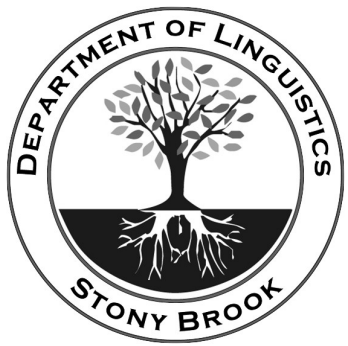
Regular Transformations in Linguistics

Jeffrey Heinz

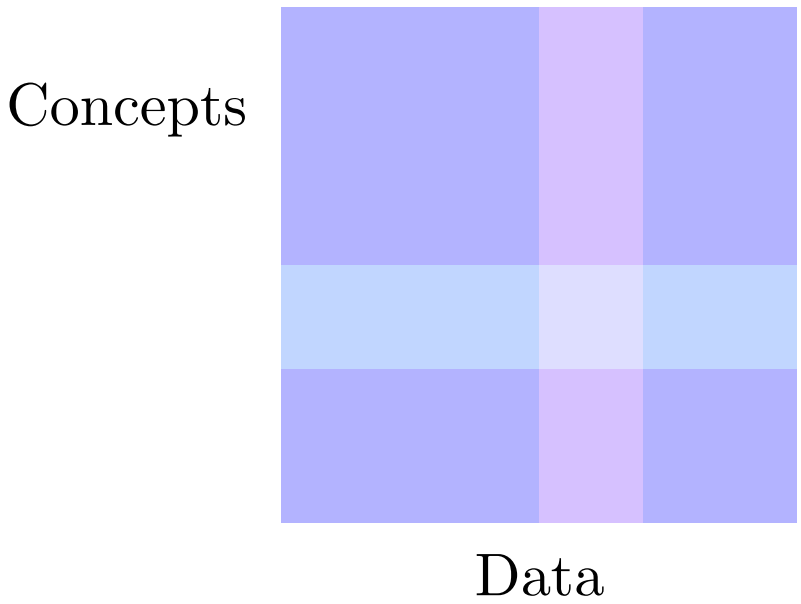


Schloss Dagsthul
16.05.2023

Motivation

1. What kind of computational resources and information are both necessary and sufficient to account for linguistic transformations?
2. How can these transformations be learned from examples?

⋆ How can we characterize successful learning, including characterizing the amount and kind of data and the kinds of target concepts for such learning to occur?



Overview of Answers

1. Regular transformations provide an upper bound.

2. Particular subclasses appear to be sufficient, and these subclasses are generally (much) less than what FO(<) provides.

3. These subclasses are learnable with relatively low time and data complexity in ways the full class of regular transformations is not.

Quiztime! Possible Linguistic Processes

- Prefixation
  abbb $\mapsto$ cabbb
- Affixation Mod 2
  abbb $\mapsto$ cabbb
  abb $\mapsto$ abbc
- Bounded Spreading
  abbb $\mapsto$ aabb
- Unbounded Spreading
  abbb $\mapsto$ aaaa
  abbbdb $\mapsto$ aaaadb
- Projected Unbounded Spreading
  $c^{10}ac^{10}bc^{10}bc^{10}bc^{10}$
  $\mapsto$ $c^{10}ac^{10}ac^{10}ac^{10}ac^{10}$
- Sour Grapes
  abbbb $\mapsto$aaaaa
  abbcb $\mapsto$abbcb
- Two-sided Unbounded Spread
  abba $\mapsto$aaaa
  abbb $\mapsto$abbb
- Majority Rules
  abbaa $\mapsto$aaaaa
  abbab $\mapsto$bbbbb
- Partial Copying
  abcd $\mapsto$ ababcd
- Full Copying
  abcd $\mapsto$ abcdabcd
- Triplication
  abcd $\mapsto$ abcdabcdabcd
- Squaring
  abcd $\mapsto$abcdabcdabcdabcd
- Iterated Prefix Copying
  abcd $\mapsto$ a ab abc abcd

Quiztime! Possible Linguistic Processes

✓ Prefixation
abbb $\mapsto$ cabbb

✗ Affixation Mod 2
abbb $\mapsto$ cabbb
abb $\mapsto$ abbc

✓ Bounded Spreading
abbb $\mapsto$ aabb

✓ Unbounded Spreading
abbb $\mapsto$ aaaa
abbbdb $\mapsto$ aaaadb

✓ Projected Unbounded Spreading
$c^{10}ac^{10}bc^{10}bc^{10}bc^{10}$
$\mapsto$ $c^{10}ac^{10}ac^{10}ac^{10}ac^{10}$

✓✗ Sour Grapes
abbbb $\mapsto$aaaaa
abbcb $\mapsto$abbcb

✓ Two-sided Unbounded Spread
abba $\mapsto$aaaa
abbb $\mapsto$abbb

✗ Majority Rules
abbaa $\mapsto$aaaaa
abbab $\mapsto$bbbbb

✓ Partial Copying
abcd $\mapsto$ ababcd

✓ Full Copying
abcd $\mapsto$ abcdabcd

✓ Triplication
abcd $\mapsto$ abcdabcdabcd

✗ Squaring
abcd $\mapsto$abcdabcdabcdabcd

✗ Iterated Prefix Copying
abcd $\mapsto$ a ab abc abcd

Examples 1: Phonology

In Navajo, sibilant sounds like {s, z} are pronounced {ʃ, ʒ} if they are followed anywhere by {ʃ, ʒ} (Sapir and Hoijer 1967).

a.	/sì-ʔá/	↦	sì-ʔá	'a round object lies'
b.	/sì-tí/	↦	sì-tí	'he is lying'
c.	/sì-ɣìʃ/	↦	**ʃ**ì-ɣì**ʃ**	'it lies bent, curved'
d.	/sì-teːʒ/	↦	**ʃ**iteː**ʒ**	'we (dual) are lying'

Projected Unbounded Spreading from Right to Left

Examples 1: Phonology

In Luganda, every toneless vowel between two high-toned vowels is pronounced with a high tone forming a high-tone 'plateau' (Hyman and Katamba 2010).

