On The Computational Complexity of Syntactic Dependencies

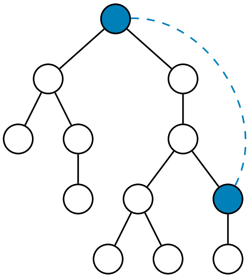
Kenneth Hanson

Stony Brook University

February 27, 2025

Language as a computational problem

1. What kinds of computations are needed to build linguistic structures?
2. What can we gain from this knowledge?



What kinds of computations are needed?

With a few caveats, linguistic patterns are **tier-based strictly local (TSL)**:

- Local and long-distance **phonotactics** (McMullin 2016; Heinz 2018)
- Local and long-distance **phonological maps** (Chandlee and Heinz 2018; Burness et al. 2021)
- Most **syntactic dependencies**

Selection	(Graf 2018)
Functional hierarchies	(Hanson 2023b)
Adjunction	(Hanson under review)
Movement	(Graf 2018, 2022b)
Case	(Vu et al. 2019; Hanson 2023a)
Agreement	(Hanson to appear)

Computational complexity: The old view

Modeled using **surface strings**, syntactic patterns are fairly **complex**

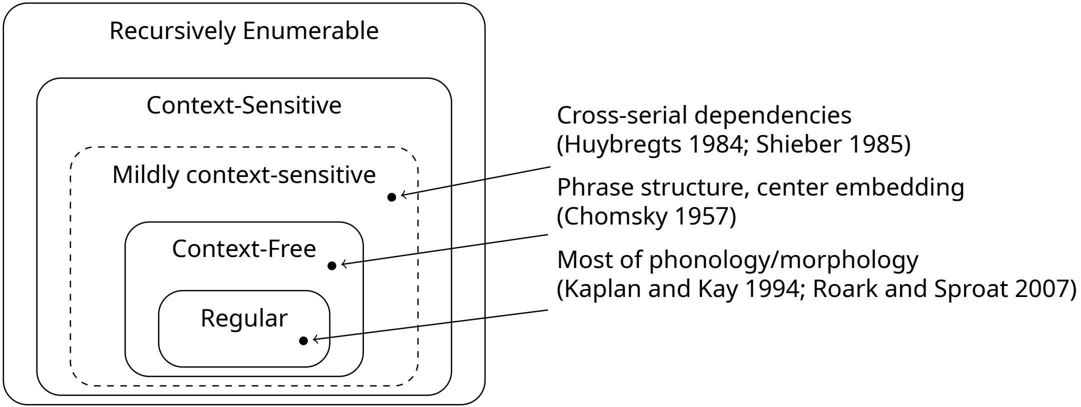


Figure 1: The Chomsky Hierarchy (Chomsky 1959), simplified

Computational complexity: The new view

Modeled using **trees**, syntactic patterns are **subregular**, along with most of phonology and morphology

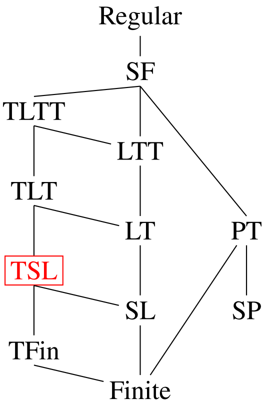


Figure 2: The Subregular Hierarchy (Heinz 2018; Lambert 2023), simplified

What can we gain from this knowledge?

- Syntax and phonology extremely similar (Graf 2022a)
  - $\longrightarrow$ Similar locality profile, similar parameters of variation **(Hanson to appear)**
- The mathematical properties of TSL patterns are well understood (Heinz et al. 2011; Lambert and Rogers 2020; Lambert 2023)
  - $\longrightarrow$ Connections to corpora **(Swanson et al. under review)** and experimental data **(Torres et al. 2023)**
- TSL patterns require simple inference mechanisms (Lambert et al. 2021) and are efficiently learnable (Jardine and McMullin 2017; Lambert 2021)
  - $\longrightarrow$ Integration with distributional approaches to learning (Belth 2023; **Hanson 2024**)

Overall theme:
By focusing on the computational properties of language, we can discover new generalizations and build connections within and beyond linguistic theory.

Roadmap

What is a TSL computation?

A TSL model of agreement

Consequences for typology

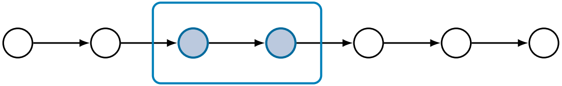
Related and ongoing work

- Gradience in syntactic islands
- Other syntactic dependencies
- Future research

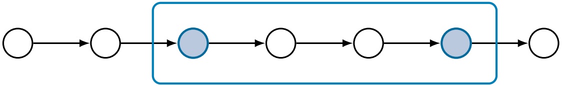
What is a TSL computation?

What does it mean to be local?

Local ⟶ finitely bounded



(window size 2)



(window size 4)

Long-distance (non-local) ⟶ no finite bound

Not just local, *strictly local*

Strictly local (SL): permitted/forbidden substrings of fixed size

Ex. Local assimilation

✓ a m p a p a n d a

✗ a n p a p a n d a

✗ a m p a p a m d a

Window:	2
Good:	nt, nd, mp, mb, ...
Bad:	*np, *nb, *mt, *md, ...

Cognitive interpretation: moving window of attention

McNaughton and Papert (1971) and Rogers et al. (2013)

Long-distance, but *local over a tier*

Tier-based strictly local (TSL): like SL, but *irrelevant elements are ignored*

Ex. Samala sibilant harmony (Heinz 2018)

✓ p i $\boxed{\text{s}}$ o t o n o $\boxed{\text{s}}$ i k i w a t

✗ p i $\boxed{\text{s}}$ o t o n o $\boxed{\text{ʃ}}$ i k i w a t

Window:	2
Visible elements:	s, ʃ
Constraints:	*sʃ, *ʃs

Also see: Heinz et al. (2011) and Lambert and Rogers (2020)

More about TSL

- SL is the **special case** of TSL where the tier contains everything
- TSL is **distinct** from autosegmental phonology (Goldsmith 1976)
  - Autosegmental tier: true multistratal representation
  - TSL: extra arcs in the basic string/tree representation

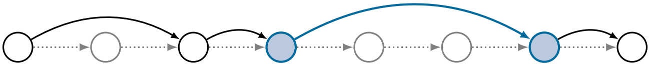


Figure 3: Two adjacent elements on a tier

Model-theoretic view of TSL: see Lambert et al. (2021) and Lambert (2023).

Invisibility and blocking

TSL with a window size of 2 (TSL-2) can handle both invisibility and blocking

Ex. Slovenian sibilant harmony (McMullin and Hansson 2016)

✓ ʃ p i ʃ

✓ n a s i t i ʃ

Window:	2
Visible elements:	s, ʃ, **t**
Constraints:	*sʃ, *ʃs

Also see McMullin (2016), Graf (2022b), and Hanson (to appear).

A TSL model of agreement

Locality of syntactic agreement

(1) Minimality (Rizzi 1990)

a. The cat$_{SG}$ **chases$_{SG}$** the rats.

b. * The cat **chase$_{PL}$** the rats$_{PL}$.

(2) Horizons (Keine 2019)

a. There **seem$_{PL}$** [$_{TP}$ to be some ducks$_{PL}$ in the garden].

b. * It **seem$_{PL}$** likely [$_{CP}$ **for** there to be some ducks$_{PL}$ in the garden].

