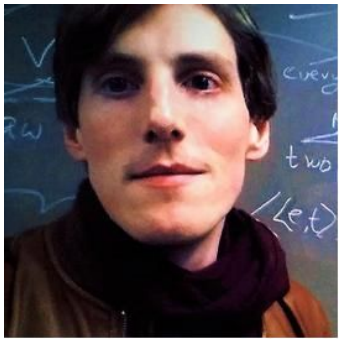
Regular Reduplication Across Modalities

Amy Li

MIT

SJSU

Jon Rawski



Reduplication: A Morphological Copying Process

- **Cross-linguistically robust** (Rubino 2013)

Attested:	Total (85%)	*buku* → *buku buku* (Indonesian ‘book’ → ‘books’)
	Partial (75%)	*buku* → *bu buku*
	Triplication (rare)	*buku* → *buku buku buku*
Unattested:	Squaring	*buku* → *buku buku buku buku*
	Iterated prefix	*buku* → *b bu buk buku*
	Exponential prefix	*buku* → *b bubu bukbukbuk bukubukubukubuku*
	Even-Odd Reverse	*buku* → *buku ukub buku ukub*

Reduplication: A Morphological Copying Process

- **Cross-linguistically robust**
- **Computationally complex?**
  - **Recognition** $f: \Sigma^* \rightarrow \mathbb{R}$

    $\{ww \mid w \in \Sigma^*\}$ - *well-nested 2-MCFL* (a TAL, Joshi 1985), triplication - *3-MCFL*

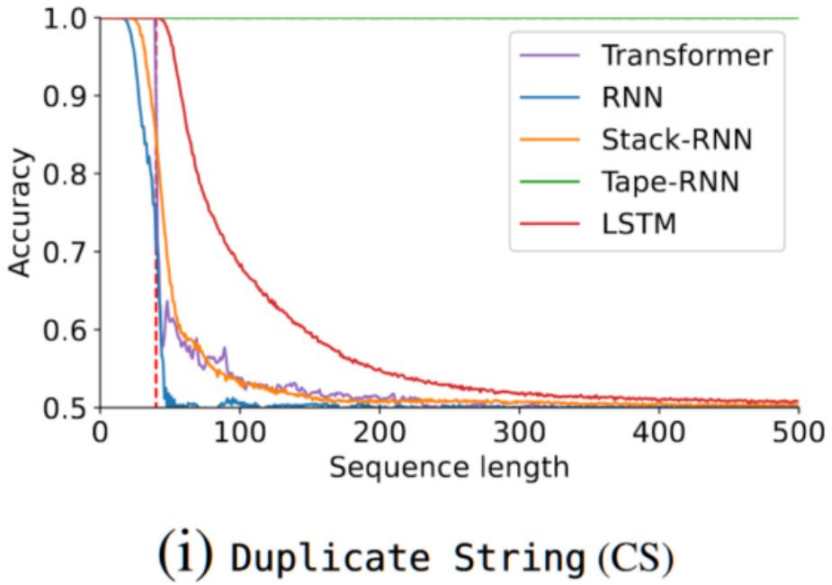
  - **Transduction** $f: \Sigma^* \rightarrow \Sigma^*$

    $w \rightarrow ww^n$ is a *Regular function* (Dolatian & Heinz 2020, Rawski et al 2023)

  - Not discussed today: syntactic copying (Yoruba relatives, Georgian case)

Reduplication: A Morphological Copying Process

- **Cross-linguistically robust**
- **Computationally complex?**
- **Unlearnable** by most neural sequence models
  - Nelson et al 2020
    Deletang et al 2022
    Zhou et al 2024
    a.o.



