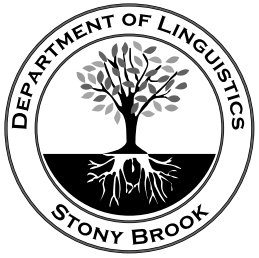
Grammar Interpretations and Learning TSL Online

Dakotah Lambert

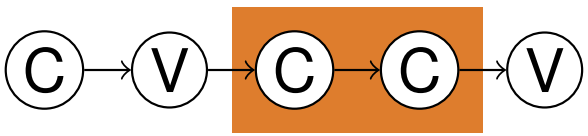
24 August 2021



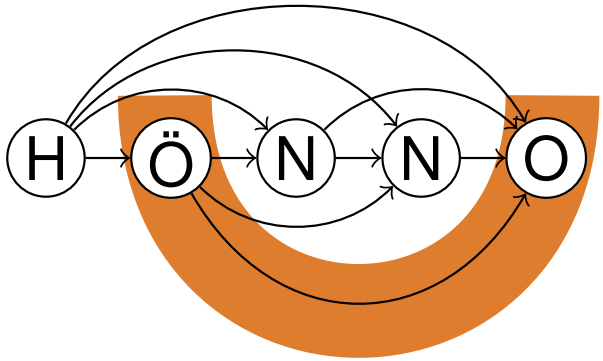
iACS
INSTITUTE FOR ADVANCED COMPUTATIONAL SCIENCE

Word Models and Constraints

(SL)
Local



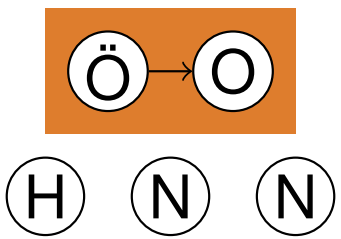
(SP)
Long-Distance



Sometimes mark boundaries with $\rtimes$ and $\ltimes$

Another Type of Long-Distance Representation

(TSL)
Ignore symbols that don't matter



The Learning Problem

- Finite data
- Positive and/or negative
- Goal: make the right generalizations

Restrictions

Gold's (1967) ILPD Framework

- No attention to negative data (Yang, 2015)
- Exact convergence
- Requirement: every sufficiently large text contains a characteristic sample

Online Learning

- Constant bound on memory requirements
- Adjust hypothesis without reviewing all prior data
- Requirement: words can be parsed into finitely many categories of informational content

A General Online ILPD-learner

String Extension Learning (Heinz, 2010)

$$\varphi(G, w) = \begin{cases} G & \text{if } w = \circledcirc \\ G \oplus f(w) & \text{otherwise} \end{cases}$$

G	One of finitely many possible grammars
w	Either $\circledcirc$ or a word over some finite alphabet Σ
f	Map words to informational content
$\oplus$	Insert informational content into grammar

Learning SL or SP

$$\varphi(G, w) = \begin{cases} G & \text{if } w = \circledcirc \\ G \oplus f(w) & \text{otherwise} \end{cases}$$

G	Set of permitted k-factors
f	The k-factors in a word
$\oplus$	Set union

The Challenge for TSL

- k-factors are dependent on the tier
- Deciding which symbols are salient

Examples of TSL Constraints

Constraint	**Tier**	**Forbidden Factors**
Liquid Dissimilation	$\{\text{k}, \text{l}, \text{r}\}$	ll, rr
One-Stress	$\{\acute{\sigma}\}$	$\rtimes\ltimes$, $\acute{\sigma}\acute{\sigma}$
Finnish Vowel Harmony	$\{\textit{round}\}$	*front back*, *back front*

Prior Methods: 2-TSL

Jardine and Heinz (2016)

t	**a**	**l**	**o**	
t	a			$\emptyset$
t		l		$\{a\}$
t			o	$\{a, l\}$
	a	l		$\emptyset$
	a		o	$\{l\}$
		l	o	$\emptyset$

Finding the Tier Alphabet

- Gather these paths (augmented subsequences)
- Find free elements (unrestricted distribution)
- Find exclusive blockers (unrestricted but on tier)
- **Result**: nonfree elements and exclusive blockers

Prior Methods: Wider but Batch

Jardine and McMullin (2017)

- Find tier first
  1. Find factors of widths $k-1$ through $k+1$
  2. Insert σ into $(k-1)$-factors: if bad, on tier
  3. Delete σ from $(k+1)$-factors: if bad, on tier

  (Also: attend to smaller factors for short words)
- Project the input and learn with a k-SL learner

Augmented Subsequences for Wider Factors

basketball	sbl	{a,e,k,t}
baseball	sbl	{a,e}

- Maintain only smallest subsets: supersets are implied
  $\langle \text{sbl}, \{\text{a}, \text{e}\}\rangle$ implies $\langle \text{sbl}, \{\text{a}, \text{e}, \text{k}, \text{t}\}\rangle$
- Word accepted iff all factors included or implied
  $\langle \text{tal}, \emptyset\rangle$ and $\langle \text{ail}, \{\text{s}\}\rangle$ would conspire to reject “tail”
- Never bother learning T
- Structure can be large (bounded by a constant though)

Learning the Tier

Back to Jardine and McMullin (2017):

1. Collect the local factors of width up to $k + 1$
2. Insert σ freely: if bad, on tier
3. Delete σ freely: if bad, on tier

Running both learners in parallel provides:

- Information needed to determine acceptance
- And the set of salient symbols

Simplifying the Structure

- Characteristic sample for finding T:
  all attested factors of width up to $k + 1$.
- Includes all those over T alone
- So that's all we need!

Some Issues

- Characteristic sample is unrealistic
- SL-based learner is broken by any SL or SP constraints
- Augmented subsequence learning differs, doesn't find tier

References

Edward Mark Gold. 1967. Language identification in the limit. *Information and Control*, 10(5):447–474.

Jeffrey Heinz. 2010. String extension learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 897–906, Uppsala, Sweden. Association for Computational Linguistics.

Adam Jardine and Jeffrey Heinz. 2016. Learning tier-based strictly 2-local languages. *Transactions of the Association for Computation in Linguistics*, 4:87–98.

Adam Jardine and Kevin McMullin. 2017. Efficient learning of tier-based strictly k-local languages. In Frank Drewes, Carlos Martín-Vide, and Bianca Truthe, editors, *Language and Automata Theory and Applications: 11th International Conference*, volume 10168 of *Lecture Notes in Computer Science*, pages 64–76. Springer, Cham.

Charles Yang. 2015. Negative knowledge from positive evidence. *Language*, 91(4):938–953.