Computational Phonology Workshop

Introduction & Tutorial

Thomas Graf

Stony Brook University
mail@thomasgraf.net
http://thomasgraf.net

Dec 12, 2016

Outline

Computational View of Language

In formal language theory, string sets are classified according to their formal complexity.

$$\text{regular} < \text{context-free} < \text{mildly context-sensitive} < \cdots$$

Phonology

Morphology

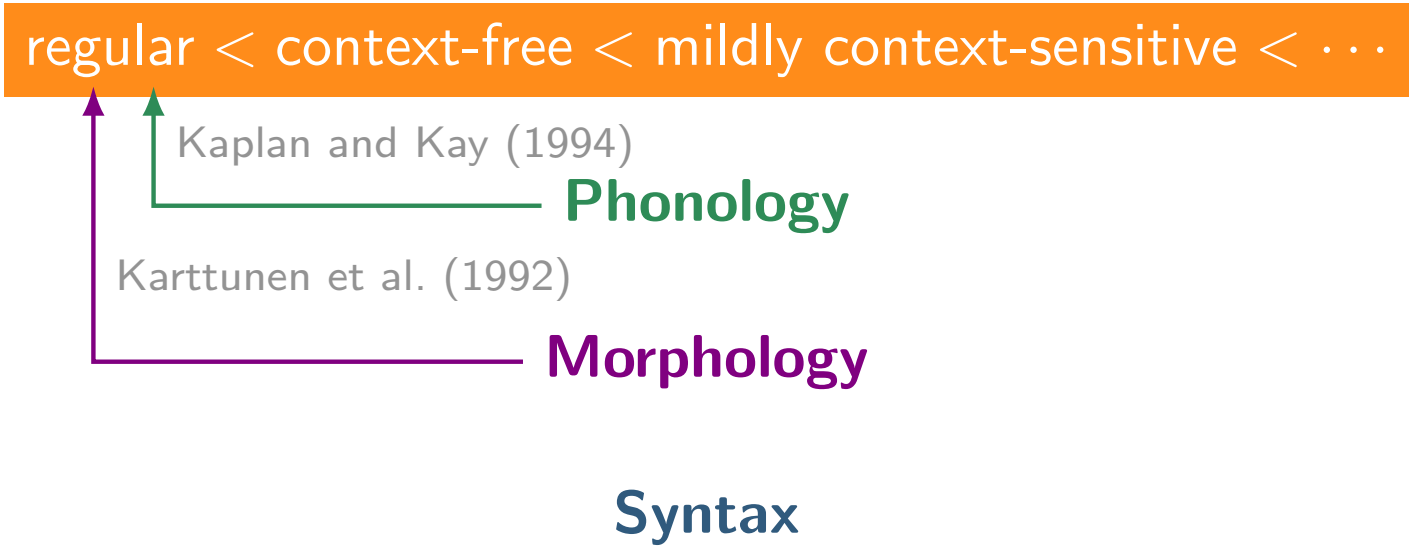
Syntax

Computational View of Language

In formal language theory, string sets are classified according to their formal complexity.

