

Class 5: How to develop a rule analysis; The duplication and conspiracy problems

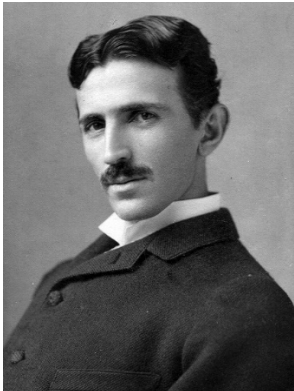
Overview: First, we practice developing a rule analysis. Then: Sometimes it looks like multiple parts of the grammar are doing the same thing. Is this bad, and if so can we do anything about it?

How this fits in: My ulterior motive is to make you yearn for constraints today, then next time experience the agony of trying to make them work in a rule theory, so that you can understand why phonologists so enthusiastically embraced OT (all constraints, no rules!).

1. Developing a rule analysis

- In real life, you are probably not starting with transcribed, simplified, arranged data like this
 - You may be starting from a reference grammar, or elicitation notes/recordings (your own or someone else's), or the ability to ask a speaker questions about their language, or your intuitions about your own language
- So we are cheating quite a bit by starting with a “problem set”
 - If you want to make things more interesting, take Ling 210A/B, Field Methods!
- Data: Adjectives and verbs in “Serbo-Croatian,” a concept that has become politically outdated
 - Now more often called Bosnian-Croatian-Montenegrin-Serbian or BCMS
 - Or Bosnian, Croatian, etc. are now talked about as separate languages
 - “Serbo-Croatian” referred to the main language spoken in what was then Yugoslavia
 - Indo-European, about 16 million speakers
 - English words that may be from BCMS: *cravat* (via French), *dinar*, *paprika*, *hussar* (both via Hungarian), *slivovitz* (via German and Bulgarian), *vampire* (via French and Hungarian)

Nikola Tesla, inventor



Novak Djokovic, tennis player



Marina Abramovich, artist



- K&K explain various ways that these data are simplified
- The ´ mark shows that a vowel is “accented” (falling or rising pitch)
 - We can call these [+accent]
 - This is the aspect of the data is the most simplified
 - We’re told not to worry about accent in the words that don’t have one marked

¹ Wikimedia commons: Tesla_circa_1890.jpeg, Novak_Djokovic_(35488012711).jpg, Marina_Abramović.jpg

Our procedure

- We'll start from the top row, make a guess, then test our guess against the next row
- Whenever we have to make an analytic choice, we'll keep track of it in a **decision tree**

Here's our data, from the Kenstowicz & Kisseberth 1977 textbook, p. 74. I'll give you some extra paper for working on your analysis. The top row and left column are for writing underlying forms.

	<i>masculine</i>	<i>fem.</i>	<i>neuter</i>	<i>plural</i>	
	bogat	bogata	bogato	bogati	'rich'
	mlád	mladá	mladó	mladí	'young'
	púst	pustá	pustó	pustí	'empty'
	béo	belá	beló	belí	'white'
	mío	milá	miló	milí	'dear'
	zelén	zelená	zelenó	zelení	'green'
	križan	križana	križano	križani	'cross'
	jásan	jasná	jasnó	jasní	'clear'
	ledan	ledna	ledno	ledni	'frozen'
	dóbar	dobrá	dobro	dobrí	'kind'
	bodar	bodra	bodro	bodri	'kind'
	veseo	vesela	veselo	veseli	'gay'
	ustao	ustala	ustalo	ustali	'tired'
	múkao	muklá	mukló	muklí	'hoarse'

	<i>1st person singular present</i>	<i>feminine past</i>	<i>neuter past</i>	<i>masculine past</i>	
	tepém	teplá	tepló	tépao	'wander'
	skubém	skublá	skubló	skúbao	'tear'
	tresém	treslá	tresló	trésao	'shake'
	vezém	vezlá	vezló	vézao	'lead'
	pletém	plelá	pleló	pléo	'plait'
	kradém	kralá	kraló	kráo	'steal'
	metém	melá	meló	méo	'sweep'
	vedém	velá	veló	véo	'lead'

2. How does ChatGPT do on the Malagasy assignment, as of October 2023?

To account for the alternations in the Malagasy verb paradigms, we need to propose underlying forms for the roots and affixes, as well as the rules and their order. We will follow the given instructions and ensure that our analysis is clear and straightforward.

