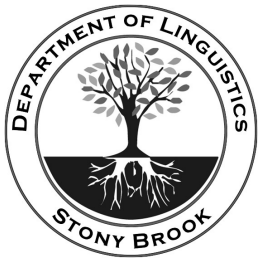
FINITE-STATE ANALYSIS OF SHUPAMEM REDUPLICATION

Magdalena Markowska, Jeffrey Heinz, Owen Rambow



SIGMORPHON 2021
AUG 5–6

In this presentation we propose a computational model of nominal and verbal reduplication in Shupamem using finite-state automata. We suggest that in order to account for that, a synthesis of two types of finite-state transducers (FSTs) is required:

In this presentation we propose a computational model of nominal and verbal reduplication in Shupamem using finite-state automata. We suggest that in order to account for that, a synthesis of two types of finite-state transducers (FSTs) is required:

- 2-way FST is required to model the reduplicative process itself because it is total reduplication (Dolatian and Heinz, 2020)

In this presentation we propose a computational model of nominal and verbal reduplication in Shupamem using finite-state automata. We suggest that in order to account for that, a synthesis of two types of finite-state transducers (FSTs) is required:

- 2-way FST is required to model the reduplicative process itself because it is total reduplication (Dolatian and Heinz, 2020)
- multiple-tape FST (MT-FST) is required to separate segmental from tonal melodies (Wiebe, 1992; Dolatian and Rawski, 2020b)

OUTLINE

1. Shupamem
   - Nominal reduplication
   - Verbal reduplication
2. Extending 1-way FSTs
   - to account for full reduplication – 2-way FST
   - to separate tone from segements – MT-FST
3. Proposed model: 2-way MT-FST
4. Conclusion

Language background

Shupamem is an understudied Grassfields Bantu language of the Western Province of central Cameroon spoken by approximately 420,000 speakers (Eberhard et al., 2021)

LANGUAGE BACKGROUND

Shupamem is a tonal language that exhibits four contrastive surface tones:

- **kà** 'onion' (L)
- **ká** 'fry.IMP (H)
- **kǎ** 'remove.IMP' (LH)
- **kâ** 'may' (HL)

The data presented here are original and have been collected from one native speaker of Shupamem, Abdoulaye Laziz Nchare, for the past two years in NYC.

Reduplication in Shupamem

Reduplication is a very common morphological process cross-linguistically. Approximately 75% of world languages exhibit either partial reduplication or full reduplication (Rubino, 2013).

In Shupamem, both nouns and verbs undergo full reduplication. The former reduplicate to create plurals, and the latter to introduce semantic contrast.

NOUN REDUPLICATION

- Nouns:
  là ‘trap’ – **lǎ là** ‘traps’
  kám ‘crab’ – **kâm kàm** ‘crabs’

	là	→	**lǎ là**	**kám**	→	**kâm kàm**
Tones:	L		LH L	H		HL L
Segments:	la		la la	kam		kam kam

VERB REDUPLICATION

- Verbs:
  fí ‘froth.IMP’ – **fí f$^{\downarrow}$í** ‘froth.IMP (as opposed to stirring)’
  lăp ‘heat.IMP’ – **lăp l$^{\downarrow}$áp** ‘heat.IMP (as opposed to freezing)’

	fí	→	**fí f$^{\downarrow}$í**	**lăp**	→	**lăp l$^{\downarrow}$áp**
Tones:	H		H H	LH		LH H
Segments:	fi		fi fi	lap		lap lap

TONES OPERATE SEPARATELY FROM SEGMENTS

Tones on a reduplicated phrase can be affected by neighboring tones.

- **ndáp** ‘house’ → **ndâp ndàp** ‘houses’

ndâp	**ndáp**	**ɣàʔʃɨ̀**
house	RED	be-big.PL

‘Houses are big’

Tones operate separately from segments

Tones on a reduplicated phrase can be affected by neighboring tones.

