Complexity and the Phonological Turing Machine

Louis Van Steene

10/04/22

University of Cambridge

Information

- Slides and source code available at lvs.me.uk
- Contact: louis@lvs.me.uk
- This presentation is based on a paper currently under review at JoULAB. You can download a preprint at the link above.

Outline

Summary

1. A good theory is not only descriptive but explanatory.
2. The theoretical similarities and differences between RBP, NP, OT and SFP have important implications for the respective explanatory adequacy of these theories.
3. The Turing programme for linguistic theory is a productive approach for looking at phonology.
4. Big-O notation is a useful metric of complexity, and in turn explanatory adequacy.

1. Introduction

1. Introduction

1.1. Data, phenomena, theory

Data in phonology

- Different kinds of evidence that we can use to inform theories about phenomena.
- (Corpus-)internal and external evidence (Kenstowicz & Kisseberth, 1979; Ohala, 1986):
    - Internal evidence is usually considered most important.
    - External evidence: language acquisition, language games, artificial language learning...

Bogen and Woodward (1988)

- Phenomena are separate from both data on the one hand, and theory on the other.
- The theory itself can influence (the collection of) the data which can then inform the postulation of phenomena.
- Thus there is a tension between accounting for the data directly and accounting for the phenomena as they actually are, abstracting from the noise (cf. Chomsky, 2017; Galilei, 1632/1967)

1. Introduction

1.2. Descriptive and explanatory adequacy

Descriptive adequacy

- Observational adequacy: restating the data (not in fact trivial).
- Descriptive adequacy: specify the data “in terms of significant generalisations that express underlying regularities in the language” (Chomsky, 1964, p. 28).
- In traditional terms, coming up with rules which generate the data.

Explanatory adequacy

- Most fundamentally, how does one select between different descriptively adequate theories?
- How do we account for the actual knowledge of the speaker?
- External evidence?
- Learnability and evolvability (cf. Chomsky, 1995, i.a.).
    - Hence discussion of 'complexity'.

1. Introduction

1.3. Competence and performance

Competence and performance (Chomsky, 1986)

- Competence: the specifically linguistic knowledge of the speaker (I-language).
- Performance: the actual linguistic productions of the speaker, mediated by numerous factors beyond the knowledge of language itself.

2. Phonological Theory

Goals

1. What is a theory of phonology?
2. What makes a theory of phonology *explanatory*?

2. Phonological Theory

2.1. Some theories

SPE (Chomsky & Halle, 1968) I

- Classic rules-based analyses.

(1) (a) $/\text{-z}/ \rightarrow [\text{s}] \ / \ \{ \text{p t k} \ldots \} \ \_$

$/\text{-z}/ \rightarrow [\text{z}] \ / \ \{ \text{b d g} \ldots \} \ \_$

$/\text{-z}/ \rightarrow [\text{əz}] \ / \ \{ \text{s z ʃ ʒ} \ldots \} \ \_$

(b) Epenthesis: $\emptyset \rightarrow [\text{ə}] \ / \ \_ \ \langle C \rangle \#$

Voicing Assimilation: $[\text{-son}] \rightarrow [\alpha\text{voice}] \ / \ [\text{-son}, \alpha\text{voice}] \ \_ \ ]_\sigma$

SPE (Chomsky & Halle, 1968) II

Central hallmarks:

1. The grammar consists of rewrite rules which operate on sequences of phonemes, potentially transforming them into a different sequence.
2. These rules apply cyclically.
3. The input to the phonological component is a sequence of base forms known as underlying representations, originally stored in the lexicon and possibly manipulated by morphosyntactic processes.
4. The output of the phonological component is a surface representation, representing phonetic instructions of some sort.

Natural Phonology (Stampe, 1979) I

- Epilogue of *SPE*: *markedness*.
  - Potential source of theory evaluation?
  - Something (phoneme, rule) is 'marked' if it is somehow less plausible or optimal than something unmarked.
  - Originates from Prague school (Trubetzkoy, 1958/1969), see also Hume (2011).
- *Processes*: automatic, parallel, cyclic; must be suppressed.
- *Rules*: exceptions that must be learned.

