

Advantages of Constituency: Computational Perspectives on Samoan Word Prosody

Kristine M. Yu^(✉)

University of Massachusetts Amherst, Amherst, MA 01003, USA
krisyu@linguist.umass.edu

Abstract. In this paper, we computationally implement and compare grammars of Samoan stress patterns that refer to feet and that refer only to syllables in Karttunen’s finite state formalization of Optimality Theory, and in grammars that directly state restrictions on surface stress patterns. The grammars are defined and compared in the high-level language of *xfst* to engage closely with specific linguistic proposals. While succinctness (size of the grammar) is not affected by referring to feet in the direct grammars, in the OT formalism, the grammar with feet is clearly more succinct. Moreover, a striking difference between the direct and OT grammars is that the OT grammars suffer from scaling problems.

Keywords: Phonology · Stress · Finite state · Optimality theory

1 Introduction

A substantial body of work in theoretical phonology suggests that referring to phonological constituents—units such as feet, prosodic words and higher-level constituents—can capture generalizations in phonological patterns [15, 35, 43, 45]. But other work in theoretical phonology offers alternative ways to capture the same kinds of generalizations [28, 29, 47]. And strikingly, computational descriptions of phonological patterns have revealed strong structural universals without referring to constituents at all [17, 18]. Thus, while “we [phonologists] tend to help ourselves to prosodic domains without further comment” [46], there is in fact a puzzle here—*Do constituents make phonological grammars more succinct?*—which hasn’t yet been carefully investigated computationally.¹ The work here is an initial effort to begin to fill this gap.

In this paper, we implement and compare phonological grammars of stress patterns in Samoan monomorphs to assess whether grammars referring to feet are more succinct than grammars that refer only to syllables. Succinctness comparisons between grammars of the same ilk have appeared in [5, 8, 14, 33, 38, 44], among other work. We define all grammars in *xfst* [4], software for computing with finite

¹ Some computational work has defined phonological patterns in terms of tiers [6, 9, 30] from autosegmental theory [12], but tiers aren’t properly nested like constituents, e.g., it’s generally assumed that feet don’t straddle prosodic words.

state machines, and our comparison is a 2×2 experimental design because we compare grammars with and without feet in two formalisms: (a) Karttunen’s finite state Optimality Theory (OT) [27], which maps underlying forms to surface forms without an intermediate mapping to violations, by using a special composition operator “lenient composition”, and (b) a “direct” approach which directly regulates the surface patterns, e.g., [17, 22]. Our work thus addresses not only the role of constituents in phonology, but also the advantages and disadvantages of different formalisms. Our method of implementation and comparison is designed to engage closely and concretely with linguistic analyses and empirical data. The goal here is not to make general, abstract claims, but to question if a particular phenomenon motivates prosodic constituents, cf. case studies that examine where phonological and syntactic patterns fall in the Chomsky hierarchy. In doing this, we carefully state very specific assumptions and claims guided by proposals by phonologists about the phenomenon to gain a clearer understanding of the phenomenon and the proposals. The code implementing and testing the grammars is available at <https://github.com/krismyu/smo-constituency-feet>.

The rest of the paper is structured as follows: the remainder of this introductory section operationalizes *Do constituents make phonological grammars more succinct?* (Sect. 1.1) and describes the simplified language of stress patterns of Samoan that our grammars are designed to capture (Sect. 1.2). We describe the four grammars in Sect. 2: the direct account with feet (Sect. 2.2), the direct account with syllables only (Sect. 2.3), the OT account with feet (Sect. 2.4), and the OT account with syllables (Sect. 2.5). The discussion follows in Sect. 3, and we conclude with Sect. 4.

1.1 Operationalizing the Research Question

In this section, we define the concepts in our research question: phonological grammars, constituents, and succinctness.

Phonological Grammars and Constituents. In Formal Language Theory, a grammar is defined by a finite alphabet of terminal symbols, a finite set of non-terminal categories (which shares no members in common with the alphabet), a finite set of rewrite rules, and a start symbol that initiates the derivation [7, 41]. The set of strings that can be derived by the grammar is defined to be the language derived by the grammar. In a tree derived by the grammar, nodes are labeled with categories, and a set of nodes form a constituent if they are exhaustively dominated by a common node.

The nature of restrictions on the structure of rewrite rules determines the structural complexity of the grammar. A standard definition of a (right) regular grammar says that it is a grammar where the rules are restricted to the form $A \rightarrow aB$, and $A \rightarrow \epsilon$, where A and B are non-terminal categories and a is a terminal symbol. This restriction results in grammars where the only constituents of length > 1 are suffixes. But prosodic constituents in phonological theory include both prefixes and suffixes. For instance, given an alphabet of light and heavy

syllables $\Sigma = \{L, H\}$, suppose we defined a (right) regular grammar that could derive the string $LLHLL$. Then the suffix-constituents derived by the grammar would never be able to pick out the initial LL or the medial H as feet.

A regular language is one that can be derived by a regular grammar. It has been shown that (almost) all phonological patterns are regular [24, 26]. We are left with an apparent contradiction: if phonology has constituents that are both prefixes and suffixes, then how is it that phonology is regular? The critical point is that it's the *language*—the set of admissible surface patterns—that's been shown to be regular in phonology, not the grammar that could derive it. There are infinitely many grammars that can define the same language; supra-regular grammars—with fewer restrictions on rewrite rules than regular ones—can define regular languages. A language that can be defined by a grammar with suffix-constituents can also be defined by one with non-suffix constituents. This paper shows how we might assess which kind of grammar is, in a certain sense, better.

We define the four accounts as regular transductions. We start with a transduction that overgenerates: it marks up input sequences of light and heavy syllables with all possible stress patterns (*Gen*, defined in Sect. 2.1). From this point on, the syllable-based grammars are defined with identity transductions (i.e., acceptors). However, the foot-based “grammars”² require an additional (non-identity) transduction because they also mark up the stressed sequences with boundary symbols indicating left and right foot edges. Here we are coding constituents into the state, not in the derivation tree. An approach that codes constituents into the state can provide an exact account if the bound on tree depth required is finite, and here we are only coding feet—constituents up to depth 1.

We implement (non-suffix) constituents into the state (with regular transducers), rather than into the derivation (with supra-regular acceptors) to keep the definition of the grammars close to those that phonologists use. Linguists have characterized a wealth of phonological patterns with SPE-style rewrite rules, introduced in [8], and also with optimality-theoretic (OT) constraints [36]. [26] showed that the expressivity of SPE-style phonological characterizations is equivalent to that of regular transductions (provided that cyclic application of rules is not permitted). And [10] showed that an OT characterization has the expressivity of a regular transduction if the mapping from input to possible output forms (*Gen*) and the constraints (*Con*) are regular, and the number of violations a constraint can assign is finitely bounded; moreover, OT limited in such a way is sufficient for capturing analyses proposed by phonologists, except for analyses with gradient constraints (which can assign unboundedly many violations). Thus, the analytical tools that phonologists use—keeping the hedges mentioned above in mind—have the expressivity of regular transductions.