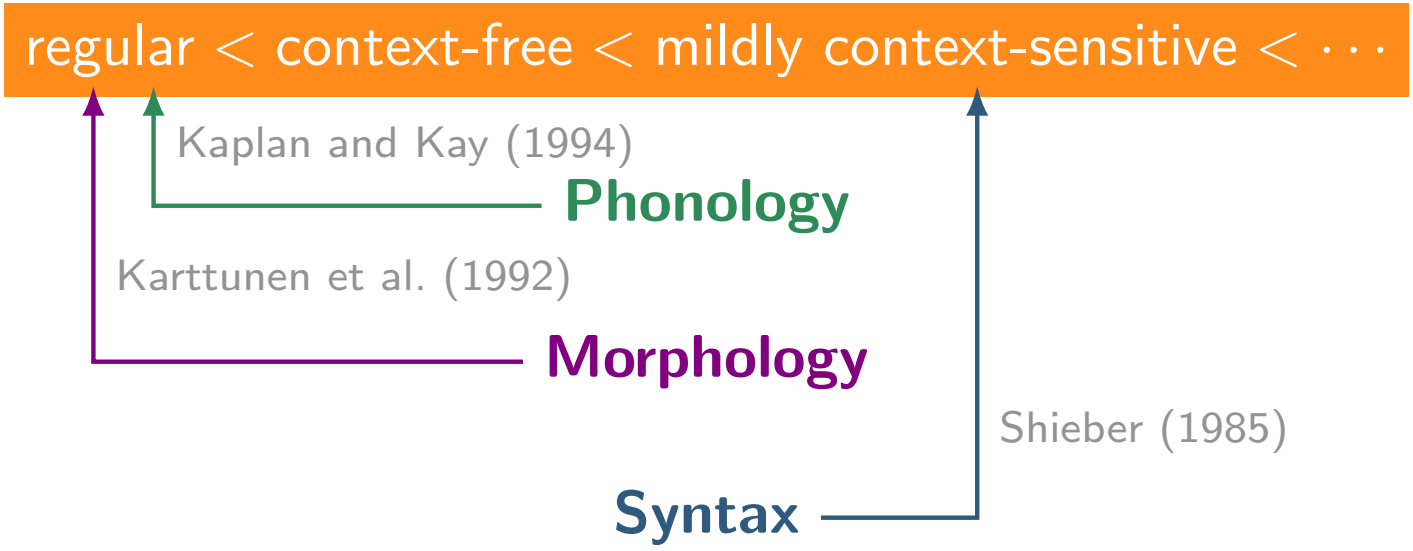
Computational View of Language

In formal language theory, string sets are classified according to their formal complexity.



Computational View of Language

In formal language theory, string sets are classified according to their formal complexity.



Implications of Formal Complexity

Heinz and Idsardi (2011, 2013) highlight the implications:

- different typology
  center embedding, crossing dependencies
- different memory architecture
  flat & finite VS unbounded nested stacks
- different learning algorithms
  much harder for syntax

Too Many Patterns are Regular

- **Problem**
  - All phonological and morphological patterns are regular.
  - But not all regular patterns occur in phonology.
- Regularity is **too loose an upper bound**.

Example

- First-last consonant harmony
- Every word with a plosive contains an open syllable
- Word with at least 3 suffixes must have exactly 5 prefixes

Too Many Patterns are Regular

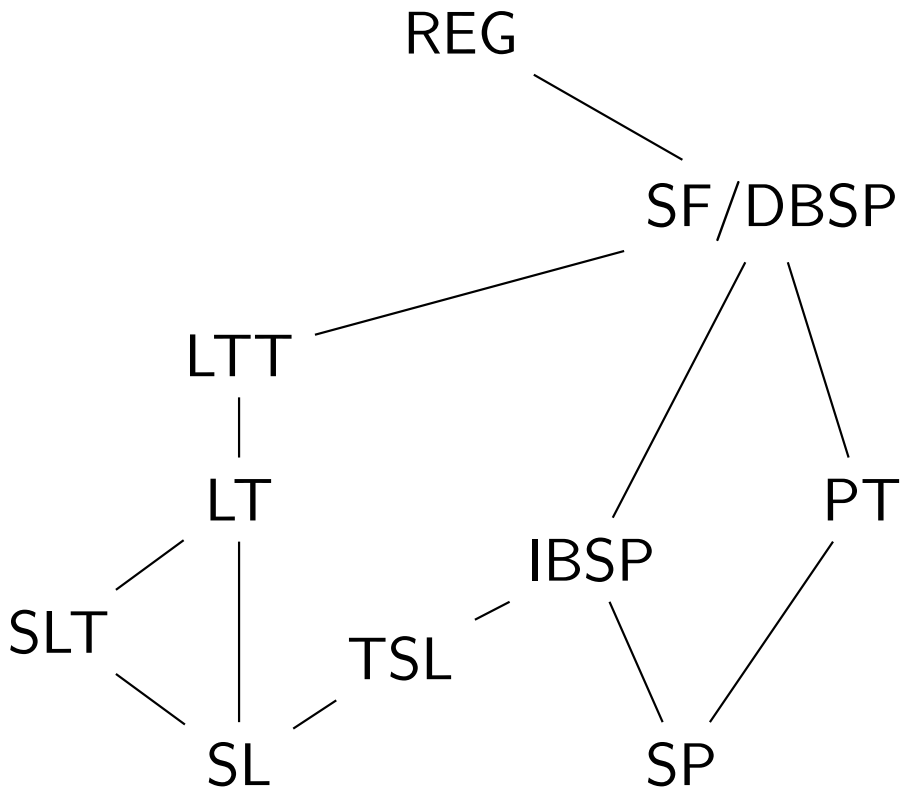
- **Problem**
  - All phonological and morphological patterns are regular.
  - But not all regular patterns occur in phonology.
- Regularity is **too loose an upper bound**.