- **ndáp** ‘house’ → **ndâp ndàp** ‘houses’

ndâp	**ndáp**	**ɣàʔʃɨ̀**
house	RED	be-big.PL

‘Houses are big’

- **kă** ‘fry.IMP’ → **kă k↓á** ‘fry.IMP (as opposed to boiling)’

ḯ	**tã̀ː**	**ŋ-kâ**	**kă**
3SG	PROG	IMPF-peel	RED

‘He is peeling (as opposed to cutting).’

Understudied languages are important

Shupamem exhibits two important linguistics phenomena that are interesting from the perspective of finite-state automata:

UNDERSTUDIED LANGUAGES ARE IMPORTANT

Shupamem exhibits two important linguistics phenomena that are interesting from the perspective of finite-state automata:

1. full reduplication

UNDERSTUDIED LANGUAGES ARE IMPORTANT

Shupamem exhibits two important linguistics phenomena that are interesting from the perspective of finite-state automata:

1. full reduplication → whose output is not a regular language, and consequently cannot be modelled with 1-way FST

UNDERSTUDIED LANGUAGES ARE IMPORTANT

Shupamem exhibits two important linguistics phenomena that are interesting from the perspective of finite-state automata:

1. full reduplication → whose output is not a regular language, and consequently cannot be modelled with 1-way FST
2. independence of segmental and tonal tiers

UNDERSTUDIED LANGUAGES ARE IMPORTANT

Shupamem exhibits two important linguistics phenomena that are interesting from the perspective of finite-state automata:

1. full reduplication → whose output is not a regular language, and consequently cannot be modelled with 1-way FST
2. independence of segmental and tonal tiers → which requires a 2-level (autosegmental) representation

2-WAY FST FOR FULL REDUPLICATION

Dolatian and Heinz (2020) addresses the challenge of modelling full reduplication with 2-way FST. 2-way FSTs are allowed to move back and forth on the input tape but not on the output tape.

2-way FST:

- reads the input tape from left to right and outputs the first copy of the input string
- once it reaches the end of the word ($\ltimes$), it rewinds and iterates the string from right to left outputting nothing and stops when it sees the beginning of the word symbol ($\rtimes$)
- reads the input tape again outputting second copy of a word

AN EXAMPLE OF 2-WAY FST

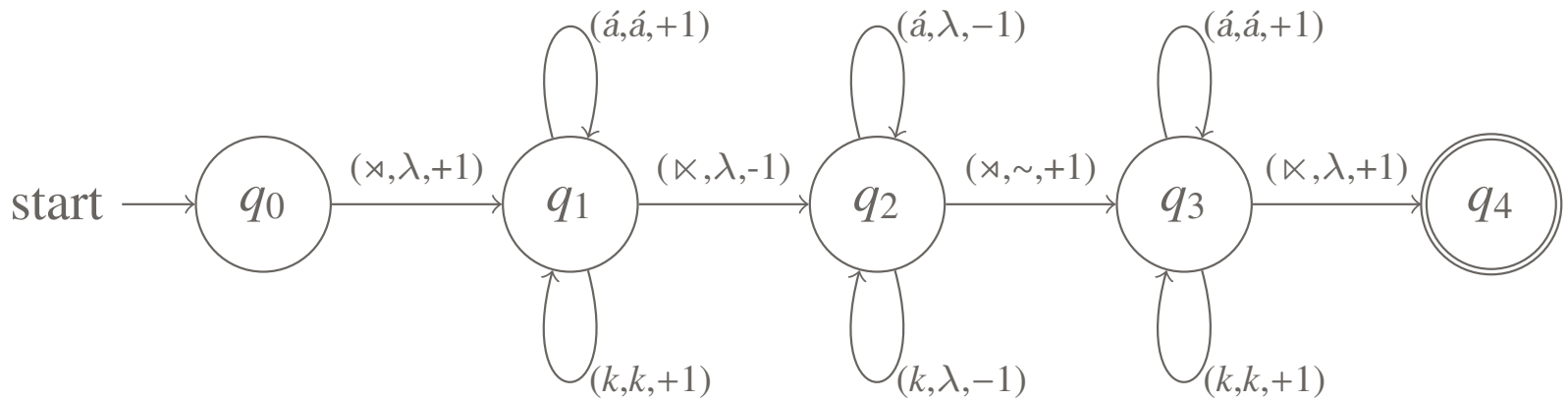


Figure 1: 2-way FST for total reduplication of *ká* 'fry.IMP → *ká ká* 'fry.IMP as opposed to boiling'