While phonologists may work with the power of regular transductions, they do not define phonological grammars by specifying transductions in the standard

² Because they are defined with non-identity transductions, the foot-based “grammars” are not grammars as defined by Formal Language Theory. But the phonological literature calls phonological transductions—input-output mappings from underlying forms to surface forms—like these “grammars”, and we'll follow that convention.

way, by listing states and transitions. Instead, they specify them at a high-level. We can do this too, by writing the grammars in `xfst` [4]. This language was carefully designed by linguists to make it easy for us to express generalizations in the high-level language which are hard to express or detect at the level of a regular grammar or a finite state machine. It includes pre-defined complex operators to give us a high-level notation for regular transducers. For example, we can write SPE-style rules using replacement rules with the syntax $A \rightarrow B \parallel L _ R$, where we simply need to specify the focus, change, and the structural description: this rule clearly does not meet the form $A \rightarrow aB$. `xfst` allows us to define our own operators and units, too, e.g., `feet`, and it compiles our high-level grammars to machine-level finite state transducers. Since `xfst` grammars are compiled into standard finite state representations, an `xfst` definition establishes that the phenomenon described is regular (see [11]) and gives us a common formalism in which we can define all four grammars. It is not possible to define “standard OT” [37] fully in `xfst`: we could define `Gen`, which generates the set of candidate outputs, and the assignment of violation marks according to the set of violable constraints in `Con`—but not the non-regular `Eval` relation for computing the optimal candidate, since the number of states required to define `Eval` can’t be bounded [10]. However, we can define Karttunen’s finite state formulation of OT [27] in `xfst`, as it avoids `Eval` by mapping underlying forms directly to surface forms with “lenient composition”.

Succinctness. Having `xfst` as a common formalism for defining all four grammars allows us to make a controlled comparison of the succinctness of the grammars. We define the succinctness of a grammar as its size—the number of symbols it takes to write it down (in `xfst`), under the conventions specified in Sect. 2. We define size over the high-level `xfst` grammar rather than at the machine level because it’s the high-level language that we can express and detect generalizations in; the machines that `xfst` compiles are big and redundant by comparison. Defining size in this way over a high-level language follows other linguistic work, e.g., [5, 8, 33, 44]. And `xfst` is a reasonable choice for the high-level language because it was designed by linguists to make it easy to state linguistic generalizations, and not, say, tailor-made to prefer feet over non-feet.

Our metric for succinctness can be thought of a special case of minimum description length (MDL) [39], relativized to the descriptive notation provided by `xfst`. MDL as a metric for succinctness balances the minimization of the size of the grammar, which favors simple grammars that often overgenerate, with minimization of the size of the data encoded by the grammar, which favors restrictive but often overly memorized grammars. The alphabet over which the grammars are defined (primary, secondary, and unstressed light and heavy syllables) is constant across grammars.³ Since all grammar definitions are expressed in the same language, we don’t need to translate them into some common language like bit

³ The foot-based accounts also introduce $(,)$, and X as symbols, where X is an unparsed syllable, but: (i) it’s not clear these should be included in the alphabet since they come in only in the calculation of stress, (ii) if they are included, they make a negligible difference.

representations; we can simply measure *xfst* grammar size with symbol counting. The data is also the same across the comparisons (the set of stress patterns in words elicited from linguistic consultants, plus some predicted ones up to 5 syllables; we have empirical data for monomorphs only up to 5 syllables). Moreover, as we show in Sect. 3, all the grammars admit exactly the same set of stress patterns up to 5 syllables (with one exception). Thus, the MDL metric, relativized to the descriptive resources of *xfst*, reduces to the size of the grammar. That is, the size of their encodings of the sequences up to 5 syllables is exactly the same, since the possibilities allowed by the grammars in that range is identical. We accordingly consider just the size of the four grammars, all expressed in the common *xfst* formalism.

1.2 Description of the Language Little Samoan

Samoan stress presents a good case study for a first comparison of the succinctness of grammars with and without feet. We can engage closely with the literature and empirical data because a recent detailed foot-based OT analysis of Samoan stress based on a rich set of elicited words is available [47]. [47]’s analysis also extends to morphologically complex words parsed into multiple prosodic words. In future work, we plan to extend the work here by comparing grammars with higher-level prosodic constituents than feet and modeling parsing at the syntax-phonology interface.

We define Little Samoan (LSmo), a language of strings of syllables marked for stress and weight, as a simplified version of the description of Samoan stress in monomorphs provided in [47]. LSmo is defined over light and heavy syllables rather than segments, and thus ignores complications from diphthongization and the interaction of stress with epenthesis. In [47]’s description, Samoan outputs LL for HL-final words to avoid heavy-light (HL) feet, and also presumably for L (content) words, to satisfy minimal word constraints. Since our OT grammars don’t change the syllable weights in the input, we model this in OT by mapping LL and HL-final inputs to a special Null symbol denoting a null output [37]. Our direct models simply define transductions that don’t accept HL and LL.

The basic primary stress pattern in LSmo (and Samoan) is moraic trochees at the right edge [47, (4)], e.g., la(‘va:) ‘energized’, (‘manu) ‘bird’, i(‘ŋoa) ‘name’; exactly like the well-known stress system of Fijian [15, Sect. 6.1.5.1], “if the final syllable is light, main stress falls on the penult; if the final syllable is heavy, main stress falls on the final syllable”. Secondary stress in LSmo is almost like in Fijian, where “secondary stress falls on the remaining [non-final] heavy syllables, and on every other light syllable before another stress, counting from right to left” [15, Sect. 6.1.5.1, p. 142], e.g., (,ma:)(,lo:)(‘lo:) ‘rest’ [47, (7)]. However, LSmo has an initial dactyl effect: initial LLL sequences are initially stressed, e.g., (‘mini)si(‘ta:) ‘minister’ (cf. Fijian mi(,nisi)(‘ta:)), (‘temo)ka(‘lasi) (cf. Fijian pe(,resi)(‘tendi) ‘president’).⁴

⁴ This ignores [47]’s evidence from LLLLL loan words showing that an initial weak-strong-weak (WSW) pattern can occur if the first vowel in the word is epenthetic.

[47] provides data on only two monomorphs that are longer than 5 moras: [(,ma:)(,lo:)('lo:)] and a 7-mora all light loanword for Afghanistan. As noted in [47, p. 281, fn. 2] the consultant produced (,ʔafa)(,kani)si('tana), while a true initial dactyl pattern would yield [(,ʔafa)ka(,nisi)('tana)]. Little data is available for longer words in Fijian, too, and there is variability [te(,reni)('sisi)(ta:)] ‘transistor’ vs. [(,ke:)(,misi)ti('ri:)] ‘chemistry’, which may be due to faithfulness to stress in the source word and a dispreference for stressing epenthetic vowels, as in Samoan [15, p. 144, (44b, 47)]. Due to the lack of data on longer words, we limit testing empirical coverage of LS_{mo} to monomorphs of 5 syllables (although of course our grammars can accept input strings of arbitrary length). Also, we allow a more general distribution of dactyls, which permits both initial or medial dactyls. This allows both (,LL)(,LL)L(LL) and (,LL)L(LL)(LL), and predicts, for example, that HLLLH may be (,H)(LL)L('H) or (,H)L(LL)('H).