Natural Phonology (Stampe, 1979) II

An example (cf. Stampe's 'divinity fudge' example):

(5) /kaɹnɪvɫ̩/

Syllabification:	*kaɹ.nɪ.vɫ̩
Flapping:	*kaɹ.ɾɪ̃.vɫ̩
Nasal assimilation:	kãɹ̃.ɾɪ̃.vɫ̩
Flap deletion:	kãɹ̃.ɪ.vɫ̩

[…vowel assimilation, syllabification, vowel shortening …]

[kãɚ̰.vɫ̩]

Optimality Theory (McCarthy & Prince, 1993; Prince & Smolensky, 1993)

- Ranked constraints (markedness; faithfulness) instead of rules.
- Similar to Stampe's processes.

(2)

Candidates	*VOICED-CODA	IDENT-IO(voice)
a. ☞ [bɛt]		*
b. [bɛd]	*!	

Key: *VOICED-CODA can be read as 'voiced obstruent in coda is marked; IDENT-IO(voice) means 'corresponding input and output segments should match for [±voice]. The candidates are produced by GEN, the workings of which are not considered relevant for the evaluation (but cf. Section 4 below). Asterisks indicate violations of constraints. Candidates are in effect in a race: (b) loses, because *VOICED-CODA outranks IDENT-IO(voice), as indicated by the exclamation mark. See Kager (1999) for a comprehensive introduction to Classic OT.

See Kager (1999) for an introduction to Classic OT. I'm ignoring more recent versions of the theory in this presentation.

Substance-Free Phonology (Hale & Reiss, 2008; Samuels, 2009)

- Similar to *SPE* minus the epilogue.
- Features are *substance-free*.
- Samuels (2009): three operations
    1. Search
    2. Copy
    3. Delete

2. Phonological Theory

2.2. Some points of comparison

Three problems (Hale & Reiss, 2008)

1. **The (weak) AI problem**: to create a model that is 'weakly equivalent' to language—i.e. a model that generates the same output strings as language.
2. **The Human's problem**: the learner acquires only one grammar, rather than any number of other extensionally equivalent grammars.
3. **The Linguist's problem**: to work out how learners solve the Human's problem.

Language acquisition I

- ‘Child Phonology’.
- Do children have adult competence?
    - What does UG provide the child?
    - How much is learned?

Child Phonology: Problem I

- Smith (1973): opaque chain shifts.
- No constraint order can give these effects.

(4) 'puzzle' /pʌzəl/ -> [pʌdəl]
'puddle' /pʌdəl/ -> [pʌgəl]

Child Phonology: Problem II

- Variability.
- Ferguson (1986): /pɛn/ →/buã/

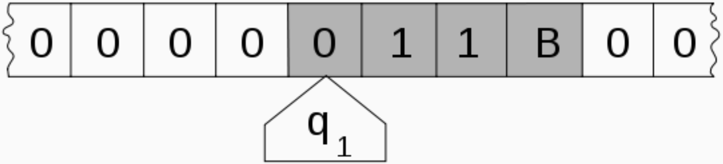
3. The Phonological Turing Machine

3. The Phonological Turing Machine

3.1. Simple introduction to Turing machines

Turing Machine I

- Turing (1936)
- The Turing machine is an *abstract mathematical device* that encodes the idea of computation.
- Every computational procedure has a corresponding Turing machine.



- A derivation is like lines of a proof.

3. The Phonological Turing Machine

3.2. How would a Turing machine look w.r.t. phonology?

Rules and representations

- McCarthy (1988, p. 84): “if the representations are right, then the rules will follow”
- My proposal: if the derivation is correct, the representation will follow.

The Turing programme for linguistic theory

- Watumull (2012, 2015): mind as Turing machine (cf. Gallistel & King, 2010), language as Turing machine. Mainly focused on syntax.
- Vaux and Watumull (2012): paper applying to phonology that never got past the draft stage.

Consequences

- Does this approach provide an evaluation procedure that allows us to select between theories, even if said theories are empirically equivalent?
- Does this get us any closer to explanatory adequacy?

4. Complexity

4. Complexity

4.1. Why look at complexity?

Complexity as evaluation metric

- Generally, the simpler the theory the better.
- However, the how ‘simple’ or ‘optimal’ a particular theory is isn’t always clear.

4. Complexity

4.2. A brief introduction to Big-O notation

Big-O notation I

- A way of quantifying time and space complexity.
- From computer science.