a.	/mu-tund-a+bi-kópo/	↦	mutunda-bikópo	'cup-seller'
	L L L L H L L	↦	L L L L H L L	
b.	/mu-tém-a+bi-sikí/	↦	mutémá-bísíkî	'log-chopper'
	L H L L H L L	↦	L H H H H L F	

Two-sided Unbounded Spread

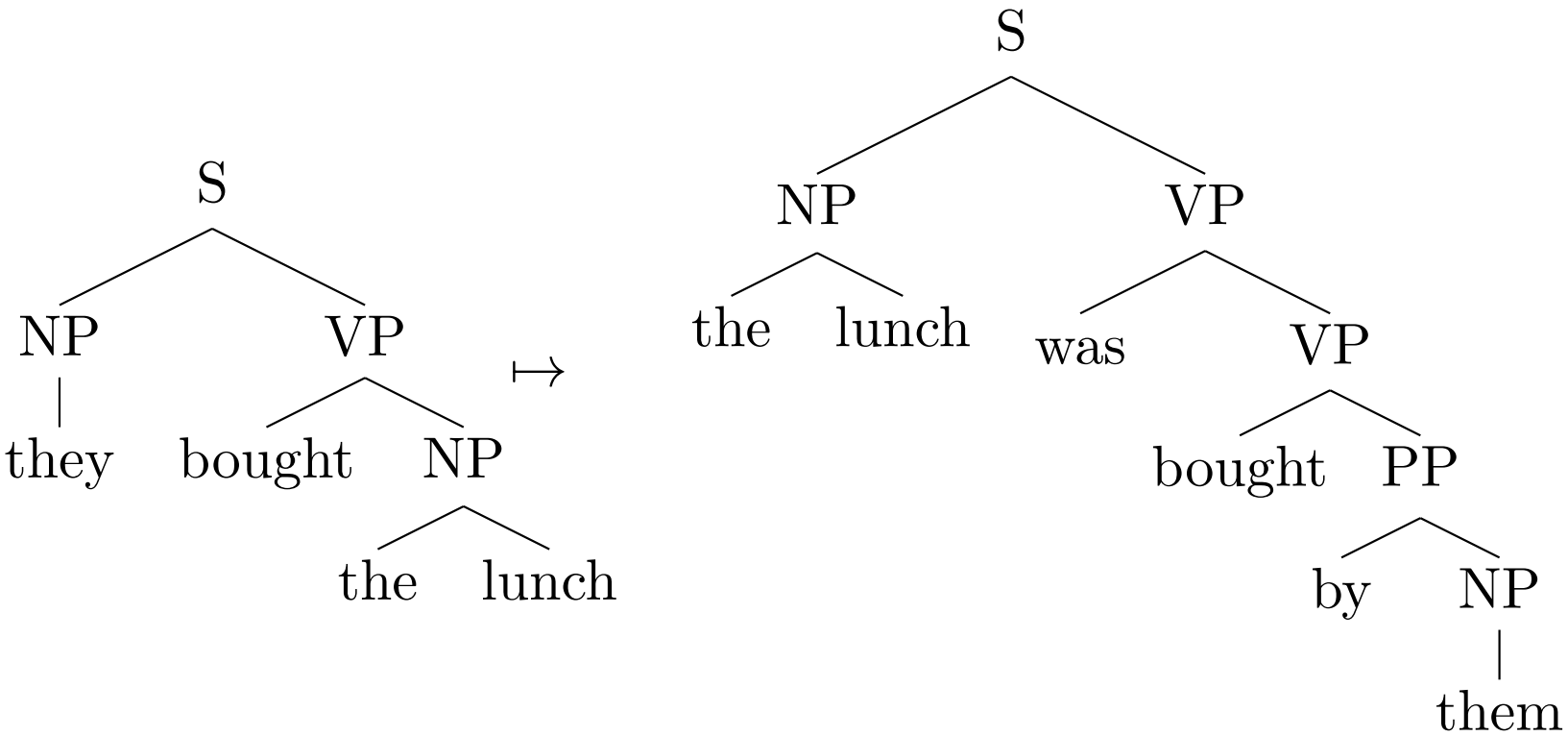
Examples 2: Morphology

Morphological copying, "Reduplication", (83% of languages have full copying, 74% have partial bounded copying)

