Tensor Product Representations of Subregular Constraints

Jon Rawski
Dept. of Linguistics
Institute for Advanced Computational Science
Stony Brook University

A Theory Digestif

- Formal languages define **necessary and sufficient conditions** on (phonological) well-formedness
  - it's not modeling!
  - Regular class (bounded memory): sufficient, unnecessary
- Problem: Translate **subregularity** to distributed computation

Geometric characterization (vector spaces) of subregular languages (Rawski 2019 IJCAI)

- Relational Structures as tensors
- Locally Threshold Testable & Star-Free constraints as multilinear maps via first-order formulas

Tensors: Quick and Dirty Overview

- Order 1 — vector:

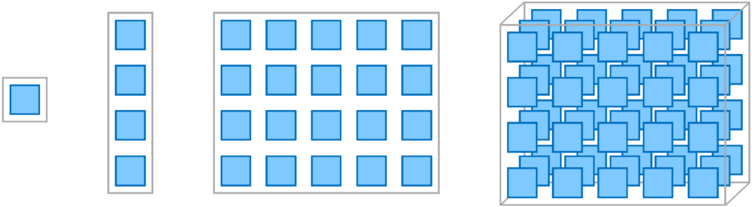
$$\vec{v} \in A = \sum_i C_i^v \overrightarrow{a_i}$$

- Order 2 — matrix:

$$M \in A \otimes B = \sum_{ij} C_{ij}^M \overrightarrow{a_i} \otimes \overrightarrow{b_j}$$

- Order 3 — Cuboid:

$$R \in A \otimes B \otimes C = \sum_{ijk} C_{ijk}^R \overrightarrow{a_i} \otimes \overrightarrow{b_j} \otimes \overrightarrow{c_k}$$



Tensors: Quick and Dirty Overview

Tensor contractions:

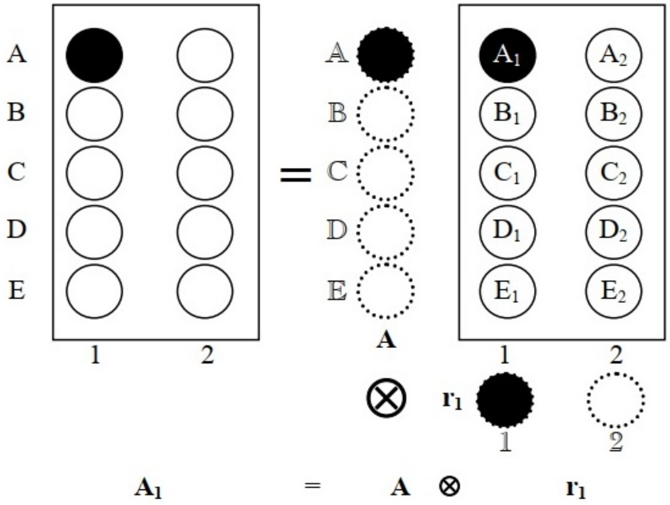
- Order 1 $\times$ order 1: inner product (dot product)
- Order 2 $\times$ order 1: matrix-vector multiplication
- Order 2 $\times$ order 2: matrix multiplication

Tensor contraction is nothing fancier than a generalization of these operations to any order.

- Order n $\times$ order m: sum through shared indices.

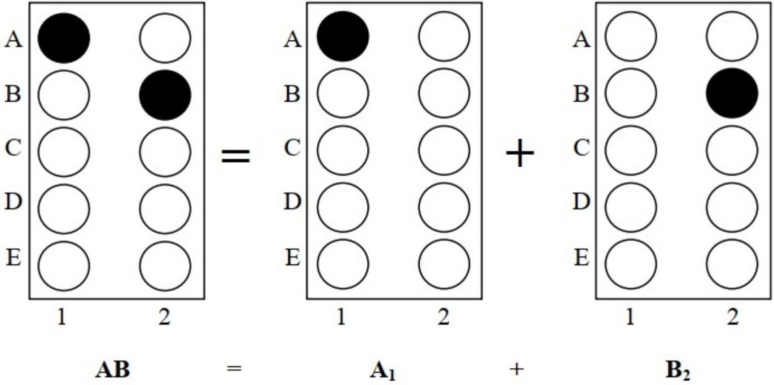
Order n $\times$ order m contraction yields tensor of order $n+m-2$.