2 Four Grammars for Little Samoan

In this section, we describe our grammars for LS_{mo}: the foot-based direct account (Sect. 2.2), the syllable-based direct account (Sect. 2.3), the foot-based OT account (Sect. 2.4), and the syllable-based OT account (Sect. 2.5). We define the transductions in the grammar in *xfst*, with symbol counts in square brackets to the right of the command. All of the *xfst* expressions we use are definitions, which get compiled into transducers associated with a variable. These have the syntax *define variable-name xfst-expression*. Our conventions for writing *xfst* expressions and counting symbols are as follows: (a) *xfst* expressions are delimited by square brackets in a *define* command. (b) Auxiliary terms are defined for any expression that appears more than once in the grammar. (c) A conjunct or disjunct longer than one symbol is enclosed in brackets. (d) A semicolon ending a line counts as a symbol; spaces do not count. (e) A number or a character enclosed in double quotes, e.g., "(", counts as one symbol. (f) Each variable name, command, operator, and atomic expression counts as a single symbol, i.e., *define*, *Heavy*, *WeakLight*, etc., *?*, ***, *+*, *^*, *##*, *[]*, *()*, *|*, *&*, *~*, **, *\$*, *->*, *..*, *=>*, *->@ _*, *...*, *.o.*, *.O*. (g) Symbols used to define *Gen* (constant across the grammars) don't contribute to the symbol count.

Each of the four grammars is implemented as a cascade of transducers and all four share the same basic schematic architecture:

```
addMarkup .o. defineSyllableAndFootTypes .o. enforceSurfaceRestrictions
```

All accounts begin with the transduction *Gen* to mark up input sequences of light and heavy syllables with stress, as described in Sect. 2.1. The foot-based accounts additionally add markup to indicate foot edges. These markup transductions are the only non-identity transductions in the implementations. They overgenerate stress patterns.⁵ These are then filtered in subsequent transductions that define

⁵ We set up the initial generation of stress patterns like *Gen* in Standard OT [36, Sects. 2.2, 5.2.3.3] for the direct accounts as well as the OT accounts. We do this for convenience; we could also generate in some other way for the direct accounts.

sub-types of syllables (and feet in foot-based accounts) and then restrict stress patterns in terms of these syllable (and foot) types. The difference between direct and OT accounts shows up in the way the surface restrictions are expressed in the grammar: the OT accounts are limited to define the restrictions as OT constraints used by phonologists, while the direct accounts are not.

2.1 Preliminaries: Adding Stress Markup with Gen

Let us call the transduction that generates all possible sequences of syllables marked with degree of stress and weight *Gen*. We define *Gen* (4) as the composition of *Input* (1), *SWParse* (2), and *ElevateProm* (3). *Input* generates Σ^* over the alphabet of light (L) and heavy (H) syllables, $\Sigma = \{L, H\}$. Then, *SWParse* marks the degree of stress on a syllable by inserting labeled brackets around each syllable,⁶ (see *Parse* in [27] and also [4, p. 68]), e.g., the input L has the output S[L] (strong/stressed), W[L] (weak/unstressed). Finally, *ElevateProm* optionally replaces any strong syllable S[] with a primary stressed syllable, P[], so that S[] now stands for secondary stress. As an example, input LL is mapped to { P[L]P[L], P[L]W[L], P[L]S[L], W[L]P[L], W[L]W[L], W[L]S[L], S[L]P[L], S[L]W[L], S[L]S[L]}.

define Input ["L" | "H"]*; [9] (1)

define SWParse [? -> ["S" "[" | "W" "["] ... "]"]; [15] (2)

define ElevateProm ["S"(->)"P"]; [10] (3)

define Gen [Input .o. SWParse .o. ElevateProm]; [10] (4)

2.2 Direct Account with Feet

ParseFoot (6) parses the output from *Gen* into feet, marking it up with boundary symbols; it refers to auxiliary terms for heavy [H] and light [L] syllables, *Heavy* and *Light* (5). We restrict a foot to being bimoraic: either a LL or a H, so *ParseFoot* wraps parentheses pairs around any LL or H, e.g., (P[L]W[L]), regardless of the stress pattern. We then define types of feet to express restrictions on stress patterns. *Foot* defines a foot as string of non-parentheses enclosed in parentheses (7); *PrimaryFoot* defines a primary stressed foot as a string accepted by *Foot* that also includes P (8), and *WeakLight* defines a weak light syllable (9) using *Light*. We define trochaic feet with *Trochee* (12), which accepts a strong-weak LL sequence *LLFoot* (10), or a strong H *HFoot* (11). The sequence \backslash "W" in (10) and (11) indicates the negation of the character W, i.e., any character but W such as P or S, which mark strong syllables.

With parsing the input into feet and definitions of types of feet behind us, the payoff comes as we can express *LSmo*'s restrictions on stress patterns in terms of feet. *TrocheesOnly* (13) forces feet to be trochaic: it only accepts strings

⁶ We assume that syllable splitting feet do not occur [15, Sect. 5.6.2, p. 121].

that are sequences of trochees that may be interspersed with unparsed syllables (weak lights), e.g., it winnows down the parses for LL to just (S[L]W[L]) and (P[L]W[L]). Additionally, `PrimaryFootRight` accepts only strings which terminate in a foot bearing primary stress and whose final foot is not preceded by any other primary stresses (14), e.g., eliminating (S[L]W[L]). Note that this transduction eliminates lone Ls and HL-final strings in the language, since they do not have parses with final primary stressed feet. Finally, we implement the “initial dactyl effect” in `LSmo` with `InitialDactyl` (15), which forbids a sequence of a weak light followed by a LL foot at the beginning of the word, if the LLL sequence is non-final, i.e., followed by at least one character (?+). This yields the SWW-initial output (S[L]W[L])W[L](P[H]) for LLLH sequences ([i(ˈmini)si(ˈta:)], [47, (8)]), but a WSW pattern W[L](P[L]W[L]) for LLL sequences [i(ˈŋoa)], [47, (4)]. This also allows outputs for HLLLH and 7Ls with medial dactyls. The final transduction going from all possible sequences of stressed and weighted syllables to only those in `LSmo` composes the foot parser with the restrictions on words in terms of feet (16).⁷ All together, excluding `Gen`, the grammar (5–16) costs 141 symbols.

```

define Heavy [ "[" "H" " "]; define Light [ "[" "L" " "]; [16] (5)
define ParseFoot [ [ [ ? Light ] ^2 | [ ? Heavy ] ] ->(" ... ")]; [22] (6)
    define Foot [ "(" \. [ "(" | " ] ] * " "]; [16] (7)
    define PrimaryFoot [ Foot & $["P" ]]; [11] (8)
        define WeakLight [ "W" Light ]; [7] (9)
    define LLFoot [ "(" \ "W" Light WeakLight " "]; [11] (10)
        define HFoot [ "(" \ "W" Heavy " "]; [10] (11)
            define Trochee [ LLFoot | HFoot ]; [8] (12)
                define TrocheesOnly [ Trochee | WeakLight ] *; [9] (13)
                    define PrimaryFootRight [ \ "P" * PrimaryFoot ]; [9] (14)
                        define InitialDactyl ~ [ WeakLight LLFoot ?+ ]; [10] (15)
                            define LSmoDirFt [ ParseFeet .o. TrocheesOnly
                                .o. PrimaryFootRight .o. InitialDactyl ]; [12] (16)

```

2.3 Direct Account Referring to Syllables only

The definition of `Gen` in this account is the same as in Sect. 2.2, but then we state restrictions on stress patterns in terms of syllables, rather than over feet. In addition to the previously defined auxiliary terms `Heavy` and `Light` (5), and `WeakLight` (9), we also define: `PrimaryLight` and `SecondaryLight`, a primary P[L] and secondary stressed light syllable S[L] (17); `StressedSyll`, a syllable of any weight that is not weak (18); and `W2`, a sequence of two weak lights W[L]W[L] (a lapse) (19).