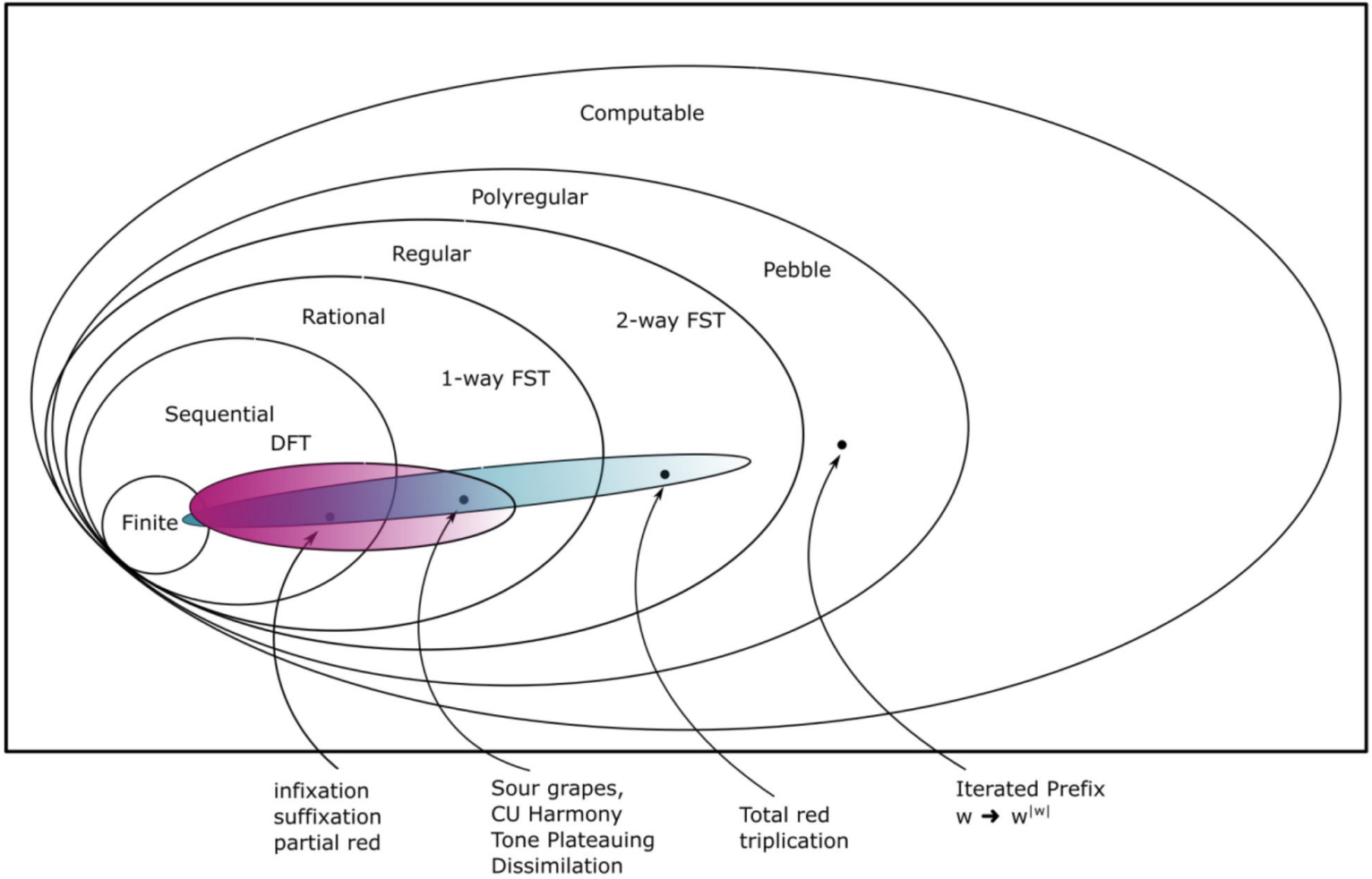
Reduplication: A Morphological Copying Process

- **Cross-linguistically robust**
- **Computationally complex?**
- **Unlearnable** by most neural sequence models
- **Computable** by (average) hard-attention transformers w/ position encodings
    - (Strobl et al *in review*)

Reduplication: A Morphological Copying Process

- **Cross-linguistically robust**
- **Computationally complex?**
- **Unlearnable** by most neural sequence models
- **Computable** by (average) hard-attention transformers w/ position encodings

Today: Do **complexity bounds** hold across modalities (i.e. in sign)?