Example

- First-last consonant harmony
- Every word with a plosive contains an open syllable
- Word with at least 3 suffixes must have exactly 5 prefixes

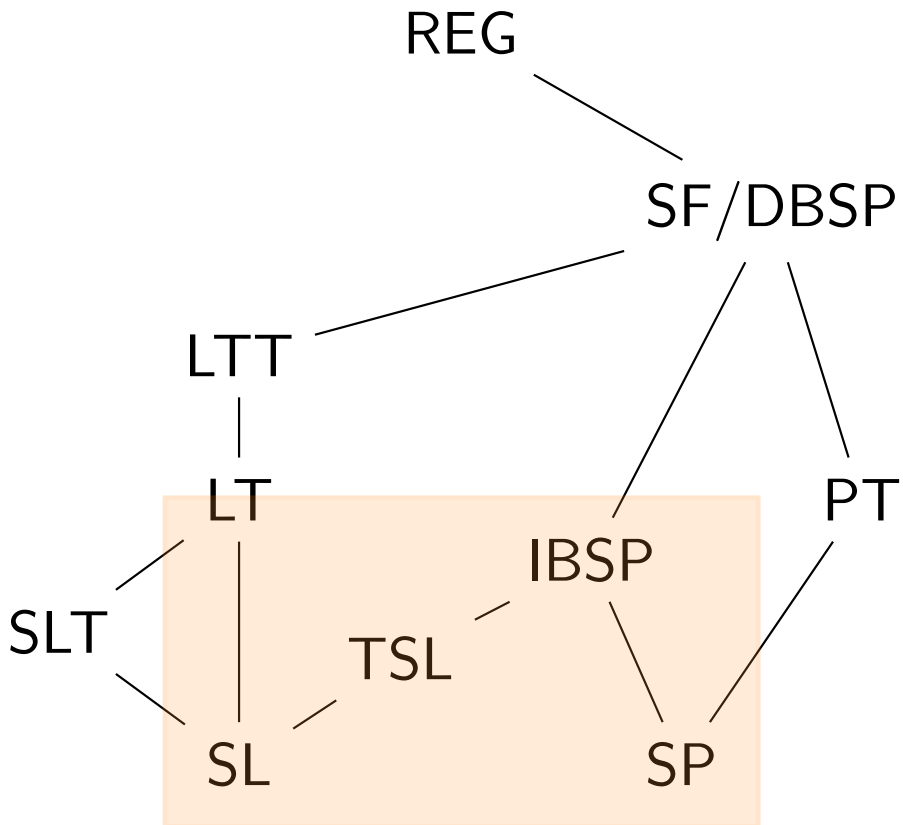
Subregular Languages

Often forgotten: hierarchy of **subregular languages**
(McNaughton and Papert 1971; Rogers et al. 2010; Ruiz et al. 1998; Rogers and Pullum 2011; Heinz et al. 2011; Graf 2016)



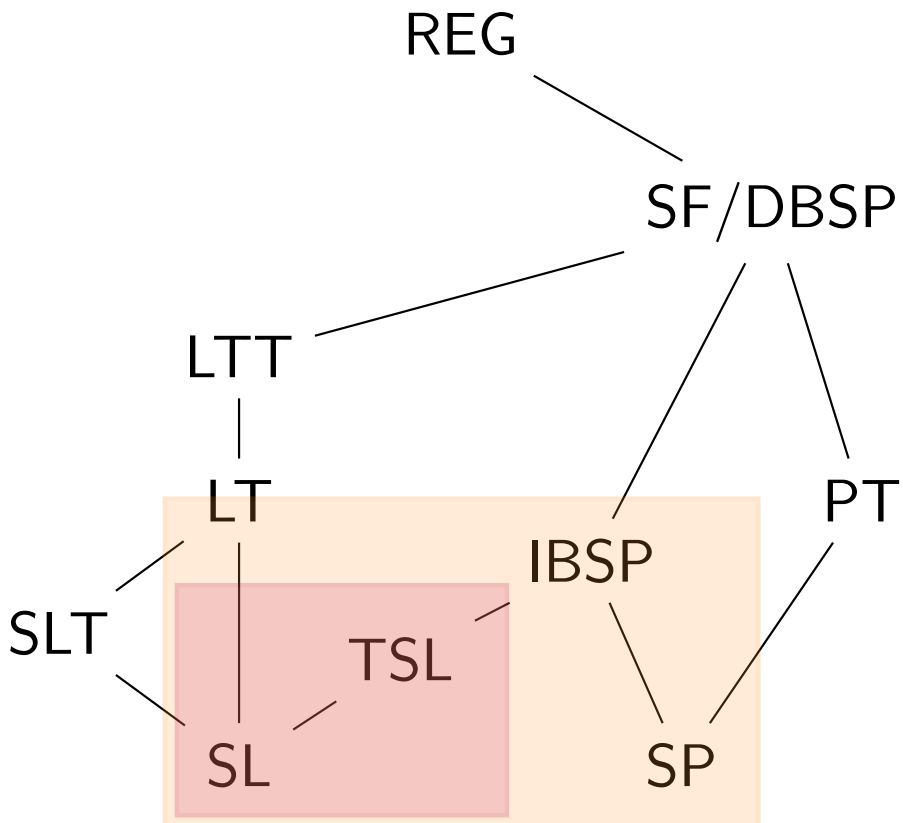
Subregular Languages

Often forgotten: hierarchy of **subregular languages**
(McNaughton and Papert 1971; Rogers et al. 2010; Ruiz et al. 1998; Rogers and Pullum 2011; Heinz et al. 2011; Graf 2016)



