Towards a Factorization of String-Based Phonology

Thomas Graf
tgraf@ucla.edu
tgraf.bol.ucla.edu

University of California, Los Angeles

ESSLLI Student Session 2009

1. Setting the Scene: Motivation & Basics of Phonology
   - Our Goal: Theory Comparison
   - Theory 1: SPE
   - Theory 2: Government Phonology (GP)
2. Formalization of GP
   - Definition of our Modal Logic
   - Axiomatization — Simple Cases
   - Axiomatization — Empty Categories and Feature Spreading
3. Parameters and Their Effects on Expressivity
   - Feature Systems
   - Syllable Template
   - Distance of Feature Spreading
4. Conclusion

What is Phonology?

Phonologists study the regularities between sounds and their interactions within and across languages.

Example

- [pl] is good at the beginning of an English word ("*plum*"), but bad at the end. [lp] is bad at the beginning but good at the end ("*scalp*").
- The voicing of the English plural marker -*s* depends on the preceding consonant ("*hats*" versus "*woods*").

Selection of Phonological Theories

Our Ultimate Goal

To develop a metatheory that allows us to predict the behavior of a phonological theory from the properties of its parts

- SPE
- Optimality Theory
- Natural Phonology
- Declarative Phonology
- Dependency Phonology
- Government Phonology (GP)
- Strict CV
- Big Tree Phonology
- . . .

Selection of Phonological Theories

Our Immediate Goal

Formalize GP and compare it to SPE to determine which parameters impact generative capacity

- SPE
- Optimality Theory
- Natural Phonology
- Declarative Phonology
- Dependency Phonology
- Government Phonology (GP)
- Strict CV
- Big Tree Phonology
- ...

SPE (Chomsky and Halle 1968)

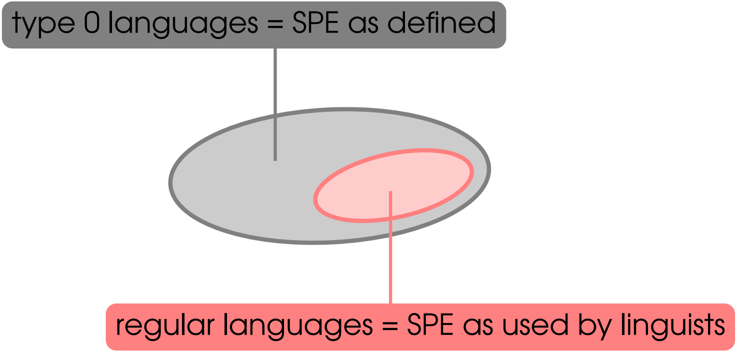
- Words are strings of matrices of $+/-$ **valued features**
- Surface forms derived from lexicon via **rewriting rules**

Example

- $\mathrm{i} = \begin{bmatrix} -\mathit{back} \\ +\mathit{high} \\ -\mathit{low} \\ +\mathit{tense} \end{bmatrix}$ $\mathrm{u} = \begin{bmatrix} +\mathit{back} \\ +\mathit{high} \\ -\mathit{low} \\ +\mathit{tense} \end{bmatrix}$ $\mathrm{a} = \begin{bmatrix} -\mathit{back} \\ -\mathit{high} \\ +\mathit{low} \\ +\mathit{tense} \end{bmatrix}$
- $\begin{bmatrix} +\mathit{cons} \\ +\mathit{voice} \end{bmatrix} \rightarrow [-\mathit{voice}] \mid \_\_ \#$ word-final devoicing

The Expressivity of SPE

Even though the rewriting rules are unrestricted, SPE as used by linguists generates only regular languages (Kaplan and Kay 1994).



type 0 languages = SPE as defined

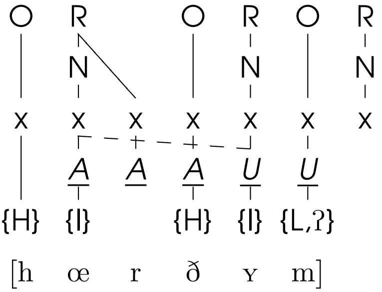
regular languages = SPE as used by linguists

GP (Kaye et al. 1990)

GP aims to be a maximally restricted theory of phonology. It differs significantly from SPE, making it difficult to compare the two ⇒ the ideal object for a case study!