The Computational Landscape (Rawski et al 2023)

Regular vs Polyregular (Bojanczyk 2022)

Properties of regular functions

- Linear growth: $|f(w)| = O(|w|)$
- Closed under composition (if $f : \Gamma^* \to \Sigma^*$ and $g : \Sigma^* \to \Pi^*$ are regular then so is $g \circ f$)
- L regular $\implies f^{-1}(L)$ regular

Properties of polyregular functions

- *Polynomial* growth: $|f(w)| = O(|w|^k)$
- L regular $\implies f^{-1}(L)$ regular; closed under composition

Questions

Reduplication in Signed Languages - is it regular?

At issue: lower bounds on morphological computation

Non-regularity implies a modality effect in either of 2 directions (Sandler 1993)

- Regular boundary may be an artifact of the spoken modality
- Signed modality may allow supra-regular operations on top of amodal ones

Our claim: it’s still regular

Reduplication Typology (Signed)

- Partial, total, triplication (widespread), two-handed, etc (Pfau & Steinbach 2006)
- pluralization, reciprocals, **aspectual modification**, etc (Wilbur 2005)

ASL GIVE

(a) GIVE [distrib]

‘give to each of them’

(b) GIVE [iterative]

‘give over and over again’

a. GIVE+dur

‘keep on giving’

‘Embedded’ Reduplication (Klima & Bellugi 1979, Wilbur 2009)

b. [GIVE + dur] + distrib]

‘keep on giving to each person, one after another’

(c) [[GIVE+distrib] iterative]

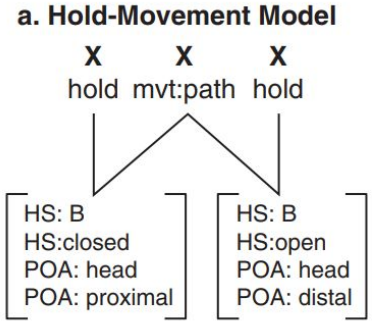
‘give to each of them over and over again’

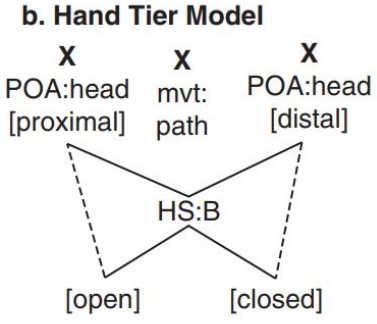
c.[[GIVE+dur]+distrib]+ iterative]

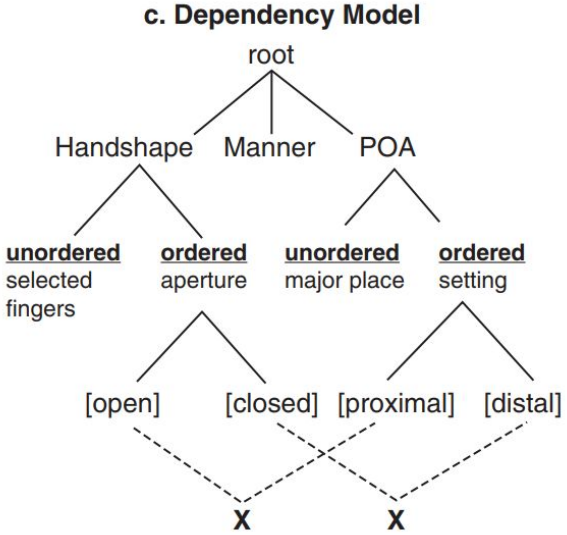
‘keep on giving to each person, one after another, this event sequence recurring regularly over expanses of time’

