Tiers, Paths, and Syntactic Locality: The View from Learning

Kenneth Hanson • Dept. of Linguistics, IACS • Stony Brook University
kenneth.hanson@stonybrook.edu • https://www.kennethhanson.net

Link to Paper

Overview

Grammar Most long-distance linguistic patterns are **tier-based strictly 2-local (TSL-2)** (McMullin 2016; Graf 2022; Hanson 2024).

Learning Pure TSL string languages are **efficiently learnable**, but multiple tiers present various difficulties. Heuser et al. (2024) propose an algorithm which learns syntactic blockers as **local constraints on paths**.

This Work Adapt the insights of Heuser et al. (2024) to produce an efficient TSL-2 learner which **factors out local constraints**.

Bonus We can derive a version of the **Height-Locality Connection** (cf. Keine 2019) which is not an inherent prediction of TSL-2.

Tiers Over a-Strings

TSL in a nutshell

1. Ignore the irrelevant items; the remainder form a **tier**.
2. Items on the tier are subject to **local constraints**.
3. **TSL-2:** constraint window contains only two items.

Modeling movement We use **a[ncestor]-strings** over MG dependency trees (Shafiei and Graf 2020). For each movement type:

- The tier contains only (i) movers, (ii) landing sites, and (iii) blockers.
- On the tier, a mover must be immediately followed by its landing site.

Example: EPP-movement

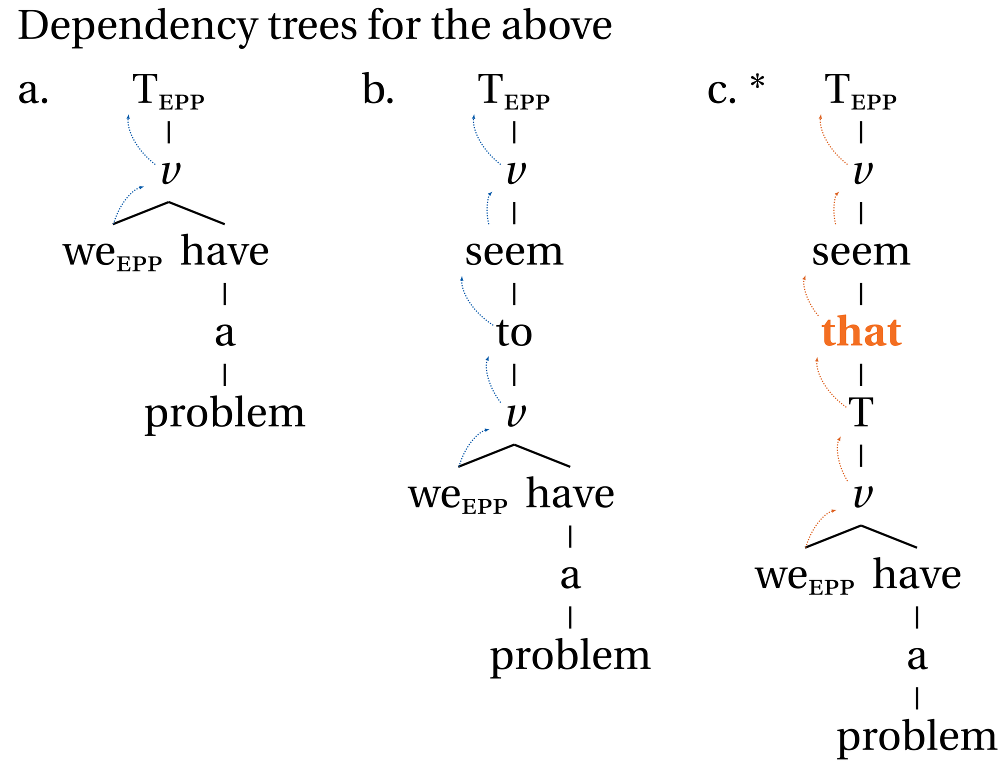
(1) EPP movement: C is a blocker, T is not