⁷ `LSmoDirFt` can also be composed with a transduction that replaces W in unparsed syllables with X, to match notation for the OT footed account in Sect. 2.4.

Restrictions on the distribution of secondary and weak lights are more complex and are expressed as a series of cases. `StressSLight` (22) restricts a secondary light to be followed by a non-final weak light (so a penult light cannot receive secondary stress). In addition, a secondary light must be string-initial, preceded by a lapse `W[L]W[L]`, or preceded by a `S[L]W[L]`, i.e., in terms of feet, start a new foot. We must further restrict the position of weak lights, because the transducer that is the intersection of (20), (21), and (22) admits strings with lapses anywhere, e.g., it accepts both `P[L]W[L]` and `W[L]W[L]` from the set of sequences generated from input `LL`. `RestrictLapse` (23) restricts a lapse `W[L]W[L]` to be preceded by a secondary light and followed by a stressed syllable, allowing a lapse just in case it is in a dactyl `S[L]W[L]W[L]` and not string-final. The intersection of `StressPLight`, `StressSLight` eliminates a lone stressed L, since `StressPLight` and `StressSLight` restrict stressed lights to being non-final. But `RestrictLapse` does not eliminate a lone weak L. Moreover, no transduction thus far eliminates HL-final sequences.⁸ Thus, we must define transducers to ban HL-final sequences and lone Ls: `NoFinHL` and `NoLoneL` (24). All together the grammar costs 145 symbols.

```

define PrimaryLight ["P" Light]; define SecondaryLight ["S" Light]; [14] (17)
      define StressedSyll [ \ "W" [ Heavy | Light ] ]; [12] (18)
            define W2 [ WeakLight WeakLight ]; [7] (19)
                  define StressHeavy [ Heavy => "P" _ .#, "S" _ ]; [13] (20)
                        define StressPLight [ PrimaryLight => _ WeakLight .#. ]; [10] (21)
                              define StressSLight [SecondaryLight =>
                [.#. | W2 | [StressedSyll WeakLight] | Heavy] _ WeakLight ? ]; [19] (22)
                        define RestrictLapse [ W2 => SecondaryLight _ StressedSyll ]; [10] (23)
define NoFinHL ~[?* ? Heavy ? Light]; [10] define NoLoneL ~[Light] ; [7] (24)
      define LSmoDirSyll [ Gen .o. [ StressHeavy & StressPLight &
                StressSLight & RestrictLapse ] .o. NoFinHL .o. NoLoneL ]; [20] (25)

```

2.4 Karttunen OT with Feet

Our constraint set for this account is a subset of the constraints used in [47]; we have removed constraints that are only relevant for segments, morphologically complex words and multiple prosodic words. The partial ranking was computed with `OTSoft`⁹ [16] based on monomorphemic candidates used in [47] and is: *Stratum 1* (i) `FootBinarity` (`FtBin`) A foot must contain exactly two moras. (ii) `RhythmType=Trochee` (`RhType=Trochee`) A foot must have stress on its initial

⁸ HL-final sequences are allowed in [17]’s acceptor for Fijian stress (<http://phonology.cogsci.udel.edu/dbs/stress/language.php?id=109>), based on [15]’s basic description of Fijian stress, but [15, p. 145, Sect. 6.1.5.2]’s more detailed description suggests that they should not be accepted.

⁹ All `OTSoft` input and output files are in the github repository.

mora, and its initial mora only. (iii) *Align(PWd,R; 'Ft,R)* (*Edgemost-R*) The end of the prosodic word must coincide with the end of a primary-stressed foot; *Stratum 2* (i) *Parse- σ* Every syllable must be included in a foot. (ii) *Align(Pwd;L,Ft,L)* The beginning of the prosodic word must coincide with the beginning of a foot. Constraints in a earlier stratum are ranked higher than those in a later one, but constraints within a stratum are not ranked with respect to one another.

```

define MarkUnparsed [ "W" (-> "X" ]; [10] (26)
define FtParse [ [ \ "X" [ Heavy | Light ] ]+ -> "(" ... ")" ]; [19] (27)
define GenFt [ Gen .o. MarkUnparsed .o. FtParse]; [10] (28)
define Culminativity [ $.P ]; [8] (29)
define NullFinHL [ [?* Heavy [?]^<4 Light ("") ] ] ->@ "Null" ]; [27] (30)
define NullLoneL [ [ ("") ? Light ("") ] ] -> "Null" ]; [21] (31)
define Unparsed [ "X" "[" ? "]" ]; [8] (32)
define FtBinH [ "(" ? Heavy ")" ]; [7] define FtBinLL [ "(" [ ? Light ] ^2 ")" ]; [13] (33)
define FtBin [ FtBinH | FtBinLL | Unparsed ]*; [11] (34)
define Stressed [ "S" | "P" ]; [8] (35)
define RhTypeTrocheeH [ [ Heavy ")" ] ] => "(" Stressed _ ]; [13] (36)
define RhTypeTroLL [ "[" "L" => "(" Stressed _ , "]" "W" _ , "X" _ ]; [18] (37)
define RhTypeTrochee [ RhTypeTrocheeH & RhTypeTrocheeLL ]; [8] (38)
define ParseSyll ~["$X"]; [8] (39)
define ParseSyll1 ~["$X"]^>1]; [13] define ParseSyll2 ~["$X"]^>2]; [13] (40)
define AlignWdLfTl [ "(" ?* ]; [8] (41)
define LSmoMonoFtOT [ GenFt .O. NullFinHL .O. NullLoneL .O. Culminativity
.O. FtBin .O. RhTypeTrochee .O. EdgemostR .O. ParseS .O. ParseS1 .O. ParseS2
.O. ParseS3 .O. ParseS4 .O. ParseS5 .O. AlignWdLfTl ]; [28] (42)

```

We compute constraints referring to a prosodic word edge with respect to the edge of the input string since the input never contains more than a single prosodic word. With the exception of *Edgemost-R*, our constraint definitions are identical to those in [47, (5), (12)]. Some constraints, called categorical, assign multiple violations to a candidate iff there are multiple places where the constraint is violated in the candidate, e.g., *Parse- σ* (39). Other constraints, called gradient, “measure the extent of a candidate’s deviance from some norm”, and can assign multiple violations even if there is a single locus of violation in the input [31, p. 75]. [47] computes *Align(Pwd;L,Ft,L)* as a categorical constraint, e.g., 1 violation for *[sika(‘lamu)] ‘scrum’. However, *Edgemost-R* in [47] is computed gradiently, e.g., 2 violations for *[('pi:)niki]. We define it instead as categorical: since it’s undominated in our constraint set, whether it is assessed categorically or gradiently makes no difference. We can paraphrase the constraint definition as “assign a violation for every PWD where there exists a primary-stressed foot such

that the right edge of the PWd and the right edge of the primary-stressed foot do not coincide” [32]. As we’ll discuss in Sect. 3, whether a constraint is categorical vs. gradient markedly impacts the succinctness of the grammar in Karttunen’s formalism, as does whether a constraint can be multiply or only singly violated.