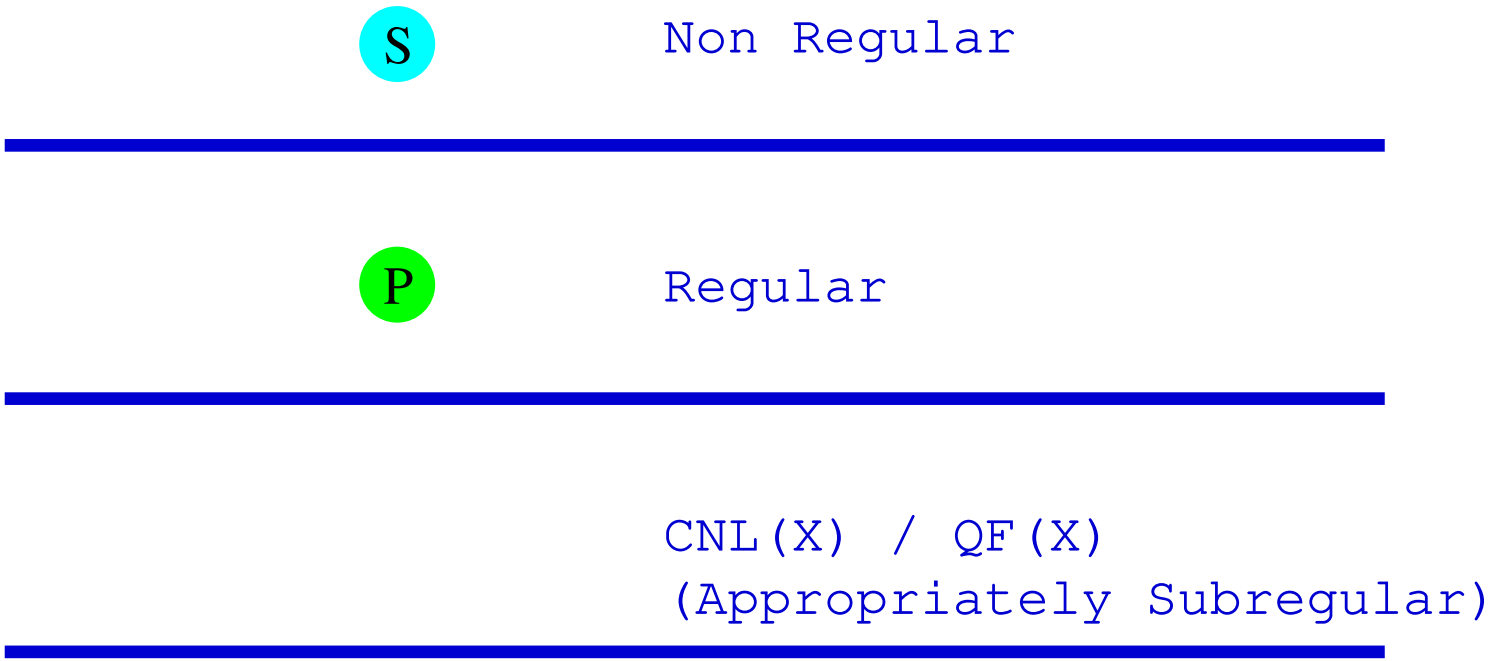
- **Full copying** (Indonesian, Cohn 1989)
  buku $\mapsto$ buku $\sim$ buku — 'book' $\mapsto$ 'books'
  wanita $\mapsto$ wanita $\sim$ wanita — 'woman' $\mapsto$ 'women'
- **Initial-CVC copying** (Panganisan, Rubino 2005:11)
  baley $\mapsto$ bal $\sim$ baley — 'town' $\mapsto$ 'towns'
- **Initial-CVC copying and opposite-edge placement** (Koryat, Riggle 2004:3)
  qanga $\mapsto$ qanga $\sim$ qan — 'fire' $\mapsto$ 'fire (ABS)'
- **Root-final CVC copying and word-initial placement** (Madurese, Brown 2017:964)
  pa-jalan-an $\mapsto$ lan $\sim$ pa-jalan-an
  'pedestrian' $\mapsto$ 'pedestrians'

Examples 3: Syntax

The active to passive transformation can be analyzed in terms of transformations over syntactic trees.