a. We [$_{vP}$ ___ have a problem].
A-string: $\rtimes \cdot \mathrm{D_{EPP}} \cdot v \cdot \mathrm{T_{EPP}} \cdot \ltimes$
Tier: $\rtimes \cdot \mathrm{D_{EPP}} \cdot \mathrm{T_{EPP}} \cdot \ltimes$

b. We seem [$_{TP}$ to ___ have a problem].
A-string: $\rtimes \cdot \mathrm{D_{EPP}} \cdot v \cdot \mathrm{T} \cdot \mathrm{V} \cdot v \cdot \mathrm{T_{EPP}} \cdot \ltimes$
Tier: $\rtimes \cdot \mathrm{D_{EPP}} \cdot \mathrm{T_{EPP}} \cdot \ltimes$

c. *We seem [$_{CP}$ that ___ have a problem].
A-string: $\rtimes \cdot \mathrm{D_{EPP}} \cdot v \cdot \mathrm{T} \cdot \mathrm{C} \cdot \mathrm{V} \cdot v \cdot \mathrm{T_{EPP}} \cdot \ltimes$
Tier: $\rtimes \cdot \mathrm{D_{EPP}} \cdot \mathrm{C} \cdot \mathrm{T_{EPP}} \cdot \ltimes$

(2) Dependency trees for the above



(3) Target TSL-2 grammar for EPP-movement
$T = \{\mathrm{D_{EPP}}, \mathrm{T_{EPP}}, \mathrm{C}\}$
$G^{-} = \{\mathrm{D_{EPP}} \cdot \mathrm{D_{EPP}}, \mathrm{D_{EPP}} \cdot \mathrm{C}, \mathrm{D_{EPP}} \cdot \ltimes\}$

Important points

- A TSL grammar applies to **the entire structure** (here, every a-string).
- Each dependency gets its own grammar. The intersection of several TSL grammars is an **MTSL** grammar.
- This works for **agreement**, too (cf. Hanson 2024).

Distributional Learning of Syntactic Blockers

Heuser et al (2024) algorithm

- **Memorizes** bigrams along attested movement paths → approximately an SL learner (Heinz 2010).
- **Generalizes** from individual items to categories when permitted by the Tolerance Principle (Yang 2016).
- **Missing bigrams** correspond to blockers.

Example Given data like (1a) and (1b), learns that EPP movement is possible out of VP/vP/TP but not CP.

(4) Sample output for EPP-movement
$G^{+} = \{\mathrm{D_{EPP}} \cdot \mathrm{V}, \mathrm{D_{EPP}} \cdot v, \mathrm{V} \cdot v, v \cdot \mathrm{T_{EPP}}, v \cdot \mathrm{T}, \mathrm{T} \cdot \mathrm{V}\}$