Differences GP vs SPE

- Feature system
- Syllable template
- Empty categories
- Feature spreading



GP vs SPE — Feature System

GP

- 4-12 **privative** features (called **elements**)
- sound = pair of a **head** and a set of **operators**

Example

r	$(\{\_\}, \underline{A})$	a	$(\{\_\}, \underline{A})$
g	$(\{\text{ʔ}\}, \_)$	ə	$(\{\_\}, \_)$
s	$(\{A, H\}, \_)$	e	$(\{A\}, \underline{I})$
n	$(\{L, \text{ʔ}\}, \underline{A})$	ɛ	$(\{I\}, \underline{A})$

SPE

- 20-24 binary valued features
- sound = matrix of features

Example

$$\mathrm{a} = \begin{bmatrix} -back \\ -high \\ +low \\ +tense \\ \vdots \end{bmatrix} \qquad \mathrm{r} = \begin{bmatrix} +cons \\ +son \\ +cont \\ +ant \\ \vdots \end{bmatrix}$$

GP vs SPE — Syllable Template

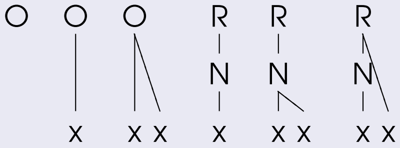
GP

- sequence of **onset-rhyme pairs** built from six basic structures
- vowels occupy nucleus (N), consonants all other positions

SPE

- sequence of feature matrices without further syllable annotation

The 6 Basic Structures of GP



GP vs SPE — Empty Categories

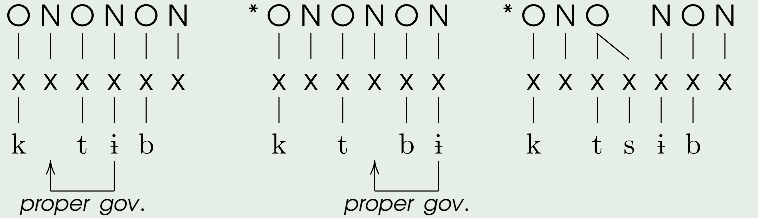
GP

- Nuclei may remain unpronounced iff they are **p-licensed**.

SPE

- Only word and morpheme boundaries remain unpronounced.

Example



Definition of p-Licensing

A nucleus is a **government licensor** iff

- its onset governs a preceding rhymal complement (i.e. a coda), or
- if its onset is binary branching.

Nucleus α **properly governs** nucleus β iff

- α and β are adjacent on the relevant projection, and
- α is not itself p-licensed, and
- neither α nor β are government licensors.

A nucleus is **p-licensed** iff

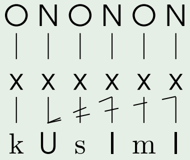
- it is domain-final, or
- it is immediately followed by an s+C sequence, or
- it is properly governed

GP vs SPE — Feature Spreading

GP

- Since there are no negatively valued features, only positive values can spread.

Example



$\Rightarrow$ [kusymy], *[kusumu]

SPE

- Any feature can affect another segment arbitrarily far away.

Example

- $i \longrightarrow y \mid u\ C_0$ __
  kusimi → kusymi → kusymy
- $i \longrightarrow u \mid u\ C_0$ __
  kusimi → kusumi → kusumu

Outline

1. Setting the Scene: Motivation & Basics of Phonology
   - Our Goal: Theory Comparison
   - Theory 1: SPE
   - Theory 2: Government Phonology (GP)
2. Formalization of GP
   - Definition of our Modal Logic
   - Axiomatization — Simple Cases
   - Axiomatization — Empty Categories and Feature Spreading
3. Parameters and Their Effects on Expressivity
   - Feature Systems
   - Syllable Template
   - Distance of Feature Spreading
4. Conclusion

Elements and Substructural Information

GP-elements may take on various roles:

- head versus operator
- local versus spread

We capture this explicitly by multiplying out the features.
Given some set of **basic elements** E, let the set of **elements** $\mathscr{E} := E \times \{head, onset\} \times \{local, spread\}$.
We further introduce three diacritic features.

- μ . . . for unpronounced segments
- $\checkmark$ . . . for p-licensed segments
- `fake` . . . for onsets without a skeleton node

Definition of the Modal Logic for GP

The set of well-formed formulas is built up from

- the connectives $\neg$ and $\wedge$,
- the set of propositional variables $\mathscr{M} := \mathscr{E} \cup \{\mu, \surd, \texttt{fake}\}$,
- three propositional constants N, O, C to mark nuclei, onsets and codas,
- two modal diamond operators $\lhd$ and $\rhd$ that move us along the string,
- $\blacktriangleleft$ and $\blacktriangleright$, the duals of $\lhd$ and $\rhd$.