Tensor Product Representations (Smolensky 1990)



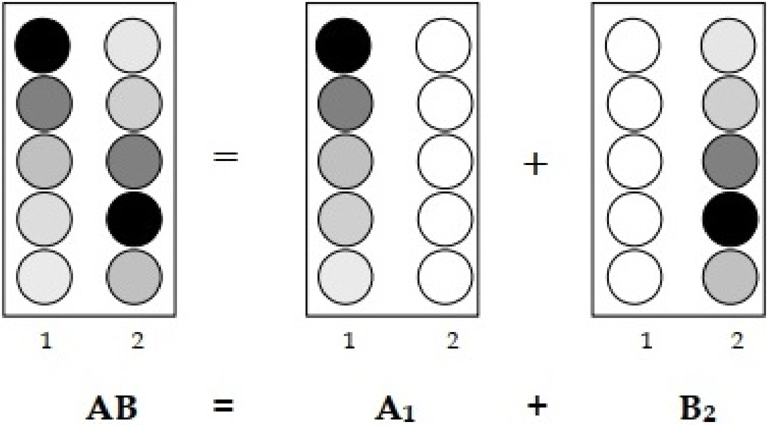
pics: Smolensky & Legendre 2006

Tensor Product Representations (Smolensky 1990)



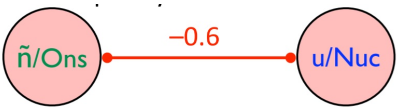
pics: Smolensky & Legendre 2006

Tensor Product Representations (Smolensky 1990)



pics: Smolensky & Legendre 2006

Tensor Product Representations (Smolensky 1990)

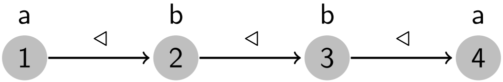


- Smolensky (and many others): grammar optimization (OT/HG) over tensors
- Hale and Smolensky: Strictly 2-Local HG for recursive tree tensors.
- beim Graben and Gerth: EEG dynamics and minimalist parsing with tree tensors

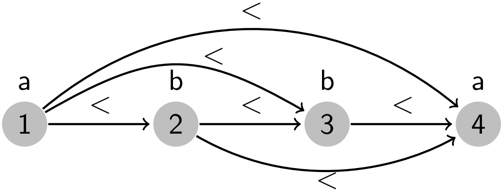
Relational Structures

Domain + Labeling Relation(s) + Ordering Relation(s)

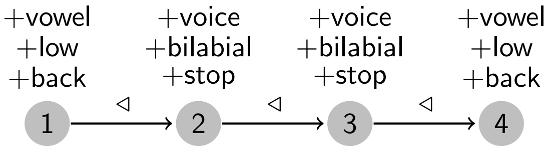
Relational Structures



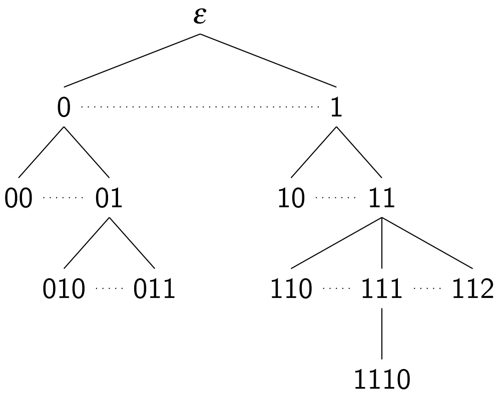
Relational Structures



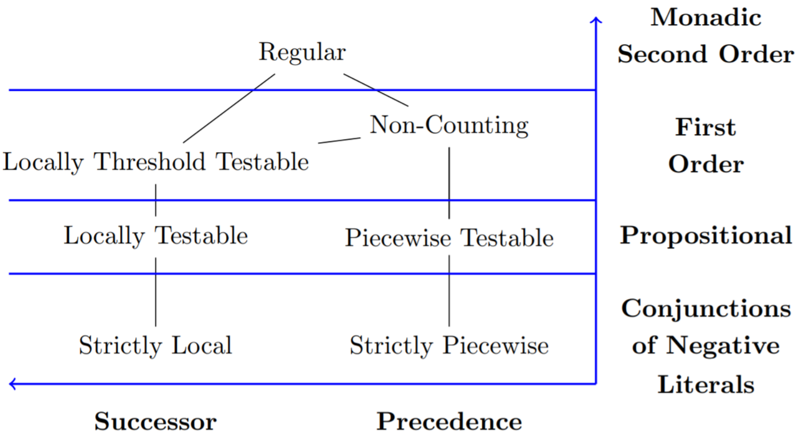
Relational Structures



Relational Structures



Subregular Hierarchy



pic: Heinz 2018

Tensors as Functions

Tensor-multilinear map isomorphism (Bourbaki, 1989; Lee, 1997)

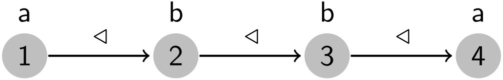
For any multilinear map $f : V_1 \to \ldots \to V_n$ there is a tensor $T^f \in V_n \otimes \ldots \otimes V_1$ such that for any $\overrightarrow{v_1} \in V_1, \ldots, \vec{v}_{n-1} \in V_{n-1}$, the following equality holds

$$f\,(\overrightarrow{v_1}, \ldots, \overrightarrow{v_{n-1}}) = T^f \times \overrightarrow{v_1} \times \ldots \times \overrightarrow{v_{n-1}}$$

Tensors therefore act as functions, with tensor contraction as function application.