We define the candidate set in the OT footed account with GenFt (28) as the composition of Gen (4), MarkUnparsed (26), and FtParse (27). MarkUnparsed optionally replaces $W[]$ with $X[]$ to mark syllables unparsed into feet. FtParse parses the input into feet by wrapping parentheses around a non-empty sequence of syllables that are not unparsed, e.g., $X[L](P[L]W[L])$. We define three undominated constraints not included in the OTSoft ranking: Culminativity (29), NullFinHL (30), and NullLoneL (31). We define Culminativity (every word must contain exactly one primary stress) as a constraint rather than a property of Gen to keep Gen constant across grammars. NullFinHL and NullLoneL map HL-final and L inputs to Null. These definitions allow the HL-final and L inputs to be footed or unfooted and use directed replacement operators [4, p. 73]. For the $-@>$ operator, replacement strings are selected right to left, and only the longest match is replaced.

We then define constraints on feet. FtBin (34) restricts footed sequences to be any sequence of heavy feet and LL feet (FtBinH, FtBinLL, (33)), and unparsed syllables (Unparsed, (32)). We define RhyTypeTrochee (38) as the conjunction of RhyTypeTrocheeH (36) and RhyTypeTrocheeLL (37). RhyTypeTrocheeH restricts a heavy syllable followed by a parentheses, i.e., a footed H, to be preceded by a parentheses and S (secondary) or P (primary): a footed H must be stressed. RhyTypeTrocheeLL restricts a light syllable to be foot-initial and stressed (defined with Stressed (35)), or non-initial in a foot and weak, or unparsed. It accepts SWW dactyls. EdgemostR (not shown) is identical to PrimaryFootRight (14).

ParseSyll (39) must be implemented as a family of constraints because it can be multiply violated; we discuss this further in Sect. 3. Each ParseSyllN constraint in the family restricts the input string to have no more than N substrings containing X, e.g., $N = 1$ in ParseSyll1 (40). We follow the implementation in [27, pp. 10–11]¹⁰ We must impose some finite k -bound on the family; here we set $k = 5$ since the range of patterns we want to account for are only as long as five syllables. AlignWdLFtL (41) states that the beginning of the input string must coincide with the beginning of a foot and then may be followed by any string. The final transduction LSmoMonoFtOT is defined as a “lenient composition” (42).¹¹ [27] defines this (.O.) to be a special form of composition where input strings are held back from being eliminated to keep the set of output candidates from becoming empty. The order of “lenient” composition is important: the higher ranked a constraint is, the earlier it must enter the composition. Within a stratum, order of composition doesn’t matter. In total the OT footed grammar costs 306 symbols, including 73 from the ParseSyll family (as well as 16 from previously defined Light and Heavy (5)).

¹⁰ But there’s an inconsistency in [27]’s definitions of Parse; it should be defined as $\mathcal{T}[X[]]$; and not $\mathcal{T}\$X[]$; in Figs. 8 and 16.

¹¹ We abbreviate ParseSyll as ParseS for space; see github repository for definitions of ParseSyllN for $N > 2$.

2.5 Karttunen OT with Syllables only

The grammar using Karttunen OT with syllables only is by far the biggest grammar: it includes not only categorical constraints that can be multiply violated, but also gradient alignment constraints. There are few OT analyses of stress patterns that are not based on feet, i.e., “grid-based” [3, 13, 25], and we drew on constraints from them, but failed to generate only the allowed stress patterns up to 5 syllable words without introducing ad-hoc constraints that referenced feet without naming them. The partial ranking was computed with OTSoft [16] and is: *Stratum 1.* (i) **WeightToStress (WSP)** A heavy syllable must be stressed. (ii) **NonfinalityL** A word-final light syllable must be unstressed. (iii) **NoLapseFollowingHeavy** A heavy syllable musn’t be followed by two unstressed syllables. (iv) **NoInitialWS** A word musn’t begin with an unstressed-stressed sequence. *Stratum 2.* **Align(x2,R,x0,PWd)** Assign a violation for every grid mark of level 2 that doesn’t coincide with the right edge of level 0 grid marks in a prosodic word. *Stratum 3.* ***Clash** Assign a violation for every sequence of two stressed syllables. *Stratum 4.* ***Lapse** Assign a violation for every sequence of two unstressed syllables. *Stratum 5.* **Align(x1;L,x0,PWd)** Assign a violation for every grid mark of level 1 that doesn’t coincide with the left edge of level 0 grid marks in a prosodic word.

The *xfst* grammar for syllable-based Karttunen OT is shown in (43)–(57). After first introducing some auxiliary terms, we define the undominated constraints: **WeightToStress** (46), **NonfinalityL** (nonfinality restricted to light syllables) (47) [23], and two ad-hoc constraints—**NoLapseFollowingHeavy** and **NoInitialWS**. **NoLapseFollowingHeavy** (48) essentially enforces that a LL sequence after a heavy should be footed, and thus receive stress and allows the general ***Lapse** constraint to be ranked lower. **NoInitialWS** (49) essentially bans iambic feet word-initially. To achieve the initial dactyl effect, we use grid-based gradient **Align-x** constraints drawn from the schema in [13, (2)]. **Align(x2,R,x0,PWd)** enforces primary stress towards the right, while **Align(x1;L,x0,PWd)** is necessary to promote SWW initial candidates. While **WSP** (46) and **NoLapseFollowingHeavy** may have multiple loci of violation, because they are undominated, we were able to implement them as if they could only be singly violated. However, we had to implement a family of constraints to effectively count multiple violations for ***Clash**, ***Lapse**, and **Align(x2,R,x0,PWd)**. We use **Culminativity** (29) to filter out strings with multiple primary stresses, so we could implement **Align(x2,R,x0,PWd)** similarly to the **ParseSyll** family. But the implementation of **Align(x1;L,x0,PWd)** requires doing arithmetic because the same number of violations could be incurred by multiple stress patterns. We present a selection of the grammar below (see the github repository for the full grammar); **Gen** (4) is the same as before. Previously defined transductions repeated here (not shown) are: **Culminativity** (29), **Heavy** and **Light** (5), and **W2** (19). We also defined **Weak** (43) and **Stressed** (44) syllables, and clash **S2** (45), two adjacent stressed syllables. Transductions similar to **NullFinHL** (30) and **NullLoneL** (31) map HL-final and L inputs to Null (not shown).