Big-O notation II

Table 1: Common Big-O complexities

Notation	Description	Evaluation
$O(1)$	Constant	Excellent
$O(\log n)$	Logarithmic	Good
$O(n)$	Linear	Fair
$O(n \log n)$	Log-linear	Poor
$O(n^2)$	Polynomial	Very poor
$O(2^n)$	Exponential	Extremely poor
$O(n!)$	Factorial	Extremely poor

A polynomial is always represented by the degree of said polynomial.

4. Complexity

4.3. Why apply this to phonology?

Big-O and phonology

Co-opting this approach into phonological theory means is that we do not need to decompose the operations that we take as primitive, assuming that we can estimate the order of magnitude of their time complexity. This is the benefit of the abstraction: we do not necessarily need to deal with neurophysiologically plausible operations, which are far from being fully understood (although the stronger argument, supported by Watumull [p.c.], is that this abstraction is in fact very close to the reality). Instead, we deal with an abstraction of the architecture, the Turing machine, and build our theory of economy on top of this architecture, using time and space complexity in the form of Big-O notation. As a result, our phonology can contain any computational operation (since it is a Turing machine), and we have a means of evaluating which (constraints on) operations are more plausible, based on time and space complexity.

4. Complexity

4.4. An example: SFP vs OT

SFP

- COPY and DELETE are both $O(1)$
- SEARCH is worst case $O(n^2)$
- But the search space should be greatly limited by phases and other principles of locality.

- For m outputs and n constraints the algorithm takes $O(mn)$—effectively quadratic, which is very inefficient.
- The learning algorithm fares even worse, because of factorial typology. $O(n!)$ complexity?

5. Conclusions

Conclusions

- It is productive to consider carefully the computational properties of phonological theories, both in terms of acquisition and in terms of synchronic complexity.
- Computational complexity, measured in terms of Big-O notation can be a useful way of quantifying explanatory adequacy.
- Classic OT faces numerous problems from a computational, acquisitional and typological perspective.

References i

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, *97*(3), 303–352.

Chomsky, N. (1964). *Current issues in linguistic theory*. Mouton.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press.

Chomsky, N. (2017). The Galilean challenge. *Inference*, *3*(1).

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. MIT Press.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 36–57). Lawrence Erlbaum Associates, Inc.

Galilei, G. (1967). *Dialogo sopra i due massimi sistemi del mondo* [Dialogue concerning the two chief world systems] (S. Drake, Trans.; 2nd ed.). University of California Press. (Original work published 1632)

Gallistel, C. R., & King, A. P. (2010). *Memory and the computational brain: Why cognitive science will transform neuroscience*. Wiley-Blackwell.

- Hale, M., & Reiss, C. (2008). *The phonological enterprise*. Oxford University Press.
- Hume, E. (2011). Markedness. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 79–106). Wiley-Blackwell.
- Kager, R. (1999). *Optimality Theory*. Cambridge University Press.
- Kenstowicz, M. J., & Kisseberth, C. (1979). *Generative phonology*. Academic Press.
- McCarthy, J. J. (1988). Feature geometry and dependency: A review. *Phonetica*, *45*(2-4), 84–108.

McCarthy, J. J., & Prince, A. (1993). *Prosodic Morphology: Constraint interaction and satisfaction* (Technical Report RuCCS-TR-3). Rutgers University Center for Cognitive Science.

Ohala, J. J. (1986). Consumer's guide to evidence in phonology. *Phonology*, *3*, 3–26.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Technical Report RuCCS-TR-2/CU-CS-696-93). Rutgers University Center for Cognitive Science. Rutgers University, Piscateway, NJ.

References v

- Samuels, B. D. (2009). *The structure of phonological theory* (Doctoral dissertation). Harvard University. Cambridge, MA.
- Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge University Press.
- Stampe, D. (1979). *A dissertation on natural phonology*. Garland.
- Trubetzkoy, N. (1969). *Grundzüge der Phonologie* [Principles of phonology] (C. A. M. Baltaxe, Trans.). University of California Press. (Original work published 1958) original-language: German original-place: Göttingen

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 2nd series 42*(1), 230–265.

Vaux, B., & Watumull, J. (2012). *The phonological Turing machine*.

Watumull, J. (2012). A Turing program for linguistic theory. *Biolinguistics*, *6*(2), 222–245.

Watumull, J. (2015). *The linguistic Turing Machine* (Doctoral dissertation). University of Cambridge. Cambridge, UK.