Embedding Structures: Domain

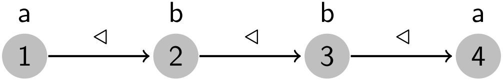
Domain elements D as the set of basis vectors in $\mathcal{D} \cong \mathbb{R}^{|D|}$.



$$D = \{1, 2, 3, 4\} \Rightarrow \mathbf{d}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{d}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{d}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{d}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Embedding Structures: Relations

k-ary relation r computed by an order-k tensor $\mathcal{R}$
truth value $[\![r(d_{i_1},\ldots,d_{i_k})]\!] = \mathcal{R}(\mathbf{d}_{i_1},\ldots,\mathbf{d}_{i_k}) = \mathcal{R} \times \mathbf{d}_{i_1} \times \cdots \times \mathbf{d}_{i_k}$



$$\mathcal{R}_a = \begin{bmatrix}1\\0\\0\\1\end{bmatrix} \qquad \mathcal{R}_b = \begin{bmatrix}0\\1\\1\\0\end{bmatrix} \qquad \mathcal{R}_\triangleleft = \begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{bmatrix}$$

$$\mathcal{R}_b(\mathbf{d}_2) = \begin{bmatrix}0\\1\\1\\0\end{bmatrix}^T \begin{bmatrix}0\\1\\0\\0\end{bmatrix} = 0+1+0+0 = 1$$

Logical Connectives (Sato 2017)

$$[\![\neg F]\!]' = 1 - [\![F]\!]'$$

$$[\![F_1 \wedge \cdots \wedge F_h]\!]' = [\![F_1]\!]' \cdots [\![F_h]\!]'$$

$$[\![F_1 \vee \cdots \vee F_h]\!]' = \min_1([\![F_1]\!]' + \ldots + [\![F_h]\!]')$$

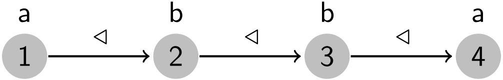
$$[\![\exists y F]\!]' = \min_1(\sum_{i=1}^{N} [\![F_{y \leftarrow d_i}]\!]')$$

$$[\![\forall y F]\!]' = [\![\neg \exists y \neg F]\!] = 1 - \min_1(\sum_{i=1}^{N} 1 - [\![F_{y \leftarrow d_i}]\!]')$$

$\min_1(x) = \min(x, 1) = x$ if $x < 1$, otherwise 1,

Easy Example: Words must contain a b

$$F_{\text{one-}b} = \exists x(R_b(x)) \qquad \mathcal{T}_{\text{one-}B} = \min_1\left(\sum_{i=1}^{N} \mathcal{R}_b(\mathbf{d}_i)\right)$$

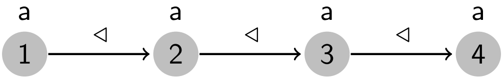


$$\min_1\left(\begin{bmatrix}0\\1\\1\\0\end{bmatrix}^T \begin{bmatrix}1\\0\\0\\0\end{bmatrix} + \begin{bmatrix}0\\1\\1\\0\end{bmatrix}^T \begin{bmatrix}0\\1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\1\\1\\0\end{bmatrix}^T \begin{bmatrix}0\\0\\1\\0\end{bmatrix} + \begin{bmatrix}0\\1\\1\\0\end{bmatrix}^T \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \right)$$

$$= \min_1(0+1+1+0) = \min_1(2) = 1$$

Easy Example: Words must contain a b

$$F_{\text{one-}b} = \exists x(R_b(x)) \qquad \mathcal{T}_{\text{one-}B} = \min_1 \left(\sum_{i=1}^{N} \mathcal{R}_b(\mathbf{d}_i) \right)$$



$$\min_1 \left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}^T \begin{bmatrix}1\\0\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\\0\end{bmatrix}^T \begin{bmatrix}0\\1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\\0\end{bmatrix}^T \begin{bmatrix}0\\0\\1\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\\0\end{bmatrix}^T \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \right)$$

$$= \min_1(0+0+0+0) = \min_1(0) = 0$$

Going Under the Hood

- tableaux are basically a graphical user interface
  - nice for converting descriptive generalizations
  - obscure the guts of computation
  - restrictiveness becomes baroque
- subregularity and optimization requires going under the hood
  - Tensor decomposition is flexible and powerful
    - Kolda/Bader 2009 review
  - fast algebraic operations to use for subregularity
    - projection, PCA, SVD, etc
  - Sato 2018: abducing relations & transitive closure in $\mathcal{O}(n^3)$

Optimization and Subregularity

- vanilla optimization & mods don't play well with subregularity
  - lots of evidence global optimization over- and undergenerates
  - Hao 2019: Serial optimization generates non-regular relations
  - Koser & Jardine 2019: SL constraints not closed under optimization
- ML theory: optimization insufficient/wrong language for neural nets (all constraint interaction is a special case)
  - Zhang et al 2017: explicit regularizers, early stopping, gradient noising tricks (batch sizes/learning rates) cant prevent algorithms from attaining low training objective even on data with random labels
  - Arora ICM/ICML plenary: optimization "may imply nothing about generalization, obscures important properties of architecture".