterministic finite automaton $\mathcal{A} = (V, Z, z_0, F, \delta)$ where (Z, \preceq) is a totally ordered set and, for any $a \in V$, the relation $z \preceq z'$ implies $\delta(z, a) \preceq \delta(z', a)$,
- *commutative* if and only if it contains with each word also all permutations of this word,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *ordered* if and only if the language is accepted by some deterministic finite automaton $\mathcal{A} = (V, Z, z_0, F, \delta)$ where (Z, \preceq) is a totally ordered set and, for any $a \in V$, the relation $z \preceq z'$ implies $\delta(z, a) \preceq \delta(z', a)$,
- *commutative* if and only if it contains with each word also all permutations of this word,
- *circular* if and only if it contains with each word also all circular shifts of this word,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *ordered* if and only if the language is accepted by some deterministic finite automaton $\mathcal{A} = (V, Z, z_0, F, \delta)$ where (Z, \preceq) is a totally ordered set and, for any $a \in V$, the relation $z \preceq z'$ implies $\delta(z, a) \preceq \delta(z', a)$,
- *commutative* if and only if it contains with each word also all permutations of this word,
- *circular* if and only if it contains with each word also all circular shifts of this word,
- *non-counting* (or *star-free*) if and only if there is a natural number $k \geq 1$ such that, for any words $x \in V^*$, $y \in V^*$, and $z \in V^*$, it holds $xy^kz \in L$ if and only if $xy^{k+1}z \in L$,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *power-separating* if and only if, there is a natural number $m \geq 1$ such that for any $x \in V^*$, either $J_x^m \cap L = \emptyset$ or $J_x^m \subseteq L$ where $J_x^m = \{x^n \mid n \geq m\}$,

Subregular Language Families

Let L be a language over an alphabet V .

We say that the language L – with respect to the alphabet V – is

- *power-separating* if and only if, there is a natural number $m \geq 1$ such that for any $x \in V^*$, either $J_x^m \cap L = \emptyset$ or $J_x^m \subseteq L$ where $J_x^m = \{x^n \mid n \geq m\}$,
- *union-free* if and only if L can be described by a regular expression which is only built by product and star.

Further Subregular Language Families

$$REG_n^Z = \{ L \mid L \in REG \text{ with } State(L) \leq n \},$$

where

$$State(L) = \min \{ State(A) \mid A \text{ is a det. finite automaton accepting } L \},$$

with

$$State(A) = |Z|$$

Further Subregular Language Families

$$REG_n^Z = \{ L \mid L \in REG \text{ with } State(L) \leq n \},$$

$$RL_n^V = \{ L \mid L \in REG \text{ with } Var_{RL}(L) \leq n \},$$

where

$$State(L) = \min \{ State(A) \mid A \text{ is a det. finite automaton accepting } L \},$$

$$Var_{RL}(L) = \min \{ Var(G) \mid G \text{ is a right-linear grammar generating } L \},$$

with

$$State(A) = |Z|, Var(G) = |N|$$

Further Subregular Language Families

$$REG_n^Z = \{ L \mid L \in REG \text{ with } State(L) \leq n \},$$

$$RL_n^V = \{ L \mid L \in REG \text{ with } Var_{RL}(L) \leq n \},$$

$$RL_n^P = \{ L \mid L \in REG \text{ with } Prod_{RL}(L) \leq n \},$$

where

$$State(L) = \min \{ State(A) \mid A \text{ is a det. finite automaton accepting } L \},$$

$$Var_{RL}(L) = \min \{ Var(G) \mid G \text{ is a right-linear grammar generating } L \},$$

$$Prod_{RL}(L) = \min \{ Prod(G) \mid G \text{ is a right-linear grammar generating } L \}.$$

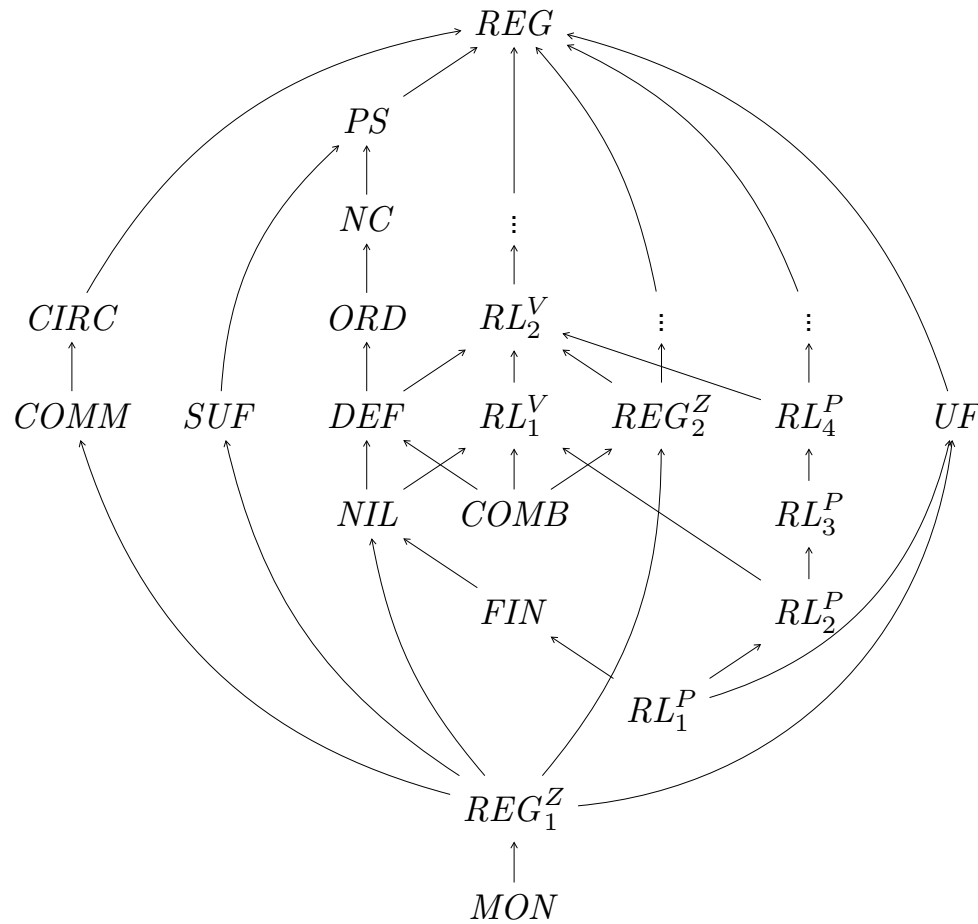
with

$$State(A) = |Z|, Var(G) = |N|, Prod(G) = |P|.$$

Families Under Consideration

$$\begin{aligned} \mathcal{F} \in & \{FIN, MON, NIL, COMB, DEF, SUF, ORD\} \\ & \cup \{COMM, CIRC, NC, PS, UF, REG\} \\ & \cup \{REG_n^Z \mid n \geq 1\} \cup \{RL_n^V \mid n \geq 1\} \cup \{RL_n^P \mid n \geq 1\} \end{aligned}$$

Hierarchy of Subregular Language Families



EC Grammars with Subregular Selection – Previous Work

Jürgen Dassow: *FI* 2005.

$$\mathcal{F} \in \{FIN, MON, COMB, NIL, DEF, COMM, SUF\}$$

(finite, monoidal, combinational, nilpotent, definite, commutative, suffix-closed)

Jürgen Dassow, Florin Manea, Bianca Truthe: *TCS* 2012 (DCFS 2011)

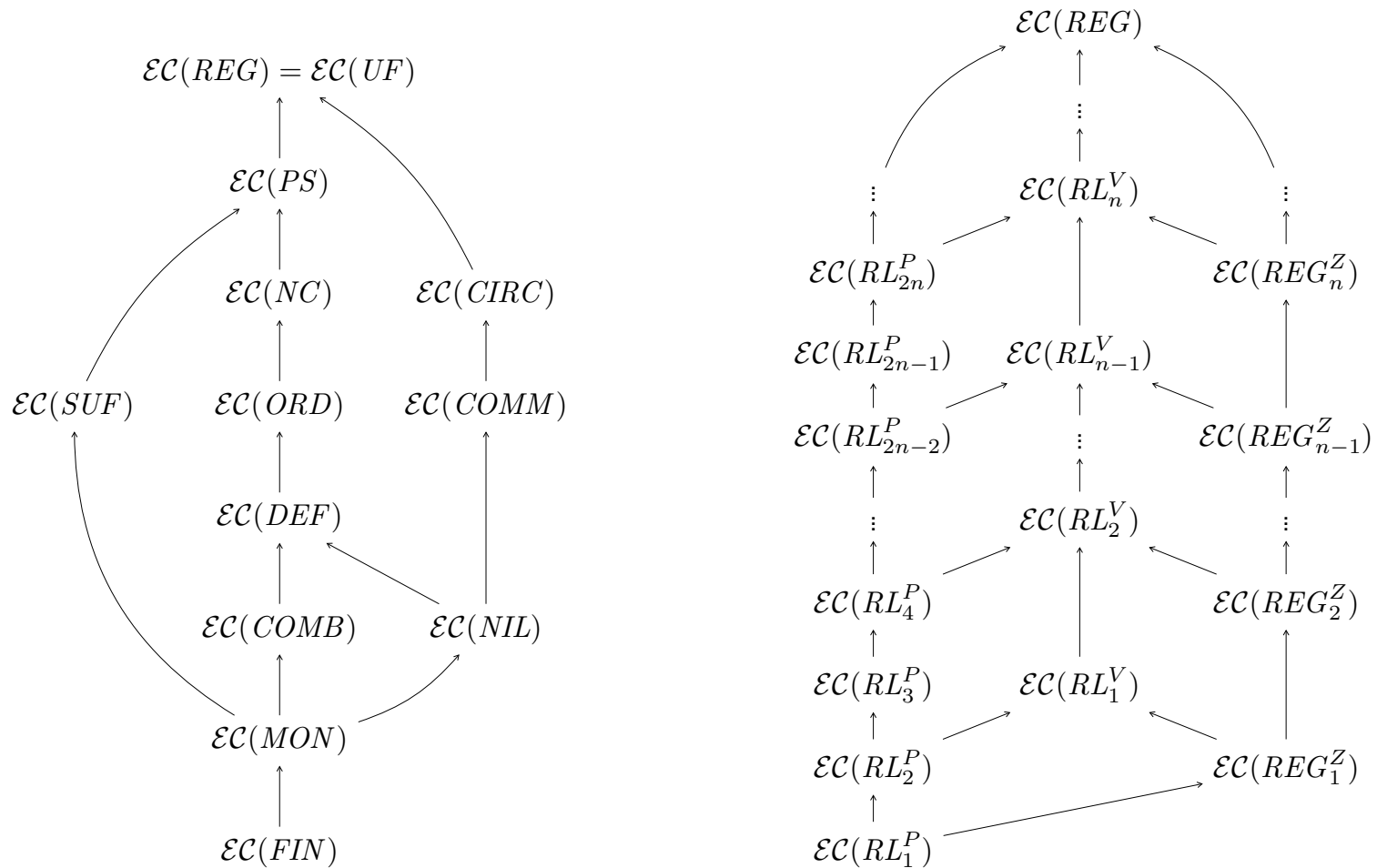
$$\mathcal{F} \in \{CIRC, ORD, UF\} \cup \{REG_n^Z, RL_n^V, RL_n^P \mid n \geq 1\}$$

(circular, ordered, union-free)

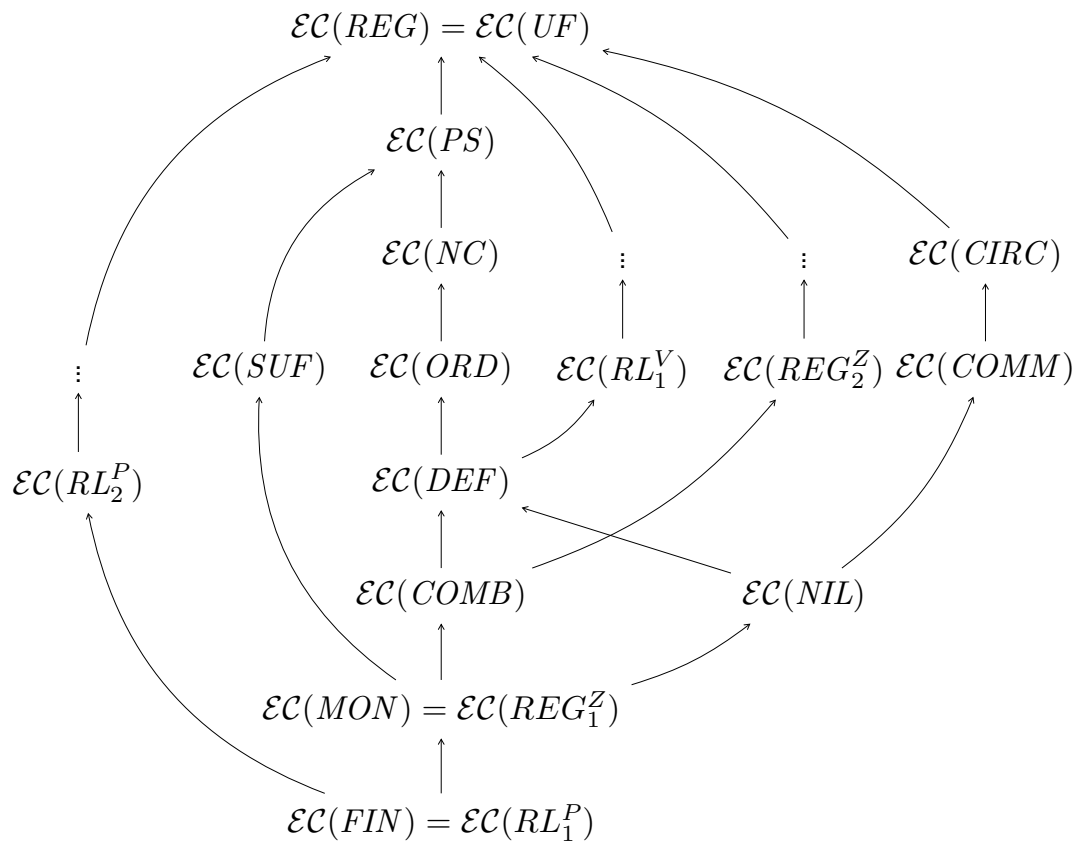
Bianca Truthe: *FI* 2021 (NCMA 2017)