Intended Models

Our intended models are (finite) strings which can be traversed by $\lhd$ and $\rhd$ and whose nodes are labeled by elements and constituency information.

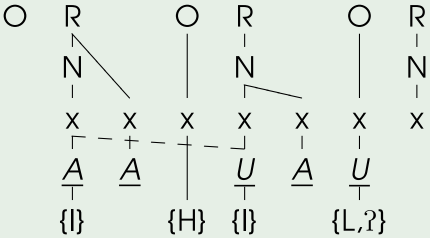
Formal Definition of the Intended Models

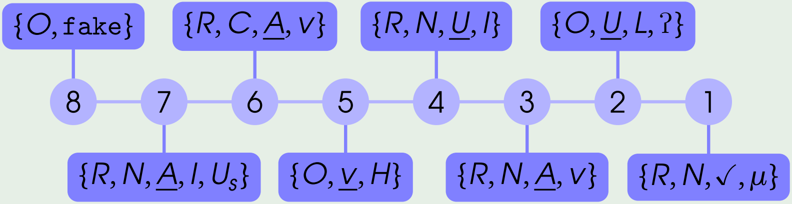
$\mathfrak{M} := \langle \mathfrak{F}, V \rangle$, where

- $\mathfrak{F} := \langle D, R_i, R_\lhd \rangle_{i \in \{N,O,C\}}$ a bidirectional frame,
- D an initial subset of $\mathbb{N}$,
- $R_i \subseteq D$ for each $i \in \{N, O, C\}$,
- $R_\lhd$ is the successor function over $\mathbb{N}$,
- $V : \mathscr{M} \rightarrow \wp(D)$ the valuation function

Intended Models — Example

Example





8: $\{O, \texttt{fake}\}$

7: $\{R, N, \underline{A}, I, U_s\}$

6: $\{R, C, \underline{A}, v\}$

5: $\{O, \underline{v}, H\}$

4: $\{R, N, \underline{U}, I\}$

3: $\{R, N, \underline{A}, v\}$

2: $\{O, \underline{U}, L, \text{ʔ}\}$

1: $\{R, N, \surd, \mu\}$

Skeleton

S1	$\bigwedge_{i\in\{N,O,C\}}(i \leftrightarrow \bigwedge_{i\neq j\in\{N,O,C\}} \neg j)$	Unique constituency
S2	$(\blacktriangleleft \bot \rightarrow O) \wedge (\blacktriangleright \bot \rightarrow N)$	Word edges
S3	$R \leftrightarrow (N \vee C)$	Definition of rhyme
S4	$N \rightarrow \lhd O \vee \lhd N$	Nucleus placement
S5	$O \rightarrow \neg \lhd O \vee \neg \rhd O$	Binary branching onsets
S6	$R \rightarrow \neg \lhd R \vee \neg \rhd R$	Binary branching rhymes
S7	$C \rightarrow \lhd N \wedge \rhd O$	Coda placement

Observation

All seven axioms have modal quantifier depth 1. In particular, we never need to consider any nodes but the immediately adjacent ones.

Feature Calculus

F1	$\bigwedge(h_n \rightarrow \bigwedge_{h_n \neq h'_n} \neg h'_n)$	Exactly one head
F2	$\neg \underline{v} \rightarrow \bigwedge(h_n \rightarrow \bigwedge_{\pi_1(h)=\pi_1(o)} \neg o_n)$	No basic element (except v) twice
F3	$v \rightarrow \bigwedge_{o \neq v} \neg o$	v excludes other operators
F4	$\bigwedge(e_2 \rightarrow \bigvee h_1 \wedge \bigvee o_1)$	Pseudo branching implies first branch

Observation

As is to be expected, the feature calculus can be axiomatized in propositional logic.

Melody — Additional Terminology

- A propositional formula ϕ over a set of variables $x_1, \dots, x_k$ is called *exhaustive* iff it denotes a unique phonological expression.
- The function *lic* maps every exhaustive formula to its set of melodic licensors.
- The set of all exhaustive formulas consistent with the feature calculus axioms is denoted by *PH*.