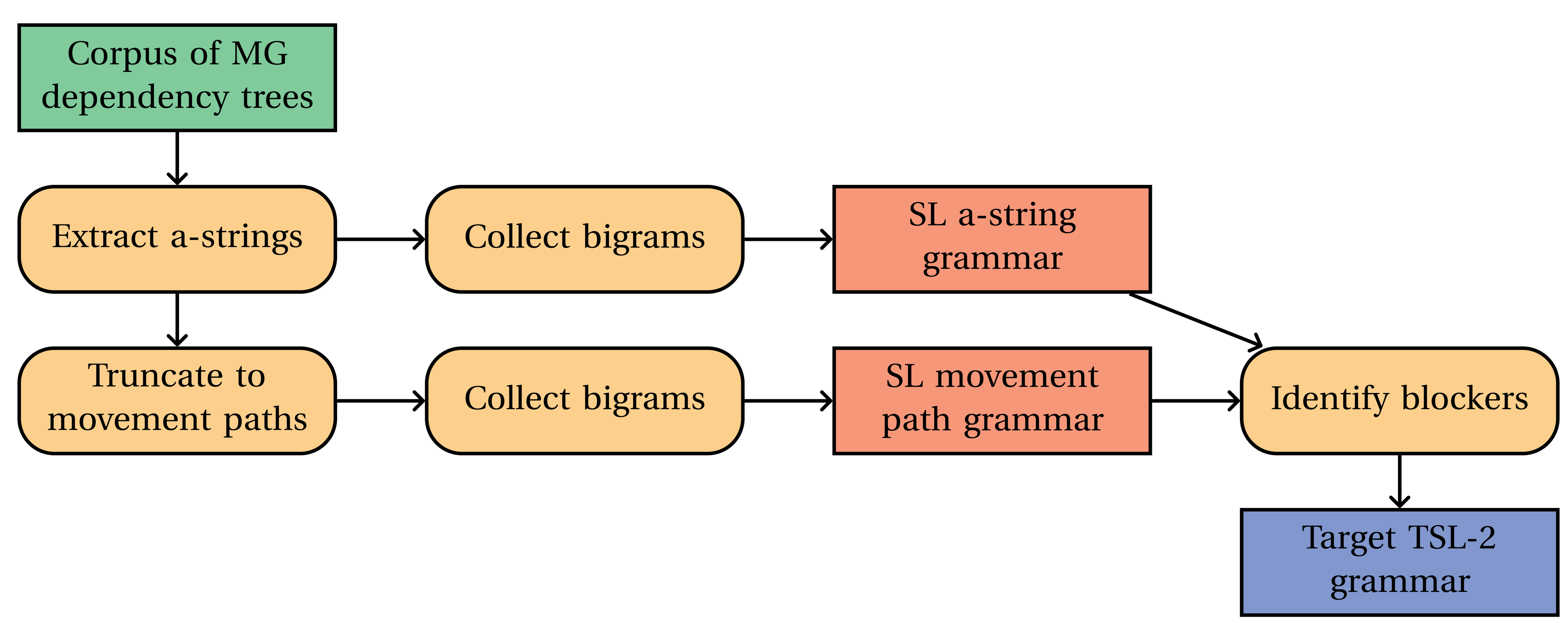
Advantages • Efficient, online learner. • Makes correct generalizations based on a realistic input distribution. • Works for a variety of syntactic blockers.

Disadvantages
Resulting grammar looks like an SL grammar, but requires a-strings to be truncated to movement paths.
→ Implicitly increases computational complexity to IBSP (Graf 2017).
→ Neither SL nor IBSP matches the typology of long-distance dependencies as well as TSL-2.

Learning Tiers from Paths

Proposal

- **Truncate** a-strings to movement paths **only during learning**.
- **Construct** movement path grammar G_M as above, **plus** a grammar G_L constructed for arbitrary a-strings, which reflects local constraints.
- **Compare** G_M and G_L to identify blockers, then construct the desired TSL-2 grammar.



Implications for Locality

Height-Locality Connection "Higher" categories in a functional hierarchy are subject to fewer locality restrictions (Keine 2019). This follows from our learner:

- DPs can start in Comp-VP, so every category from V to the landing site must be transparent.
- Will allow extraction from higher categories only if there is positive evidence.

English

Dependency	Location	Transparent? CP	TP	vP
ϕ-agreement	T	*	✓	✓
A-movement	T	*	✓	✓
Wh-movement	C	✓	✓	✓
Rel-movement	C	✓	✓	✓

Hindi

Dependency	Location	Transparent? CP	TP	vP
ϕ-agreement	T	*	*	✓
A-movement	T	*	*	✓
Wh-licensing	C	*	✓	✓
A′-movement	C	✓	✓	✓

Insights from Learnability Theory

Learning TSL string languages

- TSL allows free insertion/deletion of non-tier elements → heuristic for tier detection (Jardine and McMullin 2017).
- Pure TSL-k (string) grammars can be learned in the limit from positive data, online, by collecting all $(k{+}1)$-grams (Lambert 2021).
- An older TSL-2 learner tracks the set of possible interveners for each pair of symbols (Jardine and Heinz 2016).