$$\mathcal{F} \in \{REG_n^Z, RL_n^V, RL_n^P \mid n \geq 1\}, \text{ Survey}$$

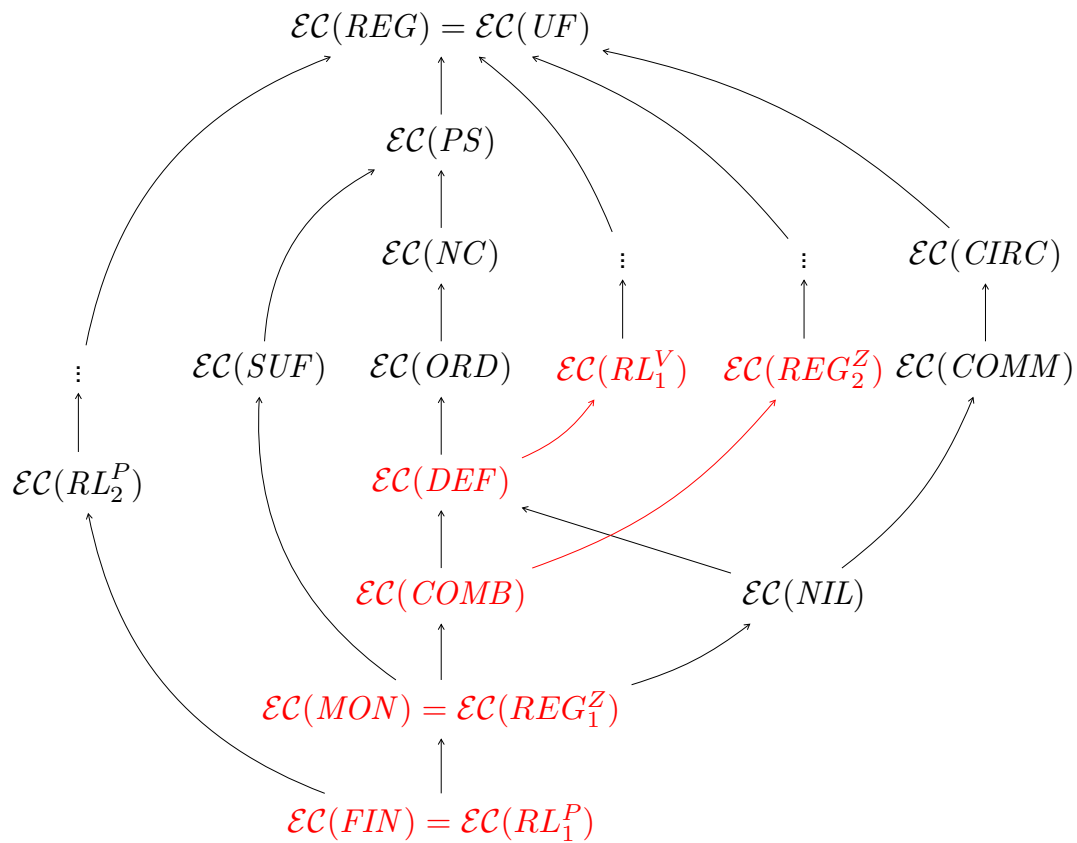
Previous Work – Two Hierarchies



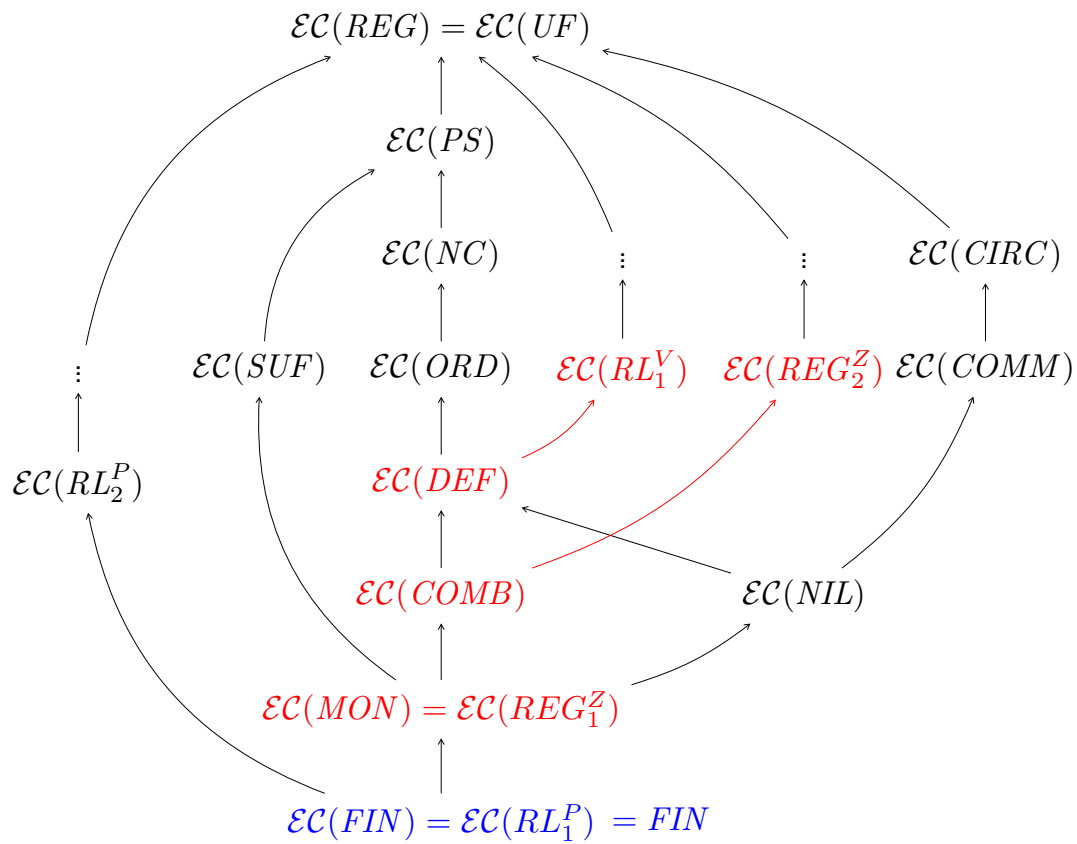
Results - One Hierarchy



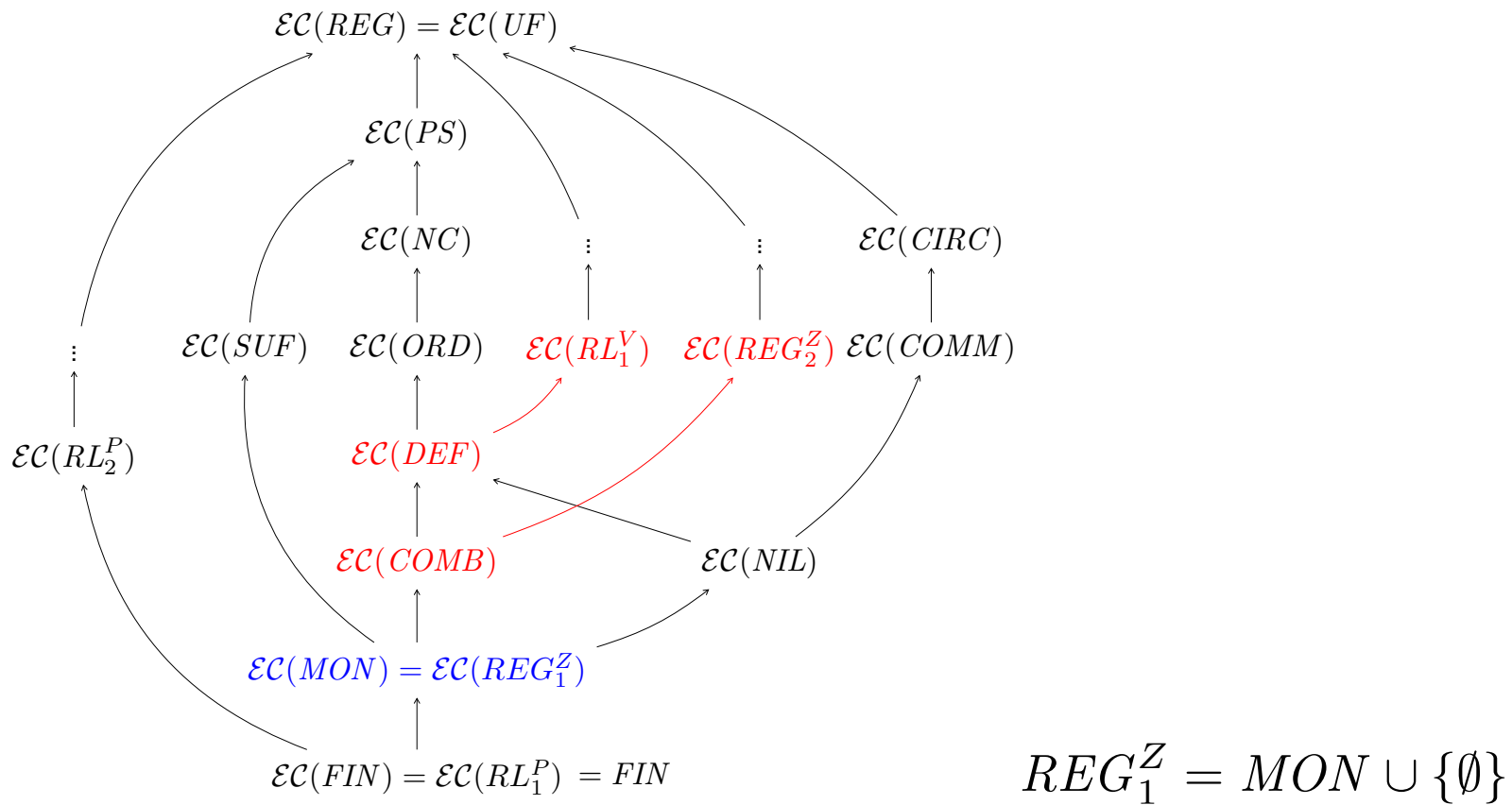
Results - One Hierarchy



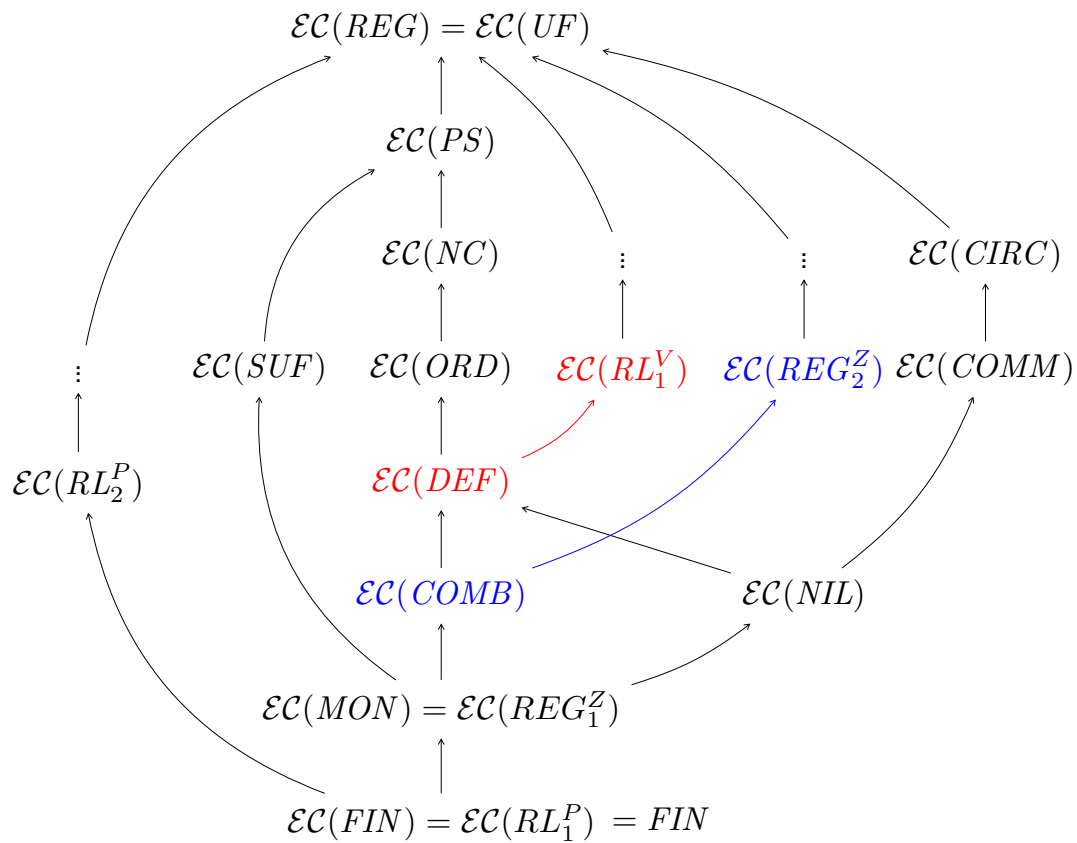
Results - One Hierarchy



Results - One Hierarchy

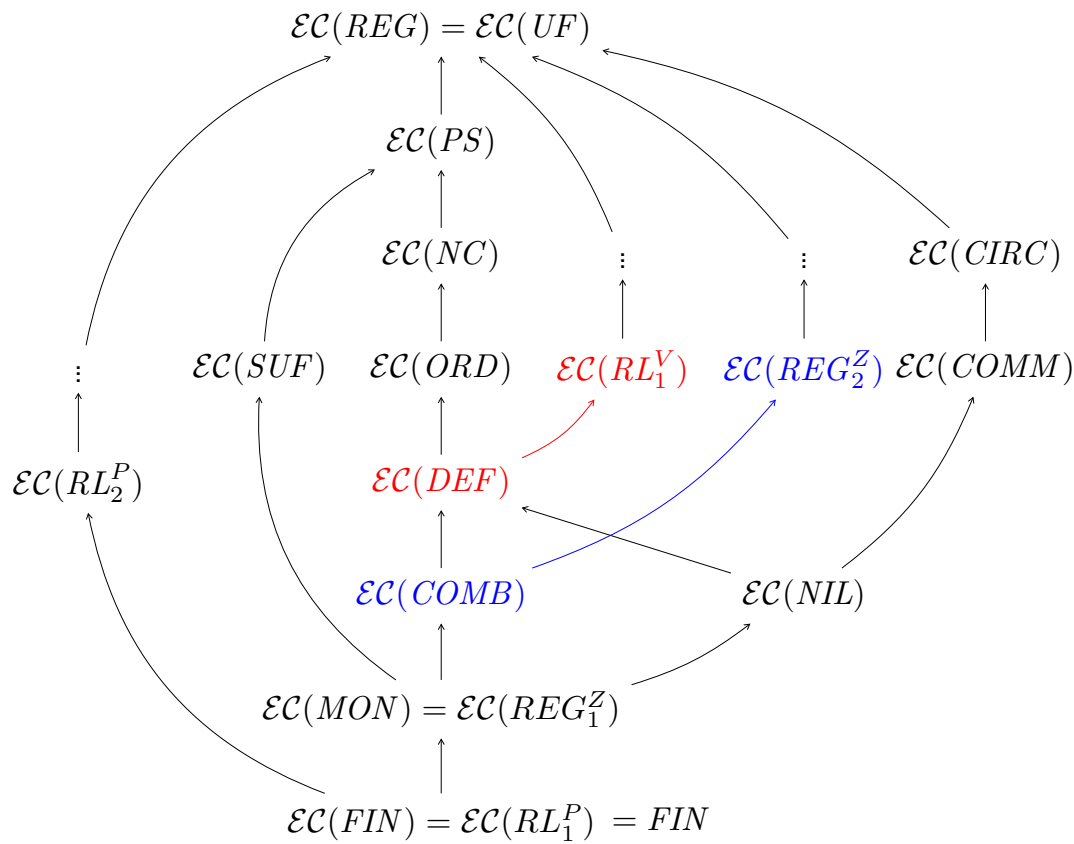


Results - One Hierarchy

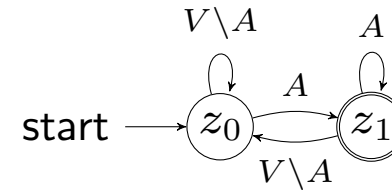


Comb. Lang.: V^*A with $A \subseteq V$

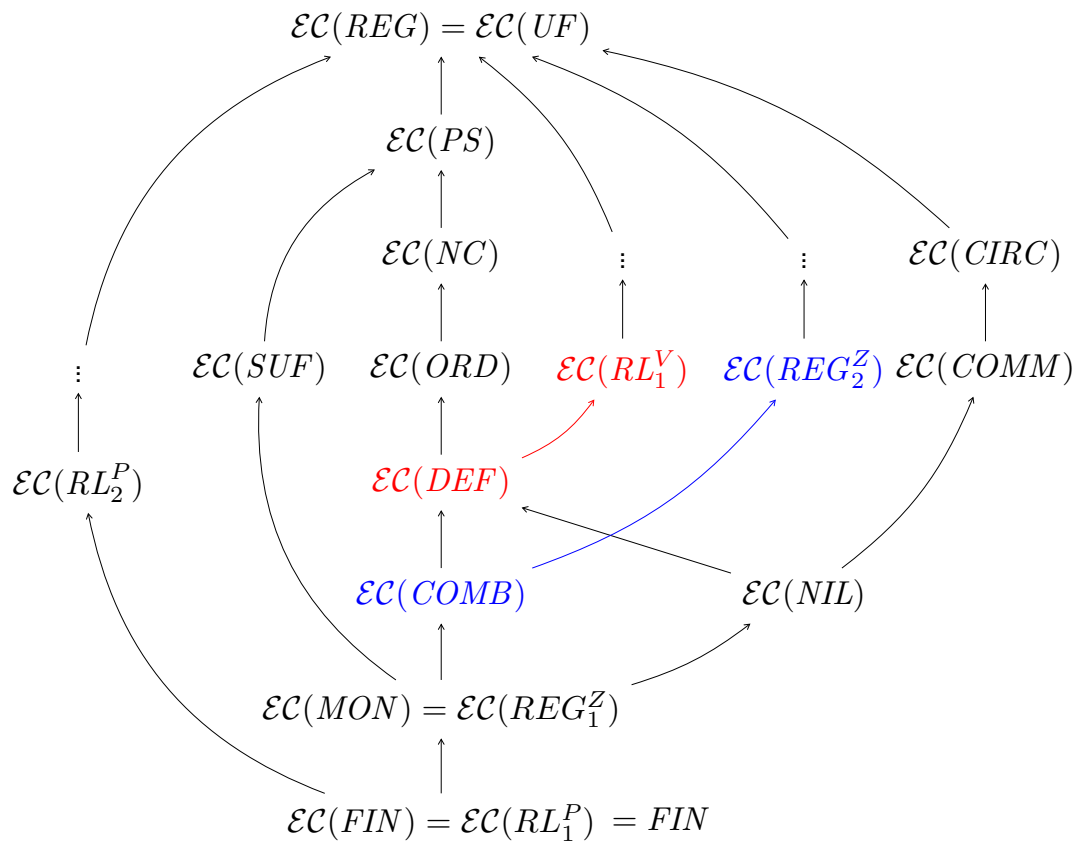
Results - One Hierarchy



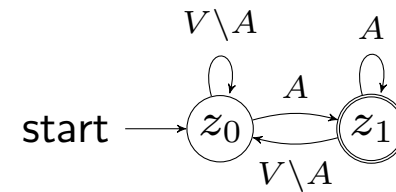
Comb. Lang.: V^*A with $A \subseteq V$