Autosegmental representations with MT-FST

Autosegmental representations can be mimicked using finite-state machines, in particular, MT-FSTs (Wiebe, 1992; Dolatian and Rawski, 2020a; Rawski and Dolatian, 2020).

MT-FST:

- can operate on more than one level (tape)
- can take a single string as input and output multiple string on different tapes
- can do the reverse: take a multiple strings as the input and output a single string
- can operate on different tapes simultaneously or it can operate in an asynchronous fashion

AN EXAMPLE OF MT-FST

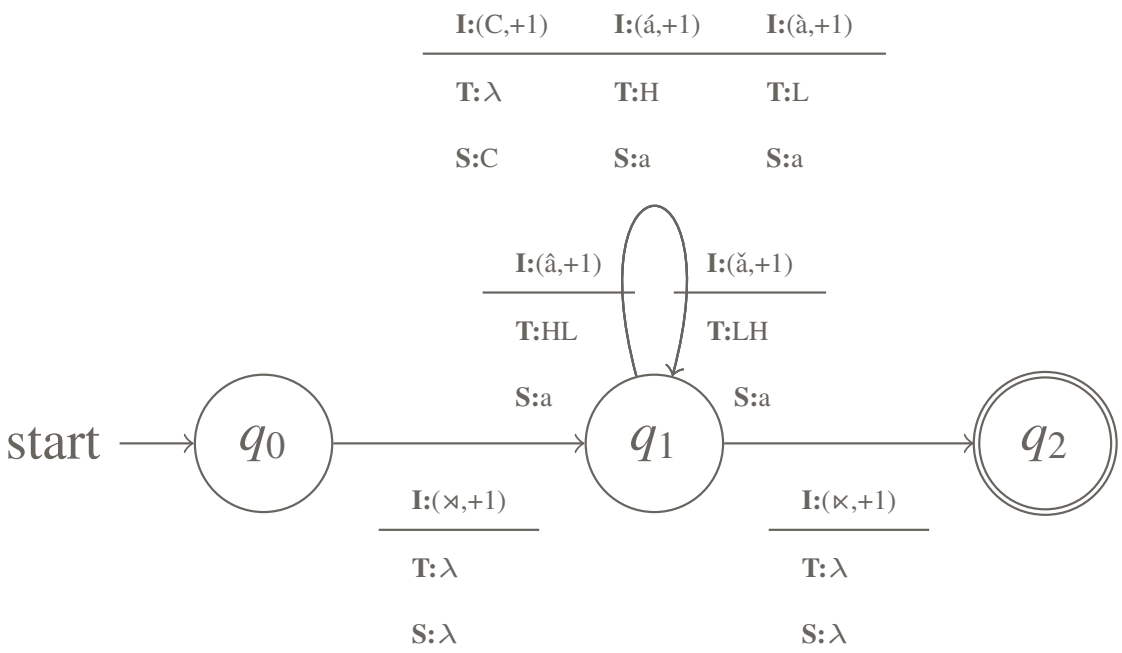
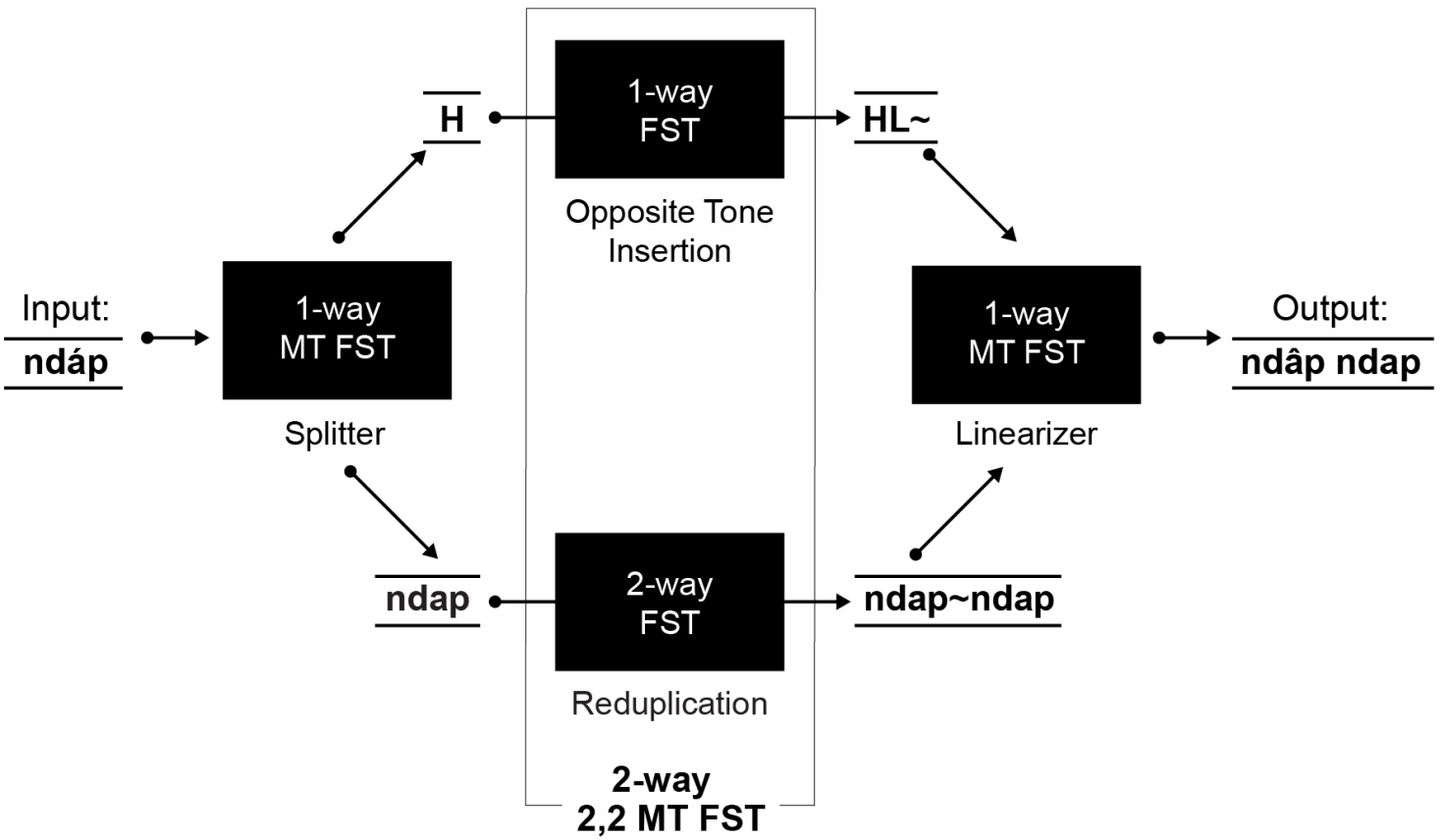


Figure 2: MT-FST that splits a linearized input (*ndáp* 'house') into tones and segments (*ndap*, *H*)

SYNTHESIS OF 2-WAY AND MT-FST

In order to model full reduplication and to faithfully represent tonal processes separate from the segmental ones, we propose a synthesis of 2-way and MT-FST.