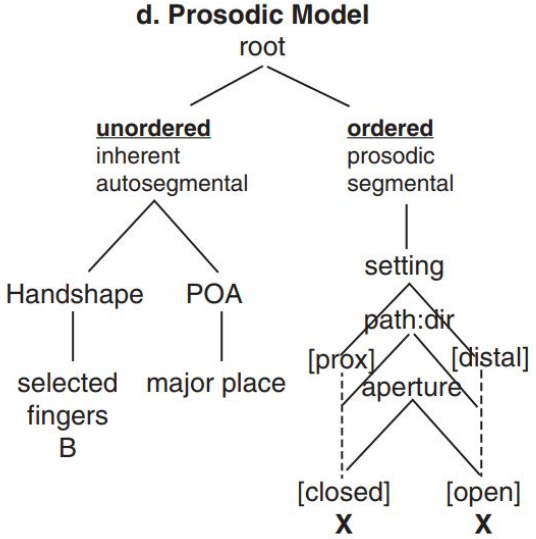
What's a model of a sign?

Brentari (2019)









Sandler's Hand Tier Model as a Finite Relational Model

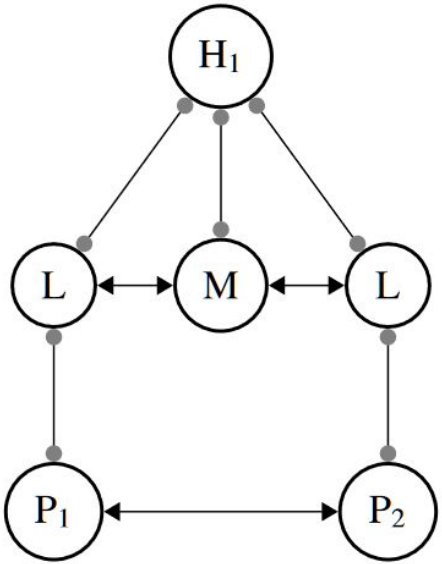
Finite model signature: $\langle D, R \rangle$

D - domain: finite set of nodes

R - relations: finite set

- Unary: node labels
- Binary: successor, association

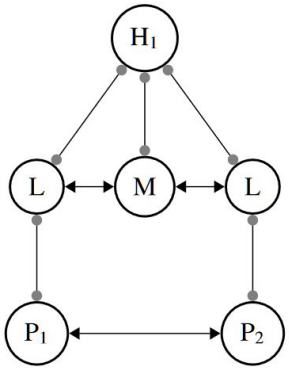
ASL 'GIVE'

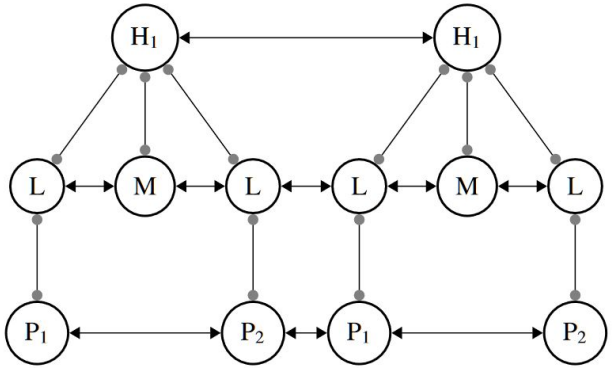


MSO Transductions & Total Reduplication

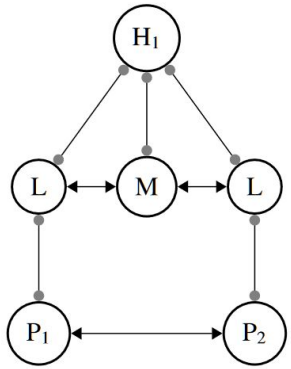
MSO Transducer: define output model *in terms of* input model

- Finite # of **copies** of the domain
- **Relation Formulas** define the output labels, successor, and association, for all copies
  - $\varphi^{c}_{L}(x) \stackrel{\text{def}}{=} L(x)$ “position is *L* in copy *c* if it’s *L* in input”
- **Domain formula**: which nodes survive (or are deleted)
- MSOT iff all formulas are MSO (set quantification)