Hanson (to appear):

The agreeing items must be adjacent on the relevant tier.

A unified model of locality

Long-distance dependencies are TSL-2 over their respective structures.

LD phonotactics	s … **t** … ʃ
Minimality (Rizzi 1990)	T … [**DP** … DP …]
Horizons (Keine 2019)	T … [$_{CP}$ … DP …]

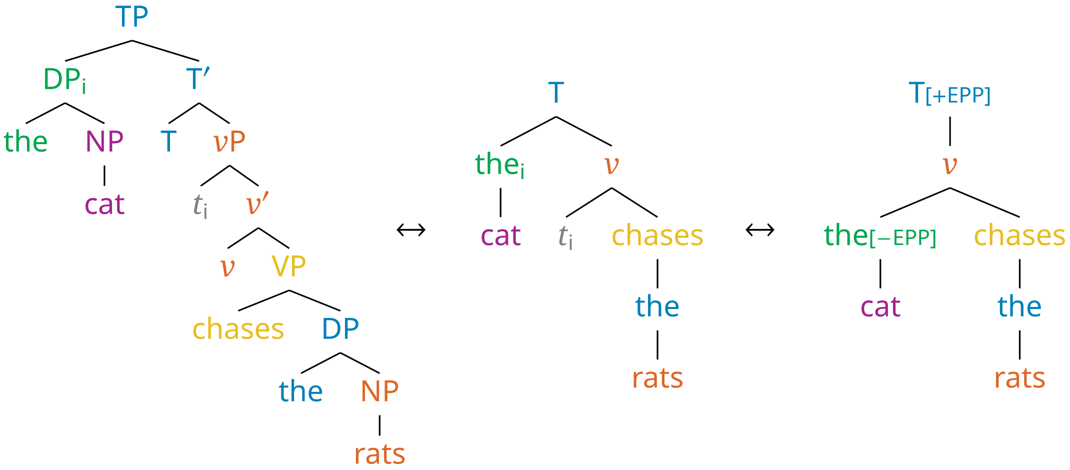
Setup

Key ingredients:

- Hierarchical representation
- Way to indicate long-distance dependencies
- Way to pick out the paths (=strings) which matter

Derivation trees

'The cat chases the rats.'



See Graf and Kostyszyn (2021). Related: Brody (2000).

Modeling long-distance dependencies

Diacritics mark items which *actually* move/agree/etc. in the current derivation, as in Minimalist Grammars (MGs, Stabler 1997, 2011).

Operation	Feature Type	Example
Move	Landing site	+EPP
	Mover	−EPP
Agree	Probe (valuee)	$+\phi$
	Goal (valuer)	$-\phi$

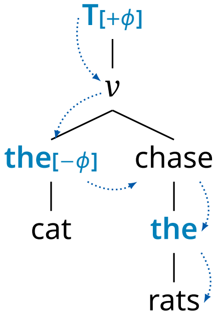
Agreement diacritics: see Hanson (to appear). Related: Ermolaeva and Kobele (2022).

Paths, spines

Individual syntactic dependencies don't make use of the entire tree at once.

What path (=string) do we need for agreement?

- Nodes ordered by **derivational prominence**
  $\approx$ order of last external Merge
  (Graf and Shafiei 2019)
  - Head < (Spec) < Comp
- At each branching point, follow the **complement spine** (Graf and De Santo 2019)



Also see: Kayne (1984), Uriagereka (1999), Adger (to appear), and Graf and Hanson (under review).

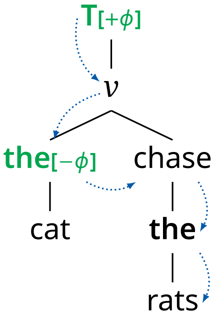
Tiers over spines

TSL grammar for English subject-verb agreement:

- **Tier contents:**
  all agreement participants (finite T, D) and blockers (C)
- **Tier constraints:**
  every ϕ-probe is immediately followed by a ϕ-goal, and vice versa

Minimality

✓ <u>The cat</u> **chases** the rats. (subject agreement)



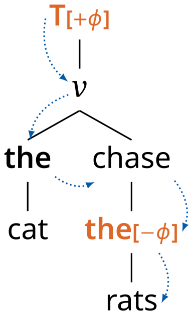
$\longrightarrow$ $\mathbf{T}_{[+\phi]} \cdot v \cdot \mathbf{the}_{[-\phi]} \cdot \text{chase} \cdot \mathbf{the} \cdot \text{rats}$

$\downarrow$

$\mathbf{T}_{[+\phi]} \cdot \mathbf{the}_{[-\phi]} \cdot \mathbf{the}$

Minimality (2)

✗ The cat **chase** the rats. (object agreement)



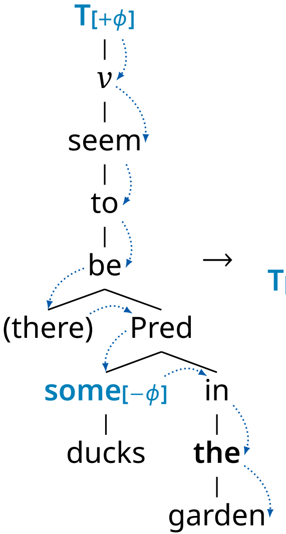
$$\longrightarrow \quad \mathbf{T}_{[+\phi]} \cdot v \cdot \mathbf{the} \cdot \text{chase} \cdot \mathbf{the}_{[-\phi]} \cdot \text{rats}$$

$$\downarrow$$

$$\mathbf{T}_{[+\phi]} \cdot \mathbf{the} \cdot \mathbf{the}_{[-\phi]}$$

Invisibility

✓ There **seem** to be <u>some ducks</u> in the garden. (no blockers)



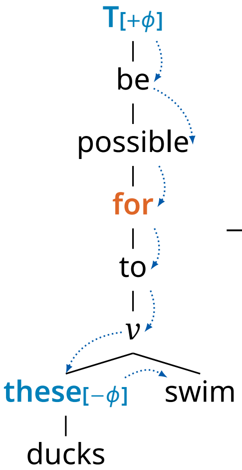
$\rightarrow$ $\mathbf{T_{[+\phi]}} \cdot v \cdot \ldots \cdot \mathrm{Pred} \cdot \mathbf{some_{[-\phi]}} \cdot \mathrm{ducks} \cdot \mathrm{in} \cdot \mathbf{the} \cdot \mathrm{garden}$

$\downarrow$

$\mathbf{T_{[+\phi]}} \cdot \mathbf{some_{[-\phi]}} \cdot \mathbf{the}$

Blocking

✗ It **are** possible **for** these ducks to swim. (C is a blocker)



$\rightarrow$ $\mathbf{T}_{[+\phi]} \cdot$ be $\cdot$ possible $\cdot$ **for** $\cdot$ to $\cdot$ v $\cdot$ $\mathbf{these}_{[-\phi]} \cdot$ ducks $\cdot$ swim

$\downarrow$

$\mathbf{T}_{[+\phi]} \cdot$ **for** $\cdot$ $\mathbf{these}_{[-\phi]}$

TSL grammar for subject-verb agreement

Window:	2
Tier:	finite T, all D, all C
Constraints:	$\left\{ \begin{array}{ll} \mathrm{X}_{[+\phi]} \cdot \mathrm{Y}_{[+\phi]} & \mathrm{X}_{[-\phi]} \cdot \mathrm{Y}_{[-\phi]} \\ \mathrm{X}_{[+\phi]} \cdot \mathrm{Y} & \mathrm{X} \cdot \mathrm{Y}_{[-\phi]} \\ \mathrm{X}_{[+\phi]} \cdot \ltimes & \rtimes \cdot \mathrm{Y}_{[-\phi]} \end{array} \right\}$

Consequences for typology

A unified model of locality

TSL-2 captures the pattern of invisibility and blocking which is characteristic of long-distance linguistic dependencies.