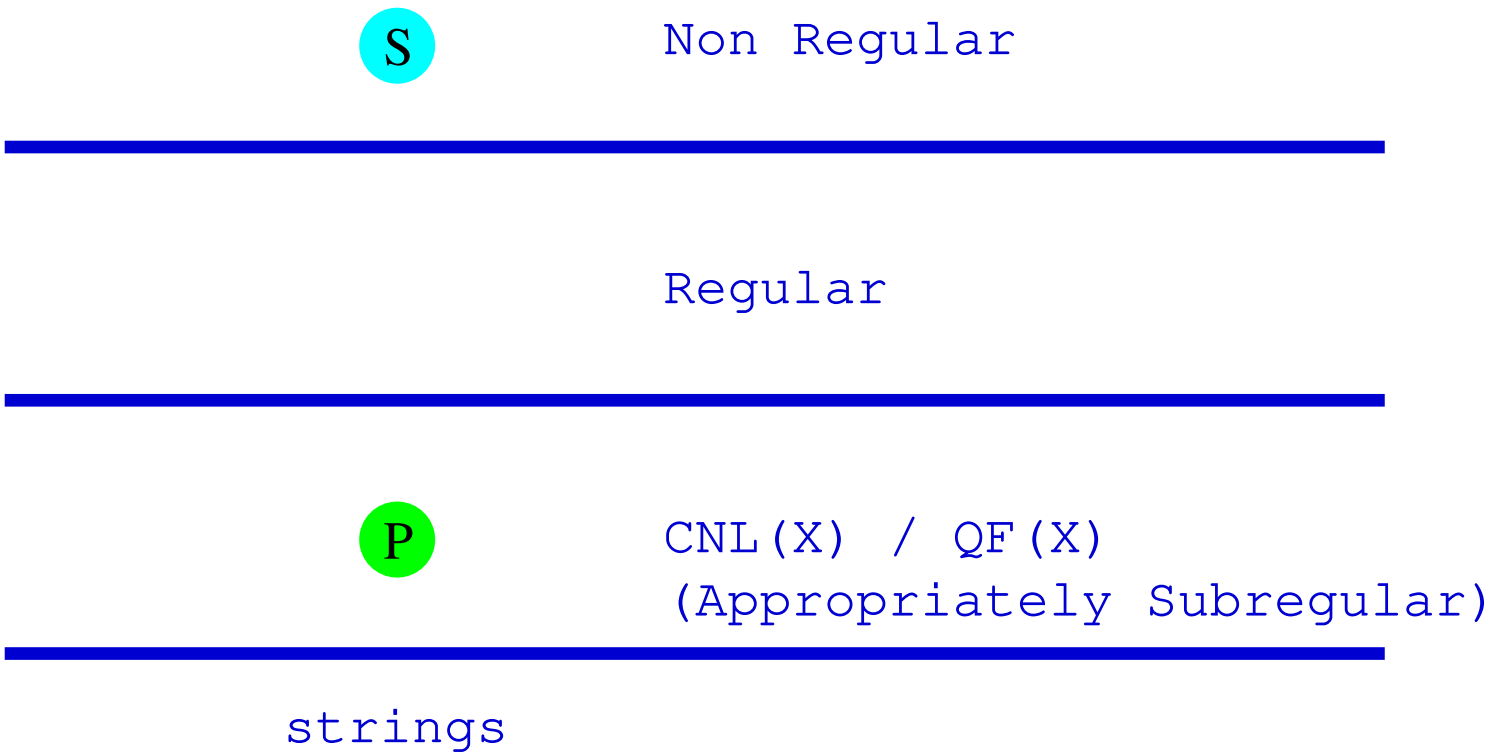


Sub-regularity in Language



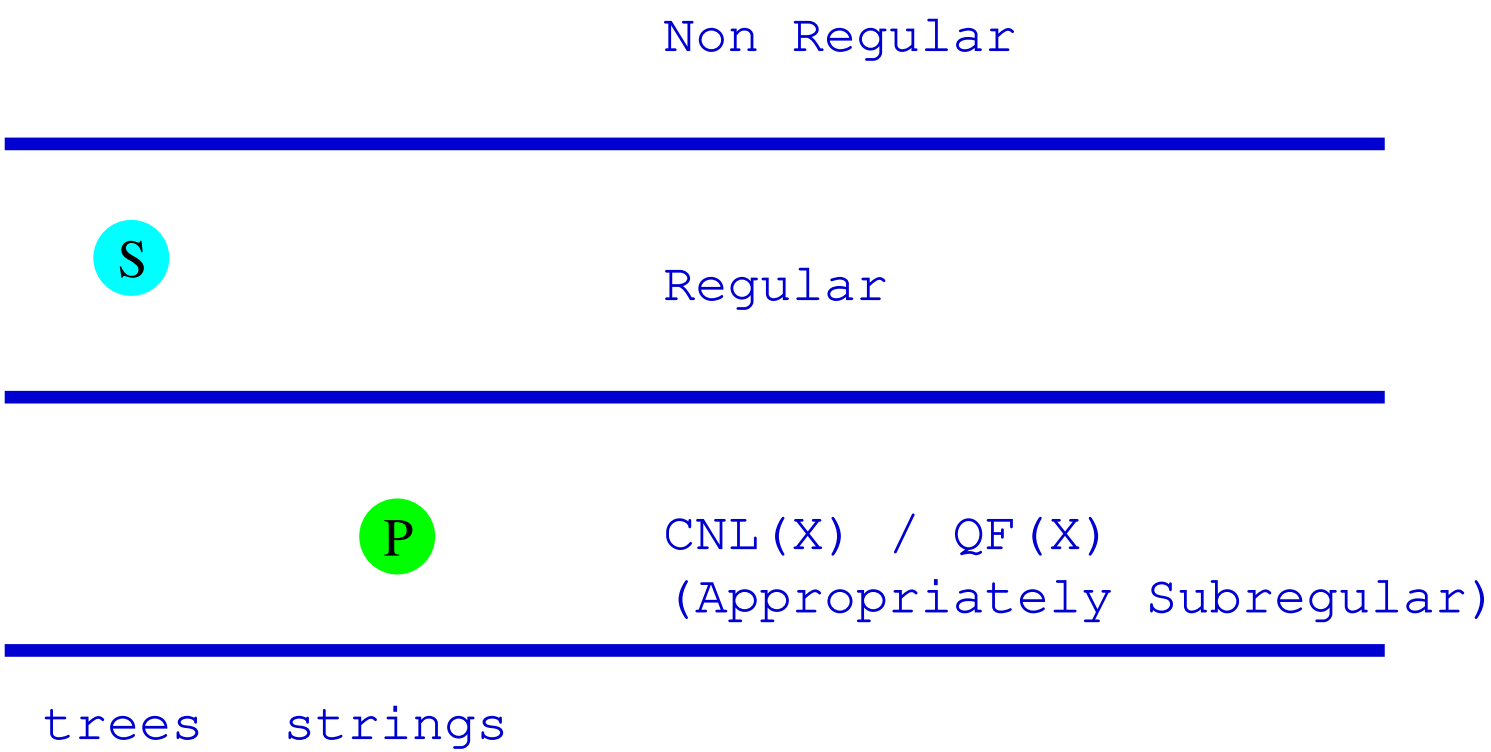
20th century view

Sub-regularity in Language



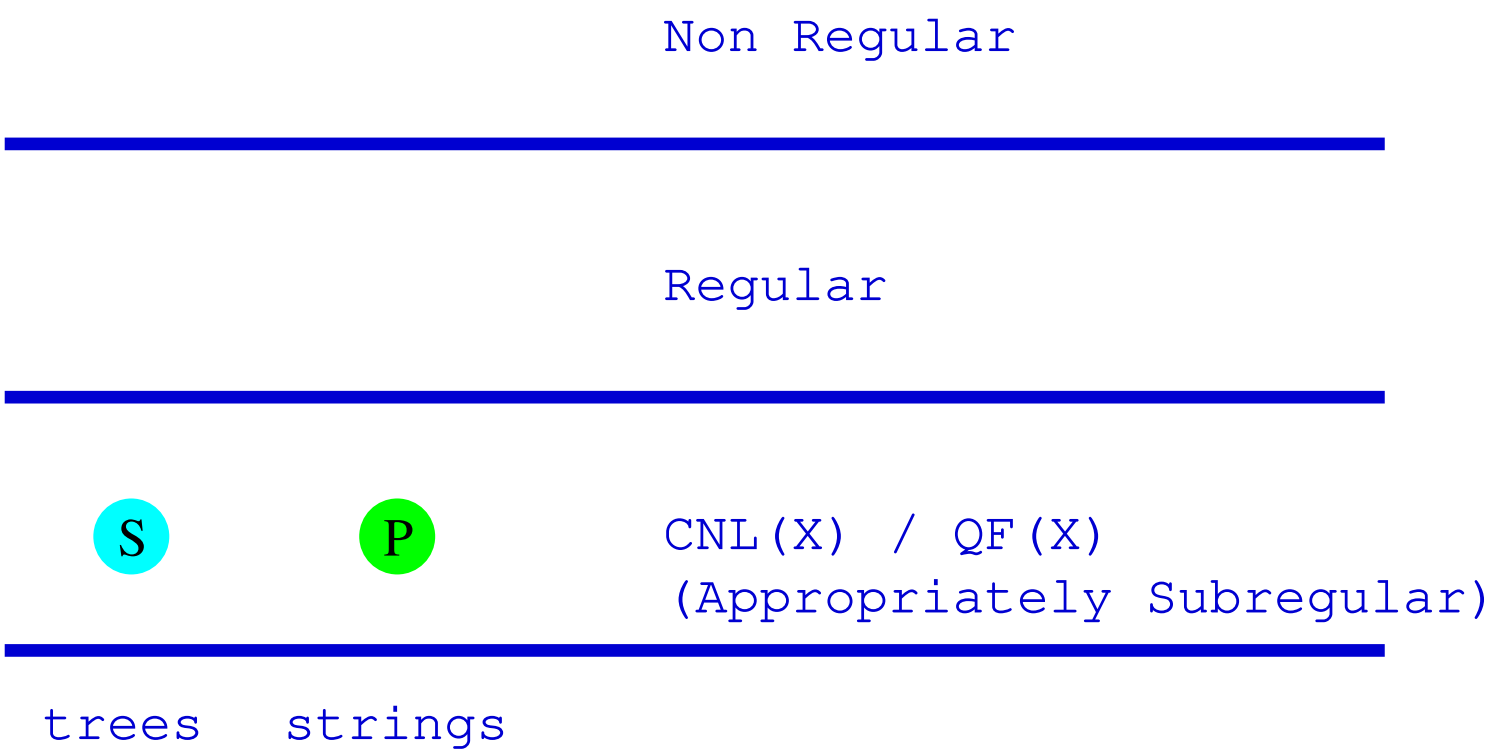
(Heinz 2018)