2-WAY 2,2 MT-FST

- 2,2 MT → both input and output are represented on two tapes: Tonal and Segmental

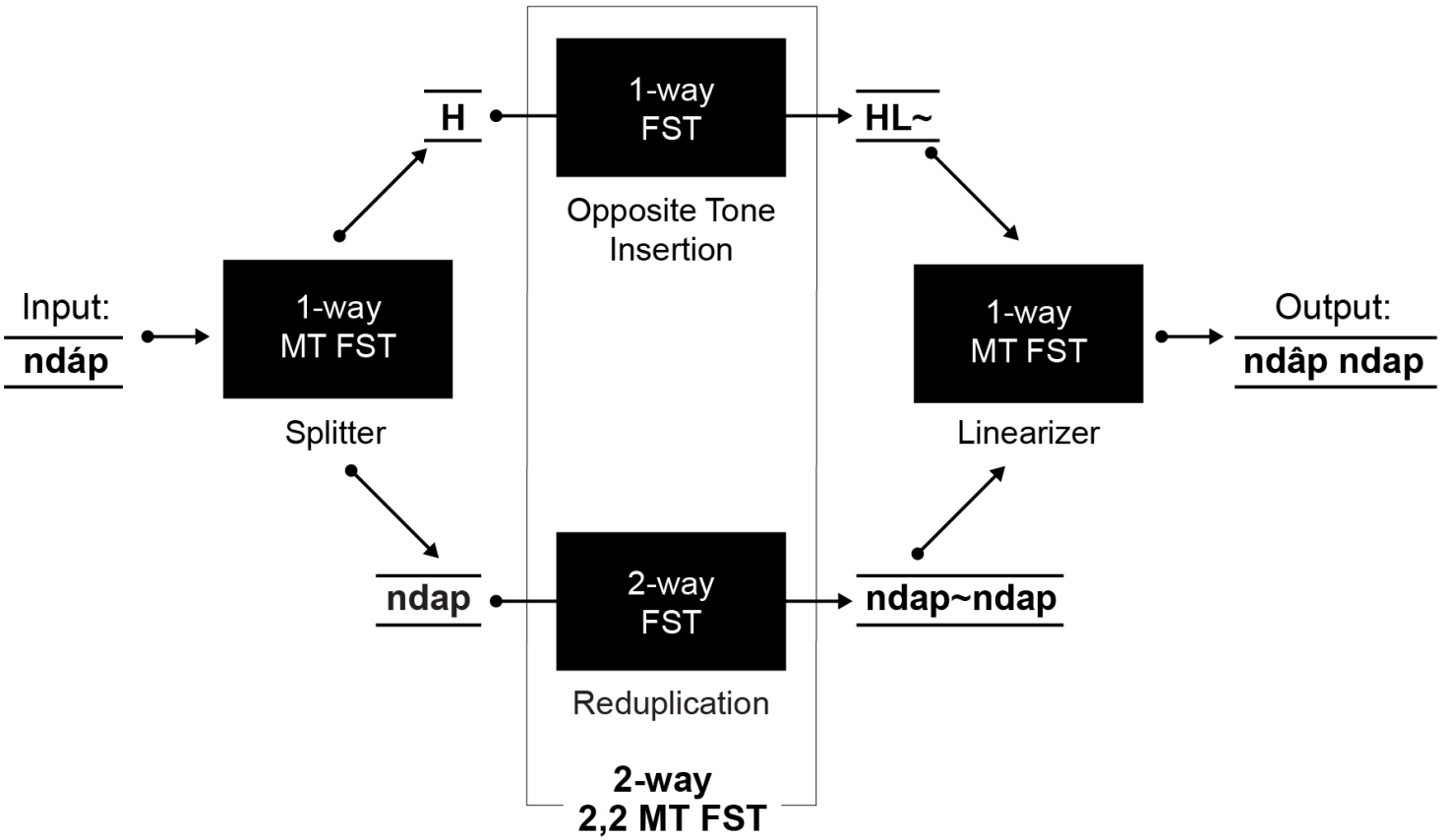
2-WAY 2,2 MT-FST

- 2,2 MT → both input and output are represented on two tapes: Tonal and Segmental
- 1-way FST → Tonal Tape: Opposite Tone Insertion (Markowska, 2020) inserts an opposite to the lexical tone (H → HL~)