Example

Set of variables $:= \{\underline{A}, A, I, v\}$

$$\begin{aligned} PH := \{&\underline{A} \wedge A \wedge \neg I \wedge \neg v,\ \underline{A} \wedge \neg A \wedge I \wedge \neg v, \\ &\underline{A} \wedge \neg A \wedge \neg I \wedge v,\ \underline{A} \wedge A \wedge I \wedge \neg v\} \end{aligned}$$

Melody

M1	$\bigwedge_{i \in \{N,O,C\}} (i \to (\bigvee h_1 \land \bigvee o_1) \lor \mu \lor \mathtt{fake})$	Universal annotation
M2	$((O \lor \lhd N \lor \rhd N) \to \bigwedge \neg e_2)$	No pseudo branching for O, C & branching N
M3	$O \land \lhd O \to \bigwedge_{\phi \in PH} (\phi \to \bigvee_{\psi \in lic(\phi)} \lhd \psi)$	Licensing within branching onsets
M4	$C \land \bigwedge_{i \in S} \neg i \to \lhd \neg \mu \land \bigwedge_{\phi \in PH} (\phi \to \bigvee_{\psi \in lic(\phi)} \rhd \psi)$	Melodic coda licensing
M5	$\mathtt{fake} \to O \land \bigwedge_{m \neq \mathtt{fake}} \neg m$	Fake onsets

Observation

The modal quantifier depth can still be limited to 1 and we only need to consider adjacent nodes.

Empty Categories

L1	$\mu \rightarrow \neg C \wedge (N \rightarrow \checkmark) \wedge \underline{v} \wedge v$	Empty categories
L2	$N \wedge \lhd N \rightarrow (\mu \longleftrightarrow \lhd \mu)$	Licensing of branching nuclei
L3	$O \wedge \lhd O \rightarrow \neg \lhd \mu \wedge \neg \mu \wedge \neg \rhd \mu$	Licensing of branching onsets
L4	$N \wedge \checkmark \longleftrightarrow \text{(special configurations)} \vee \underbrace{((\neg \lhd N \rightarrow \lhd (\lhd N \vee \blacktriangleleft \bot)) \wedge (\neg \rhd N \rightarrow \rhd \rhd (N \wedge \neg \mu)))}_{\text{Proper Government}}$	p-licensing

Observation

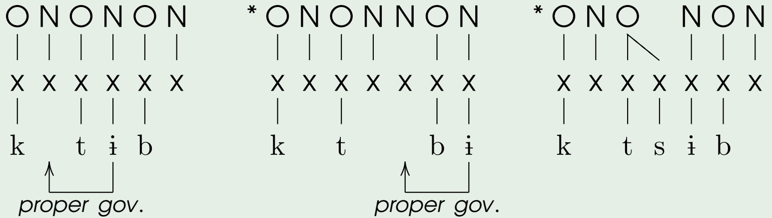
The Proper Government condition finally takes us to quantifier depth 2.

Empty Categories — L4 en detail

$$\underbrace{((\neg \lhd N \rightarrow \lhd (\lhd N \vee \blacktriangleleft \bot)) \wedge (\neg \rhd N \rightarrow \rhd\rhd (N \wedge \neg\mu)))}_{\text{Proper Government}}$$

“*N* is exactly one/two nodes away from the left word edge/a preceding nucleus and exactly two nodes away from a pronounced nucleus following it.”

Example

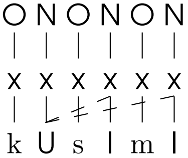


Feature Spreading — Properties

Unbounded spreading is usually assumed to arise from the iteration of local spreading steps (cf. cyclic wh-movement in syntax). It is left open in the literature

- whether spreading is always obligatory,
- how its directionality is restricted
  (only left, right, zig-zag . . .),
- what qualifies as source or target of spreading.

All variants can be implemented in our modal logic.



Feature Spreading as a Formula Scheme σ

For *i* and *j* elements derived from the same basic element and *min* and *max* the minimum and maximum range of spreading, respectively:

$$\sigma := i \wedge \omega \rightarrow \bigvee_{min \leq n \leq max} \Diamond^{n}(j \wedge \rotatebox{180}{ω})$$

Settings for Different Types of Spreading

Mode	Direction	i	ω	$\rotatebox{180}{$\omega$}$	$\Diamond$
optional	left	spread	target	source	$\rhd$
optional	right	spread	target	source	$\lhd$
mandatory	left	local	source	target	$\lhd$
mandatory	right	local	source	target	$\rhd$

Example

Optional spreading of U to the right from nucleus operator into nucleus operator positions

$$\sigma := U_s \wedge N \rightarrow \bigvee_{2\leq n\leq 7} \lhd^n (U_l \wedge N)$$

Settings for Different Types of Spreading

Mode	Direction	i	ω	ϖ	$\Diamond$
optional	left	spread	target	source	$\rhd$
optional	right	spread	target	source	$\lhd$
mandatory	left	local	source	target	$\lhd$
mandatory	right	local	source	target	$\rhd$