Subregular Languages

Often forgotten: hierarchy of **subregular languages**
(McNaughton and Papert 1971; Rogers et al. 2010; Ruiz et al. 1998; Rogers and Pullum 2011; Heinz et al. 2011; Graf 2016)



SL: Strictly Local

- SL formalizes **local dependencies**.
- SL grammars are collections of markedness constraints that are
  - hard/non-violable,
  - locally bounded.

Strictly Local Grammars & Languages

SL_n grammar finite set of forbidden n-grams

SL_n language all strings except those with forbidden n-grams

Example: SL Constraints

Process	Constraint	Forbidden n-grams
Word-final devoicing	*[+voice]⋉	z⋉, v⋉, . . .
Intervocalic voicing	*V[-voice]V	asa, asi, . . . , isa, isi, . . . , afa, afi, . . . , ifa, ifi, . . . ,
CV template	*⋊V	⋊a, ⋊i, . . .
	*CC	pp, pb, . . . bp, bb, . . .
	*VV	aa, ai, . . . , ia, ii, . . .
	*C⋉	p⋉, b⋉, . . .

SL is Too Weak

- SL grammars only handle unbounded dependencies.
- But some processes in phonology are unbounded.

Samala Sibilant Harmony (Heinz 2015:16)

ʃtojonowonowaʃ

SL is Too Weak

- SL grammars only handle unbounded dependencies.
- But some processes in phonology are unbounded.

Samala Sibilant Harmony (Heinz 2015:16)

ʃtojonowonowaʃ
*stojonowonowaʃ
*ʃtojonowonowas

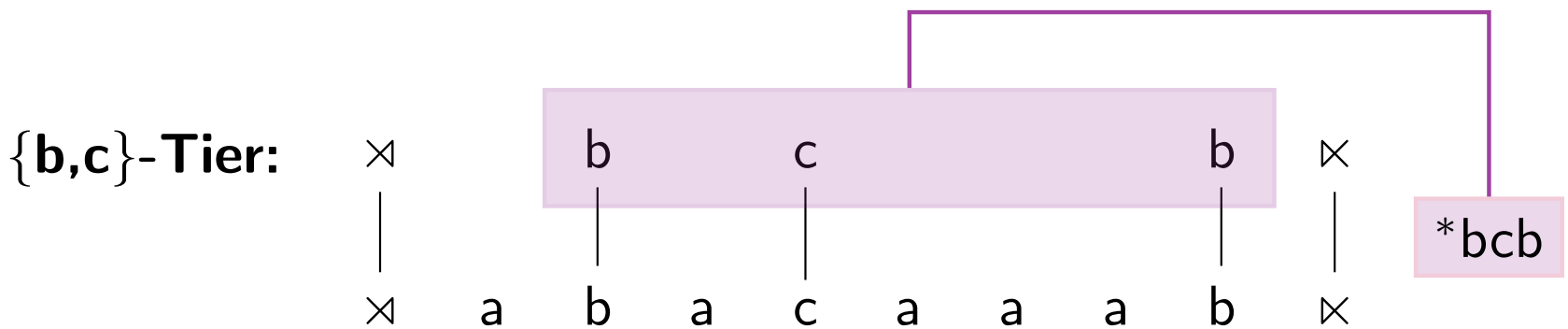
TSL: Tier-Based Strictly Local

We can make Samala SL-like if we **create new locality domains**.

Tier-Based Strictly Local Grammars & Languages

TSL_n grammar finite set of forbidden n-grams + tier alphabet

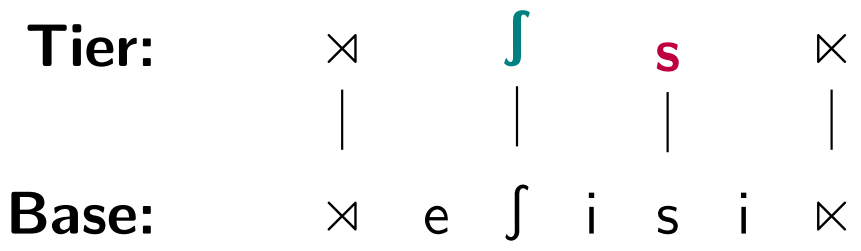
TSL_n language all strings except those with forbidden n-grams over tier