The **NoNClash** family of constraints restricts the input from having N clashes: if an input is not accepted by **NokClash**, then it is also not accepted for any

$N > k$. The constraints define languages that are in a strict subset relation, like [27]’s ParseSyll family. We implemented NoNClash constraints (for $N = 1, 2, 3, 4$) as conjunctions, because there are multiple stress patterns that can result in the same number of clashes. For instance, an input may have two clashes because it has two nonadjacent clashes, or because it has a sequence of three stressed syllables, see No2Clash (54). The higher N is, the more conjuncts in the definition; specifically, it is 2^{N-1} . Already for No3Clash, we require 4 conjuncts: strings containing a substring of 4 stressed syllables (SSSS, where S stands for stressed syllable), two substrings containing SSS sequences, a substring containing SSS followed by a substring containing SS, and a substring containing SS followed by a substring containing SSS. The definition of NoNLapse is identical to that of NoNClash, but replaces S2 with W2 and Stressed with Weak.¹²

define Weak "W" [Heavy | Light]; [9] (43)

define Stressed ["S"|"P"] [Heavy | Light]; [13] (44)

define S2 [Stressed Stressed]; [7] (45)

define WSP [Heavy => \W _]; [10] (46)

define NonfinalityL ~[* \ "W" Light]; [11] (47)

define NoLapseFollowingHeavy ~[\$[Heavy W2]]; [11] (48)

define NoNitWS ~["W" Light "S" Light ?*]; [12] (49)

define AnySyll [? "[" ? "]"]; [9] (50)

define Alignx2R0 [Primary => _ .#.]; [9] (51)

define Alignx2R1 [Primary => _ [AnySyll]^<2 .#.]; [15] (52)

define No1Clash ~[\$[S2]]; [10] (53)

define No2Clash ~[\$[Stressed]^3] & ~[\$[S2]^2]; [25] (54)

define SWStar [Stressed [Weak]*]; [10] (55)

define Alignx1L1 ~[AnySyll SWStar]; define Alignx1L2 ~[AnySyll Weak SWStar]; (56)

define Alignx1L3 ~[AnySyll W2 SWStar] & ~[AnySyll Stressed SWStar]; [16] (57)

Similarly, we implement Align(x2,R,x0,PWd) as a categorical constraint, as the family Alignx2RN. We can do this because there is only ever one grid mark of level 2 (primary stress) in our candidates, since we restrict them to be single prosodic words. Each transducer Alignx2RN restricts primary stress to be at most N syllables from the right edge, where the syllables can be of any type (AnySyll (50)). All Alignx2RN transducers for $N > 1$ have the same form as Alignx2R1 (52); the languages accepted by the transducers are in a strict subset relation, and for words only up to 5 syllables, Alignx2RN for $N > 4$ may be omitted.

Align(x1;L,x0,PWd), though, must be implemented as a gradient constraint. There can be multiple grid marks of level 1 (stressed, i.e., not weak) in our

¹² See the github code repository for definitions of No3Clash (61 symbols) and No4Clash (131 symbols) and the NoNLapse constraint family.

candidates. We do it in two parts. First we define the *Alignx1LN* family, a set of transducers where *Alignx1LN* restricts the input to have the sum of the distances that stressed syllables are away from the left edge to total N . For example, *Alignx1L3* (57) does not accept inputs that are in $?WWSW^*$ (3 violations) or $?SSW^*$ (1 + 2 violations), and *Alignx1L5* does not accept inputs that are in $?WWWSW^*$ (5 violations), $?SWWSW^*$ (1 + 4 violations), or $?WSSW^*$ (2 + 3 violations), where *W* stands for a weak syllable, *S* for a stressed syllable, and $?$ for any character. The family definition refers to auxiliary term *SWStar*, the language of a stressed syllable followed by any number of unstressed syllables (55). Definitions for $N = 1, 2, 3$ are given in (56, 57).

We then take intersections of the *Alignx1LN* languages to define languages *Alignx1LgM* that accept any number of violations less than M . For example *Alignx1Lg5* is the intersection of *Alignx1L5*, *Alignx1L6*, *Alignx1L7*, ..., *Alignx1Lk*, “don’t have 5, 6, 7, ... k violations” where k is some finite upper bound. The *Alignx1Lg5* language contains *S[H]S[L]W[L]P[H]*, which has a total of $1 + 3 = 4$ violations, but not *S[H]W[L]S[L]P[H]*, which has a total of $2 + 3 = 5$ violations. How high does k need to be? The output *S[H]W[L]S[L]S[H]P[H]* for *HLLHH* has $2 + 3 + 4 = 9$ violations in total, while the output *S[H]S[L]W[L]S[H]P[H]* has $1 + 3 + 4 = 8$ violations in total. With our constraints, these two candidates have no other difference in their violation profiles. Thus, our transduction should admit the candidate with 8 violations, and not the one with 9. The candidate with 8 violations should be in any *Alignx1LgM* language where $M > 8$, in particular, in the *Alignx1Lg9* language, while the candidate with 9 should not. However, if $k = 8$, and we define *Alignx1LgM* transducers up to $M = 7$, then *Alignx1Lg7* is the intersection of *Alignx1L7* and *Alignx1L8* “don’t have 7 or 8 violations”. Then the candidate with 9 violations is in the *Alignx1Lg7* language, while the candidate with 8 isn’t: the transduction outputs the wrong candidate. To output the correct one, we must have $k \geq 10$, with $M \geq 9$: even for only up to 5-syllable words, we must have $k \geq 10$. Minimally we must include *Alignx1L10* in the conjunction that defines *Alignx1Lg9*. In general, a n -syllable word can have a maximum of $\sum_{i=2}^{n-1} i$ violations of *Align(x1;L,x0,PWd)* and k must be greater than that sum.

The language derived by the OT syllable grammar has a small difference from the others: while all the other grammars admit two outputs for *HLLLH*: *S[H]S[L]W[L]W[L]P[H]* or *S[H]W[L]S[L]W[L]P[H]*, the OT syllable account admits only *S[H]W[L]S[L]W[L]P[H]*. But [47] doesn’t actually include elicited data for *HLLLH*, so we don’t know which pattern(s) our consultants would accept. The striking difference about the OT syllable grammar is in its size. It is much larger than any of the other grammars, and the growth of the size of the grammar increases rapidly with the length of the input string: the definition of just the gradient constraint *Align(x1;L,x0,PWd)* takes more symbols than any of the other entire grammars, and the definitions of the clash and lapse constraints alone already grow exponentially with the size of the input. Moreover, all of the constraints which can be multiply violated cannot be implemented as finite state transducers, and our implementations approximating these constraints require doing arithmetic and defining constraint families that have the effect of counting

up violations to some finite bound. Finally, to just get near-coverage of the data without feet, we needed to define ad-hoc constraints that referenced feet without revealing generalizations in the structural restrictions on stress patterns. And additional exploratory calculations from OTSoft showed that we still need gradient alignment constraints to fit the data, despite including additional categorical constraints from [25]’s Rhythmic Licensing Theory—designed to avoid gradient constraints (see github repository, `otsoft-files/syll/test-with-rlt`).

3 Discussion

We examined the set of possible stress patterns from our grammars by composing our final transducers with an identity transducer that was defined as a disjunction of elicited/expected stress patterns for LSmo up to 5 syllables. We also defined an identity transducer for all possible light-heavy inputs up to 5 syllables and composed that with our final transducers. Then we checked that the set of strings defined by these two compositions was identical for each grammar and across grammars. Our four grammars for LSmo admit exactly the same set of stress patterns up to 5 syllables, with the one exception mentioned above: the OT syllable account admits only one stress pattern for HLLLH.¹³ Thus, the MDL metric, relativized to the descriptive resources of `xfst`, reduces to the size of the grammar, although that’s not quite the case for the OT syllable account. The direct accounts were almost the same in size: 145 symbols for the syllable account and 141 symbols for the footed account; the OT footed grammar cost 306 symbols. The small differences between these is insignificant, compared to the qualitatively different character of exponential growth we saw in definition of the OT syllable grammar, with a count in the 1000s. Even if including a battery of constraints from Rhythmic Licensing Theory [25], we found that an OT syllable grammar would still need to include multiply violated clash and lapse constraints and gradient Align constraints; we’d also expect this to be true in general beyond the Samoan case study here, such that the size of OT syllable-based grammars would in general blow up.