Outline

1. Setting the Scene: Motivation & Basics of Phonology
   - Our Goal: Theory Comparison
   - Theory 1: SPE
   - Theory 2: Government Phonology (GP)
2. Formalization of GP
   - Definition of our Modal Logic
   - Axiomatization — Simple Cases
   - Axiomatization — Empty Categories and Feature Spreading
3. Parameters and Their Effects on Expressivity
   - Feature Systems
   - Syllable Template
   - Distance of Feature Spreading
4. Conclusion

The Equivalence of Feature Systems

Theorem

For every finitely valued feature system, there is an equivalent system with privative features.

Proof.

Introduce a privative feature f^i for every finitely-valued feature f and appropriate feature value i. Since both the number of features and the number of feature values is finite, the privative feature system is finite, too. ☐

Open Problem I

What is the impact of privativity (the ban against spreading of negative feature values) when the set of features is fixed?

Generalizing the Syllable Template

No current phonological theory uses more distinct constituents than GP, but many allow for more than binary branching within a constituent.
⇒ relax the branching restriction and allow for fewer constituents (licensing conditions and mapping to sounds needs to be adapted, too)

Example

	O	N	C
GP	binary	binary	unary
Strict CV	unary	unary	—
SPE	—	unbounded	—

Syllable Templates Affect Expressivity

Syllable templates can have a restrictive impact on generative capacity.

Example

The following SPE rule generates languages containing no vowels:

$$[\ ] \rightarrow [+cons]$$

"For every segment, set its value for *consonantal* to +."

This is not a GP-language, since at least every other nucleus (and hence a vowel) has to be pronounced in GP.

Open Problem II

Under what conditions does the syllable template negatively affect expressivity?

The Limits of Spreading So Far

So far, we can only implement unbounded feature spreading as iterated bounded feature spreading. We increase the power of spreading by adding new operators familiar from temporal logic.

- $\lhd^+/\rhd^+$ . . . transitive closure of $\lhd/\rhd$
- $U(\phi, \psi)/S(\phi, \psi)$ . . . ϕ holds until ψ
- ν . . . least fixed-point operator

Generative Capacity of the New Variants

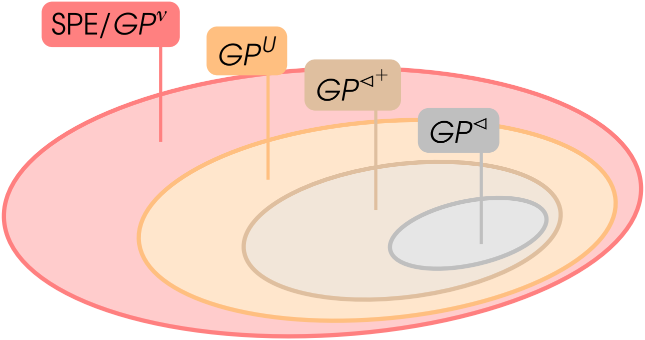
	Temporal logic	Predicate logic	Formal language
$GP^{\lhd^n}$	—	—	SL_{2n+1}
$GP^{\lhd^+}$	RTL	FO^2	—
GP^U	LTL	FO	star-free
GP^ν	RLTL/ν-LTL	MSO	regular

Empirical Motivation for Stronger Operators (Graf 2009)

GP^U: n-retroflexion in Sanskrit (aka *nati*)

GP^ν: stress assignment in Creek and Cairene Arabic

Diagram of Expressivity



Empirical Motivation for Stronger Operators (Graf 2009)

GP^{U}: n-retroflexion in Sanskrit (aka *nati*)
GP^{ν}: stress assignment in Creek and Cairene Arabic

Conclusion

- GP
    - Formalization of GP in a simple modal logic
    - Only spreading may require going beyond quantifier depth 2
- Parameters
    - Feature system type not important
    - Spreading the decisive factor in expressivity (surprising given how little is said about it in the literature)

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Evanston.

Graf, Thomas. 2009. Comparing incomparable frameworks — a model theoretic approach to phonology. Handout of a talk given at the 33rd Penn Linguistics Colloquium (PCL33), March 27 – 29, University of Pennsylvania, Philadelphia, PA. Available online: `http://tgraf.bol.ucla.edu/doc/talks/PLC33_mtp.pdf`.

Kaplan, Ronald M., and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20:331–378.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology Yearbook* 7:193–231.