 $COMB \subset REG_2^Z$:



Results - One Hierarchy

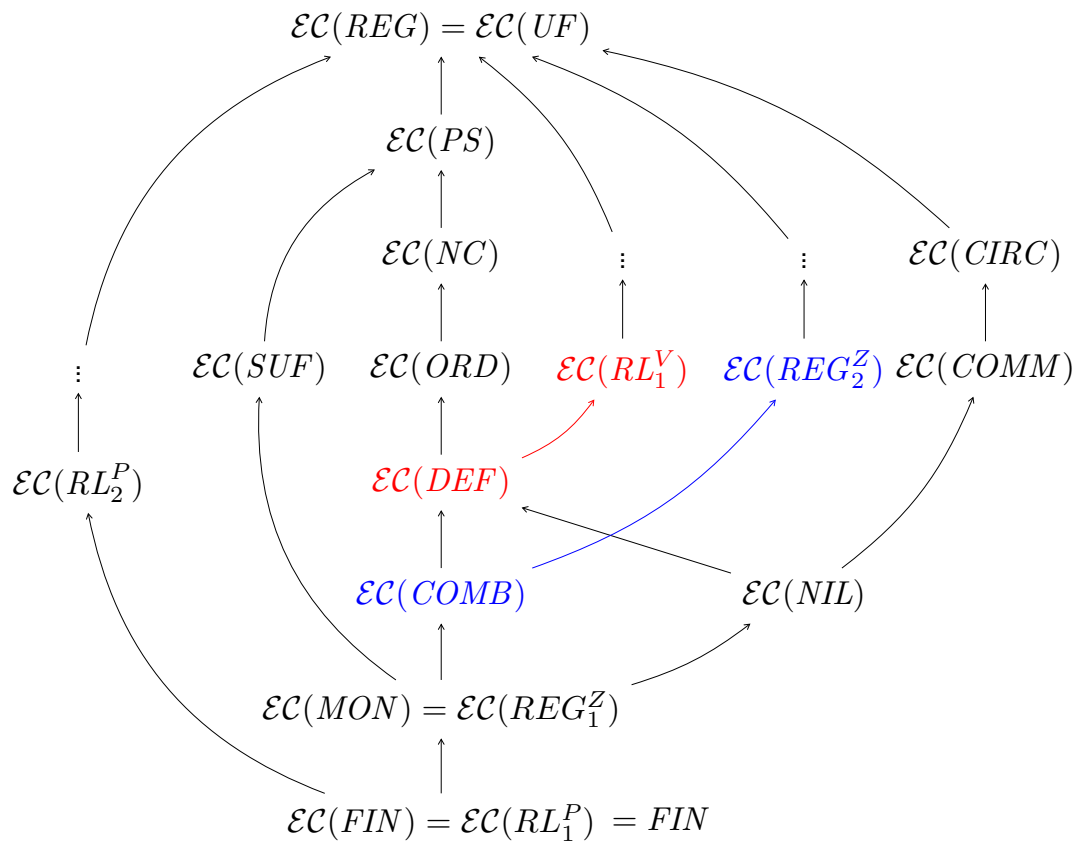


Comb. Lang.: V^*A with $A \subseteq V$
 $COMB \subset REG_2^Z$:

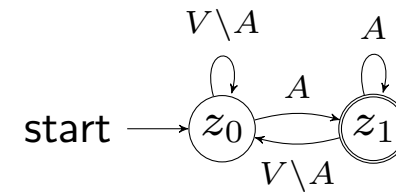


$\rightsquigarrow \mathcal{EC}(COMB) \subseteq \mathcal{EC}(REG_2^Z)$

Results - One Hierarchy



Comb. Lang.: V^*A with $A \subseteq V$
 $COMB \subset REG_2^Z$:

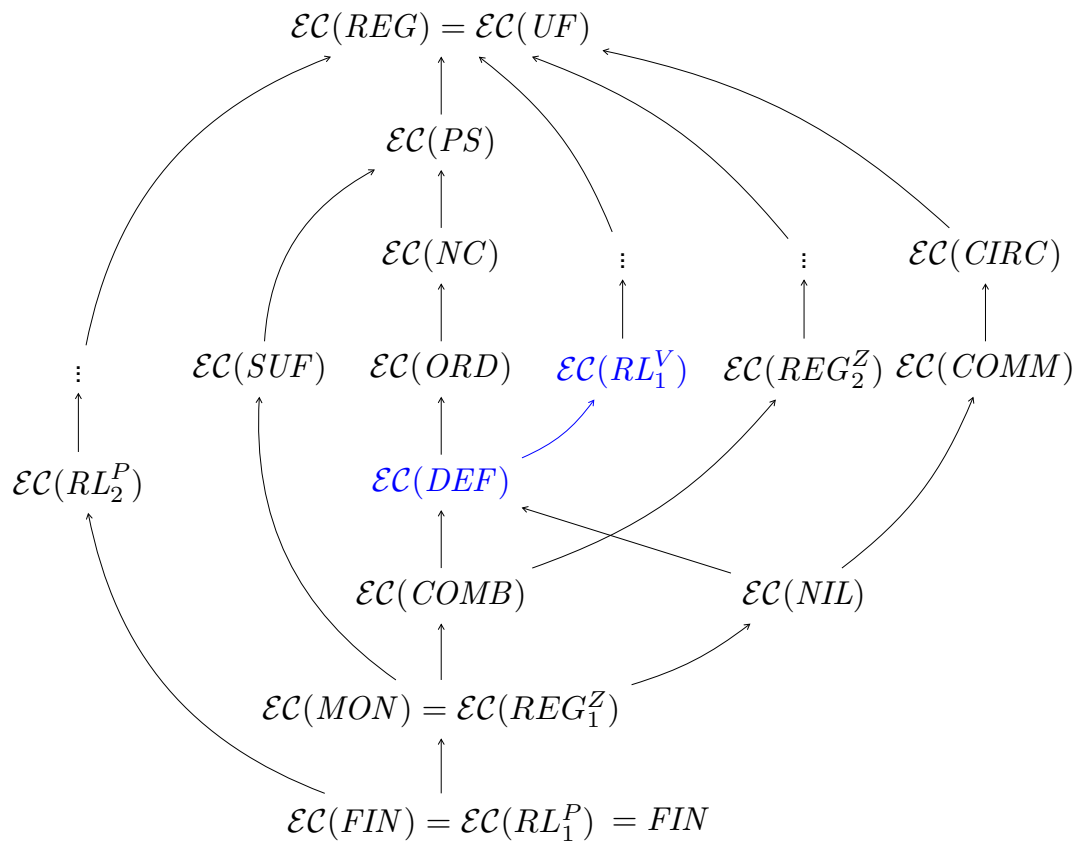


$\leadsto \mathcal{EC}(COMB) \subseteq \mathcal{EC}(REG_2^Z)$

$L = \{a\}^+ \cup \{a\}^* \{b\} \{a, b\}^*$
 $\cup \{c\} \{a\}^* \{b\} \{a, b\}^* \{c\}$

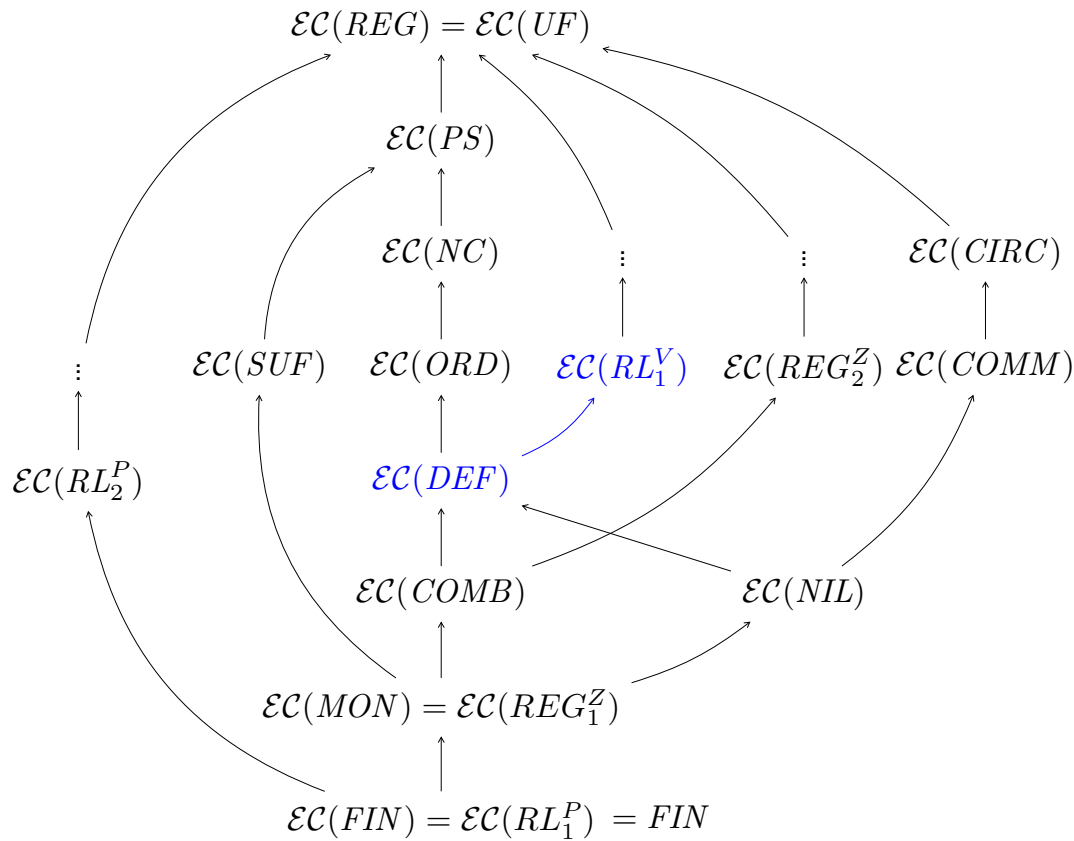
$\in \mathcal{EC}(REG_2^Z) \setminus \mathcal{EC}(COMB)$