LD phonotactics	s … **t** … ʃ
Minimality (Rizzi 1990)	$T_{[+\phi]}$ … **D** … $D_{[-\phi]}$
Horizons (Keine 2019)	$T_{[+\phi]}$ … **C** … $D_{[-\phi]}$
Islands (Ross 1967)	$C_{[+Wh]}$ … **because** … $D_{[-Wh]}$

Parameters of variation

In general, syntactic dependencies have different visibility conditions, and therefore require distinct tiers.

Dependency	Participants	Blockers
ϕ-agreement	T_{FIN}, all D	all C
EPP movement	T_{FIN}, all D, *there*	all C
wh-movement	C_{wh}, all *wh*-movers	if, because, ...

Table 1: Visibility conditions for English

See Preminger (2014), Deal (2015), and Keine (2019), a.o.

Parameters of variation (2)

The contents of a given tier also vary across languages.

Language	Participants	Blockers
English	T_{FIN}, all D	all C
Hindi	T_{FIN}, D[NOM]	all C, T_{INF}

Table 2: Verbal agreement in English vs. Hindi

Case-sensitive agreement

In Hindi, the verb agrees with the closest nominative argument, which may not be the subject.

(3) Hindi verbal agreement ignores ergatives (Mahajan 1990)

a.

Raam	roṭii	khaat-**aa**	th-**aa**.
Raam.M.**NOM**	bread.F.NOM	eat-IPFV.M	be-PST.M

‘Raam ate bread (habitually).’

b.

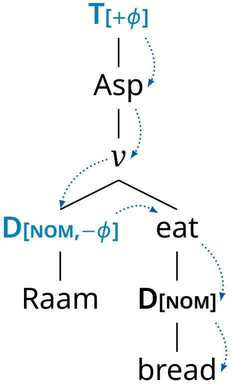
Raam-ne	roṭii	khaay-**ii**.
Raam.M-**ERG**	bread.F.NOM	eat-PFV.F

‘Raam ate bread.’

Analysis: Project D only if nominative. Tier constraints are unchanged.

Case-sensitive agreement (2)

‘Raam ate bread (habitually).’ (nominative subject, subject agrees)



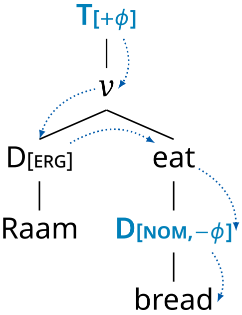
$\rightarrow$ $\mathbf{T}_{[+\phi]} \cdot \text{Asp} \cdot v \cdot \mathbf{D}_{[\text{NOM},-\phi]} \cdot \text{eat} \cdot \mathbf{D}_{[\text{NOM}]} \cdot \text{bread}$

$\downarrow$

$\mathbf{T}_{[+\phi]} \cdot \mathbf{D}_{[\text{NOM},-\phi]} \cdot \mathbf{D}_{[\text{NOM}]}$

Case-sensitive agreement (3)

‘Raam ate bread.’ (ergative subject, object agrees)



$\longrightarrow$ $\mathbf{T}_{[+\phi]} \cdot v \cdot \mathrm{D}_{[\text{ERG}]} \cdot \text{eat} \cdot \mathbf{D}_{[\text{NOM},-\phi]} \cdot \text{bread}$

$\downarrow$

$\mathbf{T}_{[+\phi]} \cdot \mathbf{D}_{[\text{NOM},-\phi]}$

Long-distance agreement

In Hindi, agreement can extend into *v*P, but not TP (Keine 2019).

(4) Hindi LDA and default agreement (Bhatt 2005)

a. Ram-ne [$_{vP}$ <u>roṭii</u> khaa-**nii**] chaah-**ii**
   Ram-ERG bread.F eat-INF.F want-PFV.FSG
   'Ram wanted to eat bread.'

b. Ram-ne [$_{TP}$ <u>roṭii</u> khaa-**naa**] chaah-**aa**
   Ram-ERG bread.F eat-INF.M want-PFV.MSG
   'Ram wanted to eat bread.'

Analysis: T_{INF} appears on the tier, but *v* does not.

Limits of variation

Ergative DPs are not necessarily invisible:

- In Nepali, both ergatives and nominatives agree (Preminger 2014)
- In some languages, only ergatives can agree (Baker 2015)

There are some cross-linguistic tendencies:

- Case-visibility hierarchy (Bobaljik 2008):
  Unmarked (Nom/Abs) < Marked (Erg/Acc) < Oblique (Dat, Gen)
- Height-Locality Connection (Keine 2019):
  Higher position of probe $\leftrightarrow$ Fewer possible horizons

When we consider the full range of variation, TSL-2 gets the right fit.

More parameters

We can modify not just the tier contents, but also the constraints.

Phenomenon	Change to baseline grammar
Invisibility	Some D's fail to project
Blocking	Some non-agreeing items project
Upward agreement	Swap order of $+\phi/-\phi$
Chain agreement	Allow sequential $+\phi$
Multiple agreement	Allow sequential $-\phi$
Independent subfeatures of ϕ	Each probe gets its own tier/constraints

Table 3: Parameters for agreement (Hanson to appear)

These are extremely similar to the parameters for long-distance harmony!

More parameters (2)

Even seemingly complex agreement patterns can be handled:

- Upward complementizer agreement (Lubukusu, Diercks 2013)
  $\longrightarrow$ Only project EPP-movers, $-\phi$ precedes $+\phi$

$$\mathrm{D}_{[-\mathrm{EPP},-\phi]} \;\ldots\; \mathrm{C}_{[+\phi]}$$

- A′-agreement (Dinka, Van Urk 2015)
  $\longrightarrow$ Only project A′-movers, let $+\phi$ iterate

$$\mathrm{C}_{[+\mathrm{A}',+\phi]} \;\ldots\; \mathrm{C}_{[+\mathrm{A}',+\phi]} \;\ldots\; \mathrm{D}_{[-\mathrm{A}',-\phi]}$$

- Parasitic agreement (Hindi, Bhatt 2005)
  $\longrightarrow$ Second "concord tier" allows parasitic elements between $+\phi$ and $-\phi$

$$\mathrm{T}_{[+\phi]} \;\ldots\; v_{[+\phi]} \;\ldots\; v_{[+\phi]} \;\ldots\; \mathrm{D}_{[\textsc{nom},-\phi]}$$

See appendix for details.

Impossible patterns

Logically possible patterns which are not TSL-2 are generally unattested.

Type	Class	Example	Visible Cs
Unbounded	TSL-2	Aari	Only sibilants
LD w/ blocking	TSL-2	Slovenian	All coronals
Transvocalic	TSL-2	Koyra	All consonants
At most 1 C intervenes	TSL-3	Unattested	—
Exactly 1 C intervenes	TSL-3	Unattested	—
At least 1 C intervenes	TLT/OTSL	Unattested	—

Table 4: Typology of consonant harmony (adapted from McMullin and Hansson 2016)

Impossible patterns (2)

Non-TSL-2 and unattested island types (Graf 2022b):

- Gang-up islands: A mover can escape n islands, but not $n + 1$.
- Cowardly islands: XP is an island iff there are at least n XPs in the same clause.
- ...

