# The Logic of Linearization: Interpretations of Trees via Strings

Vincent Czarnecki | Rutgers University | vincent.czarnecki@rutgers.edu

## Overview

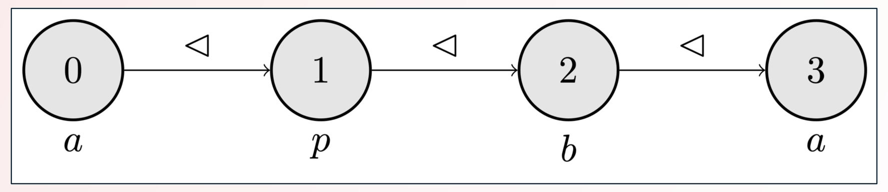
- Prior work has independently analyzed the computational properties of phonological (Strother-Garcia, 2019) and syntactic structure in terms of Model Theory (Rogers and Nordlinger, 1998).
- This work aims to understand formal properties of their *interaction*, showing that linearization can be formalized as a First-Order logical interpretation between trees and strings.

## Logical Interpretations

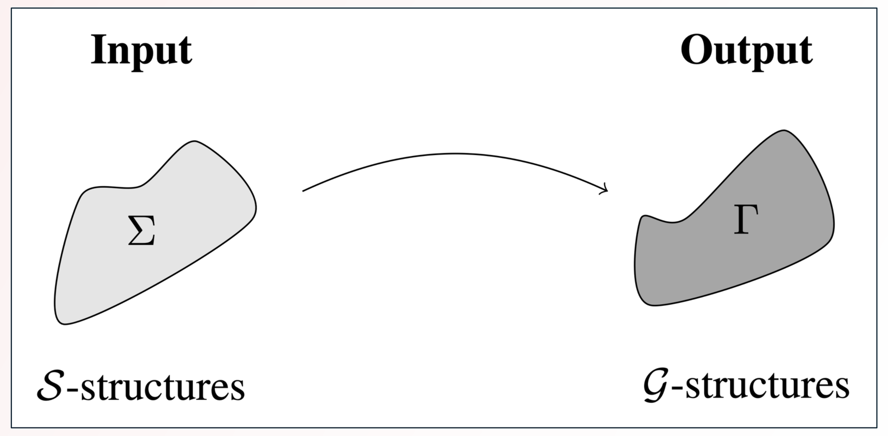
- A **signature** $\mathcal{S}$ is a collection of relations, functions, or constants:
- A **Relational Structure** $\Sigma$ is a pair of a domain $D$ with relations from $\mathcal{S}$ over elements in $D$ :

$$\Sigma = < D \mid r_1, r_2, \dots r_n >$$

- A string structure for the string *apba* is shown below, where $\lhd\ (x, y)$ is *strict precedence* and $a(x),\ b(x),\ p(x)$ are labeling relations:



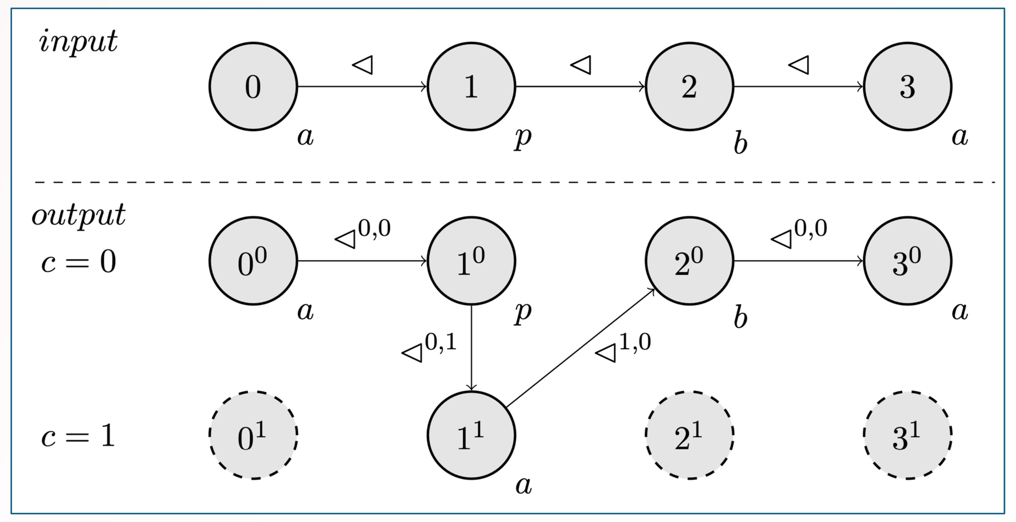

- An **interpretation** is a mapping from an input structure $\Sigma$ in $\mathcal{S}$ to an output structure $\Gamma$ in $\mathcal{G}$. See (Hodges, 1993)