Important points

- The Heuser et al. (2024) algorithm tracks similar information as existing TSL learners, but interprets it differently.
- None of the above learners deal with interacting constraints on different tiers → need to compare against *expected* k-grams, e.g. an SL grammar for selection and functional hierarchies.

Related Work

- Pearl and Sprouse (2013) track trigram probabilities in order to learn syntactic islands.
- Belth (2023) presents a Tolerance Principle-based algorithm for learning vowel harmony, encoded as a TSL-2 grammar.
- McMullin et al. (2019) develop an algorithm which learns a subclass of the MTSL-2 string languages. Swanson (2024) does the same for MTSL-2 tree languages.

Future Work

Empirical coverage Can the proposed model be used to learn dependent case, and other dependencies between >2 elements?

Typology Contra Keine (2019), the proposed model predicts that exceptions to the height-locality connection are possible under the right conditions – do these exist?

Learnability A proof is needed to show that SL constraints can be successfully factored out in all cases. Also, can we handle interactions between many tiers under realistic conditions?

References and Acknowledgments

This work was supported by the National Science Foundation under Grant No. BCS-1845344, and by the Institute for Advanced Computational Science at Stony Brook University.

References

Belth, Caleb (2023). "Towards a Learning-Based Account of Underlying Forms: A Case Study in Turkish". In: *Proceedings of the Society for Computation in Linguistics 2023*. University of Massachusetts Amherst. • **Graf, Thomas** (2017). "The Power of Locality Domains in Phonology". In: *Phonology* 34. • **Graf, Thomas** (2022). "Typological implications of tier-based strictly local movement". In: *Proceedings of the Society for Computation in Linguistics 2022*. University of Massachusetts Amherst. • **Hanson, Kenneth** (2024). "Tier-Based Strict Locality and the Typology of Agreement". Under review. • **Heinz, Jeffrey** (2010). "String Extension Learning". In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics. • **Heuser, Annika et al.** (2024). *The Learnability of Syntactic Islands*. Presentation at NELS 54, MIT. • **Jardine, Adam and Jeffrey Heinz** (2016). "Learning Tier-based Strictly 2-Local Languages". In: *Transactions of the Association for Computational Linguistics* 4. • **Jardine, Adam and Kevin McMullin** (2017). "Efficient Learning of Tier-Based Strictly k-Local Languages". In: *Language and Automata Theory and Applications*. Springer. • **Keine, Stefan** (2019). "Selective Opacity". In: *Linguistic Inquiry* 50.1. • **Lambert, Dakotah** (2021). "Grammar Interpretations and Learning TSL Online". In: *Proceedings of the Fifteenth International Conference on Grammatical Inference*. Vol. 153. Proceedings of Machine Learning Research. PMLR. • **McMullin, Kevin** (2016). "Tier-based locality in long-distance phonotactics: learnability and typology". PhD thesis. University of British Columbia. • **McMullin, Kevin et al.** (2019). "Learning phonotactic restrictions on multiple tiers". In: *Proceedings of the Society for Computation in Linguistics 2019*. University of Massachusetts Amherst. • **Pearl, Lisa and Jon Sprouse** (2013). "Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem". In: *Language Acquisition* 20.1. • **Shafiei, Nazila and Thomas Graf** (2020). "The Subregular Complexity of Syntactic Islands". In: *Proceedings of the Society for Computation in Linguistics 2020*. University of Massachusetts Amherst. • **Swanson, Logan** (2024). "Syntactic Learning over Tree Tiers". To appear in Proceedings of ESSLLI 2024 Student Session. • **Yang, Charles** (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.