Non-TSL-2 and unattested verbal agreement patterns:

- T agrees with the subject unless there is a temporal adjunct, in which case it agrees with the object.
- Only DPs which contain a relative clause which contains two PPs can agree.
- ...

The big picture

- TSL computations are **extremely restricted**
  - No boolean logic
  - No counting of violations
  - Tier contents determined solely by the element labels
- Observed typology emerges from this **highly structured** logical space
- Syntax is revealed to be **much more similar** to phonology and morphology than previously thought (Graf 2022a)
  - Different structures, different features, but same computations

$$\text{FIN} \subseteq \text{SL} \subseteq \boxed{\text{TSL}} \subseteq \ldots \subseteq \text{REG} \subseteq \ldots \subseteq \text{CFL} \subseteq \ldots \subseteq \text{CS} \subseteq \ldots \subseteq \text{RE}$$

Related and ongoing work

Related and ongoing work

Islands and gradient acceptability

- We have been treating islands and other blockers as being **categorical**
- **Experimental studies** have found that judgments of island violations are often **gradient** (Chaves 2022)
- Can this be incorporated into the TSL model?

Probabilistic TSL

Gradient blocking be captured via **probabilistic tier projection** (Mayer 2021).

Torres, Hanson, Graf, and Mayer (2023):
Applied pTSL to syntactic islands, modeling data from Sprouse et al. (2016)

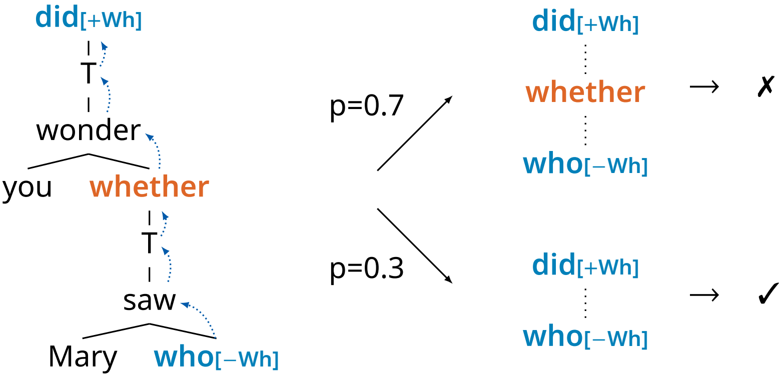
Charles Torres
UC Irvine

Connor Mayer
UC Irvine

Thomas Graf
Stony Brook

Probabilistic TSL: Example

'Who did you wonder whether Mary saw?'



$$
\begin{aligned}
Score(tree) &= p(tier1) * licit(tier1) + p(tier2) * licit(tier2) \\
&= 0.7 * 0 + 0.3 * 1 \\
&= 0.3
\end{aligned}
$$

Probabilistic TSL: Modeling Study

- Data from Sprouse et al. (2016)
- Three island types and four conditions

$$\left\{\begin{array}{c} \textit{whether}\text{ island} \\ \text{adjunct island} \\ \text{complex NP island} \end{array}\right\} \times \left\{\begin{array}{c} \text{non-island, matrix clause} \\ \text{non-island, embedded clause} \\ \text{island, matrix clause} \\ \text{island, embedded} \end{array}\right\}$$

- Provided a structural analysis for each sentence
- Fit projection probability of island nodes to approximate mean Likert rating for each sentence (z-score normalized and transformed to range $[0, 1]$)
  - other nodes fixed to 1 (movers/landing sites) or 0 (others)

Probabilistic TSL: Results

Node	Projection probability
that	.46
complex NP	.63
whether	.73
adjunct	.89

Table 5: Mean projection probabilities of fitted pTSL island model

- Projection probabilities mirror relative badness of each island type
- Replicated **superadditivity effect** – extraction from an island is worse than (i) extraction from non-island plus (ii) mere presence of an island
- Model does not capture relative badness of matrix clause movement

Probabilistic TSL: Discussion

- The TSL model is perfectly compatible with gradience in the grammar
  - categorical tier projection $\leftrightarrow$ categorical output
  - probabilistic tier projection $\leftrightarrow$ gradient output
- Further extensions:
  - Can add probabilities to the constraints to produce a stochastic grammar, which can be fit to corpus data in the same manner as a PCFG

Related and ongoing work

Other syntactic dependencies

Local	Selection	Graf (2018); Hanson (2023b)
	Functional hierarchies	Hanson (2023b)
	Adjunction	Hanson (under review)
Long-distance	**Movement**	Graf (2018, 2022b)
	Case	Vu et al. (2019); Hanson (2023a)
	Agreement	Hanson (to appear)

Functional hierarchies

Hanson (2023b): modeling local dependencies with spines

- English clausal hierarchy (Adger 2003):
  T < (Neg) < (Perf) < (Prog) < (Pass) < v
- ex. 'The cake might not have been being eaten.'
- **TSL grammar**
  Tier elements: all symbols
  Window size: 2
  Good substrings:

T Neg				
T Perf	**Neg Perf**			
T Prog	Neg Prog	**Perf Prog**		
T Pass	Neg Pass	Perf Pass	**Prog Pass**	
T v	Neg v	Perf v	Prog v	**Pass v**

might$_{\text{T}}$
|
not$_{\text{Neg}}$
|
have$_{\text{Perf}}$
|
be$_{\text{Prog}}$
|
be$_{\text{Pass}}$
|
v
|
eat
|
the
|
cake

Case dependencies

- **Hanson (2023a):** in-depth analysis of case in Japanese*
  - Three tiers: verbal domain, nominal domain, lexical case
  - Valency alternations, nominative objects, long-distance case
- **Current research:** to what extent do the parameters of a TSL-2 model account for cross-linguistic variation in case assignment?
  - Case spreading / multiple assignment
  - Dependent case
  - Split alignment
  - Differential argument marking

*See appendix for an example.

Some open questions

- To what extent do the **same patterns** occur in different dependencies?
- Are **constraint interactions** across tiers always limited to intersection? (Meinhardt et al. 2024)
- How does the computational system interact with **other factors**, such as the feature system?
- Can we view the **mappings** from syntax to morphology/semantics as TSL functions? (cf. Graf 2023)

Wrapping up

Summary

- Linguistic structures can be built with TSL computations:
  local or long-distance, phonological or syntactic
- Syntax and phonology (partly) share a common computational basis
- Many new empirical puzzles can be identified
- The underlying mathematics connect theory, experiment, and more

By separating out the computational factors underlying language, we can learn many things that we could not otherwise!

Thank you for your attention!

Appendices

More agreement case studies

Upward agreement

(5) Complementizer Agreement in Lubukusu (Diercks 2013)

a. Ba-ba-ndu ba-bolela Alfredi ba-li a-kha-khile.
c2-c2-people c2-said c1.Alfred c2-that c1-FUT-conquer
'The people told Alfred that he will win.'