- Output signature relations are defined in terms of input signature relations:

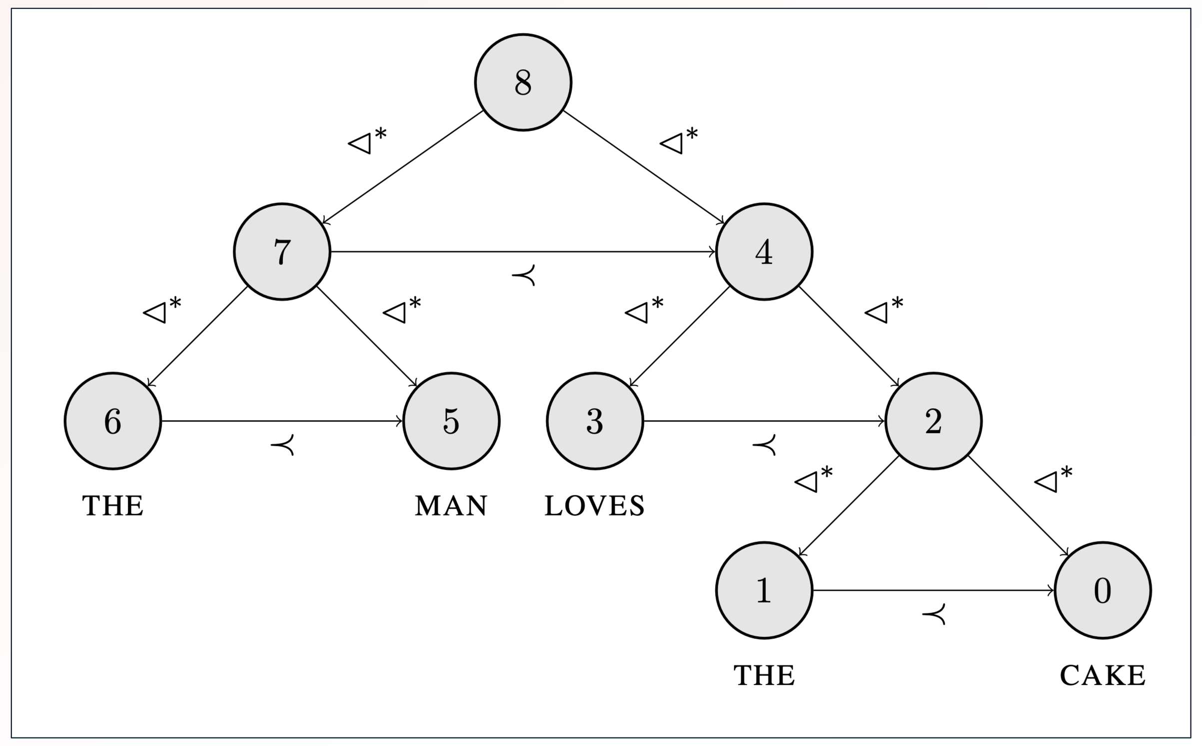
output relations in $\mathcal{G}$ := input relations in $\mathcal{S}$

- A visualization of a *string-to-string* interpretation that performs epenthesis is shown below:

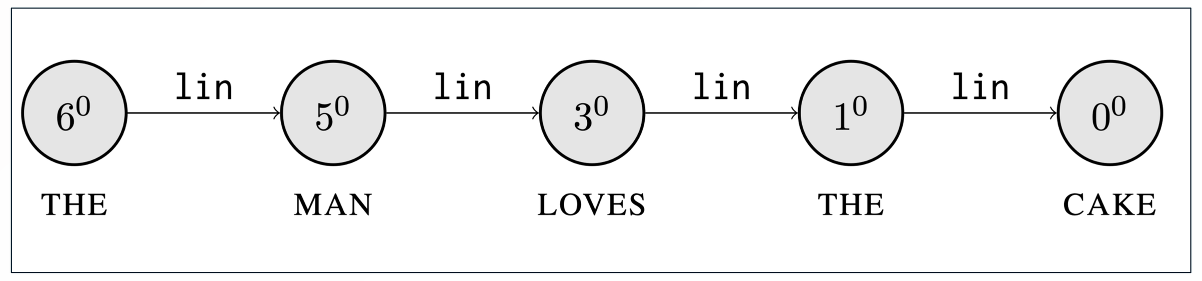



## Tree Structures and Flattening

- Tree structures are defined in the following way:
  - Binary general dominance relation: $\lhd^*\ (x, y)$
  - Binary precedence relation: $\prec (x, y)$
  - Unary labeling relations: $\sigma_i(x)$ where each $\sigma_i \in \Sigma_{syn}$ is a symbol in the abstract syntactic alphabet $\Sigma_{syn}$
- The linearization of this simplified tree (only lexical labels shown) is shown to the right:

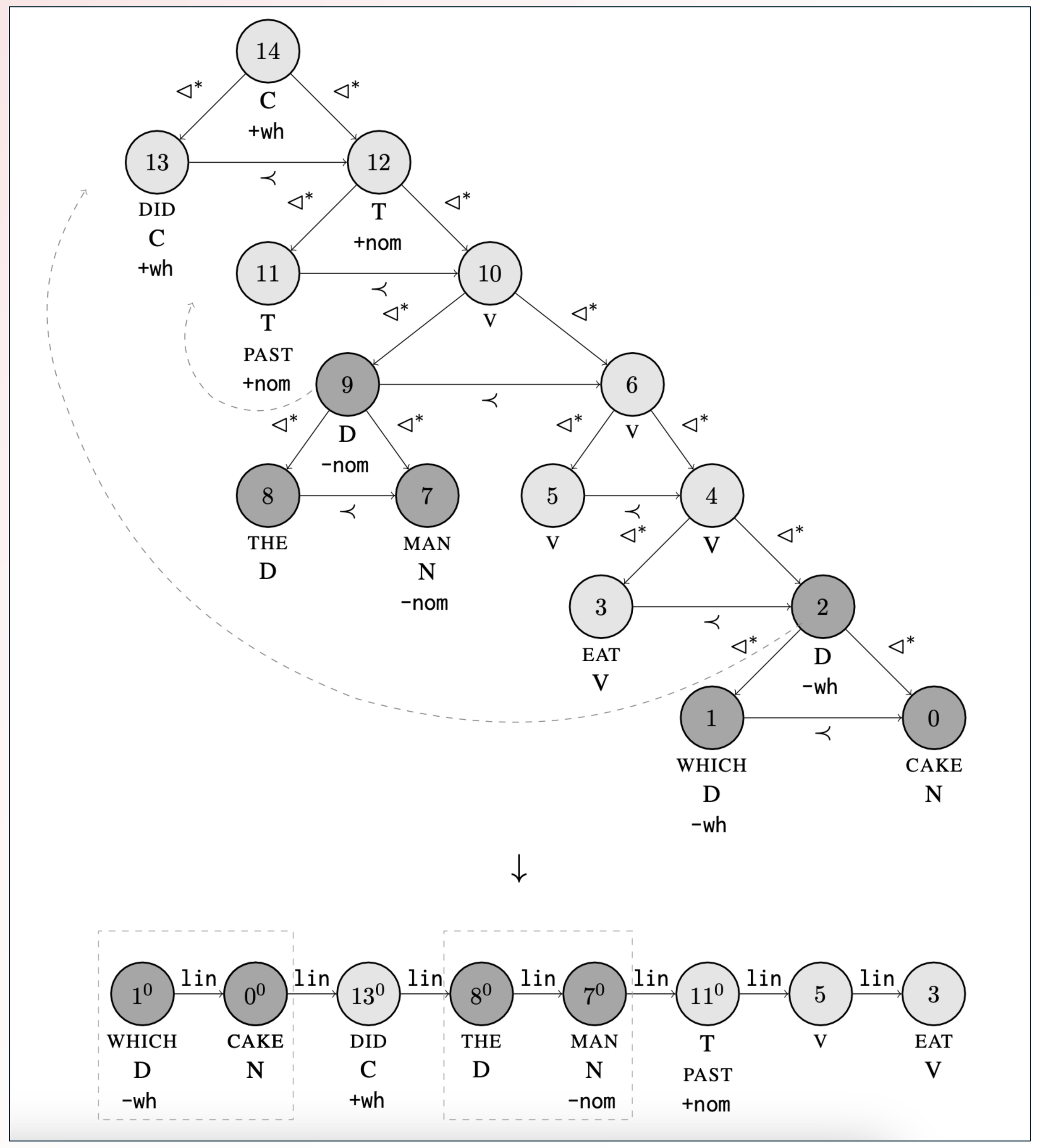



- Only leaves are relevant to the output, non-terminal nodes are forgotten.
- When will strict precedence hold *in the output string* between two input tree nodes $x, y$?
- *Leftmost-leafhood* and *rightmost-leafhood* are both relevant here.




## Incorporating Movement

- A visualization of an analysis incorporating movement to this method of linearization is sketched below:
- Note that the leaves of –f-movers (dark gray nodes) strictly precede +f-bearing heads.