Sub-regularity in Language



(Stabler 2019)

Sub-regularity in Language



(Graf 2022)

Where we started

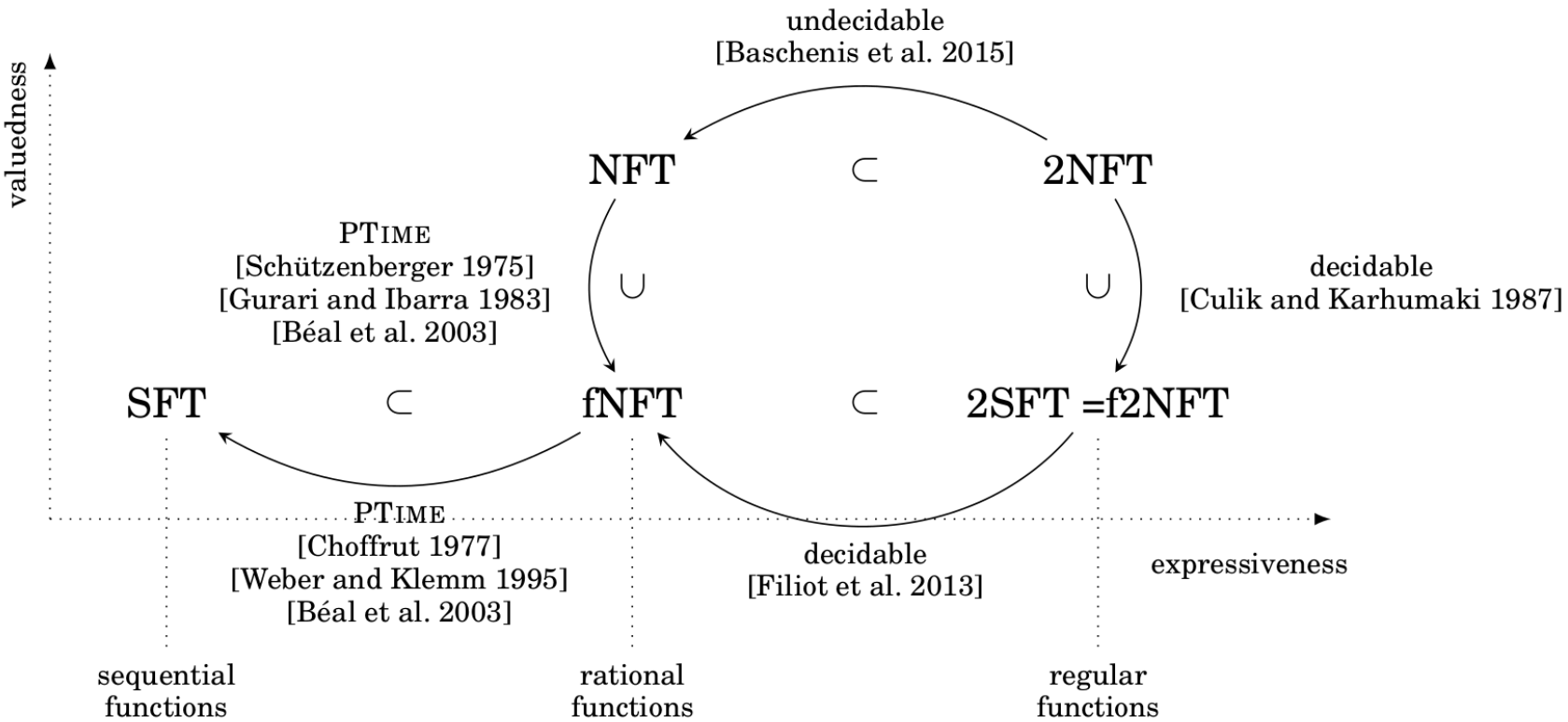


Fig. 3. A landscape of transducers of finite words.