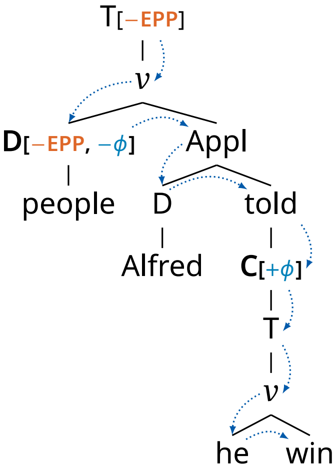
b. Alfredi ka-bolela ba-ba-ndu a-li ba-kha-khile.
c1.Alfred c1-said c2-c2-people c1-that c2-FUT-conquer
'Alfred told the people that they will win.'

Analysis:

- Agreement is upward $\longrightarrow$ allow $\mathrm{D}_{[-\phi]}$ to precede $\mathrm{C}_{[+\phi]}$
- Agreement on C is subject oriented $\longrightarrow$ project only DPs bearing –EPP

Upward agreement (2)

‘The people told Alfred that he will win.’



$$\longrightarrow \quad \mathrm{T}_{[-\mathrm{EPP}]} \cdot v \cdot \mathbf{D}_{[-\mathrm{EPP},\ -\phi]} \cdot \mathrm{Appl} \cdot \mathrm{D} \cdot \mathrm{told} \cdot \mathbf{C}_{[+\phi]} \cdot \ldots$$

$$\downarrow$$

$$\mathbf{D}_{[-\mathrm{EPP},\ -\phi]} \cdot \mathbf{C}_{[+\phi]}$$

Analysis: Project DPs only if [−EPP]. Allow $\mathrm{D}_{[-\phi]}$ to precede $\mathrm{C}_{[+\phi]}$.

Syntactic counterfeeding

Agreeing C is impossible in hyperraising structures.

(6) Agreeing complementizer incompatible with hyperraising

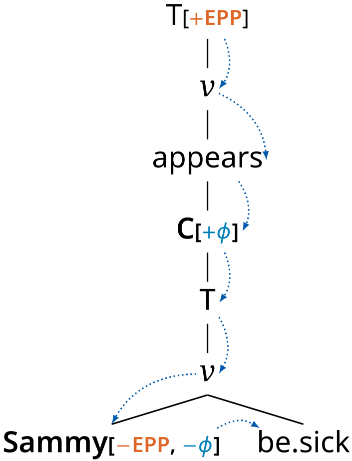
Sammy	a-lolekhana	**mbo**	(*a-li)	a-likho	a-lwala.
c1.Sammy	c1-appears	**that**	(*c1-that)	c1.PROG	c1-be.sick

‘Sammy appears to be sick.’ (lit. ‘Sammy seems that is sick.’)

This follows immediately from the TSL analysis!

Syntactic counterfeeding (2)

The hyperraised subject is below C in the derivation tree and cannot agree.



$$\longrightarrow \quad \mathrm{T}_{[+\mathrm{EPP}]} \cdot v \cdot \mathrm{V} \cdot \mathbf{C}_{[+\phi]} \cdot \mathrm{T} \cdot v \cdot \mathbf{D}_{[-\mathrm{EPP},\, -\phi]} \cdot \mathrm{V}$$

$$\downarrow$$

$$* \ \mathbf{C}_{[+\phi]} \cdot \mathbf{D}_{[-\mathrm{EPP},\, -\phi]}$$

(should be $\mathbf{D}_{[-\mathrm{EPP},\, -\phi]} \cdot \mathbf{C}_{[+\phi]}$)

A′ agreement

(7) Dinka verbal agreement with Spec-CP

a. Mòc à-cé yîin tîiŋ.
man 3SG-PRF.SV you see.NF
'The man has seen you.' (Van Urk 2015, ch. 4, 19b)

b. Yîin ∅-cíi môc tîiŋ.
you 2-PRF.OV man.GEN see.NF
'You, the man has seen.' (Van Urk 2015, ch. 4, 20a)

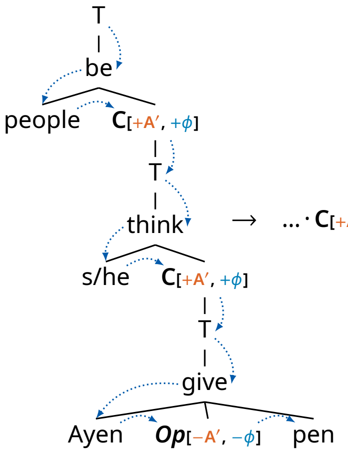
(8) Agreement in both matrix and embedded clause

Yè kɔ̂ɔc-kó [$_{\text{CP}}$ *Op* é-kè-yá ké tàak [$_{\text{CP}}$ è ___
be **people**.CS-which PST-**PL**-HAB.2SG 3PL think.NF C
é-kè-cíi Áyèn ké gâam gàlàm]]?
PST-**PL**-PRF.OV Ayen.GEN 3PL give.NF pen

'Which people did (s)he think that Ayen had given a pen to?'
(Van Urk 2015, ch. 5, 14a)

A′ agreement (2)

‘Which people did (s)he think that Ayen had given a pen to?’



→ … · **C**$_{[+A', +\phi]}$ · T · think · s/he · **C**$_{[+A', +\phi]}$ · T · give · Ayen · ***Op***$_{[-A', -\phi]}$ · pen

↓

C$_{[+A', +\phi]}$ · **C**$_{[+A', +\phi]}$ · ***Op***$_{[-A', -\phi]}$

Analysis: Project all D bearing [–A′]. Allow C$_{[+\phi]}$ to iterate.

Parasitic agreement

(9) Hindi participles and infinitives agree iff the main verb does

a. Shahrukh-ne [$_{\nu P}$ ṭehnii kaaṭ-**nii**] chaah-**ii** **thii**
   Shahrukh-ERG branch.F cut-**INF.F** want-**PFV.F** **be.PST.FSG**
   'Shahrukh had wanted to cut the branch.'

b. Shahrukh-ne [$_{TP}$ ṭehnii kaaṭ-**naa**] chaah-**aa** **thaa**
   Shahrukh-ERG branch.F cut-**INF.M** want-**PFV.MSG** **be.PST.MSG**
   'Shahrukh wanted to cut a/the branch.'

Analysis: Verbs agree iff they occur along a chain from $T_{[+\phi]}$ to $D_{[-\phi]}$.

Complication:
T-agreement and V-agreement must be regulated on separate tiers.

Parasitic agreement (2)

If T agrees with DP, then all verbs along the path also agree.

T-agreement tier: $\mathrm{T}_{\mathrm{FIN}[+\phi]}$ $\mathrm{Aux}_{[+\phi]}$ $v_{[+\phi]}$ $\mathrm{D}_{[\mathrm{ERG}]}$ V $v_{[+\phi]}$ V $\mathrm{D}_{[\mathrm{NOM},-\phi]}$

V-agreement tier: $\mathrm{T}_{\mathrm{FIN}[+\phi]}$ $\mathrm{Aux}_{[+\phi]}$ $v_{[+\phi]}$ $\mathrm{D}_{[\mathrm{ERG}]}$ V $v_{[+\phi]}$ V $\mathrm{D}_{[\mathrm{NOM},-\phi]}$

Why isn't one tier enough?

- Agreement can fail, so non-agreeing pairs $\mathrm{T} \cdot \mathrm{Aux}$, $\mathrm{Aux} \cdot v$, etc., must be allowed.
- Once you do this, agreement is incorrectly predicted to always be optional.

Learning tiers from positive data

Learnability considerations

TSL constraints are easy to learn if the tier is already known (Lambert et al. 2021)...