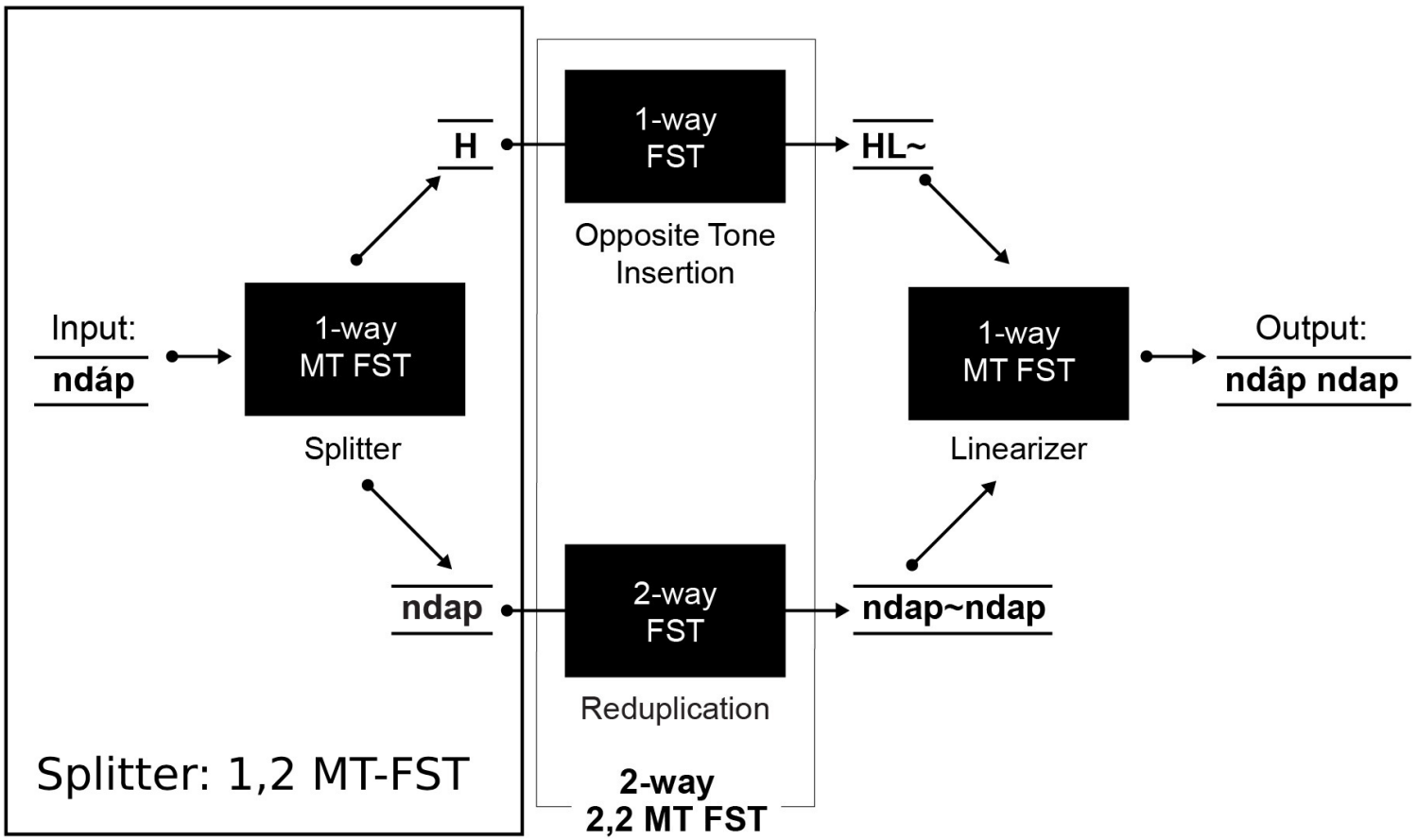
2-WAY 2,2 MT-FST

- 2,2 MT → both input and output are represented on two tapes: Tonal and Segmental
- 1-way FST → Tonal Tape: Opposite Tone Insertion (Markowska, 2020) inserts an opposite to the lexical tone (H → HL~)
- 2-way FST → Segmental Tape: full reduplication (ndap → ndap~ndap)