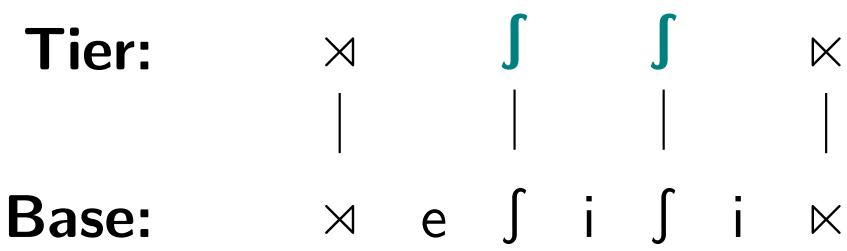


Example: Sibilant Harmony

Constraint	**Forbidden tier-n-grams**
***[α ant] $\cdots$ [$-\alpha$ ant]**	**ʃs, sʃ**

Tier contains all sibilants



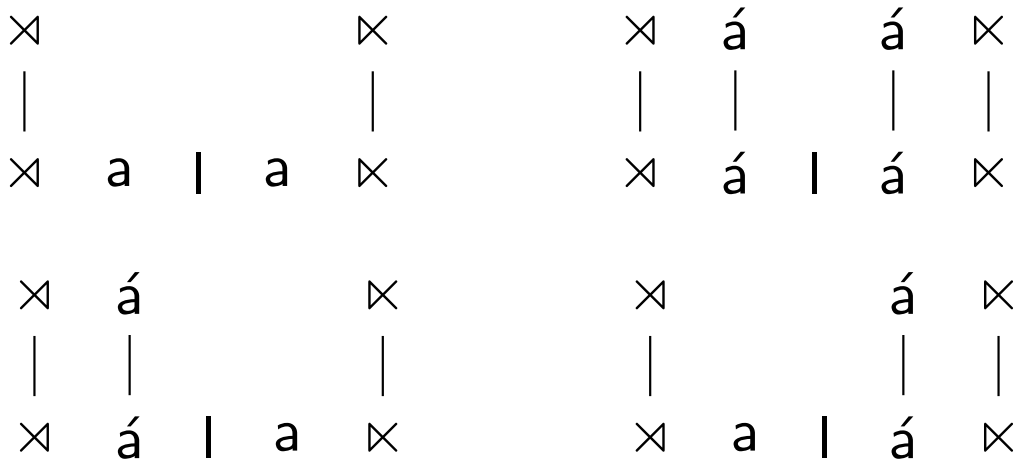


Example: Stress Assignment

Culminativity every word has exactly one primary stress

Tier contains segments with primary stress

n-**grams** **śś** and ⋊⋉



A Non-TSL Pattern: Sour Grapes Harmony

Sour Grapes vowel harmony applies only if it can apply to the whole word (i.e. there is no blocker)

Why Sour Grapes isn't TSL

- All vowels **V** must be on the vowel harmony tier.
- The blocker **B** must be on the same tier in order to block it.
- But there is no bound on the number of vowels per tier.
- The tier thus may have the shape

$$\cdots \mathbf{V}\ \mathbf{V}\ \mathbf{V} \cdots \mathbf{B} \cdots$$

- **B** can be arbitrarily far away from **VVV** $\Rightarrow$ not a local relation
- But we need to know whether **B** is on the tier in order to determine the well-formedness of **VVV**.

More Patterns Beyond TSL?

A few other patterns may go beyond TSL.

- **Non-Final RHOL**
  - discussed in **Hyunah Baek's talk**
  - a minor extension of TSL suffices
- **Multiple Harmony**
  - discussed in **Alëna Aksënova's talk**
  - is actually TSL due to important restrictions

Complexity of Phonology

- All local phonological constraints are SL.
- All segmental long-distance constraints are TSL.
- Suprsegmental constraints (tone, stress) may go beyond TSL. (Graf 2010a,b; Jardine 2015)
- TSL avoids instances of OT overgeneration:
  - cannot generate *sour-grapes* or *majority rules* patterns
  - does not allow *agreement by proxy*
  - explains why consonant harmony is unbounded or transvocalic, but never transconsonantal (McMullin and Hansson 2015)

Cognitive Implications

- SL and TSL languages are **learnable** from positive data. (Heinz et al. 2012; Jardine and Heinz 2016)
  - UG: specifies upper bound on size of n-grams
  - memorize which sequences have not been seen so far
  - induce tier (more complex)
  - learning input can be relatively small
- What cognitive resources are required?
  - Only memorization of the last n segments of a specific type
  - For most processes $n \leq 3$, and for all $n \leq 7$
  - Fits **within bounds of human working memory**