In Malagasy, stress falls primarily on the penultimate syllable of the word. To account for this, we can propose the following stress rule:

Stress Rule:

- $V \rightarrow [+stress] / _C0(VC0)\#$

This rule states that the second-to-last vowel in a word should bear stress if there is one. If the word is monosyllabic, the last vowel will bear stress.

Next, we will analyze the provided data by grouping verbs based on the behavior they exhibit in terms of affixation and stress patterns. We will also propose underlying forms for the roots and affixes.

Correct, but I gave this rule and its paraphrase in the instructions

Pattern 1: Unaffixed Noun or Adj. → Passive Imperative

- *bábu* → *babúi* (plunder)
 - Underlying forms: *babu* (root) + *i* (passive imperative suffix)
- *bári* → *baríu* (brilliant)
 - Underlying forms: *bari* (root) + *u* (passive imperative suffix)

This is impressive for a chatbot! It understood the task, and found a morpheme boundary. But not impressive for a human: the passive imperative suffix has to have the same underlying form every time.

Pattern 2: Active Imperative → Passive1 → Passive2

- *babúa* → *babúana* → *babúina* (plunder)
 - Underlying forms: *babu* (root) + *a* (active imperative suffix) → *babu* (root) + *ana* (passive1 suffix) → *babu* (root) + *ina* (passive2 suffix)
- *baría* → *bariana* → *barína* (brilliant)
 - Underlying forms: *bari* (root) + *a* (active imperative suffix) → *bari* (root) + *ana* (passive1 suffix) → *bari* (root) + *na* (passive2 suffix)

I think it's trying to derive the 'passive 2' from the 'passive 1', instead of deriving them both from an underlying root? And again, it has two different underlying forms for the 'passive 2' suffix. Also, we need an explanation of why some of these don't appear to follow the stress rule.

Pattern 3: Unaffixed Noun or Adj. → Passive Imperative

More of the same

- *bá^hgu* → *ba^hgúi* (plait)
 - Underlying forms: *ba^hgu* (root) + *i* (passive imperative suffix)

Pattern 4: Active Imperative → Passive1 → Passive2

- *ba^hgúa* → *ba^hgúana* → *ba^hgúina* (plait)
 - Underlying forms: *ba^hgu* (root) + *a* (active imperative suffix) → *ba^hgu* (root) + *ana* (passive1 suffix) → *ba^hgu* (root) + *ina* (passive2 suffix)

By analyzing these patterns, we can see that the alternations in the verb paradigms are accounted for by a combination of underlying forms for the roots and affixes, as well as a set of rules for affixation. The rules for affixation occur in the following order: active imperative, passive1, passive2. This analysis is not substantially more complicated than necessary and clearly explains how the alternations in the data are derived.

It's repeating back the grading specifications, which is impressive.
But no, the alternations were not accounted for at all. And the rules for affixation shouldn't occur in *any* order, because each word is eligible for at most one of them.

Me: What about the words that don't seem to obey the stress rule?

You're absolutely right; there are instances in the data where the stress does not appear to follow the stress rule. For example, the word "babúana" (plunder) exhibits stress on the antepenultimate syllable instead of the penultimate, as predicted by the stress rule. To account for these exceptions without changing the stress rule or adding more stress rules, we can propose an additional rule that affects stress placement in specific circumstances. Let's call this rule the "Stress Exception Rule."

Wow! I'm really impressed that it identified a counterexample and explained how it is a counterexample.

Stress Exception Rule:

- $V \rightarrow [+stress] / __C0(V\#)C(C)0$

But no, the instructions say that you're not allowed to introduced more stress rules.

This rule states that when a word has two or more syllables and ends with a vowel followed by a consonant, and the penultimate syllable contains a long vowel or a diphthong, stress falls on the antepenultimate syllable. The presence of the long vowel or diphthong in the penultimate syllable triggers this exception to the default stress rule.

There were a couple more paragraphs exemplifying how this works, but they were kind of incoherent, so I cut them for space

- It's possible we could improve ChatGPT's performance by arguing with it some more, but I left off there, in part because I don't want to ruin this problem set by making it ChatGPT-able

3. Dynamic vs. static phonology

- The ‘dynamic’ phonology of a language is the phonology that shows up in alternations. We have analyzed this with rules:

cat[s]	walk[t]
dog[z]	jog[d]

- The ‘static’ phonology is the generalizations that hold of monomorphemic words. Often analyzed with morpheme structure rules/constraints:

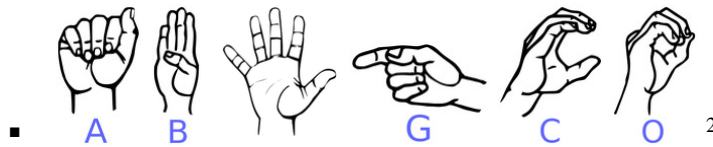
[læps], [lɪst] but no words like *[læpz], *[lɪsd]

- ⌘ Let’s try writing both a phonological rule and a morpheme structure rule for this. Then, let’s see if we can devise an “ordering solution” as you read about in (Kenstowicz & Kisseberth 1977).
 - Reminder: the “ordering solution” (p. 428) says, turn the morpheme structure rule into a normal rule, and insert it into the rule ordering

4. Side point: why *morpheme* structure constraints and not *word* structure constraints?

- English *words* can have sequences like *si*[ksθs] and *a*[skt]
 - But English *morphemes* can’t

- Two-handed ASL *morphemes* obey “Battison’s conditions” (Battison 1978)
 - Symmetry Condition: if non-dominant hand moves, must have same handshape and movement as dominant hand
 - Dominance Condition: if non-dominant hand doesn’t move, handshape is from a restricted set



- But when a root morpheme is combined with a “classifier” morpheme, Battison’s conditions can be violated in the resulting *word*
 - Example from Emmorey 2001, p. 87: ASL ‘A bicycle is near the house’
 - During BIKE, the weak hand is making a shape that is not from the usually-allowed set [*I think this doesn’t count as a C shape*], because it represents the classifier for ‘whole entity’ (refers back to HOUSE)



HOUSE

whole-entity CL + *loc*

BIKE

whole-entity CL + *loc*

5. Conceptual remarks (Stanley 1967 is an early but hard-to-read discussion of many of these issues)

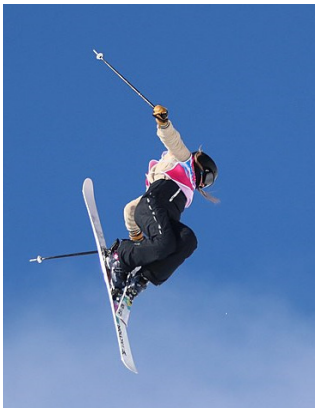
- Morpheme structure rules/constraints are weird:
 - no one is claiming that the English lexicon actually contains /ækd/, repaired by MSR to ækt
 - after all, on hearing [ækt], why would a learner construct a lexical entry /ækd/ instead of /ækt/?
- But if speakers know that ækd is bad, that must be expressed somewhere in the grammar of English:
 - e.g., if they reject ækd as a new word, or have trouble distinguishing between ækd and a legal alternative.
- Some might claim that the lexicon contains /ækd/, with a final consonant underspecified for [voice].
 - Still, if the MSR applies only to underspecified Cs, what *would* happen to hypothetical fully specified /ækd/? What prevents it from existing??

² Thanks, www.wpclipart.com/sign_language/ !

- This comes back to the ‘lexical symmetry’ idea we saw in K&K’s discussion of Russian final devoicing:
 - The grammar needs to explain, one way or another (phoneme inventory, MSRs, or rules), why certain types of underlying forms don’t occur.
- ? An even weirder case: some English speakers think that *slol* and *smæŋ* sound strange.³ But if we tried to write a rule to change them, instead of merely a constraint banning them, what would they change to??

6. Example: Estonian

- Finno-Ugric language from Estonia with 1.1 million speakers
- Official language of Estonia



Kelly Sildaru, freestyle skier



Arvo Pärt, composer



Kerli, singer/songwriter

- I’ve seen the basic data cited as being from Prince 1980, but I couldn’t find them there (?).
 - Data below are just spelling [which does not reflect all three length levels] plus some guesses about syllabification that I hope are reasonable, from this Estonian noun decliner: www.filosoft.ee/gene_et, using additional roots from Blevins 2005.

³ There are few monosyllabic words like this—here are all the examples from the CMU Pronouncing Dictionary, excluding probable proper names. Oxford English Dictionary (oed.com) has a few more but they are pretty obscure.