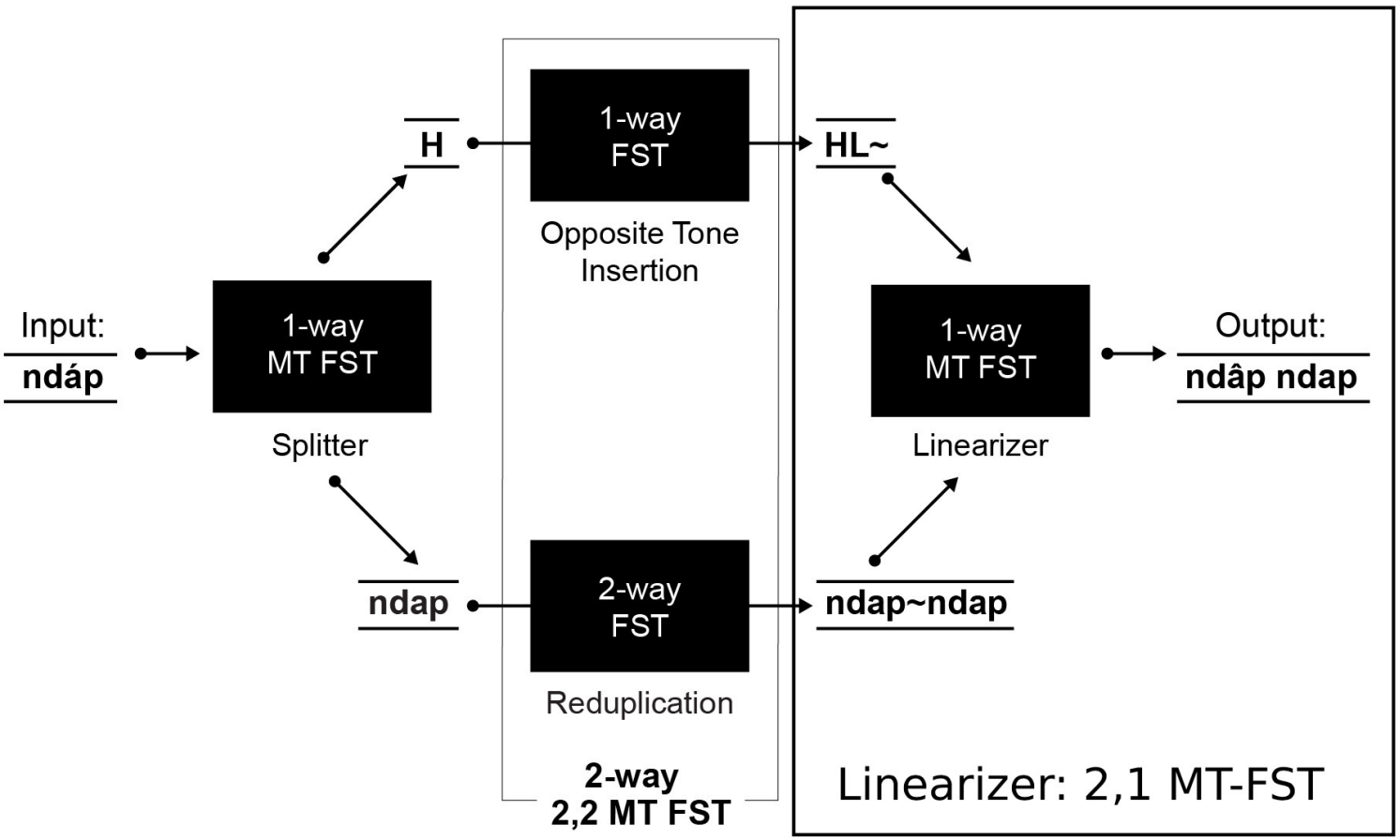
THE SPLITTER AND THE LINEARIZER



SPLITTER


Input:
ndáp
1-way
MT FST
Splitter
H
ndap
1-way
FST
Opposite Tone
Insertion
2-way
FST
Reduplication
2-way
2,2 MT FST
HL~
ndap~ndap
1-way
MT FST
Linearizer
Output:
ndâp ndap
Splitter: 1,2 MT-FST