Results - One Hierarchy



Def. Lang.: $A \cup V^*B$
 with finite $A, B \subseteq V^*$
 $A \in RL_1^V: S \rightarrow a$ for $a \in A$
 $V^*B \in RL_1^V:$
 $S \rightarrow xS$ for $x \in V$
 $S \rightarrow b$ for $b \in B$

Results - One Hierarchy



Def. Lang.: $A \cup V^*B$
 with finite $A, B \subseteq V^*$
 $A \in RL_1^V$: $S \rightarrow a$ for $a \in A$
 $V^*B \in RL_1^V$:
 $S \rightarrow xS$ for $x \in V$
 $S \rightarrow b$ for $b \in B$

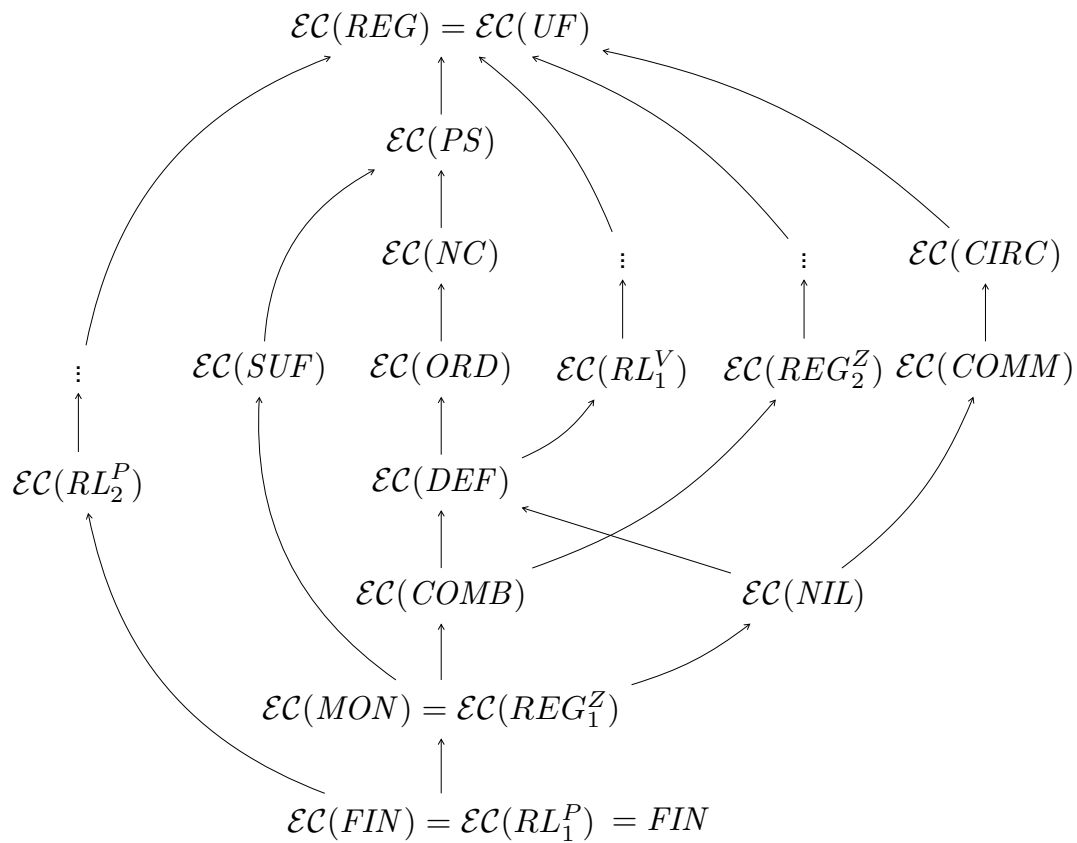
$\rightsquigarrow \mathcal{EC}(DEF) \subseteq \mathcal{EC}(RL_1^V)$

$L = \{ a^n \mid n \geq 1 \}$

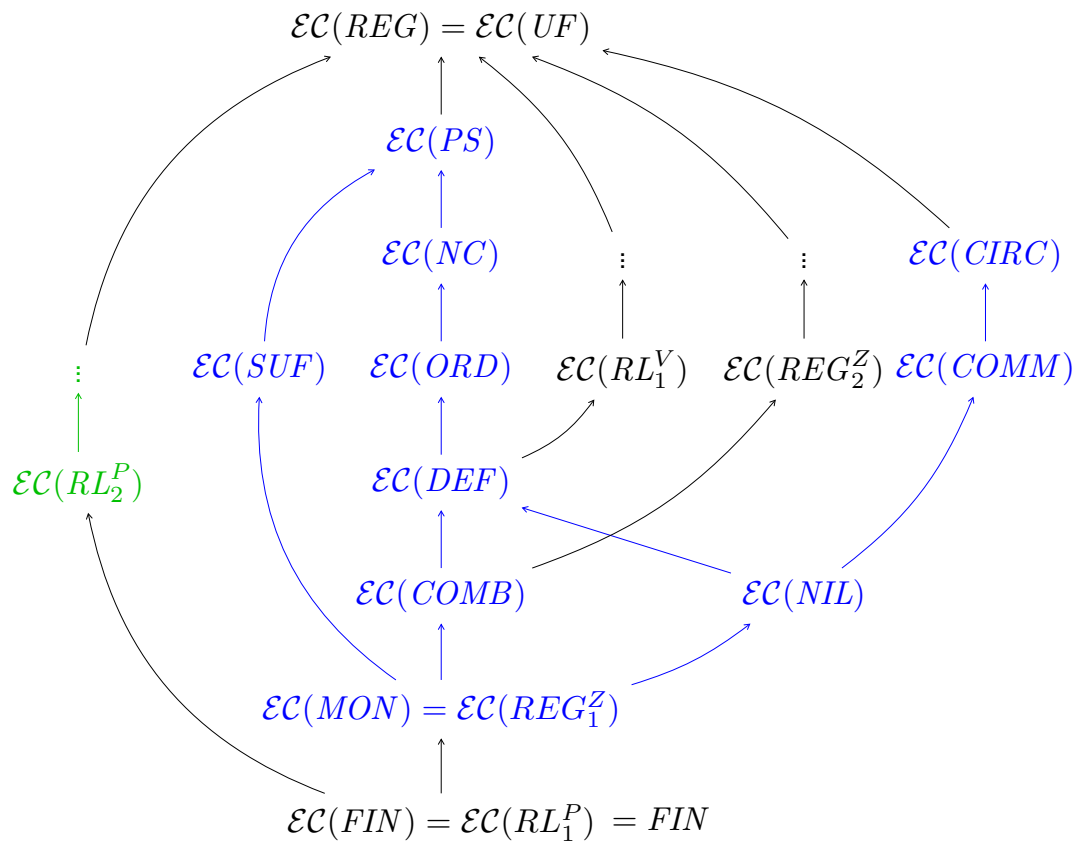
$\cup \{ ba^{2n}b \mid n \geq 1 \}$

$\in \mathcal{EC}(RL_1^V) \setminus \mathcal{EC}(PS)$

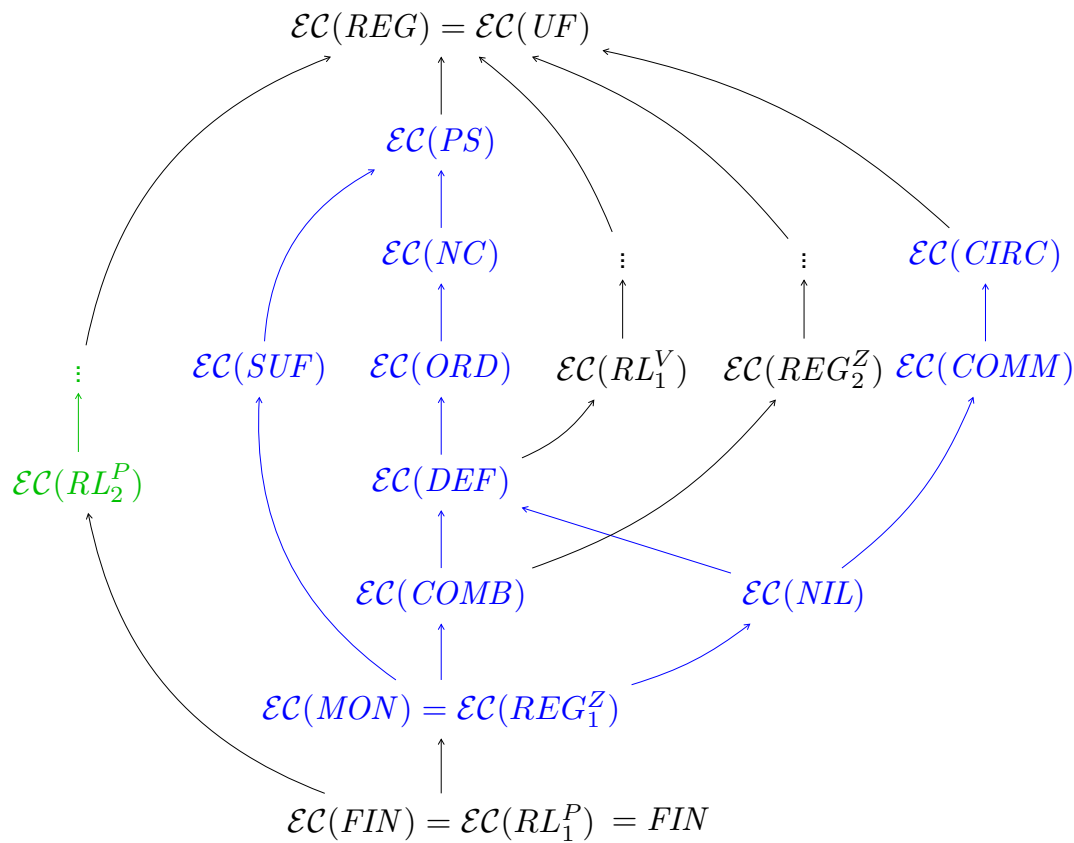
Results - One Hierarchy: Incomparabilities



Results - One Hierarchy: Incomparabilities



Results - One Hierarchy: Incomparabilities

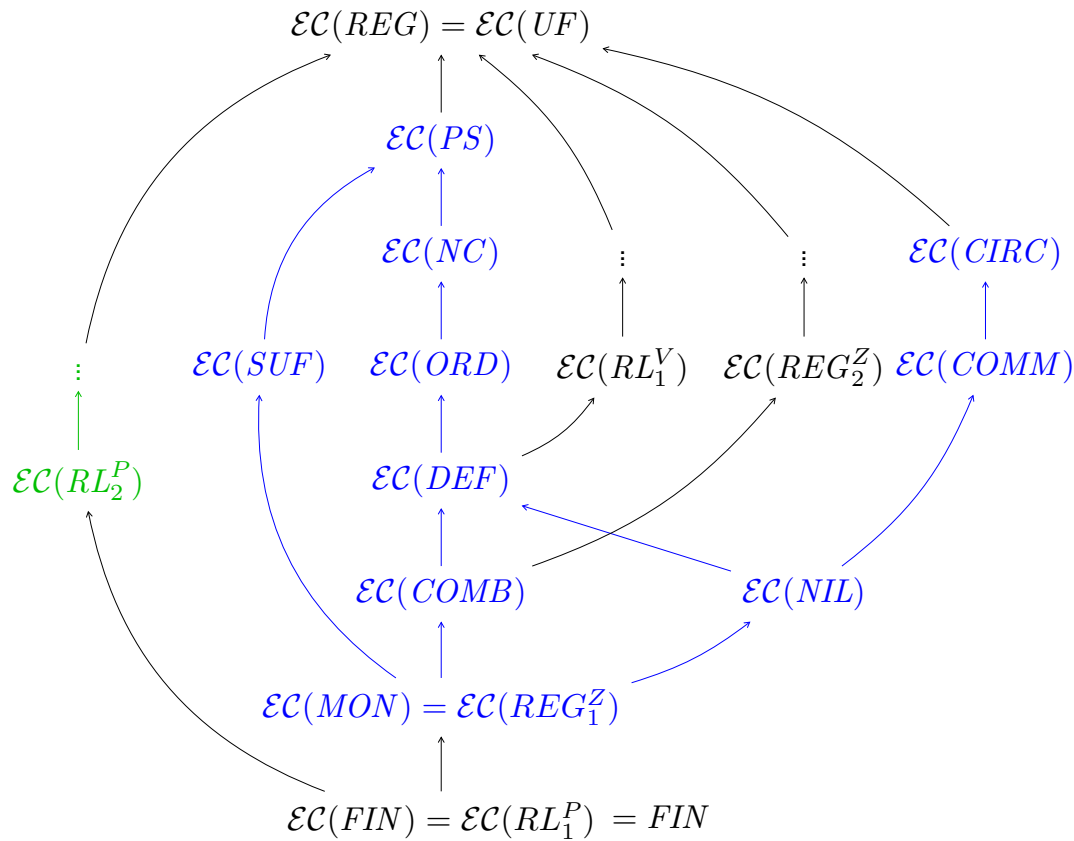


$\mathcal{EC}(RL_n^P)$ with $n \geq 2$ to
 $\mathcal{EC}(MON) \dots \mathcal{EC}(PS)$
 $\mathcal{EC}(MON) \dots \mathcal{EC}(CIRC)$

$$L_{RL_2^P, \neg PS} = \{ a^n \mid n \geq 1 \} \\ \cup \{ ba^{2n}b \mid n \geq 1 \} \\ \in \mathcal{EC}(RL_2^P) \setminus \mathcal{EC}(PS)$$

$$L_{RL_2^P, \neg CIRC} = \{ ab \}^+ \cup \{ ba \}^+ \\ \cup \{ c(ba)^n c \mid n \geq 1 \} \\ \in \mathcal{EC}(RL_2^P) \setminus \mathcal{EC}(CIRC)$$