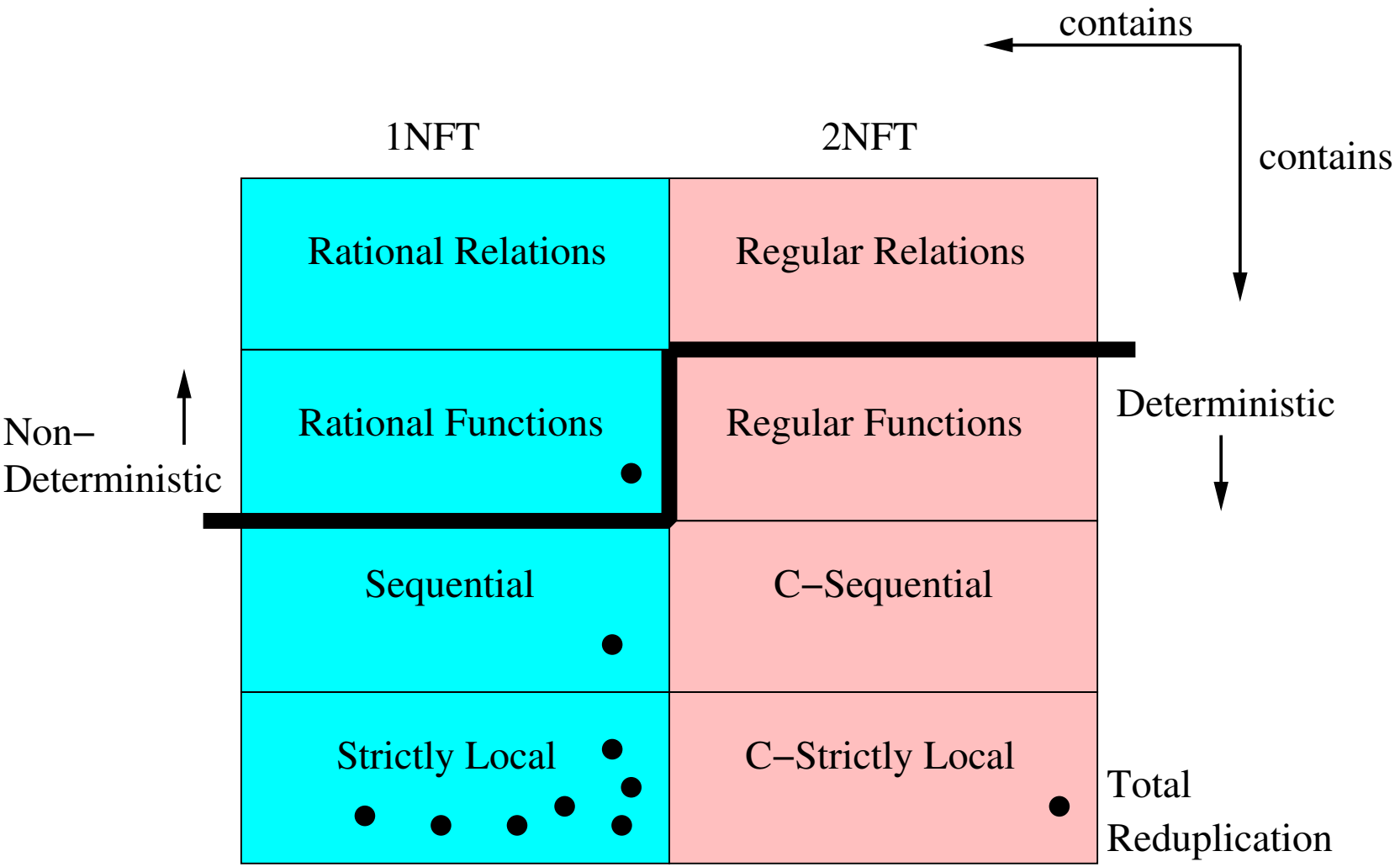
(Filiot and Reynier 2016 ACM SIGLOG)

Round 1 Summary of string functions in morphology and phonology

1. Many morphological and phonological operations are either left or right sequential, except for sour grapes, two-sided unbounded spreading and full copying
2. Natural Language processing and computational linguistics have focused on rational relations, leaving full copying as a"problem" (Roark and Sproat 2007).
3. Two reasons:
   - Those communities were unaware of 2-way transducers, Courcelle-style logical transductions, and streaming string transducers
   - Rational relations are invertible, which provides one model for both generation and analysis.

String Relations: Room at the bottom

A more articulated view



(Filiot and Reynier 2016, Chandlee 2017, Dolatian and Heinz 2020)