But identifying the tier itself is not trivial

- If there are n symbols, then there are 2^n possible tiers

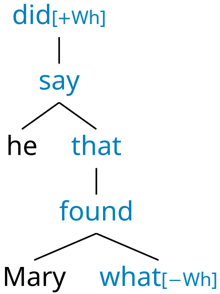
Current formal learners can do efficiently this for pure TSL-2 languages (Jardine and McMullin 2017; Lambert 2021)...

But natural language always involves intersecting tiers

Learning tiers from paths

Heuser et al. (2024): track *licit bigrams along a movement path*, and generalize to categories when allowed by the Tolerance Principle (Yang 2016)

Hanson (2024): this algorithm already tracks the information needed by a formal TSL learner → we can adapt it to produce a TSL-2 grammar!



$$\rightarrow \left\{ \begin{array}{c} \text{did} \cdot \text{say} \\ \text{say} \cdot \text{that} \\ \text{that} \cdot \text{found} \\ \text{found} \cdot \text{what} \end{array} \right\}$$

See appendix for details.

Starting point

Let's assume that:

- The learner already knows the basics of constituency and selection
- The learner can identify the dependent elements,
  e.g. mover and landing site, probe and goal

The dependent elements must appear on the tier, and must be adjacent

Next step: identify the blockers

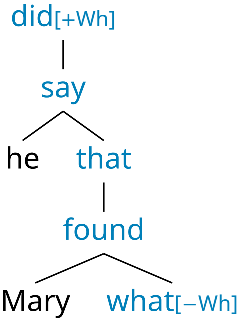
cf. Clark and Yoshinaka 2016; Liang et al. 2022; Li and Schuler 2023

Blockers for *wh*-movement

- ✓ **What** did Mary find ___?
- ✓ **What** did he say that Mary found ___?
- ✗ **What** did he wonder **whether** Mary found ___?
- ✗ **What** did he **mutter** that Mary found ___?

Learning blockers from paths

Heuser et al. (2024): learn movement blockers by tracking **licit bigrams along a movement path** (plus generalization to categories as in Yang 2016)



$$\rightarrow \left\{ \begin{array}{c} \text{did} \cdot \text{say} \\ \text{say} \cdot \text{that} \\ \text{that} \cdot \text{found} \\ \text{found} \cdot \text{what} \end{array} \right\}$$

Key point: Bigrams like {wonder · **whether**, **whether** · found, ...} will not be attested along any *wh*-movement path

Learning blockers from paths (2)

Hanson (2024): the path grammar contains everything we need to construct a TSL-2 grammar

- Extract the elements from the set of path bigrams for movement type X
- The elements which are conspicuously missing are just the tier blockers

$$\left\{\begin{array}{c}\text{that}\\\text{whether}\\\text{say}\\\text{mutter}\\\ldots\end{array}\right\} - \left\{\begin{array}{c}\text{that}\\\text{say}\\\ldots\end{array}\right\} = \left\{\begin{array}{c}\text{whether}\\\text{mutter}\\\ldots\end{array}\right\}$$

Learning blockers from paths (3)

Advantages to the combined system:

- Combines the typological merits of TSL-2 with an empirically-motivated acquisition model
- Produces a version of the Height-Locality Connection (Keine 2019), which is not inherent to the TSL-2 model
- Could be adapted for learning constraints on agreement

Summary

- The space of possible tiers is too large to search exhaustively, but we don't actually need to do this
- When learning non-local constraints, the local constraints should be factored out
- Some empirical generalizations may derive from the details learning process rather than the grammar formalism itself

Avoiding overgeneration in the category system

Collaborators

Logan Swanson
Stony Brook

Thomas Graf
Stony Brook

The problem with subcategorization

- There is a LI of category V that selects a CP
- ...only if that CP contains an AP
- ...and also contains three VPs, none of which contain one another
- ...but only if the LI does not also select a PP
- ...in which case the three VPs must instead be self-embedded
- ...and CP must also contain a non-finite TP
- ...

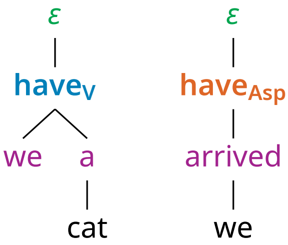
Without further restrictions, any system of category selection can do this

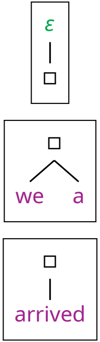
Reining in category features

- Any regular tree constraint can be indirectly implemented via selection
- Selection is SL-2 over trees
  - $\longrightarrow$ Massive overgeneration
  - $\longrightarrow$ All subregular distinctions are lost
- **Conjecture:** category and subcategorization features must be inferable from the local tree context $\longrightarrow$ **ISL-recoverable** (Graf 2020)
- **This project:** test this conjecture against MGBank (Torr 2017, 2018)

ISL-Recoverability

It should be possible to infer the category of every lexical item based only on the phonological features of its **selector** or its **selectees** (or maybe both).





$\rightarrow$ Non-disambiguating

$\rightarrow$ Context for $\textbf{have}_V$

$\rightarrow$ Context for $\textbf{have}_{Asp}$

Results

Forms with ambiguous category:	8369
ISL-2 recoverable in all contexts:	79–86%
ISL-2 recoverable in at least one context:	95–97%

- Most errors are due to empty heads or misparsed sentences (good!)
- Subcategorization features are recoverable at a similar rate (good!)
- Movement features are recoverable at a similar rate (not good!)

Discussion

- It should not be possible to accurately identify all movement features from the local context
- Simulations suggest that a Zipfian distribution of feature specifications over phonetic exponents produces similar quantitative results
  $\longrightarrow$ ISL inferrability might be epiphenomenal
- In any case, local inferrability makes a good heuristic for learning: identical items in the same context must have the same category

Miscellaneous

What makes a possible tier?

We know that the tier elements can be fairly arbitrary, but they can't be completely arbitrary.

- $T = \{\text{the}_{\text{[Nom]}}, \text{all}_{\text{[Acc]}}, \text{for}_C, \text{smile}_V, \text{already}_{\text{Adv}}, \ldots\}$
- $\bar{T} = \{\text{the}_{\text{[Acc]}}, \text{all}_{\text{[Nom]}}, \text{that}_C, \text{wash}_V, \ldots\}$

Usually, we want entire classes of items:

- all D, all $\text{D}_{\text{[NOM]}}$, all $\text{D}_{[-\text{EPP}]}$, ...

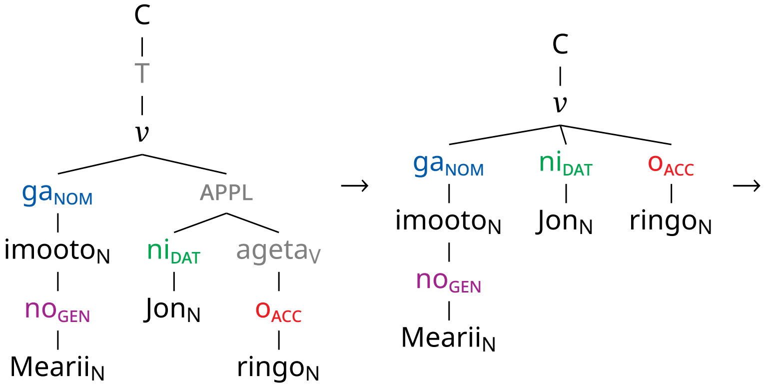
A couple of possibilities:

- a tier must be defined in terms of natural classes
  (insert your favorite theory of syntactic features here)
- unnatural tiers are never posited by the learner under realistic conditions

Case in Japanese

Hanson (2023a): in-depth analysis of case in Japanese, relativizing both dominance and precedence to a tier $\longrightarrow$ **tree tier**

(10) Mearii no imooto ga Jon ni ringo o ageta.
Mary GEN sister NOM John DAT apple ACC gave
'Mary's sister gave John an apple.'