Our results show that with a direct account, a grammar referencing feet in the description of stress patterns in LSmo is about as succinct as a grammar that does not. By this metric, one isn’t preferable to the other. Also, the size of the direct grammars is a few times smaller than even the OT footed grammar, so Karttunen OT grammars are certainly not preferable by succinctness. For the OT accounts, a grammar referencing feet is sizeably more succinct than one that references only syllables, showing the utility of feet. It’s interesting that the

¹³ Although our accounts define the same transduction, that does not mean that the transducers `LSmoDirFt`, `LSmoDirSyl`, `LSmoMonoFtOT` are identical at the machine-level. While any finite-state acceptor can be determinized and minimized to a unique, canonical acceptor [21, Sect. 4.4], the same is not true for finite-state transducers. First, not all finite-state transducers are determinizable [40, p. 587]. Second, minimization of a finite-state transducer does not in general result in a unique transducer [34, p. 29].

direct footed account wasn't notably more succinct than the direct syllable one; this could be because of the narrowness of the scope of phonological phenomena considered here. For instance, patterns of stress shift in Samoan upon affixation can be generalized on the basis of constituents [47], but here we considered only monomorphs. The more phonological processes that reference constituents in the grammar, the more the savings from those constituents.

One thing to stress about the foot-based grammars, is that although they place boundaries (parentheses) in the string language, they are very different from SPE-style “boundary symbol theory” [8, 42]. In our grammars, the use of boundary symbols is not arbitrary; rather a left parentheses signals entering into a sequence of states representing a constituent, and a right parentheses is invariably placed when that sequence of states is completed. As [42] points out, compared to a grammar which references nested units in the prosodic hierarchy, grammars with boundary symbols may be expressive enough to fit the data, but the lack of a well-defined relation between the different boundary symbols makes the grammars much too expressive. Moreover, boundary symbol theory locate the boundary symbols in the alphabet and allows their placement to be restricted only by general restrictions on possible rewrite rules. But we are coding constituency into the state: the LSmo foot-based grammars place parentheses in the string language so we can refer to the units that they enclose, and what restricts their placement is the phonological generalizations defined in the grammar.

Comparing the direct grammars to the OT grammars, a number of the transducers defined are identical or similar, e.g., *EdgemostR* appears in both footed accounts. This suggests that structural regularities we notice in phonological patterns can be well-described in both types of grammars. However, there is a striking difference between the direct grammars and the OT grammars: the OT grammars have scaling problems. The two direct accounts defined can handle syllable strings of arbitrary length. But for the OT accounts, as the syllable string gets longer, the amount of counting that needs to be done increases. In the foot-based OT account, we defined *ParseSyll* constraint family only up to 5 syllables. Add another syllable to the syllable string, and the grammar becomes inadequate. In the syllable-based OT account, the *NokClash* and *Align* constraint families also effectively count up violations, resulting in the same kind of scaling problem. As previously mentioned, it is in fact the unlimited violation counting that “pushes [standard] optimality theory out of the finite domain” [27, p. 11]. The adequacy of the direct accounts—which are similar in size to the OT grammars and even describe similar regularities—suggests that perhaps this additional power is unnecessary.

It is also an advantage that finite state transducers are sufficient to define the direct grammars. Defining phonology with finite state transducers is not only helpful as a common formalism with comparison of syntax (and other patterns), but also enables us to compose phonological transducers with syntactic ones to model the syntax-phonology interface. In contrast, the expressive power of finite state transducers is not enough to define OT grammars where underlying forms

are mapped directly to surface forms rather than violation vectors. As noted by [27], any constraint that can be multiply violated such as *Parse- σ* cannot be defined with a finite state transducer in [27]’s formalization of OT because a finite system cannot distinguish between infinitely many degrees of well-formedness [10, 31]. If, instead we do define relations that map from underlying forms to violations as in standard OT, we can easily use *xfst* to implement *Parse- σ* as: define *ParseSyllOT* ["Unparsed" -> "1", \ "Unparsed" -> "0"]; [11] The FST is trivial: a 2-state machine, where one state maps any syllable that is not unparsed to 0 and the other maps an unparsed syllable $X[?]$ to 1, e.g., it maps $X[L]P[L]X[L]$ to 101. This shows an advantage to mapping to violations in *Con*, but then the scaling problem with counting is shifted to *Eval*. In Karttunen’s formalization, both gradient and categorical constraints that can assign multiple violations cannot be defined with a FST. Moreover, as we saw with *Align(x1;L,x0,PWd)*, the implementation of gradient constraints is even more cumbersome. Not only are the machines that *xfst* compiles them into much too big and redundant to discover generalizations in; even the high-level language description are, too. When OT transduces instead to violation marks, only gradient constraints cannot be defined with a FST. While [31] argues that OT constraints are categorical, even if that is the case, *Eval* isn’t a finite state process. OT isn’t regular if the number of violations is unbounded [10].

However, one potential advantage of OT grammars vs. direct grammars in the work here is that limiting ourselves to mainstream constraints proposed by phonologists has provided some restrictions on the defined grammars—even if these restrictions might not be well-characterized in terms of structural classes in the Chomsky or sub-regular hierarchy, and even if standard OT admits supra-regular transductions. Limiting the statement of restrictions in direct grammars to the expressivity of regular transductions (as we’ve done here) is not restrictive enough: it is well-known that regular transductions include patterns that we’d never expect to see in phonology, e.g., [19]. On-going work has been tightening bounds on the expressivity of ‘direct’ transductions to sub-regular classes and making connections between restrictions on these transductions and restrictions on OT constraints, see [20] for an overview.

4 Conclusion

In this paper, we implemented and compared syllable- and foot-based grammars of Samoan stress patterns. We made this comparison in Karttunen’s finite state formalization of OT, and in grammars directly describing restrictions of the surface patterns. The definition and comparison of the grammars was done in the *xfst* language to follow linguistic practice, since *xfst* was designed to be a high-level language that makes it easy to express and detect linguistic generalizations. Such generalizations might not be revealed at the level of a regular grammar or finite state machine.¹⁴ In the OT formalism, having the prosodic constituents of

¹⁴ See the github repository for graphs of the transducers defined for each of the four accounts.

feet clearly allowed the grammar to be much more succinct. However, whether or not we have feet in the direct account did not impact succinctness of the grammar. It is striking that direct finite-state descriptions of phonological patterns have revealed strong structural universals without referring to constituents, while the advantage of having constituents is clear in the OT formalism used here. The difference in the comparison between the two types of grammars may simply be because our measure of succinctness is not appropriate, and also may not hold in general, or because the range of phonological phenomena considered here is too narrow.

A natural follow-up to the work here would be to extend grammar comparisons to a wider range of phonological phenomena that have been studied in prosodic phonology. For instance, all the dependencies in the Little Samoan language defined here are local. What if the language included non-local dependencies? Another natural follow-up would be to explore the consequences of introducing constituents in more expressive grammars. For instance, OpenFST is a finite state transducer library that offers the capability to define grammars with the expressivity of context-free languages via pushdown automata, which are finite state transducers augmented with a stack [1, 2]. It would be interesting to see if a comparison of syllable-based and foot-based grammars for Samoan stress defined with pushdown automata might yield different results from the ones here. More broadly, this paper shows a way in which we can study concrete, specific linguistic proposals and engage closely with linguistic practice, while still maintaining a rigorous approach. We hope that this proof of concept may inspire additional computational work taking this kind of approach.