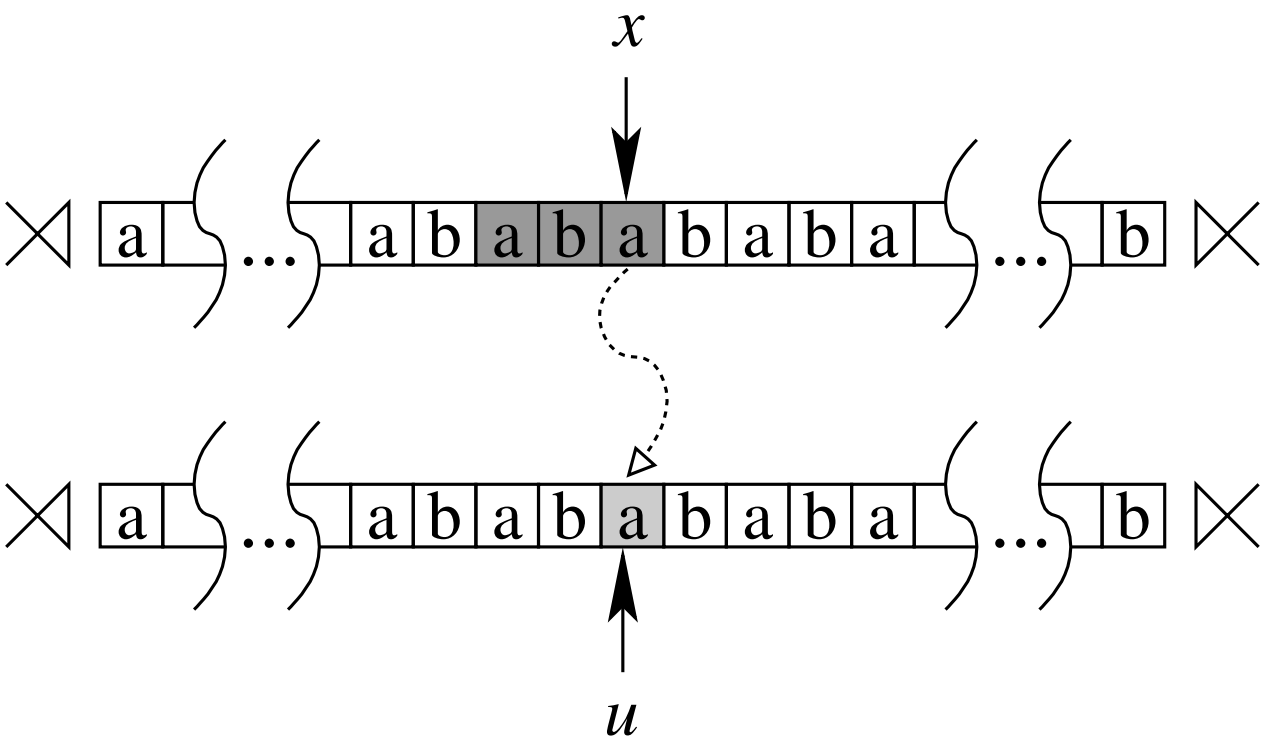
Local functions

Dans cet article, nous introduisons les fonctions p-locales et les fonctions p-sous locales (où p est un entier strictement positif) et nous les caractérisons par une propriété simple de leur semigroupe syntactique : ce semigroupe doit satisfaire l'équation $yx_1 \ldots x_p = x_1 \ldots x_p$. Nous en déduisons quelques propriétés des fonctions p- locales.

(Vaysse 1986, see also Schiffler 1973)

Local functions as scanners with outputs

Vaysse defines local functions in terms of a sliding window over the input which produces corresponding outputs.



Local functions, Nerode/transducers

1. Independently, Chandlee et al. 2014 define these functions in terms of their Nerode equivalence classes: two inputs x and y are are Nerode equivalent (that is have the same functional residuals) whenever they share the same k-sized suffix

$$\textsc{tails}_f(x) = \{(y, v) \mid f(xy) = \textsc{lcp}(f(x\Sigma^*))v\}$$

$$x_1 \sim_f^N x_2 \leftrightarrow \textsc{tails}_f(x_1) = \textsc{tails}_f(x_2)$$

$$\textsc{suff}_k(x_1) = \textsc{suff}_k(x_2) \Rightarrow \textsc{tails}_f(x) = \textsc{tails}_f(x)$$

2. Their transducer characterization is the same as Vaysse's transducer characterization.

Local functions, logically

1. Lindell and Chandlee (2016) prove that quantifier-free interpretations of strings to strings (modeled with the successor and predecessor functions) are local functions (but not quite vice versa).

2. Provides a way to define "local" transformations over any model-theoretic representations, not just strings!

3. Subsequent work has explored this with respect to linguistically motivated graph representations of words (Jardine 2016, Strother-Garcia 2018, Jardine et al. 2021)

Local functions, linguistics + learning

1. Many phonological and morphological processes can be described with local functions such as Prefixation, Suffixation, Bounded Spreading . . . but not all. (Chandlee 2014, Chandlee 2017, Chandlee and Heinz 2018, Chandlee et al. 2018)

2. Chandlee et al. 2014 provide better time and complexity bounds for learning these functions than previously known (see also Jardine et al. 2014).

3. Current work: doing better with alternative representations of strings.

Local functions, algebraically

- Vaysse 1986 showed that local functions have a *definite* algebraic structure.

$$Se = e$$

$$(\text{recalling } yx_1x_2 \dots x_p = x_1x_2 \dots x_p)$$

- Syntactic Monoids for Transducers are based on the Myhill equivalence relation:

$$\textsc{contexts}_f(x) = \{(w, y, v) \mid f(wxy) = \textsc{lcp}(f(wx\Sigma^*))v\}$$

$$x_1 \sim_f^M x_2 \leftrightarrow \textsc{contexts}_f(x_1) = \textsc{contexts}_f(x_2)$$

(Filiot, Gauwin, Lhote 2016)

Reverse Definite Structures

$$eS = e$$

$$(x_1 x_2 \ldots x_p y = x_1 x_2 \ldots x_p)$$

1. Transducers come in two kinds: left and right.
2. Unbounded Spreading in natural languages includes both left and right Reverse Definite.

(Lambert 2022)

Tier-projections (Lambert 2022)

1. A tier T is a subset of the alphabet which contain "salient" symbols.

   In Navajo T equals the sibilants like {s,z,ʃ,ʒ}.

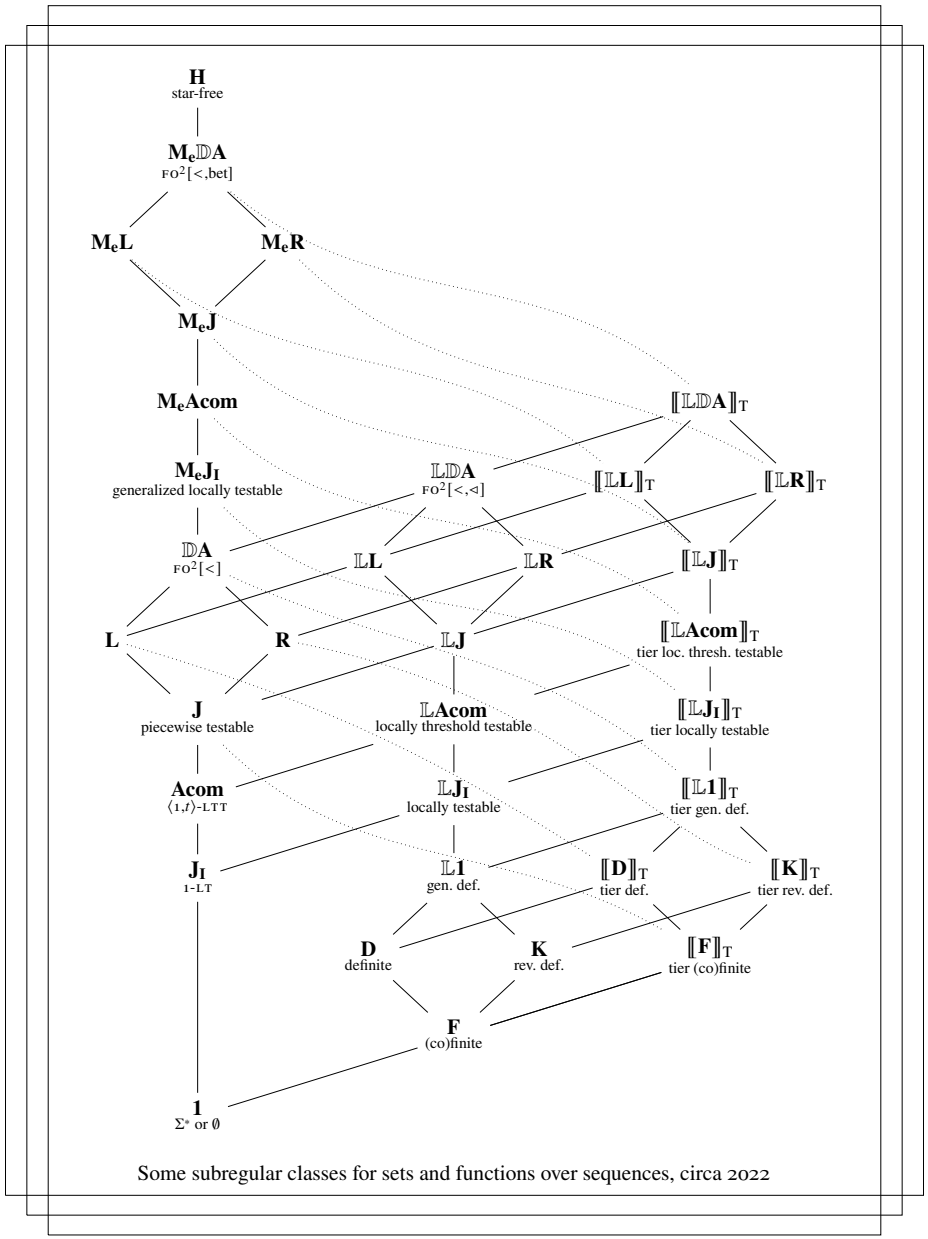
2. Projecting a string to a tier removes the non-salient symbols.