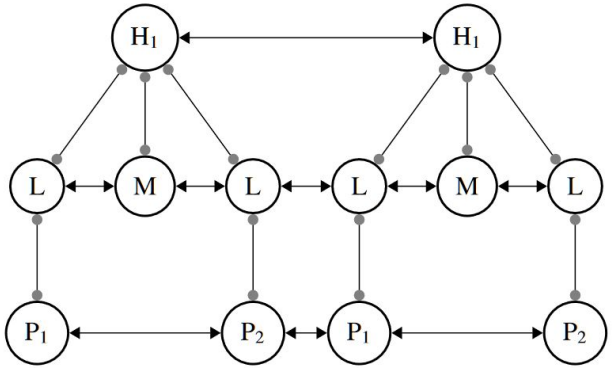


MSO Transductions & Total Reduplication



MSO Transducer: define output model *in terms of* input model

- Finite # of **copies** of the domain
- **Relation Formulas** define the output labels, successor, and association, for all copies
  - $\varphi^{c}_{L}(x) \stackrel{\text{def}}{=} L(x)$ “position is *L* in copy *c* if it’s *L* in input”
- **Domain formula**: which nodes survive (or are deleted)
- MSOT iff all formulas are MSO (set quantification)



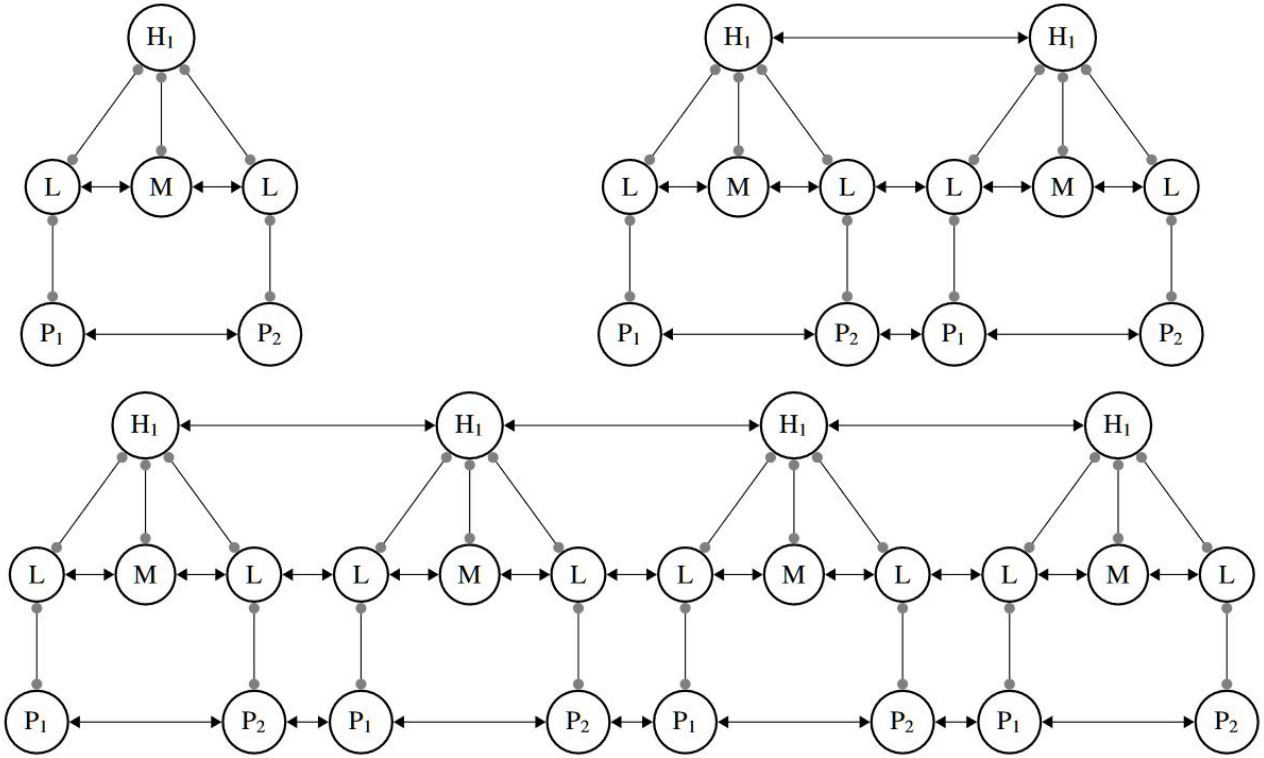
Theorem: MSO = DFA = Regular languages (BEEVT)