Results - One Hierarchy: Incomparabilities



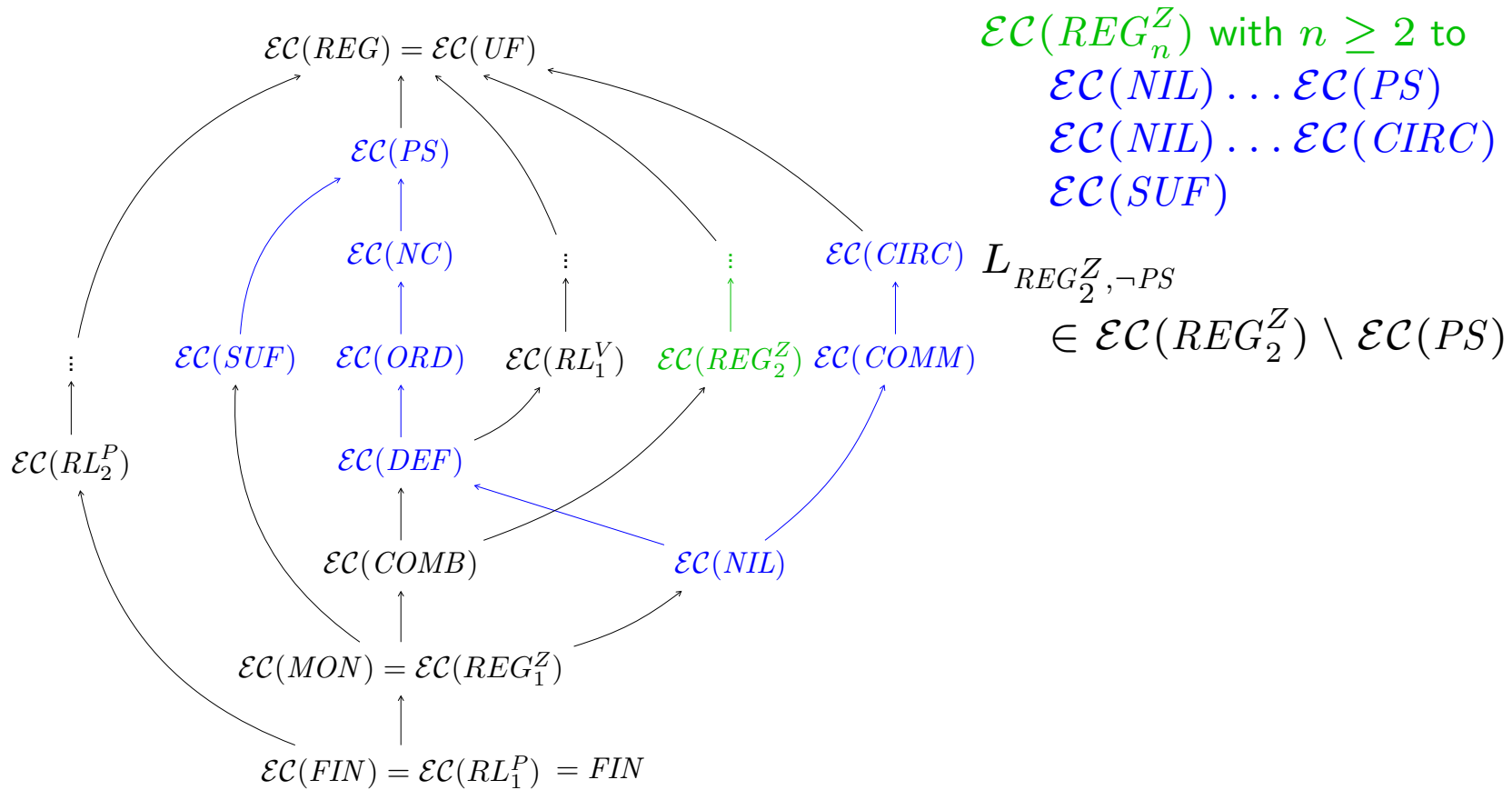
$\mathcal{EC}(RL_n^P)$ with $n \geq 2$ to
 $\mathcal{EC}(MON) \dots \mathcal{EC}(PS)$
 $\mathcal{EC}(MON) \dots \mathcal{EC}(CIRC)$

$$L_{RL_2^P, \neg PS} = \{ a^n \mid n \geq 1 \} \cup \{ ba^{2n}b \mid n \geq 1 \} \\ \in \mathcal{EC}(RL_2^P) \setminus \mathcal{EC}(PS)$$

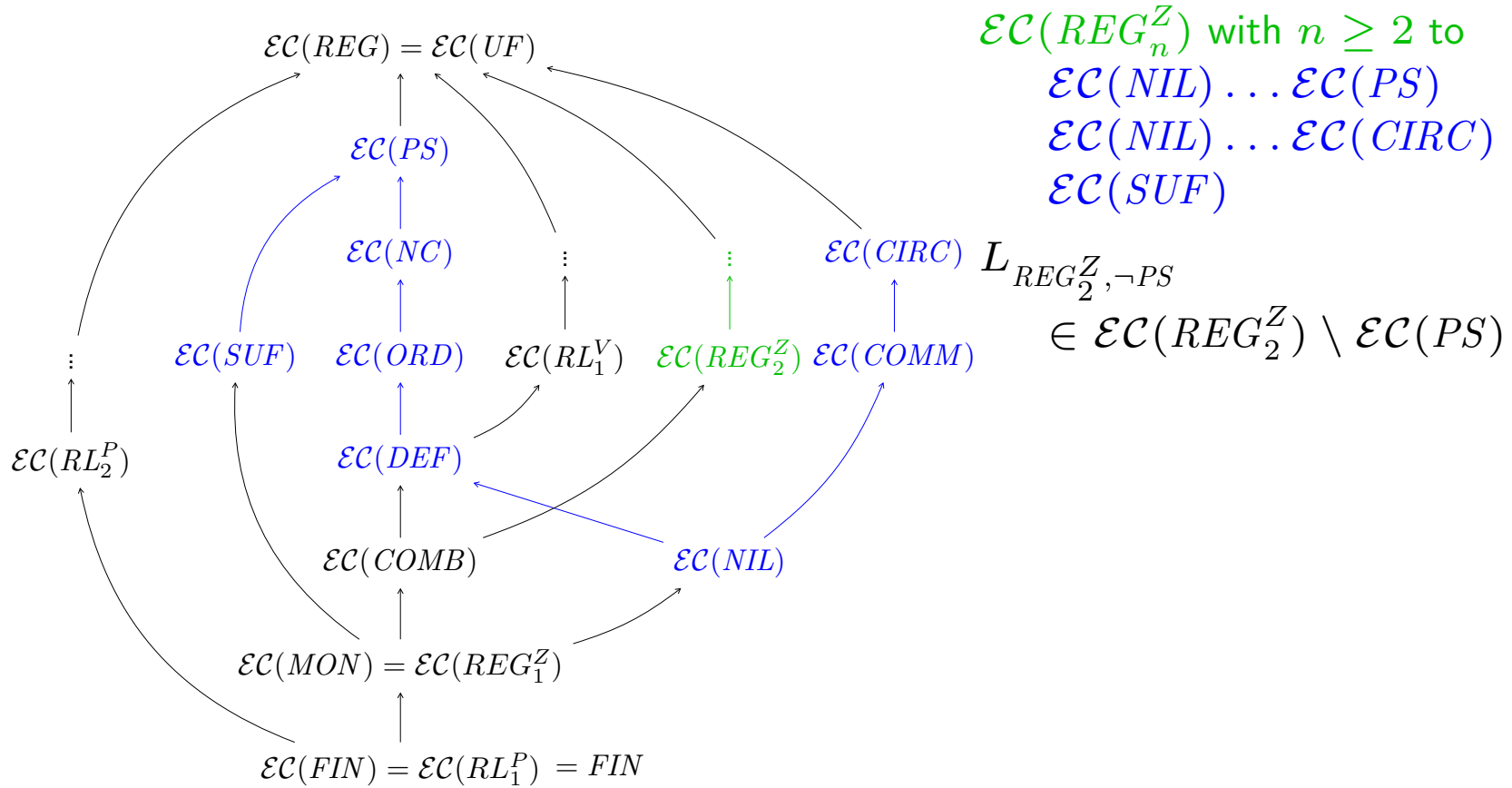
$$L_{RL_2^P, \neg CIRC} = \{ ab \}^+ \cup \{ ba \}^+ \cup \{ c(ba)^n c \mid n \geq 1 \} \\ \in \mathcal{EC}(RL_2^P) \setminus \mathcal{EC}(CIRC)$$

$$L_{MON, \neg RL_n^P} = \{ a_1, \dots, a_n \}^* \\ \in \mathcal{EC}(MON) \setminus \mathcal{EC}(RL_n^P)$$