$s\{p,m\}C_0VC_0\{p,b,m\}$: smarm(y), smurf, spam, sperm, spiff(y), spoof

$s\{m,n\}C_0VC_0\{m,n,\eta\}$: smarm(y)

$\{f,s\}\{l,r\}C_0VC_0\{l,r\}$: shrill, slur, slurp—notice none with $l\dots l$ or $r\dots r$

$skC_0VC_0\{k,g,\eta\}$: skink, skulk, skunk

- Estonian content morphemes have a **minimum size**: at least two syllables or one “heavy” syllable ((C)VV or (C)VCC):
 */ko/, */ma/, */kan/ ← no good because they would be a single “light” syllable

- Estonian also has a rule deleting final vowels in the nominative sg.:

	<i>nom. pl</i>	<i>nom. sg.</i>	
/ilma/	il. ma -d	ilm	‘weather’
/matsi/	mat. si -d	mats	‘lout’
/konna/	kon. na -d	konn	‘frog’
/tänav/	tä.na. va -d	tä.nav	‘street’
/seminari/	se.mi.na. ri -d	se.mi.nar	‘seminar’
/tuleviku/	tu.le.vi. ku -d	tu.le.vik	‘future’
/raamatu/	raa.ma. tu -d	raa.mat	‘book’

- But the rule fails to apply in certain cases:

/pesa/	pe. sa -d	pe. sa	‘nest’
/kana/	ka. na -d	ka. na	‘hen’
/koi/	koi-d	koi	‘clothes-moth’
/maa/	maa-d	maa	‘country’
/koli/	ko. li -d	ko. li	‘trash’

- ? Let’s try to write a mini-grammar for Estonian that tries to capture these facts. What’s unsatisfying about it?

7. The duplication problem (Kenstowicz & Kisseberth 1977)

= cases where phonological rules and morpheme structure constraints seem to be doing the same thing (‘duplicating’ each other’s effects).

- These troubled researchers from the late 1970s onwards, because it seems (although we don’t actually know) that a single phenomenon (e.g., avoidance of sub-minimal words) should have a single explanation in the grammar.

? Can you remember anything about how this plays out in Chamorro?

8. Another duplication case

- Many sign languages require that a content morpheme can have only one handshape (though within that handshape, fingers can open or close during the morpheme)
- When two roots are put together to form a compound word, there is often a rule that assimilates handshape, so that the resulting word obeys the one-handshape maximum.
 - Hong Kong Sign Language example and images from Tang et al. 2010
 - HKSL is related to Chinese Sign Language



TASTE handshape is

TASTE

GOOD handshape is



GOOD

- TASTE^GOOD (meaning 'tasty') takes the TASTE handshape plus the 'thumb-extended' feature



to get handshape (plus a closing movement): TASTE^GOOD

- In Estonian, a word-shape requirement **prevents** a rule from applying
- In Hong Kong Sign Language, a word-shape requirement **causes** a rule to apply

9. Shortening a grammar

- Using the curly-bracket notation to collapse $\emptyset \rightarrow V / C _ C\#$
 $\emptyset \rightarrow V / C _ CC$
 into the schema $\emptyset \rightarrow V / C _ C\{C,\#\}$ says that these rules have something significant in common.
 (Why? recall SPE's evaluation metric...)

10. Kisseberth 1970: cases where the notation doesn't allow shortening

- These rules have something in common too (what?), but they can't be collapsed using curly brackets:

$$\emptyset \rightarrow V / C _ CC$$

$$C \rightarrow \emptyset / CC _$$
- Cases of languages that have sets of rules like this are called *conspiracies*, and their widespread existence is the *conspiracy problem*.
 - (The difference between a case of the duplication problem and a case of the conspiracy problem is sometimes fuzzy and the terms are sometimes used interchangeably)

11. Constraints

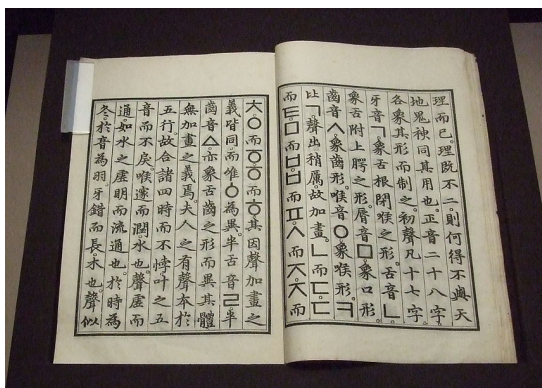
- The $\emptyset \rightarrow V$ and $C \rightarrow \emptyset$ rules both seem to be applying to get rid of CCC sequences
- Moreover, there's a rule that could be made simpler if we invoked a **constraint** *CCC
 - Kisseberth proposes...