Dtrs of C:	ε	✓
Dtrs of v:	ga · ni · o	✓
Dtrs of N:	no / ε	✓

References

Adger, David (2003). *Core syntax: A Minimalist approach*. Vol. 20. Core Linguistics. Oxford University Press.

Adger, David (to appear). *Mereological syntax: phrase structure, cyclicity, and islands*. Cambridge, MA: MIT Press.

Baker, Mark (2015). *Case*. Cambridge University Press. DOI: 10.1145/1463891.1464000.

Belth, Caleb (2023). Towards a Learning-Based Account of Underlying Forms: A Case Study in Turkish. *Proceedings of the Society for Computation in Linguistics 2023*. Amherst, MA: University of Massachusetts Amherst, pp. 332–342. DOI: 10.7275/BC8Q-VJ22.

Bhatt, Rajesh (2005). Long Distance Agreement in Hindi-Urdu. *Natural Language & Linguistic Theory* 23, pp. 757–807.

Bobaljik, Jonathan (2008). Where's Phi? Agreement as a Postsyntactic Operation. *Phi Theory*. Ed. by Daniel Harbour, David Adger, and Susana Béjar. Oxford: Oxford University Press, May 1, pp. 295–328. DOI: 10.1093/oso/9780199213764.003.0010.

Brody, Michael (2000). Mirror theory: Syntactic representation in perfect syntax. *Linguistic Inquiry* 31.1, pp. 29–56.

Burness, Phillip Alexander et al. (2021). Long-distance phonological processes as tier-based strictly local functions. *Glossa: a journal of general linguistics* 6.1. DOI: 10.16995/glossa.5780.

Chandlee, Jane and Jeffrey Heinz (2018). Strict Locality and Phonological Maps. *Linguistic Inquiry* 49.1, pp. 23–60. URL: https://muse.jhu.edu/article/683698.

References (2)

Chaves, Rui P. (2022). Sources of Discreteness and Gradience in Island Effects. *Languages* 7.4, p. 245.

Chomsky, Noam (1957). *Syntactic structures*. De Gruyter Mouton.

Chomsky, Noam (1959). On certain formal properties of grammars. *Information and control* 2.2, pp. 137–167.

Clark, Alexander and Ryo Yoshinaka (2016). Distributional learning of context-free and multiple context-free grammars. *Topics in Grammatical Inference*, pp. 143–172.

Deal, Amy Rose (2015). Interaction and satisfaction in ϕ-agreement. *Proceedings of NELS 45*. Amherst, MA: GLSA Publications, pp. 179–192.

Diercks, Michael (2013). Indirect agree in Lubukusu complementizer agreement. *Natural Language & Linguistic Theory* 31.2, pp. 357–407.

Ermolaeva, Marina and Gregory M. Kobele (2022). Agree as information transmission over dependencies. *Syntax* 25.4, pp. 466–507. DOI: `https://doi.org/10.1111/synt.12240`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/synt.12240`.

Goldsmith, John (1976). "Autosegmental phonology". PhD thesis. Massachusetts Institute of Technology.

Graf, Thomas (2018). Why movement comes for free once you have adjunction. *Proceedings of CLS 53*. Chicago, IL: The Chicago Linguistic Society, pp. 117–136. URL: `https://thomasgraf.net/output/graf18cls.html`.

References (3)

Graf, Thomas (2020). Curbing Feature Coding: Strictly Local Feature Assignment. *Proceedings of the Society for Computation in Linguistics 2020*. Amherst, MA: University of Massachusetts Amherst, pp. 224–233. DOI: 10.7275/f7y5-xz32.

Graf, Thomas (2022a). Subregular linguistics: bridging theoretical linguistics and formal grammar. *Theoretical Linguistics* 48.3–4, pp. 145–184. DOI: 10.1515/tl-2022-2037.

Graf, Thomas (2022b). Typological implications of tier-based strictly local movement. *Proceedings of the Society for Computation in Linguistics 2022*. Amherst, MA: University of Massachusetts Amherst, pp. 184–193. DOI: 10.7275/gb65-ht31.

Graf, Thomas (2023). Subregular Tree Transductions, Movement, Copies, Traces, and the Ban on Improper Movement. *Proceedings of the Society for Computation in Linguistics 2023*. Amherst, MA: University of Massachusetts Amherst, pp. 289–299. DOI: 10.7275/TK1N-Q855.

Graf, Thomas and Aniello De Santo (2019). Sensing Tree Automata as a Model of Syntactic Dependencies. *Proceedings of the 16th Meeting on the Mathematics of Language*. Toronto, Canada: Association for Computational Linguistics, pp. 12–26. URL: https://www.aclweb.org/anthology/W19-5702.

Graf, Thomas and Kenneth Hanson (under review). "Syntax with strings attached".

Graf, Thomas and Kalina Kostyszyn (2021). Multiple Wh-Movement is not Special: The Subregular Complexity of Persistent Features in Minimalist Grammars. *Proceedings of the Society for Computation in Linguistics 2021*. Amherst, MA: University of Massachusetts Amherst, pp. 275–285. URL: https://aclanthology.org/2021.scil-1.25.

References (4)

Graf, Thomas and Nazila Shafiei (2019). C-command dependencies as TSL string constraints. *Proceedings of the Society for Computation in Linguistics 2019*. Amherst, MA: University of Massachusetts Amherst, pp. 205–215. URL: https://aclanthology.org/W19-0121.

Hanson, Kenneth (2023a). A TSL Analysis of Japanese Case. *Proceedings of the Society for Computation in Linguistics 2023*. Amherst, MA: University of Massachusetts Amherst Libraries, pp. 15–24. DOI: 10.7275/xqhr-r404.

Hanson, Kenneth (2023b). Strict Locality in Syntax. *Proceedings of CLS 59*. Ed. by Kutay Serova and M. K. Snigaroff. Chicago, IL: The Chicago Linguistic Society, pp. 131–145. URL: https://www.kennethhanson.net/files/hanson-cls59-sl-in-syntax.pdf.

Hanson, Kenneth (to appear). Tier-Based Strict Locality and the Typology of Agreement. *Journal of Language Modeling*. URL: https://www.kennethhanson.net/files/hanson-jlm-tsl-agreement-preprint.pdf.

Hanson, Kenneth (2024). Tiers, Paths, and Syntactic Locality: The View from Learning. *Proceedings of the Society for Computation in Linguistics 2024*. Amherst, MA: University of Massachusetts Amherst Libraries, pp. 107–116. DOI: 10.7275/scil.2135.

Hanson, Kenneth (under review). “Adjunction in (T)SL Syntax”.

Heinz, Jeffrey (2018). The computational nature of phonological generalizations. *Phonological Typology*. Ed. by Larry M. Hyman and Frans Plank. Phonetics and Phonology 23. De Gruyter Mouton, pp. 126–195. DOI: 10.1515/9783110451931-005.