Theorem: MSOT = 2DFT = Reg (Engelfriet & Hoogeboom 2002)

Theorem: FOT = Aperiodic 2DFT = Ap. Reg (Carton & Dartois 2021)

Embedded Reduplication

g(f(w)): w → ww → wwww



First Summary

Signed reduplication is regular because…

- Each reduplication process copies a finitely bounded number of times
  - an MSO transduction with a finite copy set
- Regular functions are closed under finite composition
- Number of copies is independent of input size

Comparing to Spoken Multiple Reduplication

Guébie reduplication (Sande 2017; 2021)

- Reciprocalization: *w* → *ww*
- Nominalization: *w* → *ww*
- Nominalization applies to the output of reciprocalization: *w* → *ww* → *wwww*

Runyankore reduplication (Hyman 2020)

- Frequentativity: *w* → *ww* ‘to *w* a lot’
- Root can reduplicate further to mark greater degrees of freq: *w* → *www**
- Like English contrastive *salad salad* reduplication (Ghomeshi et al. 2004)
- Each application of reduplication is regular

Issues for Embedded Reduplication in Sign (Rawski et al 2023)

Cyclicity?

- [[[GIVE + dur] + dist] + iter] more like Guebie than Runyankore
- Typology: is there Runyankore-esque iterative reduplication in SLs?

Issues for Embedded Reduplication in Sign (Rawski et al 2023)

Cyclicity?

Bounds on copies?

Issues for Embedded Reduplication in Sign (Rawski et al 2023)

Cyclicity?

Bounds on copies?

- In citation form, each reduplication has finitely bounded copies.
- Van Boven (2021): individual variation via NGT corpus and elicitation data

	Number of repetitions						
	Total	**<One**	**One**	**Two**	**Three**	**Four**	**Unclear**
Simple reduplication	133	19	90	14	2	1	7
Sideward reduplication	197	1	78	44	4	1	69

Issues for Embedded Reduplication in Sign (Rawski et al 2023)

Cyclicity?

Bounds on copies?

[[[GIVE + dur] + dist] + iter]	okay (all unique)
[[[[GIVE + dur] + dist] + dur] + dur]	??

- Direct evidence? murky
- Indirect evidence: Wilbur et al 1983 - not all embeddings work
    - 35% spatial-spatial, 50% temporal-temporal

Issues for Embedded Reduplication in Sign (Rawski et al 2023)

Cyclicity?

Bounds on copies?

How do signers represent the semantic input?

Issues for Embedded Reduplication in Sign (Rawski et al 2023)

Cyclicity?

Bounds on copies?

How do signers represent the semantic input?

- **Feature** as mapping $w \rightarrow www$ regular (word-formation, base copying)
- **Feature** as input $wFF \rightarrow www$ polyregular (BRCT, template-filling)

Conclusion

Reduplication is regular across speech and sign (and maybe pro-tactile)

Insights:

- The number of copies per reduplication AND their composition is bounded
- The input does not contain information about number of copies

Predictions:

- no [[[[GIVE + dur] + dist] + dur] + dur] …
- no squaring, iterative prefix, exponential prefix, etc.

Signed languages are a rich laboratory to study bounds on linguistic computation

Questions?

Thanks for discussion to Mark Aronoff, Iris Berent, Kate Davidson, Hossep Dolatian, Harry van der Hulst, Diane Lillo-Martin, Tory Sampson, Wendy Sandler, Ronnie Wilbur,
MIT LingLunch, SCiL Reviewers