References

1. Allauzen, C., Riley, M.: A pushdown transducer extension for the OpenFst library. In: Moreira, N., Reis, R. (eds.) CIAA 2012. LNCS, vol. 7381, pp. 66–77. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31606-7_6
2. Allauzen, C., Riley, M., Schalkwyk, J., Skut, W., Mohri, M.: OpenFst: A general and efficient weighted finite-state transducer library. In: Holub, J., Žd’árek, J. (eds.) CIAA 2007. LNCS, vol. 4783, pp. 11–23. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76336-9_3
3. Bailey, T.: Nonmetrical constraints on stress. Ph.D. thesis, University of Minnesota (1995)
4. Beesley, K.R., Karttunen, L.: Finite State Morphology. CSLI, Stanford, CA (2003)
5. Berwick, R.C.: Mind the gap. In: Gallego, A., Ott, D. (eds.) 50 Years Later, MITWPL77, pp. 1–12. MIT, Cambridge, Massachusetts (2015)
6. Bird, S., Ellison, T.M.: One level phonology: autosegmental representations and rules as finite automata. *Comput. Linguis.* **20**, 55–90 (1994)
7. Chomsky, N.: Three descriptions of language. *IRE Trans. Inf. Theory* **2**(3), 113–124 (1956)
8. Chomsky, N., Halle, M.: *The Sound Pattern of English*. The MIT Press, Cambridge (1968)
9. Eisner, J.: Efficient generation in primitive optimality theory. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics* (1997)

10. Frank, R., Satta, G.: Optimality theory and the generative complexity of constraint violability. *Comput. Linguist.* **24**(2), 307–315 (1998)
11. Gainor, B., Lai, R., Heinz, J.: Computational characterizations of vowel harmony patterns and pathologies. In: Choi, J., et al. (eds.) WCCFL 29, pp. 63–71. Cascadia Proceedings Project, Somerville, MA (2012)
12. Goldsmith, J.A.: Autosegmental phonology. Ph.D. thesis, MIT (1976)
13. Gordon, M.: A factorial typology of quantity insensitive stress. *NLLT* **20**, 491–552 (2002)
14. Hartmanis, J.: On the succinctness of different representations of languages. *SIAM J. Comput.* **9**, 114–120 (1980)
15. Hayes, B.: *Metrical Stress Theory*. University of Chicago Press, Chicago (1995)
16. Hayes, B., Tesar, B., Zuraw, K.: Otsoft 2.4. Software package (2016). www.linguistics.ucla.edu/people/hayes/otsoft/
17. Heinz, J.: On the role of locality in learning stress patterns. *Phonology* **26**(02), 303–351 (2009)
18. Heinz, J.: Learning long-distance phonotactics. *LI* **41**(4), 623–661 (2010)
19. Heinz, J.: Computational phonology - part I: foundations. *Lang. Linguist. Compass* **5**(4), 140–152 (2011)
20. Heinz, J.: The computational nature of phonological generalizations. In: Hyman, L.M., Plank, F. (eds.) *Phonological typology*. De Gruyter Mouton, Berlin/Boston (2018)
21. Hopcroft, J.E., Motwani, R., Ullman, J.D.: *Introduction to Automata Theory, Languages, and Computation*. Pearson Education, Boston (2007)
22. Hulden, M.: Finite state syllabification. In: Yli-Jyrä, A., Karttunen, L., Karhumäki, J. (eds.) *FSMNLP 2005*, pp. 120–131. Springer, Berlin (2006)
23. Hyde, B.: Non-finality and weight sensitivity. *Phonology* **24**(2), 287–334 (2007)
24. Johnson, C.D.: *Formal aspects of phonological description*. Mouton (1972)
25. Kager, R.: Rhythmic licensing theory: an extended typology. In: *Proceedings of the 3rd Seoul International Conference on Phonology*, pp. 5–31 (2005)
26. Kaplan, R.M., Kay, M.: Regular models of phonological rule systems. *Comput. Linguist.* **20**(3), 331–378 (1994)
27. Karttunen, L.: The proper treatment of optimality in computational phonology. In: *FSMNLP 1998* (1998)
28. Kenstowicz, M.: Cyclic vs. non-cyclic constraint evaluation. *Phonology* **12**, 397–436 (1995)
29. Kiparsky, P.: Word formation and the lexicon. In: Ingemann, F. (ed.) *Proceedings of the 1982 Mid-America Linguistics Conference*, pp. 3–29. University of Kansas, Lawrence (1982)
30. Kornai, A.: *Formal phonology*. Ph.D. thesis, Stanford University (1991)
31. McCarthy, J.J.: OT constraints are categorical. *Phonology* **20**(1), 75–138 (2003)
32. McCarthy, J.J., Prince, A.S.: Generalized alignment. In: Booij, G., Van Marle, J. (eds.) *Yearbook of Morphology*, pp. 79–153. Springer, Dordrecht (1993). https://doi.org/10.1007/978-94-017-3712-8_4
33. Meyer, A., Fischer, M.: Economy of description by automata, grammars, and formal systems. In: *SWAT 1971*, pp. 188–191 (1971)
34. Mohri, M.: Finite-state transducers in language and speech processing. *Comput. Linguist.* **23**, 1–42 (1997)
35. Nespor, M., Vogel, I.: *Prosodic Phonology*. Foris Publications, Dordrecht (1986)
36. Prince, A., Smolensky, P.: *Optimality theory: Constraint interaction in generative grammar*. ROA version, 8/2002, Rutgers University Center for Cognitive Science (1993)

37. Prince, A., Smolensky, P.: *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing, Malden (2004)
38. Rasin, E., Katzir, R.: On evaluation metrics in Optimality Theory. *LI* (To appear)
39. Rissanen, J.: *Stochastic Complexity in Statistical Inquiry Theory*. World Scientific Publishing Co. Inc., River Edge (1989)
40. Sakarovitch, J.: *Elements of Automata Theory*. Cambridge University Press, Cambridge (2009). Translated by Reuben Thomas edn.
41. Salomaa, A.: *Formal Languages*. Academic Press, New York (1973)
42. Selkirk, E.: Prosodic domains in phonology: Sanskrit revisited. In: Aronoff, M., Keans, M.L. (eds.) *Juncture*. Anma Libri, Saratoga, California (1980)
43. Selkirk, E.O.: *Phonology and Syntax*. MIT Press, Cambridge (1986)
44. Stabler, E.P.: Two models of minimalist, incremental syntactic analysis. *Top. Cogn. Sci.* **5**(3), 611–633 (2013)
45. Wagner, M., Watson, D.G.: Experimental and theoretical advances in prosody: a review. *Lang. Cogn. Processes* **25**, 905–945 (2010)
46. Zuraw, K.: Prosodic domains for segmental processes?, June 2009. https://www.mcgill.ca/linguistics/files/linguistics/Handout_RevisedForMcGill.pdf
47. Zuraw, K., Yu, K.M., Orfitelli, R.: The word-level prosody of Samoan. *Phonology* **31**(2), 271–327 (2014)