Interim Summary: Phonotactics

- Natural languages have **TSL phonotactics**.
- gives tighter bound on typology
- solves poverty of stimulus by greatly simplifying learning
- reduces cognitive resource requirements

Next

- phonological mappings
- SL & TSL syntax

Interim Summary: Phonotactics

- Natural languages have **TSL phonotactics**.
- gives tighter bound on typology
- solves poverty of stimulus by greatly simplifying learning
- reduces cognitive resource requirements

Next

- phonological mappings
- SL & TSL syntax

Phonological Mappings

- So far we have only considered phonotactics.
- But mappings from underlying representations to surface forms can be studied, too.
- Regular mappings are enough.
  (Kaplan and Kay 1994)
- What about subregular mappings?

Input Strictly Local Mappings

Input Strictly Local (ISL)

- Move through string from left to right.
- Rewrite x as y based on previous n symbols in input string.
- Output is **not considered**!

A Note on TSL

Every TSL_n grammar can be decomposed into

1. an **ISL_1** function (the tier projection), and
2. an SL_n grammar.

An Interesting Puzzle

- What happens if we use an ISL_k function for tier projection?
- Addressed in **Aniello De Santo**'s talk

A Note on TSL

Every TSL_n grammar can be decomposed into

1. an **ISL_1** function (the tier projection), and
2. an SL_n grammar.

An Interesting Puzzle

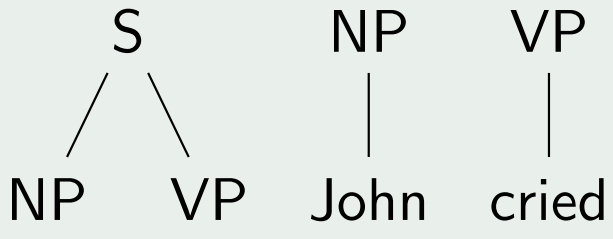
- What happens if we use an ISL_k function for tier projection?
- Addressed in **Aniello De Santo**'s talk

(Tier-Based) Strictly Local Syntax

- SL tree grammars are common in computational linguistics: **context-free grammars**
- By adding tier projection, we get TSL tree grammars.