## Linearization as a Logical Interpretation

- The *input signature* $\mathcal{S}$ consists of *tree* structures.
- The *output signature* $\mathcal{G}$ consists of *string* structures.
- The linearization condition in the output string between two input tree nodes is:

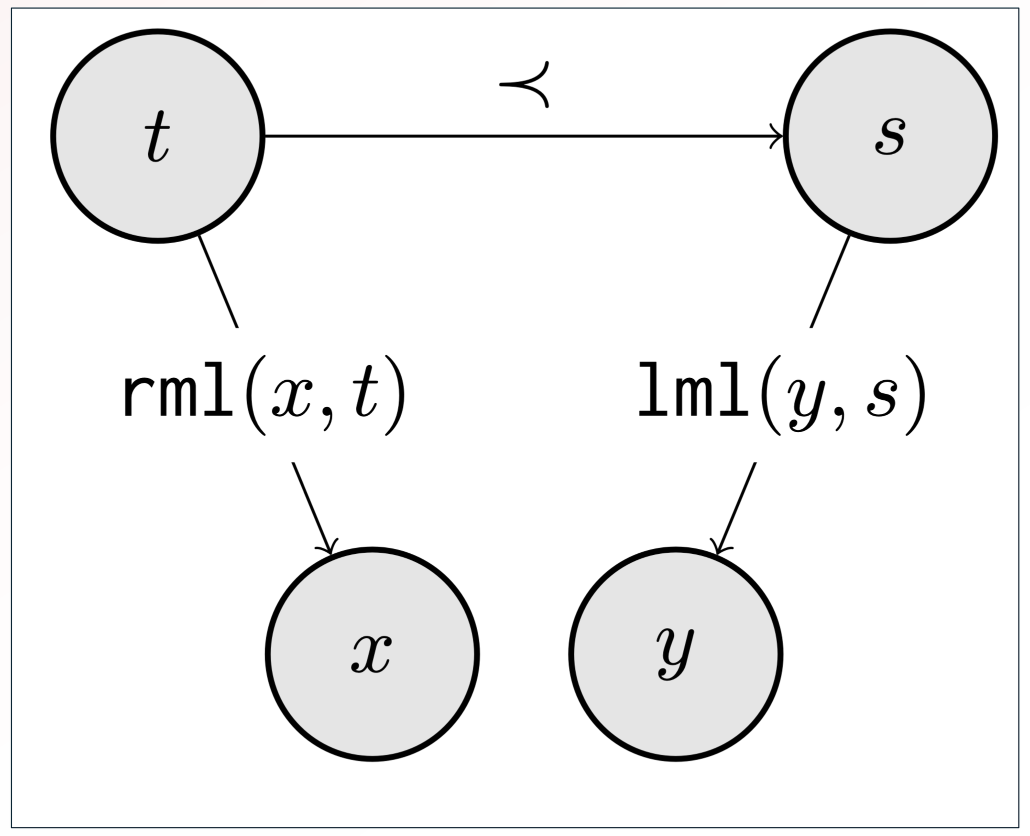
$$\mathtt{lin}(x, y) := \exists t \exists s [\prec (t, s) \wedge \mathtt{rml}(t, x) \wedge \mathtt{lml}(s, y)]$$

- A node $x$ is the left-most leaf of a node $y$ iff for all the left-leaf nodes $z$ that $y$ dominates, the only one with nothing further left is $x$:

$$\mathtt{lml}(x, y) := \forall z [(\lhd^*(y, z) \wedge \mathtt{left\text{-}leaf}(z) \wedge \forall s [\mathtt{between}(y, s, z) \wedge \neg \exists t [\prec (t, s)]]) \leftrightarrow z = x]$$

- A node $x$ is the right-most leaf of a node $y$ iff for all the right-leaf nodes $z$ that $y$ dominates, the only one with nothing further right is $x$:

$$\mathtt{rml}(x, y) := \forall z [(\lhd^*(y, z) \wedge \mathtt{right\text{-}leaf}(z) \wedge \forall s [\mathtt{between}(y, s, z) \wedge \neg \exists t [\prec (s, t)]]) \leftrightarrow z = x]$$




- This explicitly bakes locality into the post-syntactic representation. (Embick & Noyer, 1995)

## Conclusion

- Model-theoretic representations allow us to understand computational properties of semantics, syntax, phonology, morphology, and phonetics in isolation.
- It is also well-suited for understanding mappings between structures, and so, it is an *invaluable tool* for understanding the formal properties of the *interfaces between these modules*.
- This view of linearization also allows us to more closely investigate the relationship between computational complexity and typological expectations (Lambert et al., 2021), particularly for syntax-phonology interaction.

## Selected References

- Strother-Garcia, Kristina. Using model theory in phonology: a novel characterization of syllable structure and syllabification. University of Delaware, 2019.
- Rogers, James, and Rachel Nordlinger. A descriptive approach to language-theoretic complexity. Vol. 19. Stanford: CSLI Publications, 1998.
- Hodges, Wilfrid. Model theory. Cambridge university press, 1993.
- Lambert, Dakotah, Jonathan Rawski, and Jeffrey Heinz. "Typology emerges from simplicity in representations and learning." Journal of Language Modelling 9 (2021).

Find the full paper on my website!