Results - One Hierarchy: Incomparabilities

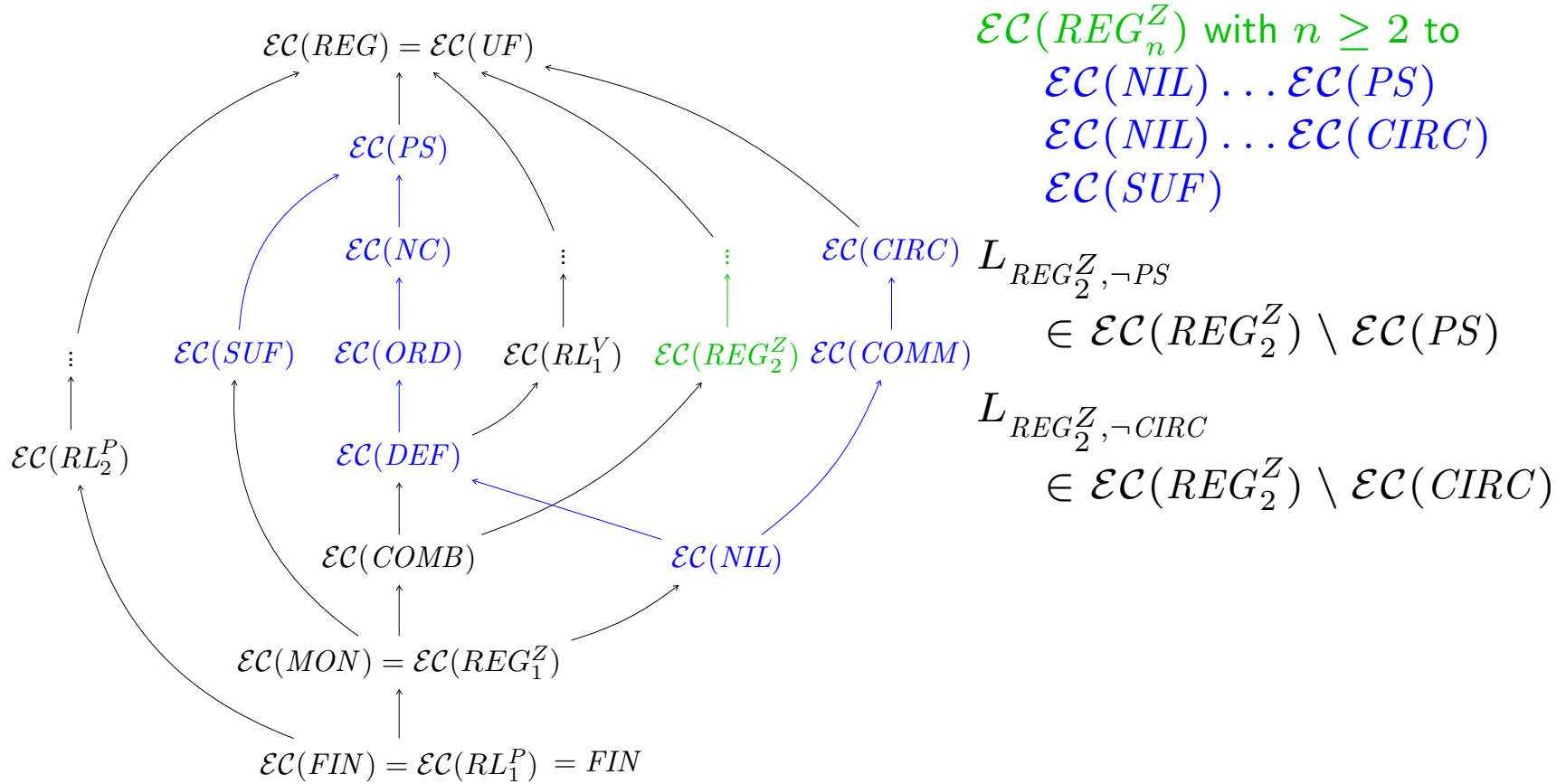


Results - One Hierarchy: Incomparabilities



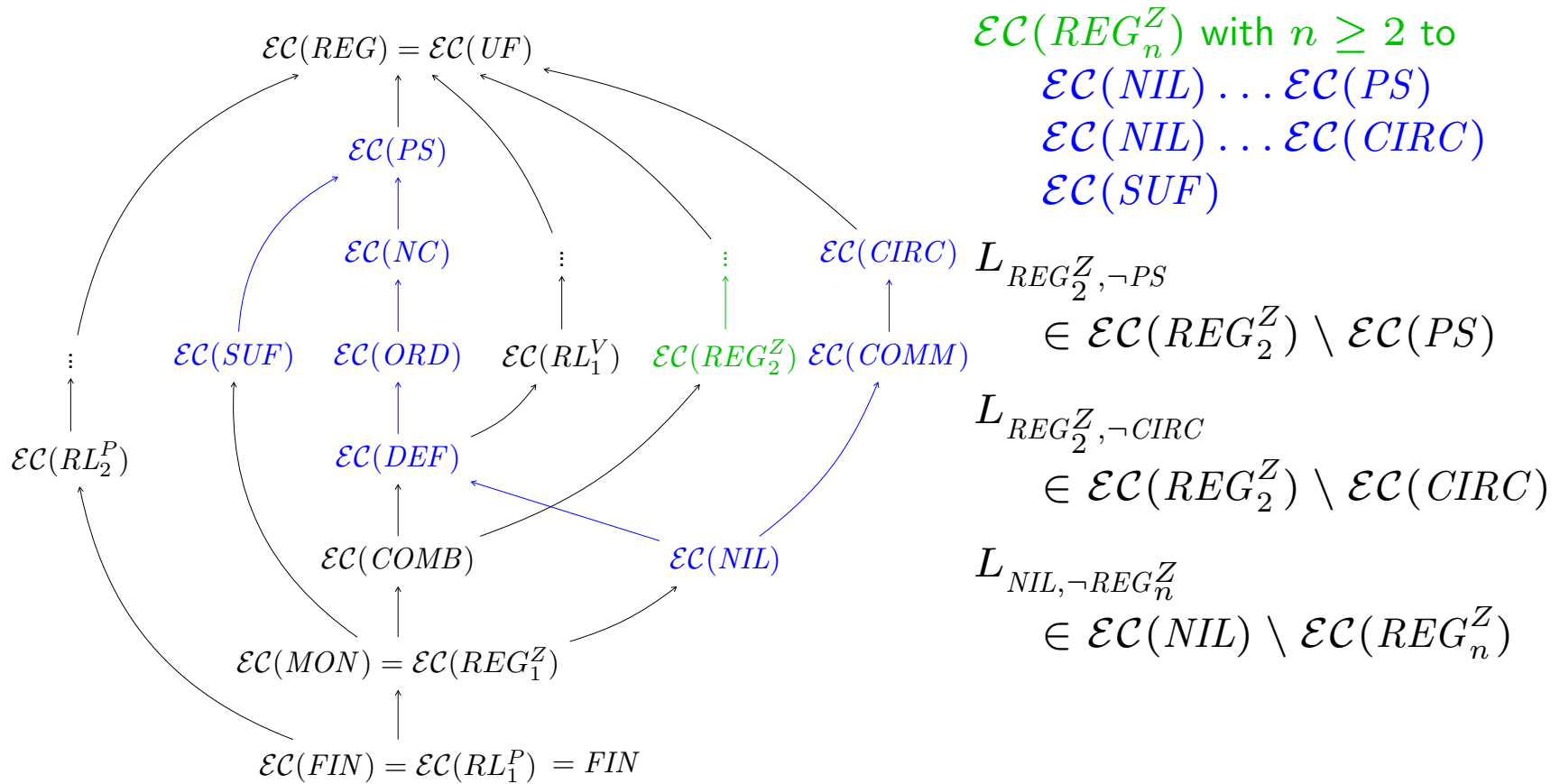
$$L_{REG_2^Z, \neg PS} = \{ a^n \mid n \geq 1 \} \cup \{ ba^{2^n}b \mid n \geq 1 \}$$

Results - One Hierarchy: Incomparabilities

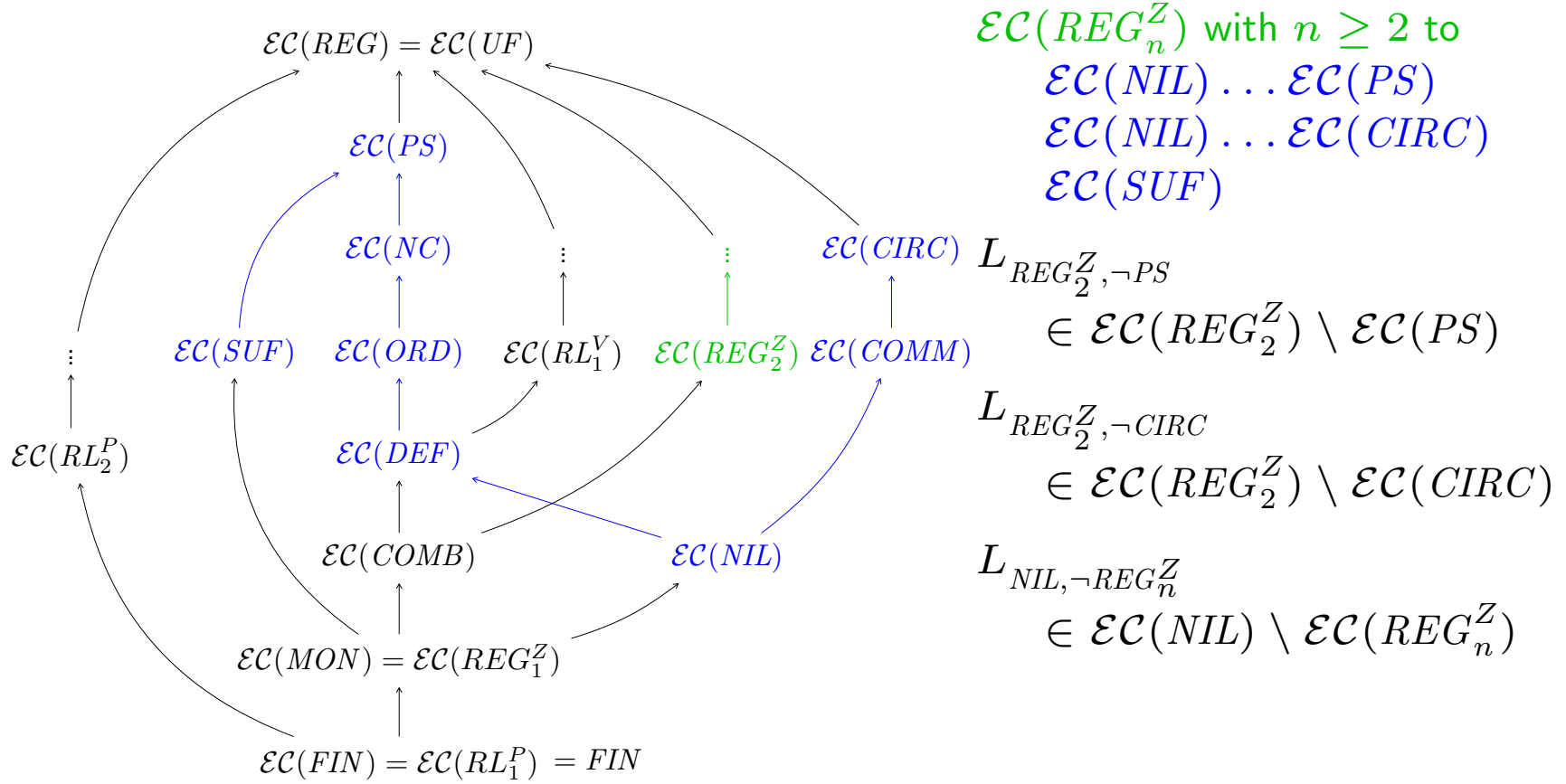


$$L_{REG_2^Z, -CIRC} = \{ab\}^+ \cup \{ba\}^+ \cup \{c(ba)^n c \mid n \geq 1\}$$

Results - One Hierarchy: Incomparabilities

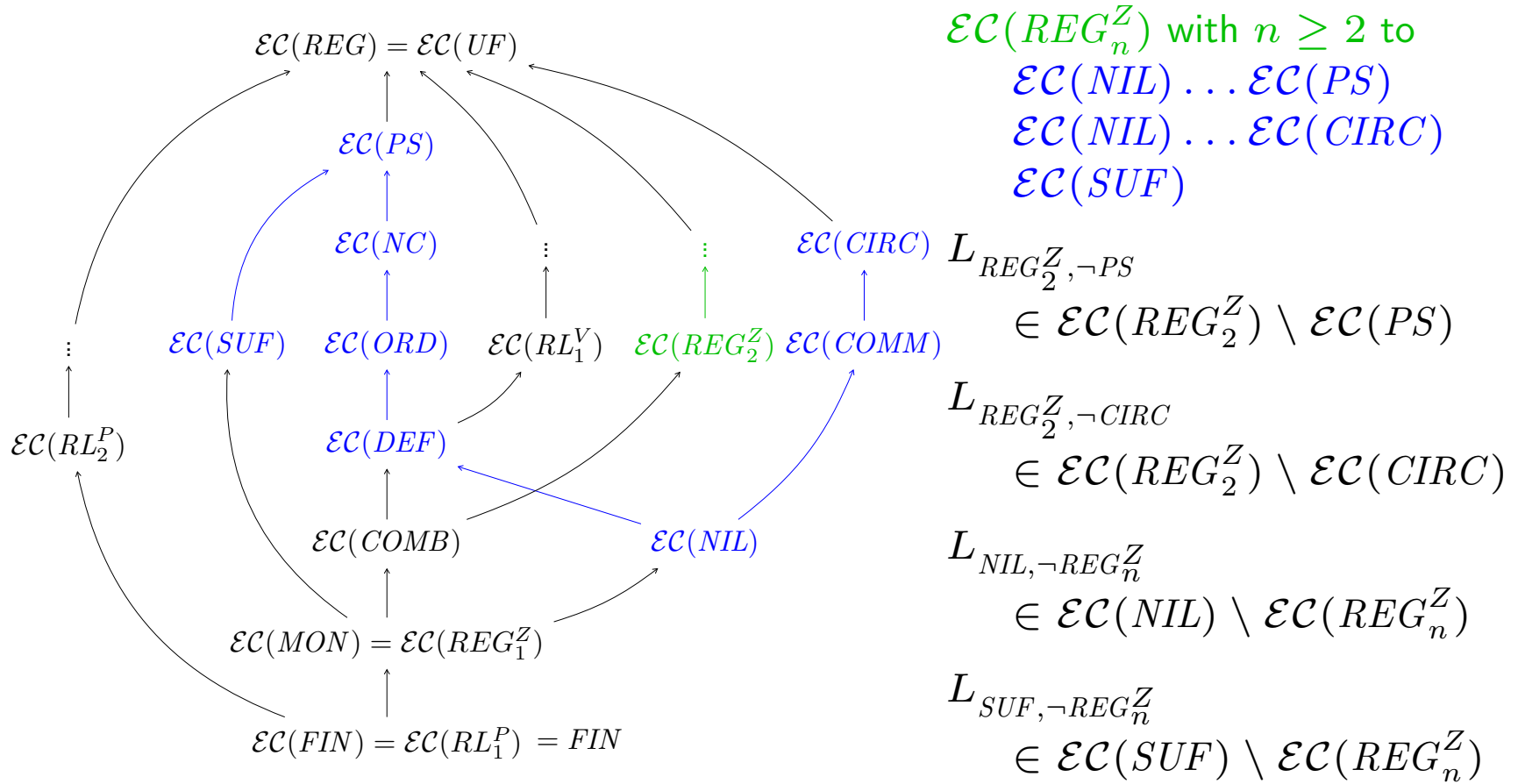


Results - One Hierarchy: Incomparabilities

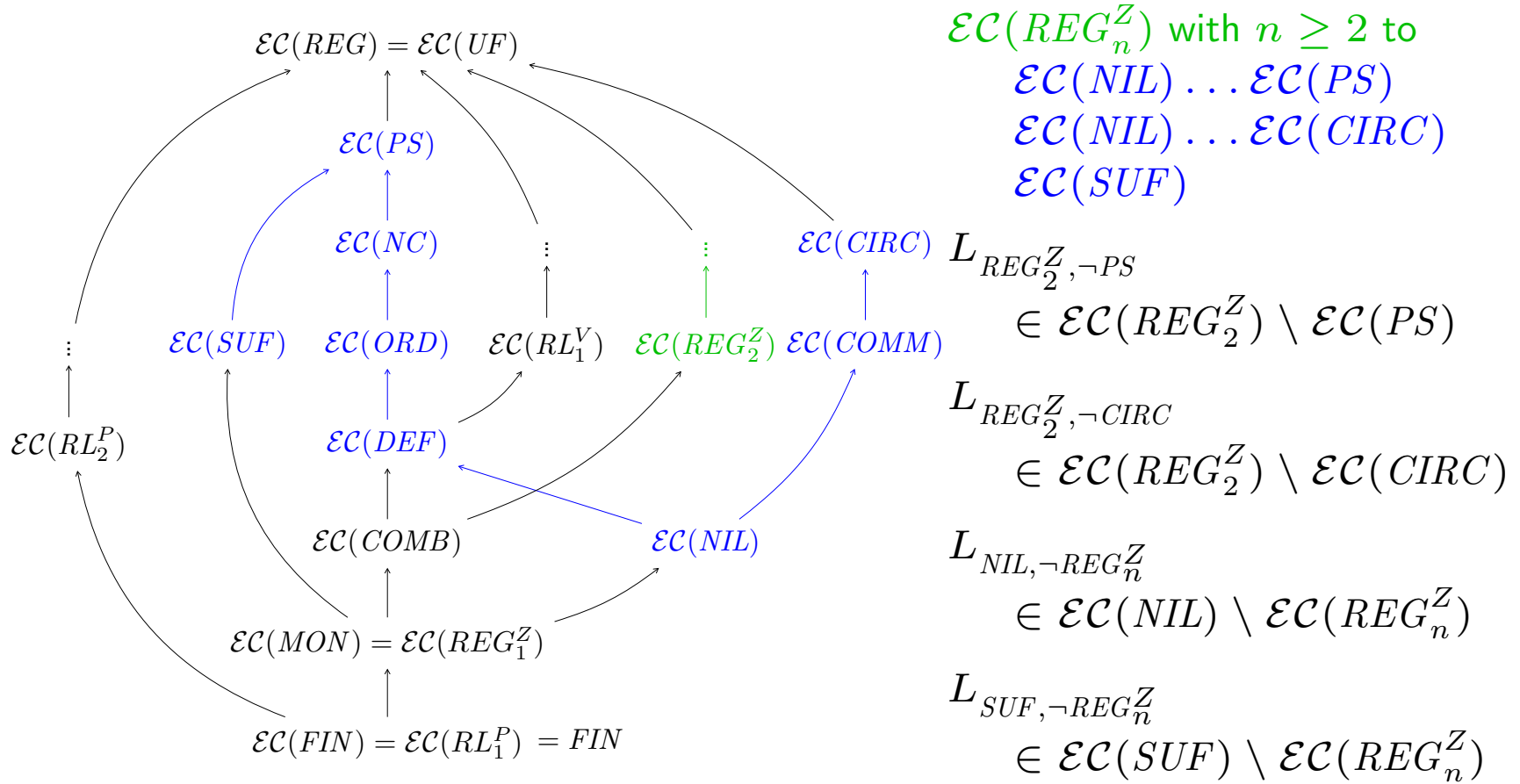


$$L_{NIL, -REG_n^Z} = \{ a^i \mid 1 \leq i \leq n \} \cup \{ a^{k(n+1)} \mid k \geq 2 \}$$