Instead of $V \rightarrow \emptyset / V C _ C V$
[−long]

use $V \rightarrow \emptyset / C _ C$ subject to the constraint *CCC (or *{C,#}C{C,#})
[−long]

12. Here's another conspiracy: Korean

- The main language of both North Korea and South Korea
- Considered to form Koreanic family together with Jejuan (from Jeju Island)
 - Relationships beyond that are more controversial
- About 80 million speakers, including around 160,000 in L.A. County
 - That makes it the #5 or #6 most-spoken language in the county
- Has own writing system



Page from Hunminjeongeum Kaerye, commentary on then-new writing system⁴



Kyung-sook Shin, author of *Please Look After Mom*



Garion, developed rhyming conventions for Korean rap⁵

⁴ en.wikipedia.org/wiki/Hunminjeongeum_Haerye#/media/File:Hunminjeongeumhaerye.jpg

⁵ twitter.com/Garionhiphop/photo

Case from Kim & Alderete 2008

- {p, t, tʃ, k} → [+spread glottis] / h ____
 - then, h → Ø / ____ {C, #}
- C → [-spread glottis] / ____ {C, #}

? Find me evidence for each of the rules above

- a. /suh-talk/ → [su.tʰak] ‘rooster’
- b. /ilh-ta/ → [il.tʰa] ‘loses’
- c. /nah.ta/ → [na.tʰa] ‘bear’
- d. /suh-pəm/ → [su.pʰəm] ‘male tiger’
- e. /coh-ke/ → [co.kʰe] ‘well’
- f. /anh / → [an] ‘in’
- g. /suh/ → [su] ‘male’
- h. /natʰ-kε/ → [nat.kʰε] ‘piece’
- i. /kipʰ-ta/ → [kip.tʰa] ‘it is deep’
- j. /apʰ-to/ → [ap.tʰo] ‘front also’
- k. /mitʰ-pa-tak/ → [mit.pʰa.dak] ‘bottom’
- l. /apʰ/ → [ap] ‘front’
- m. /patʰ/ → [pat] ‘field’
- n. /pu-əkʰ/ → [pu.ək] ‘kitchen’

? Proposals for a good constraint here?



13. Constraints as rule blockers

- $V \rightarrow \emptyset / C_C$, unless result would violate *CCC

? Let's try to lay out, step by step, what an algorithm would have to do to implement the rule and its blocking constraint

A question that came up on Perusall: how does this work if there is a sequence of rules? Such as...

- $V \rightarrow \emptyset / C_C$, unless result would violate *CCC
- $\emptyset \rightarrow g / \eta _$, unless result would violate *CCC

Try it for /salipɲa/, /tominu/, /taɲi/, /soɲte/

14. Constraints as rule triggers

- $\emptyset \rightarrow i$, only if needed to eliminate *CCC violation

? What exactly will happen, step by step?

15. Where this leaves us

- Many more conspiracies were identified, giving rise to more constraints.
- People liked constraints, because they solved the conspiracy problem and also gave clearer theoretical status to the idea of “markedness”
 - Everyone knew languages don’t “like” CCC sequences (they are “marked”), but this was not directly encoded in grammars until constraints like *CCC came along.
- On the other hand, we’ll see that it’s unclear exactly how constraints should work.
 - Next time we’ll wallow in this problem
 - Then we’ll start trying to solve it

Closing item for index cards: Write one thing you’re currently finding appealing about constraints, and one thing you’re currently finding problematic about them.

References

- Battison, Robin. 1978. *Lexical borrowing in American Sign Language*. Silver Spring: Linstok Press.
- Blevins, James P. 2005. Word-based declensions in Estonian. *Yearbook of Morphology* 2005. 1–25.
- Emmorey, Karen. 2001. *Language, Cognition, and the Brain: Insights From Sign Language Research*. Psychology Press.
- Kenstowicz, Michael & Charles Kisseberth. 1977. *Topics in Phonological Theory*. New York: Academic Press.
- Kim, Lan & John Alderete. 2008. Why [h] and aspirated consonants are different in Korean. *Toronto Working Papers in Linguistics* 28. <https://twpl.library.utoronto.ca/index.php/twpl/article/view/6140>. (9 October, 2022).
- Kisseberth, Charles. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1. 291–306.
- Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11. 511–562.
- Stanley, Richard. 1967. Redundancy Rules in Phonology. *Language*. Linguistic Society of America 43(2). 393–436. <https://doi.org/10.2307/411542>.
- Tang, Gladys, Diane Brentari, Carolina González & Felix Sze. 2010. Crosslinguistic variation in the use of prosodic cues: The case of blinks. In Diane Brentari (ed.), *Sign languages*, 519–541. Cambridge: Cambridge University Press.