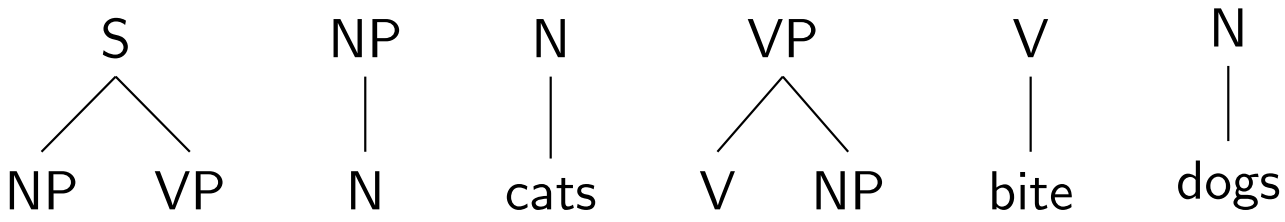
Example: CFGs as SL_2 Tree Grammars

$$\begin{aligned} S &\to NP\ VP \\ NP &\to John \\ VP &\to cried \end{aligned}$$

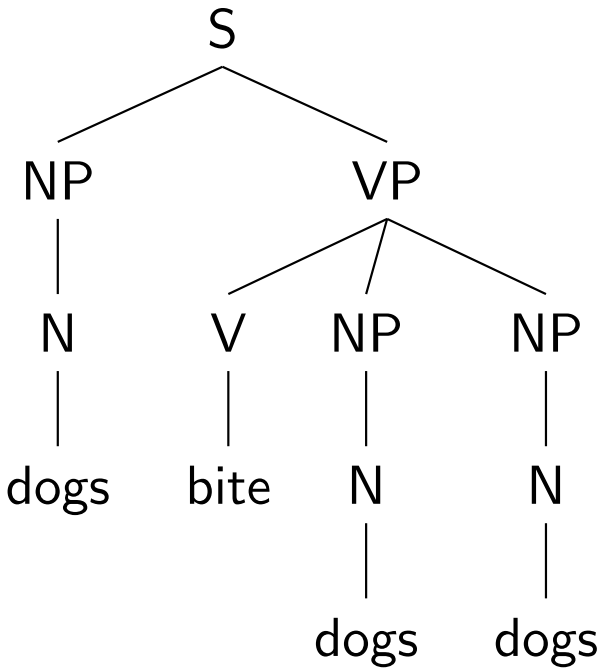


Example: An Illicit Tree

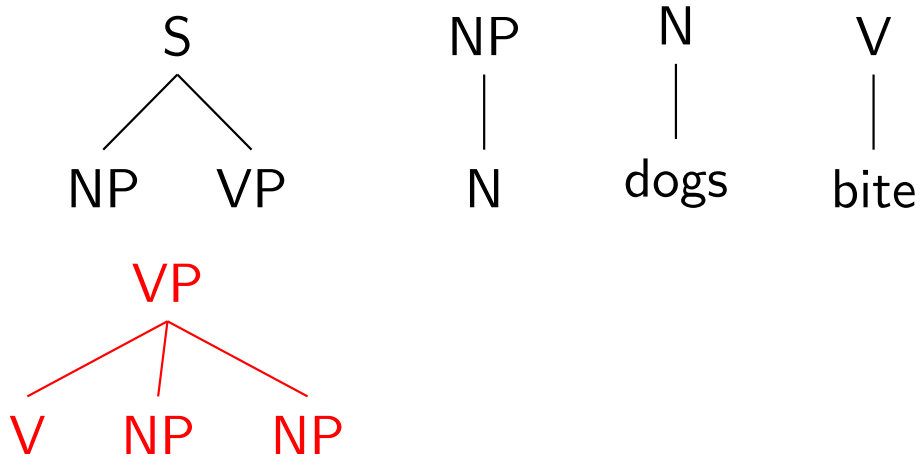
SL_2 Tree Grammar



Example Tree

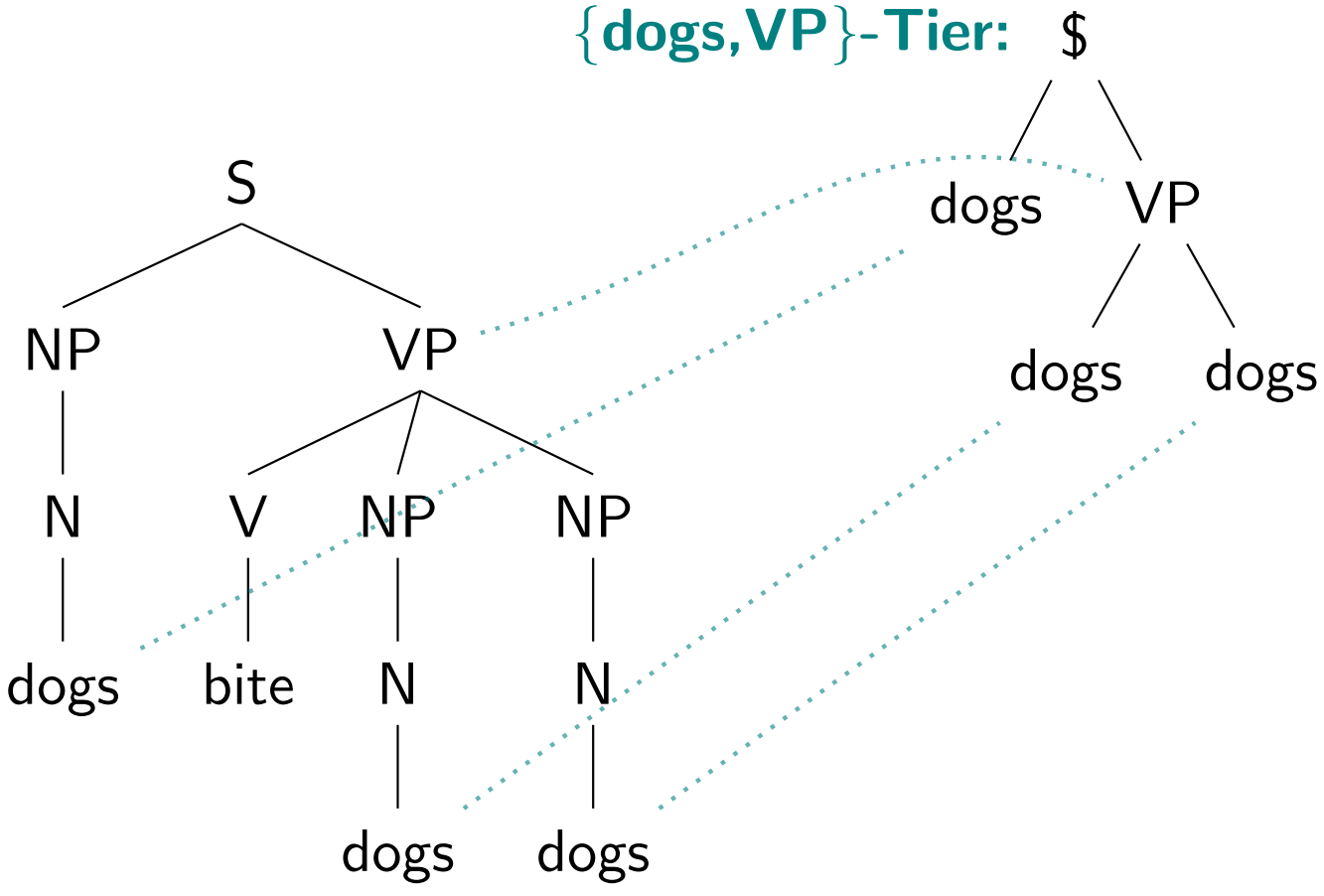


Tree Bigrams of Example Tree



Tier Projection for Trees

Just as for strings, we can project tiers for trees.
(Graf and Heinz 2016)



Towards TSL-Syntax

While TSL-Syntax is still young, it holds promise:

- movement dependencies are TSL (Graf and Heinz 2016)
- scope ambiguities in **Lei Liu's blitz talk**
- Mandarin negation in **Hongchen Wu's talk**

Conclusion

References I

Graf, Thomas. 2010a. Comparing incomparable frameworks: A model theoretic approach to phonology. *University of Pennsylvania Working Papers in Linguistics* 16:Article 10. URL `http://repository.upenn.edu/pwpl/vol16/iss1/10`.

Graf, Thomas. 2010b. *Logics of phonological reasoning*. Master's thesis, University of California, Los Angeles. URL `http://thomasgraf.net/doc/papers/LogicsOfPhonologicalReasoning.pdf`.

Graf, Thomas. 2016. The power of locality domains in phonology. Ms., Stony Brook University.

Graf, Thomas, and Jeffrey Heinz. 2016. Tier-based strict locality in phonology and syntax. Ms., Stony Brook University and University of Delaware.

Heinz, Jeffrey. 2015. The computational nature of phonological generalizations. URL `http://www.socsci.uci.edu/~lpearl/colareadinggroup/readings/Heinz2015BC_Typology.pdf`, ms., University of Delaware.

Heinz, Jeffrey, and William Idsardi. 2011. Sentence and word complexity. *Science* 333:295–297.

Heinz, Jeffrey, and William Idsardi. 2013. What complexity differences reveal about domains in language. *Topics in Cognitive Science* 5:111–131.

Heinz, Jeffrey, Anna Kasprzik, and Timo Kötzing. 2012. Learning with lattice-structure hypothesis spaces. *Theoretical Computer Science* 457:111–127.

References II