Results - One Hierarchy: Incomparabilities

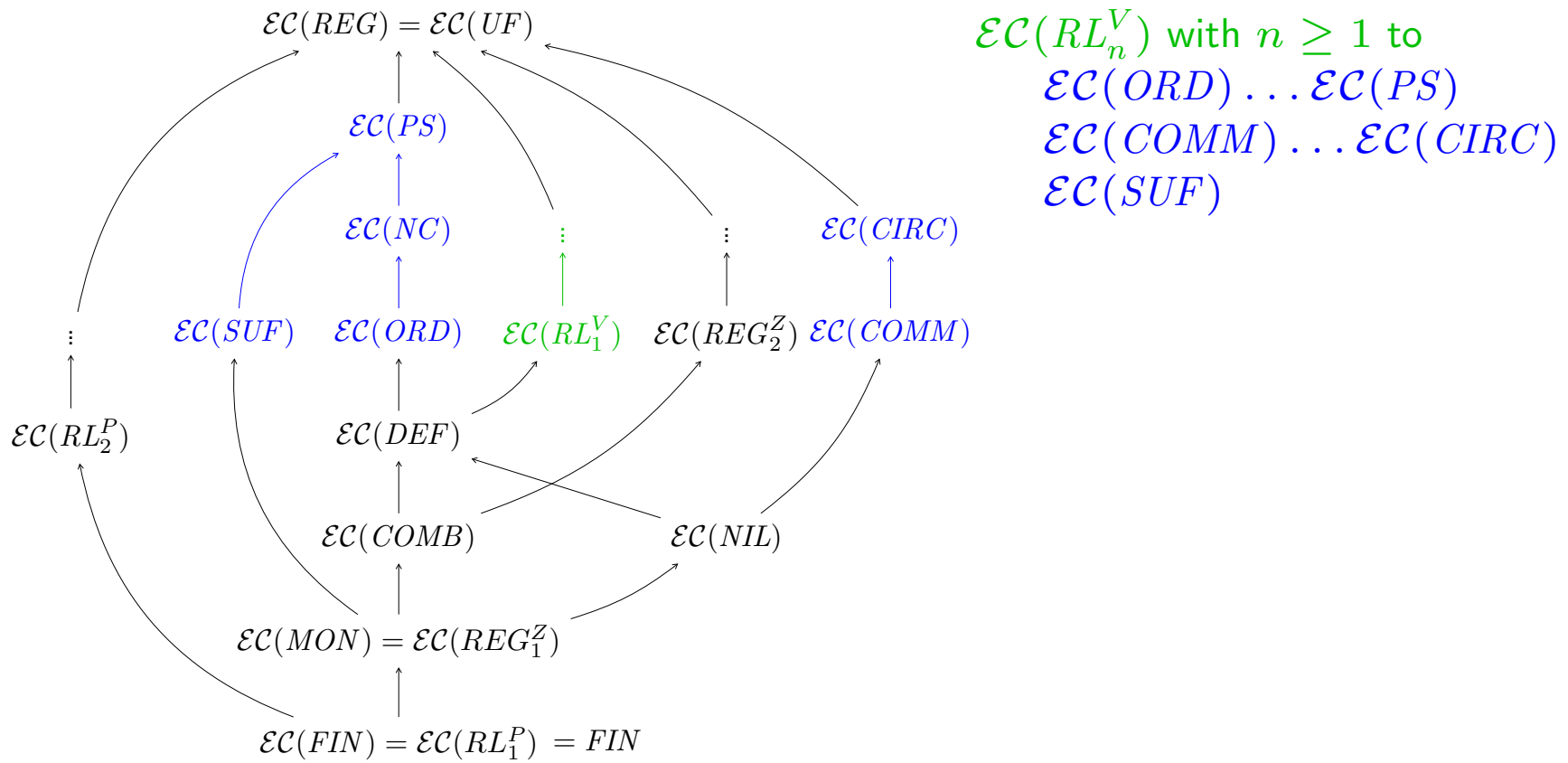


Results - One Hierarchy: Incomparabilities

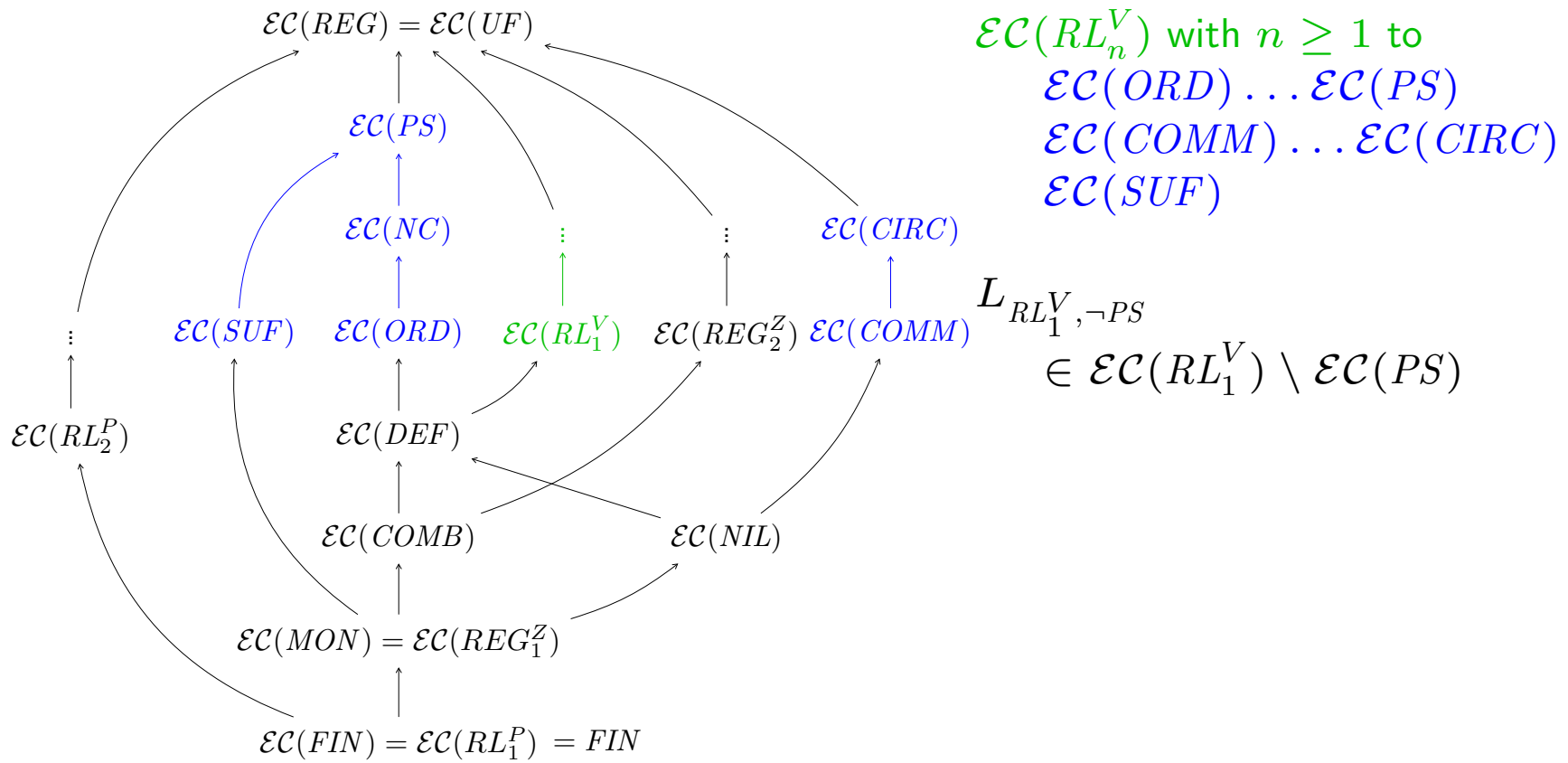


$$L_{SUF, \neg REG_n^Z} = \{ ba^i \mid 1 \leq i \leq n \} \cup \{ ba^{k(n+1)} \mid k \geq 2 \}$$

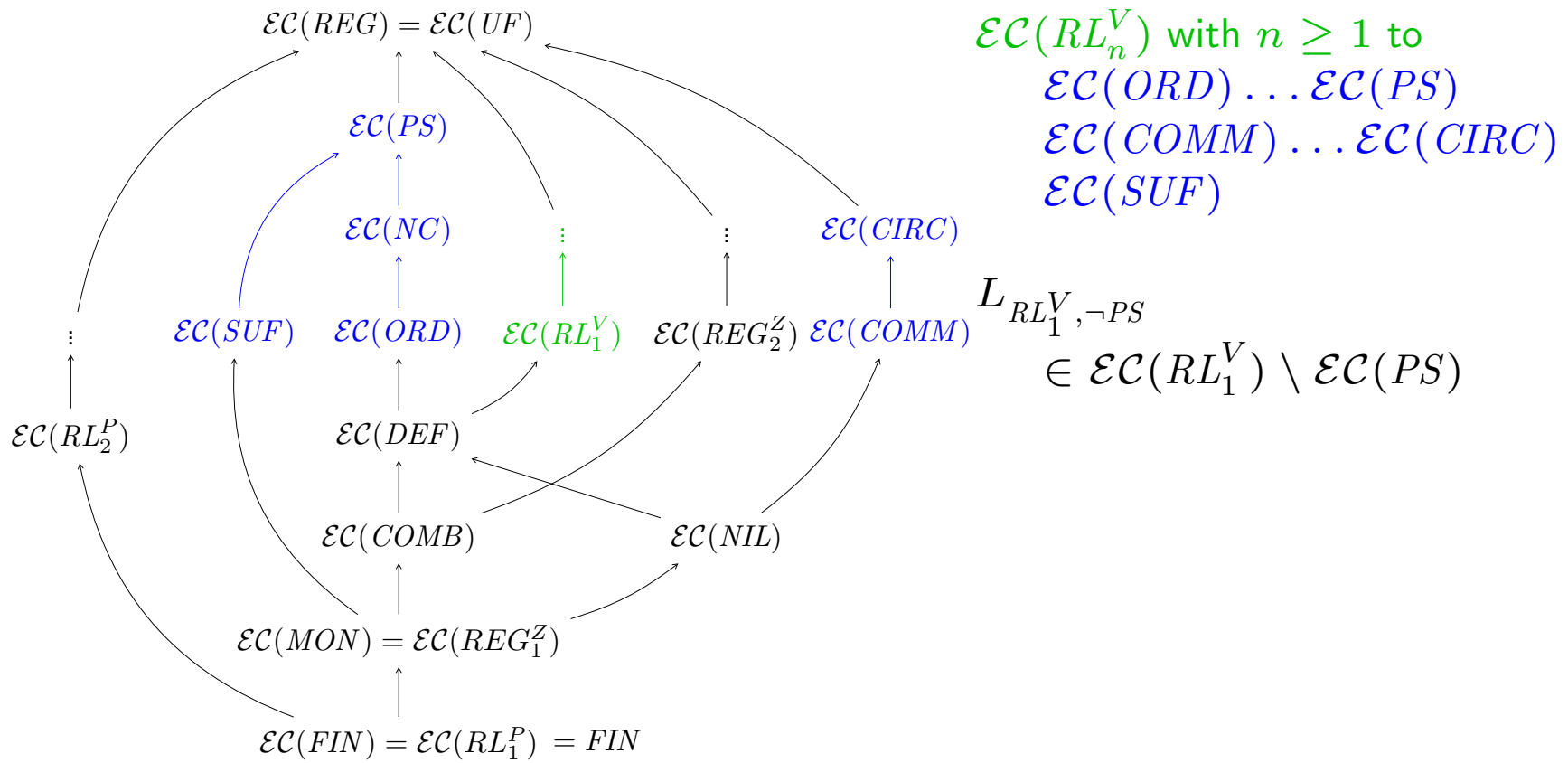
Results - One Hierarchy: Incomparabilities