   $$\text{sì-ɣìʃ} \mapsto \text{sʃ}$$

3. A function belongs to algebraic class "Tier X" if and only if its tier projection belongs to class X.
4. Lambert shows for each X that X $\subsetneq$ Tier X.

- Another flavor of Unbounded Spreading that is attested is Tier Definite.
- Projected Unbounded Spreading is Tier Reverse Definite, with both left and right directions attested..

Lambert 2022: Algebraic Hierarchy



Some subregular classes for sets and functions over sequences, circa 2022

Round 2 Summary of string functions in morphology and phonology

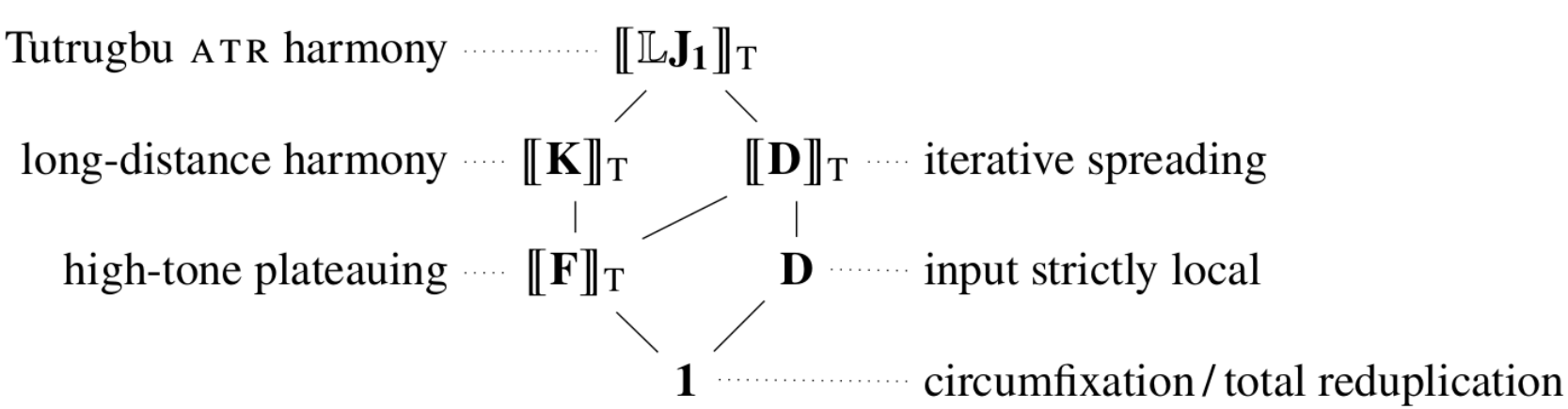


Figure 5.18: Some attested morphophonological functions.

Lambert (2022)

- Both Tutrugbu ATR Harmony (Sour Grapes) High Tone Plateauing (Two sided Unbounded Spreading) require non-deterministic or 2-way transducers.
- Their algebraic characterizations follow the methods of Carton and Dartois (2015).

Learning revisited

1. Each of the classes X $\leq$ Tier Locally Testable can be parameterized by a window size k and/or tier T.
2. Given k and T, the k-X classes can be learned from examples with relatively low time and data complexity.
3. Given only k, learning the tier T is possible for languages (Jardine and Heinz 2016, Jardine and McMullin 2017) and there is some work on functions as well (Burness and McMullin 2019) that still needs to be integrated into the algebraic perspective.

Local Tree Transductions

The states of the transducer correspond to local treelets of bounded depth.

1. Jing and Heinz (2020) restricts deterministic bottom-up tree transducers.
2. Graf (2020) takes a top-down approach.
3. Someone should look at QF interpretations of tree structures.

Conclusion

1. Regular transformations provide an upper bound on morpho-phonological processes in natural languages.

2. Particular subclasses appear to be sufficient, and these subclasses are generally (much) less than what FO($<$) provides.

3. These subclasses are learnable with relatively low time and data complexity in ways the full class of regular transformations is not.

Open Questions

1. Factoring transducers via composition and/or direct product
2. Learning factored representations of transducers
3. Subregular classes of tree transductions
4. Deterministic regular relations

Thank You