Heinz, Jeffrey, Chetan Rawal, and Herbert G. Tanner. 2011. Tier-based strictly local constraints in phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 58–64. URL http://www.aclweb.org/anthology/P11-2011.

Jardine, Adam. 2015. Computationally, tone is different. *Phonology* URL http://udel.edu/~ajardine/files/jardinemscomputationallytoneisdifferent.pdf, to appear.

Jardine, Adam, and Jeffrey Heinz. 2016. Learning tier-based strictly 2-local languages. *Transactions of the ACL* 4:87–98. URL https://aclweb.org/anthology/Q/Q16/Q16-1007.pdf.

Kaplan, Ronald M., and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20:331–378. URL http://www.aclweb.org/anthology/J94-3001.pdf.

Karttunen, Lauri, Ronald M. Kaplan, and Annie Zaenen. 1992. Two-level morphology with composition. In *COLING'92*, 141–148. URL http://www.aclweb.org/anthology/C92-1025.

McMullin, Kevin, and Gunnar Ólafur Hansson. 2015. Long-distance phonotactics as tier-based strictly 2-local languages. In *Proceedings of AMP 2014*.

McNaughton, Robert, and Seymour Papert. 1971. *Counter-free automata*. Cambridge, MA: MIT Press.

References III

Rogers, James, Jeffrey Heinz, Gil Bailey, Matt Edlefsen, Molly Vischer, David Wellcome, and Sean Wibel. 2010. On languages piecewise testable in the strict sense. In *The mathematics of language*, ed. Christan Ebert, Gerhard Jäger, and Jens Michaelis, volume 6149 of *Lecture Notes in Artificial Intelligence*, 255–265. Heidelberg: Springer. URL `http://dx.doi.org/10.1007/978-3-642-14322-9_19`.

Rogers, James, and Geoffrey K. Pullum. 2011. Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information* 20:329–342.

Ruiz, José, Salvador España, and Pedro García. 1998. Locally threshold testable languages in strict sense: Application to the inference problem. In *Grammatical inference: 4th international colloquium, ICGI-98 Ames, Iowa, USA, July 12–14, 1998 proceedings*, ed. Vasant Honavar and Giora Slutzki, 150–161. Berlin: Springer.

Shieber, Stuart M. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8:333–345.