Results - One Hierarchy: Incomparabilities

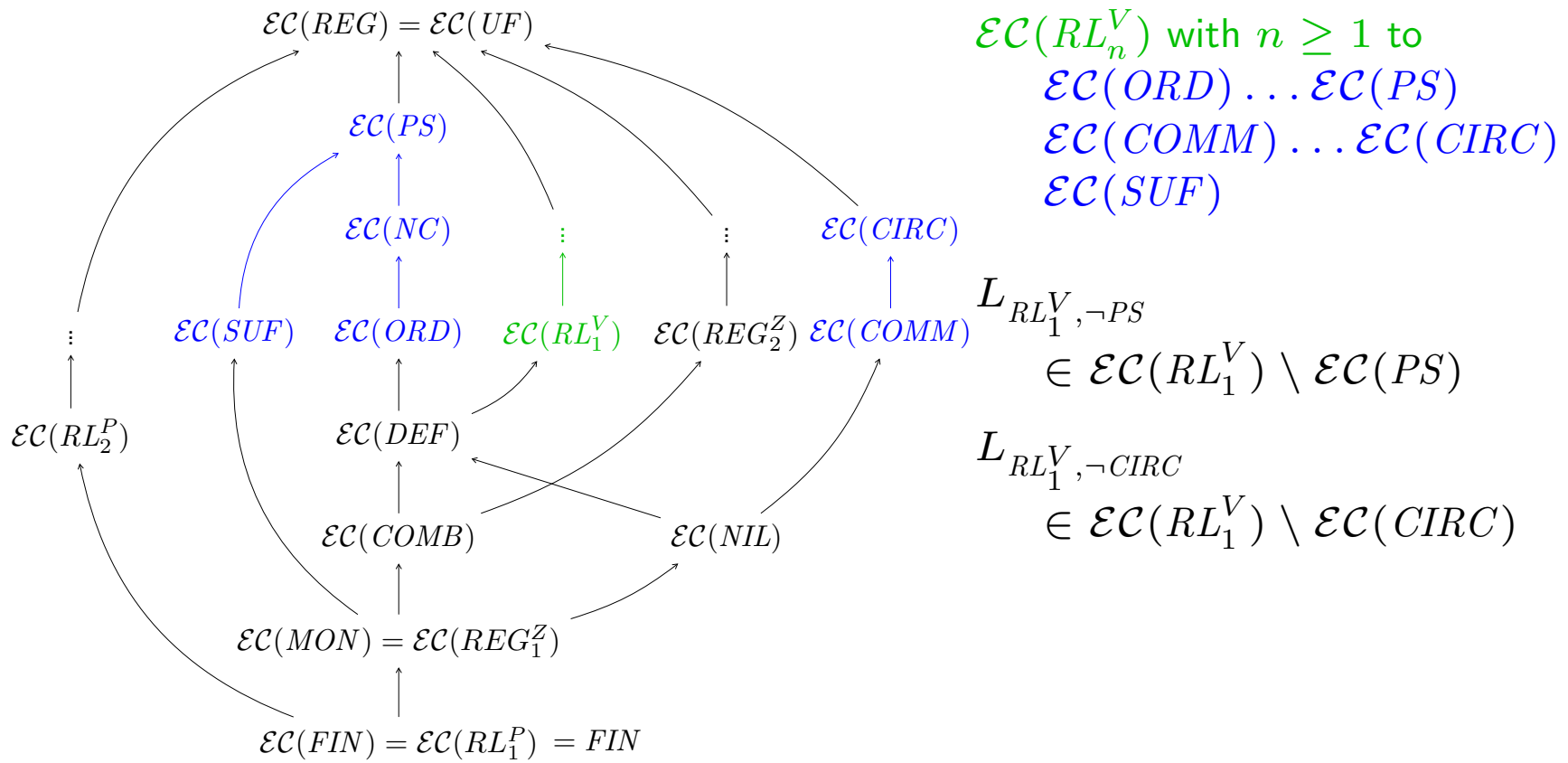


Results - One Hierarchy: Incomparabilities

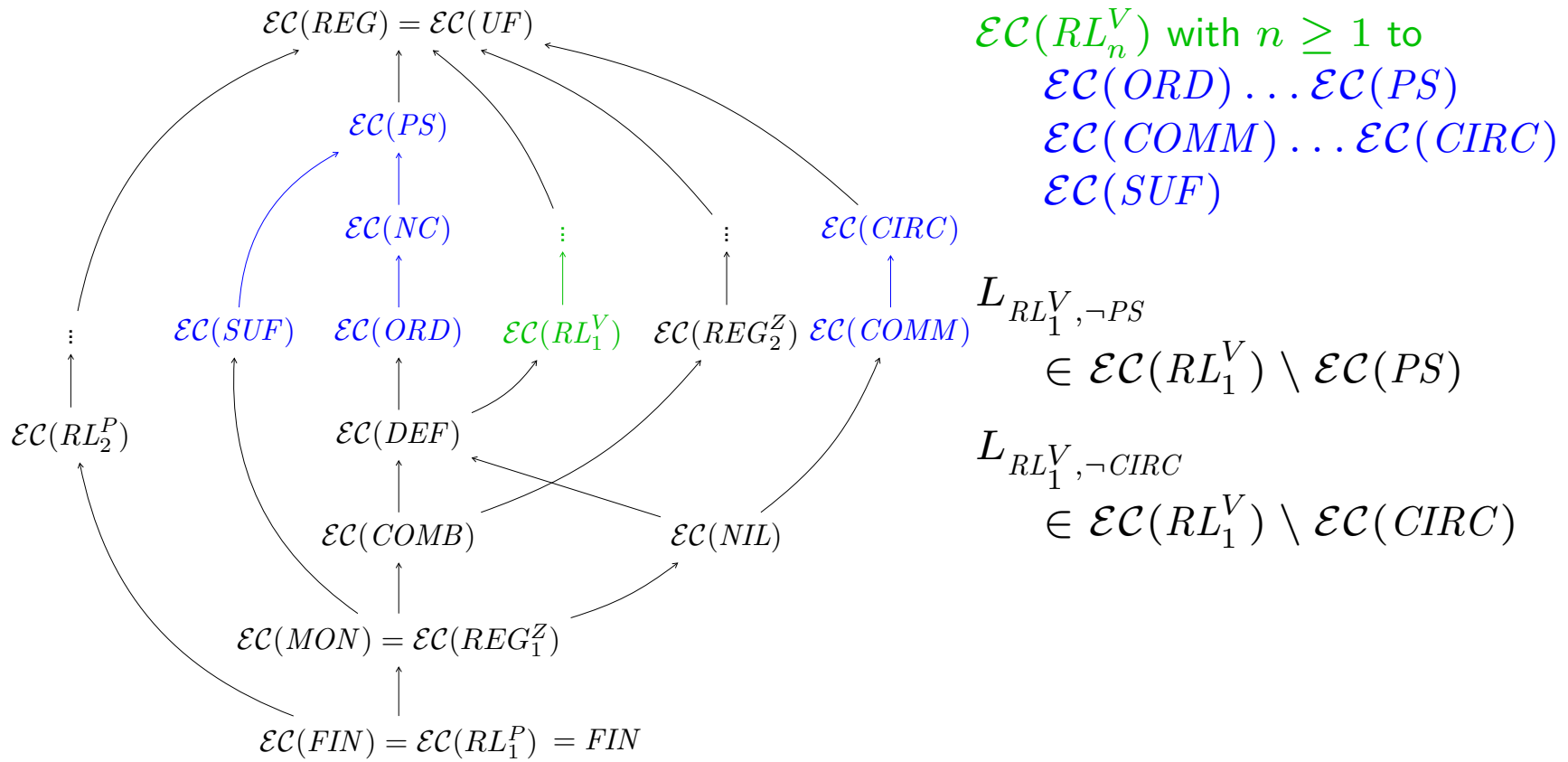


$$L_{RL_1^V, \neg PS} = \{ a^n \mid n \geq 1 \} \cup \{ ba^{2n}b \mid n \geq 1 \}$$

Results - One Hierarchy: Incomparabilities

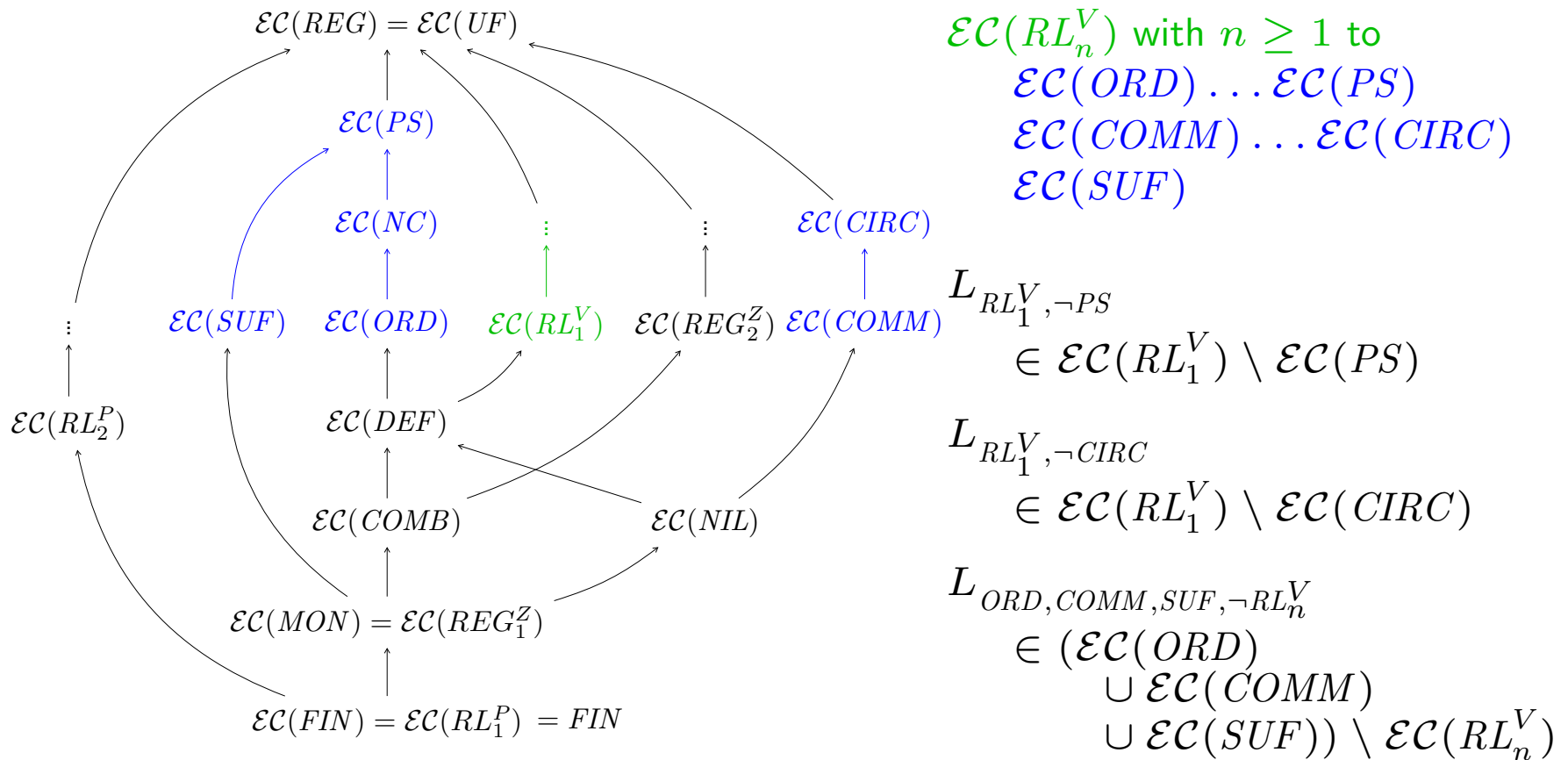


Results - One Hierarchy: Incomparabilities

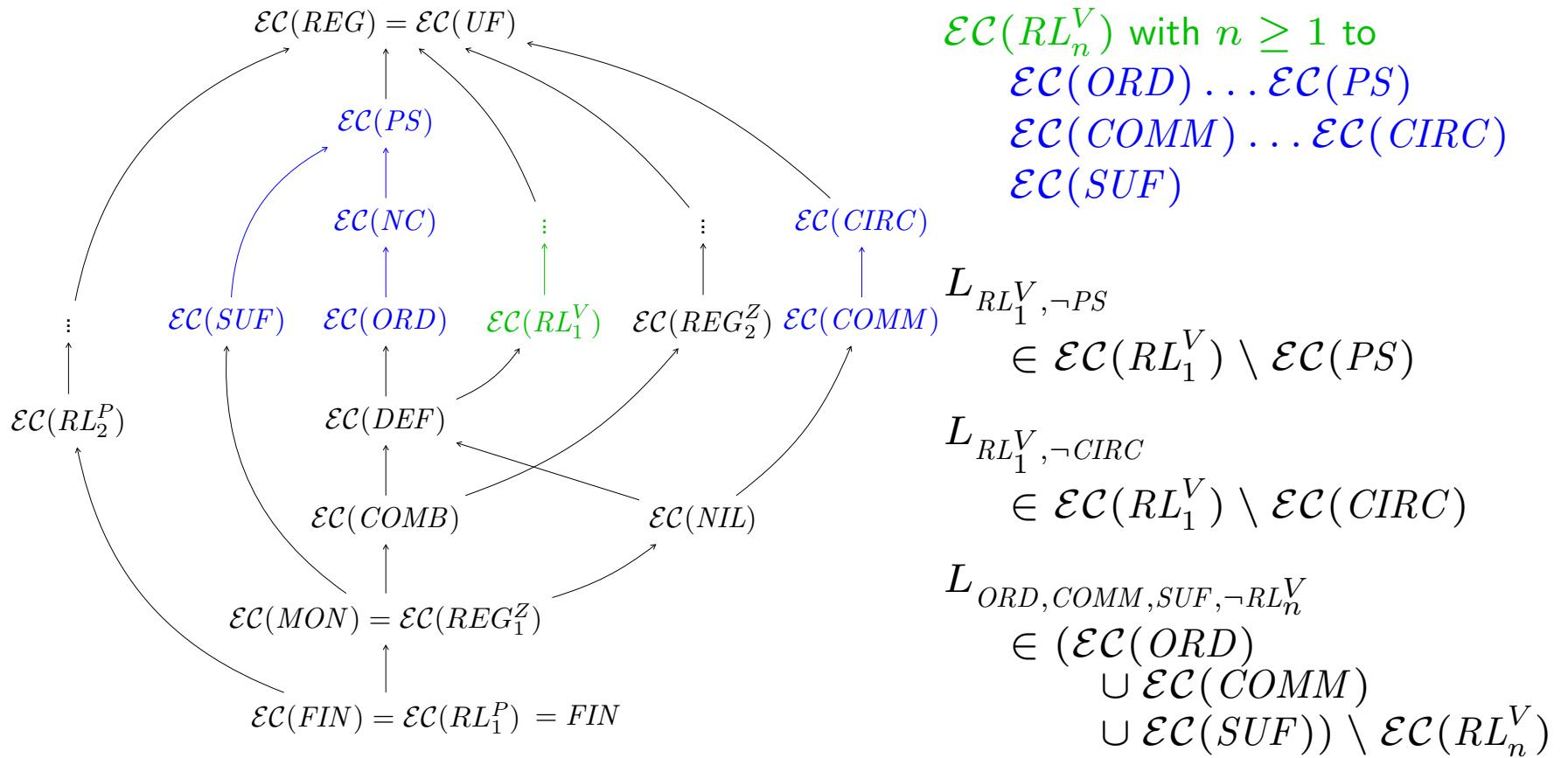


$$L_{RL_1^V, \neg CIRC} = \{ab\}^+ \cup \{ba\}^+ \cup \{c(ba)^n c \mid n \geq 1\}$$

Results - One Hierarchy: Incomparabilities



Results - One Hierarchy: Incomparabilities



$$L_{ORD, COMM, SUF, \neg RL_n^V} = \{a, b\}^+ \cup \{c\} \{ w \mid w \in \{a, b\}^*, |w|_a \leq n \} \{c\}$$

Future Work

Future Work

- Inclusion relations and incomparabilities between classes based on strictly locally testable languages and those classes here

Future Work

- Inclusion relations and incomparabilities between classes based on strictly locally testable languages and those classes here
- Other subregular families for the selection languages

Future Work

- Inclusion relations and incomparabilities between classes based on strictly locally testable languages and those classes here
- Other subregular families for the selection languages
- Further properties of the generated language families

Future Work

- Inclusion relations and incomparabilities between classes based on strictly locally testable languages and those classes here
- Other subregular families for the selection languages
- Further properties of the generated language families
- Merging the two hierarchies of internal contextual grammars (submitted)