LINEARIZER


Input:
ndáp
1-way
MT FST
Splitter
H
1-way
FST
Opposite Tone
Insertion
ndap
2-way
FST
Reduplication
2-way
2,2 MT FST
HL~
1-way
MT FST
Linearizer
ndap~ndap
Output:
ndâp ndap
Linearizer: 2,1 MT-FST

SUMMARY AND BENEFITS OF THE PROPOSED MODEL

- We proposed a synthesis of 2-way and 1-way MT-FSTs in order to account for Shupamem reduplication and the tonal processes related to it

SUMMARY AND BENEFITS OF THE PROPOSED MODEL

- We proposed a synthesis of 2-way and 1-way MT-FSTs in order to account for Shupamem reduplication and the tonal processes related to it
- We predict that such machine works for all tonal languages exhibiting reduplicative processes, e.g. Adhola (Kaplan, 2006) or Kikerewe (Odden, 1996)

SUMMARY AND BENEFITS OF THE PROPOSED MODEL

- We proposed a synthesis of 2-way and 1-way MT-FSTs in order to account for Shupamem reduplication and the tonal processes related to it
- We predict that such machine works for all tonal languages exhibiting reduplicative processes, e.g. Adhola (Kaplan, 2006) or Kikerewe (Odden, 1996)
- Such machine both represents complex morphonological processes and faithfully mimics linguistics generalizations regarding tones (Goldsmith, 1976)

SUMMARY AND BENEFITS OF THE PROPOSED MODEL

- We proposed a synthesis of 2-way and 1-way MT-FSTs in order to account for Shupamem reduplication and the tonal processes related to it
- We predict that such machine works for all tonal languages exhibiting reduplicative processes, e.g. Adhola (Kaplan, 2006) or Kikerewe (Odden, 1996)
- Such machine both represents complex morphonological processes and faithfully mimics linguistics generalizations regarding tones (Goldsmith, 1976)
- Tonal locality yields simpler computation that can be accounted for by less expressive local functions (Chandlee, 2014)

Thank you!

Chandlee, J. (2014). Strictly local phonological processes. Ph. D. thesis, University of Delaware.

Dolatian, H. and J. Heinz (2020). Computing and classifying reduplication with 2-way finite-state transducers. Journal of Language Modelling 8(1), 179–250.

Dolatian, H. and J. Rawski (2020a). Computational locality in nonlinear morphophonology. Ms., Stony Brook University.

Dolatian, H. and J. Rawski (2020b). Multi-input strictly local functions for templatic morphology. Proceedings of the Society for Computation in Linguistics 3(28).

Eberhard, D. M., G. F. Simmons, and C. D. Fenning (2021). Enthologue: Languages of the World. 24th edition. Dallas, Texas: SIL International.

Goldsmith, J. (1976). Autosegmental Phonology. Ph. D. thesis, Massachusetts Institute of Technology.

Kaplan, A. F. (2006). Tonal and morphological identity in reduplication. In Proceedings of the Annual Meeting of the Berkeley Linguistics Society, Volume 31.

Markowska, M. (2020). Tones in Shupamem reduplication. CUNY Academic Works.

Odden, D. (1996). Patterns of reduplication in kikerewe. OSU WPL 48, 111–148.

Rawski, J. and H. Dolatian (2020). Multi-input strict local functions for tonal phonology. Proceedings of the Society for Computation in Linguistics 3(1), 245–260.

Rubino, C. (2013). Reduplication. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Wiebe, B. (1992). Modelling autosegmental phonology with multi-tape finite state transducers.