References (5)

Heinz, Jeffrey, Chetan Rawal, and Herbert G. Tanner (2011). Tier-based strictly local constraints for phonology. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human language technologies*, pp. 58–64.

Heuser, Annika, Hector Vazquez Martinez, and Charles Yang (2024). *The Learnability of Syntactic Islands*. Presentation at NELS 54, MIT.

Huybregts, Riny (1984). The weak inadequacy of context-free phrase structure grammars. *Van periferie naar kern*. Ed. by G. J. de Haan, M. Trommelen, and W. Zonneveld. Dordrecht: Foris, pp. 81–99.

Jardine, Adam and Kevin McMullin (2017). Efficient Learning of Tier-Based Strictly k-Local Languages. *Language and Automata Theory and Applications*. Ed. by Frank Drewes, Carlos Martín-Vide, and Bianca Truthe. Cham: Springer, pp. 64–76.

Kaplan, Ronald M. and Martin Kay (1994). Regular models of phonological rule systems. *Computational linguistics* 20.3, pp. 331–378. URL: http://www.aclweb.org/anthology/J94-3001.pdf.

Kayne, Richard S. (1984). *Connectedness and binary branching*. De Gruyter, Dec. DOI: 10.1515/9783111682228.

Keine, Stefan (2019). Selective Opacity. *Linguistic Inquiry* 50.1, pp. 13–62. DOI: 10.1162/ling_a_00299.

References (6)

Lambert, Dakotah (2021). Grammar Interpretations and Learning TSL Online. *Proceedings of the Fifteenth International Conference on Grammatical Inference*. Ed. by Jane Chandlee et al. Vol. 153. Proceedings of Machine Learning Research. PMLR, pp. 81–91. URL: `https://proceedings.mlr.press/v153/lambert21a.html`.

Lambert, Dakotah (2023). Relativized Adjacency. *Journal of Logic, Language and Information* 32.4, pp. 707–731. DOI: `10.1007/s10849-023-09398-x`.

Lambert, Dakotah, Jonathan Rawski, and Jeffrey Heinz (2021). Typology emerges from simplicity in representations and learning. *Journal of Language Modelling* 9.1, pp. 151–194. DOI: `10.15398/jlm.v9i1.262`.

Lambert, Dakotah and James Rogers (2020). Tier-Based Strictly Local Stringsets: Perspectives from Model and Automata Theory. *Proceedings of the Society for Computation in Linguistics 2020*. Amherst, MA: University of Massachusetts Amherst, pp. 159–166. URL: `https://aclanthology.org/2020.scil-1.21`.

Li, Daoxin and Kathryn D. Schuler (2023). Acquiring recursive structures through distributional learning. *BUCLD 47: Proceedings of the 47th Annual Boston University Conference on Language Development*.

Liang, Kevin, Diana Marsala, and Charles Yang (2022). Distributional Learning of Syntactic Categories. *BUCLD 46: Proceedings of the 46th annual Boston University Conference on Language Development*.

Mahajan, Anoop Kumar (1990). "The A/A-bar distinction and movement theory". PhD thesis. Cambridge, MA: Massachusetts Institute of Technology.

References (7)

Mayer, Connor (2021). Capturing gradience in long-distance phonology using probabilistic tier-based strictly local grammars. *Proceedings of the Society for Computation in Linguistics 2021*. Amherst, MA: University of Massachusetts Amherst, pp. 39–50. URL: `https://aclanthology.org/2021.scil-1.4`.

McMullin, Kevin (2016). "Tier-based locality in long-distance phonotactics: learnability and typology". PhD thesis. University of British Columbia.

McMullin, Kevin and Gunnar Ólafur Hansson (2016). Long-Distance Phonotactics as Tier-Based Strictly 2-Local Languages. *Proceedings of the Annual Meetings on Phonology*. Vol. 2. Washington, DC: Linguistic Society of America. DOI: `10.3765/amp.v2i0.3750`.

McNaughton, Robert and Seymour A. Papert (1971). *Counter-Free Automata*. MIT Press.

Meinhardt, Eric et al. (2024). Weak determinism and the computational consequences of interaction. *Natural Language & Linguistic Theory*. DOI: `10.1007/s11049-023-09578-1`.

Preminger, Omer (2014). *Agreement and its failures*. Vol. 68. Linguistic Inquiry Monographs. Cambridge, MA: MIT Press.

Rizzi, Luigi (1990). *Relativized minimality*. Linguistic Inquiry Monographs. Cambridge, MA: MIT Press.

Roark, Brian and Richard Sproat (2007). *Computational approaches to morphology and syntax*. Oxford: Oxford University Press.

Rogers, James et al. (2013). Cognitive and Sub-Regular Complexity. *Formal Grammar*. Ed. by Glyn Morrill and Mark-Jan Nederhof. Vol. 8036. Lecture Notes in Computer Science. Springer, pp. 90–108. DOI: `10.1007/978-3-642-39998-5_6`.

References (8)

Ross, John Robert (1967). "Constraints on variables in syntax". PhD thesis. Cambridge, MA: Massachusetts Institute of Technology.

Shieber, Stuart M. (1985). Evidence against the context-freeness of natural language. *The formal complexity of natural language*. Springer, pp. 320–334. URL: `http://dx.doi.org/10.1007/BF00630917`.

Sprouse, Jon et al. (2016). Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34.1, pp. 307–344. DOI: `10.1007/s11049-015-9286-8`.

Stabler, Edward P. (1997). Derivational minimalism. *Logical Aspects of Computational Linguistics*. Ed. by Christian Retore. Springer.

Stabler, Edward P. (2011). Computational perspectives on Minimalism. *Oxford handbook of linguistic Minimalism*. Ed. by Cedric Boeckx. Oxford: Oxford University Press, pp. 617–643.

Swanson, Logan, Kenneth Hanson, and Thomas Graf (under review). "Are syntactic categories ISL-2 inferrable? A corpus study".

Torr, John (2017). Autobank: a semi-automatic annotation tool for developing deep Minimalist Grammar treebanks. *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain: Association for Computational Linguistics, Apr., pp. 81–86. URL: `https://aclanthology.org/E17-3021`.

References (9)

Torr, John (2018). Constraining MGbank: Agreement, L-Selection and Supertagging in Minimalist Grammars. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, July, pp. 590–600. DOI: 10.18653/v1/P18-1055.

Torres, Charles et al. (2023). Modeling island effects with probabilistic tier-based strictly local grammars over trees. *Proceedings of the Society for Computation in Linguistics 2023*. Amherst, MA: University of Massachusetts Amherst Libraries, pp. 155–164. DOI: 10.7275/nz4q-6b09.

Uriagereka, Juan (1999). Multiple Spell-Out. *Working minimalism*. Ed. by Samuel D. Epstein and Norbert Hornstein. Cambridge, MA: MIT Press.

Van Urk, Coppe (2015). “A Uniform Syntax for Phrasal Movement: A Dinka Bor Case Study”. PhD thesis. Cambridge, MA: Massachusetts Institute of Technology.

Vu, Mai Ha, Nazila Shafiei, and Thomas Graf (2019). Case assignment in TSL syntax: A case study. *Proceedings of the Society for Computation in Linguistics 2019*. Amherst, MA: University of Massachusetts Amherst, pp. 267–276. URL: https://aclanthology.org/W19-0127.

Yang, Charles (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.