Formalizing phonological knowledge verses usage with weighted logic

Scott Nelson – UIUC (*sjnelson@illinois.edu*)

Overview

(1) Traditionally, two types of phonological knowledge are of interest (*cf.* Anderson 1985, Heinz 2018).

a. *Representational Knowledge*: what is a possible long term memory representation? what is a possible output of the phonological grammar?

b. *Process Knowledge*: how do we relate phonological inputs to phonological outputs?

(2) Some researchers believe that phonological knowledge should also include the rate at which phonological generalizations hold based on things like *The law of frequency matching*.

a. Hayes et al. (2009: p. 826): "speakers of languages with variable lexical patterns respond stochastically when tested on such patterns. Their responses aggregately match the lexical frequencies."

b. See also Magri (2023: p. 4): "If we do categorical phonology based on fieldwork data...there are no generalizations for us to capture here. There is nothing for us to do. If we instead turn from categorical data to frequencies extracted from corpora or obtained through psycholinguistic experiments, we will now see that straightforward phonological generalizations emerge."

(3) There is no denying that variable/gradient behavior exists, but the question always becomes, "how do we include gradience in our formal phonological models?"

a. Option A: gradience goes directly into the phonological grammar (Hayes 2022)

b. Option B: gradience emerges outside the narrow phonological grammar somehow $\Leftarrow$

(4) I believe that our goal as theorists should be to formally explain what the capacities are that underly human cognition (Cummins 1983) which can be defined as mathematical functions (Marr 1982, van Rooij & Baggio 2021, Guest & Martin 2021, Nelson 2024).

a. *Capacity 1*: our phonological system can generate more than one possible output for a given input.

b. *Capacity 2*: specific outputs at any moment are selected based on a variety of different factors.

(5) In the remainder of the talk I will formalize the previous idea under the broad view of "model-theoretic phonology" which is based around mathematical logic and computation.

a. Mathematical logic allows for clarity in expressing the properties of different linguistic systems.

b. Specifically, I will use semirings and weighted logic (Mohri 1997, Goodman 1999, Droste & Gastin 2009, Roark & Sproat 2007, Chandlee & Heinz 2017, Heinz Forthcoming) to distinguish between what I will refer to as *phonological knowledge* and *phonological usage*.

(6) Nasal place assimilation in English will be the phenomena that empirically grounds the idea (Borowsky 1986, Avery & Rice 1989, Coleman et al. 2016).

Philosophical Priors

(7) Probabilities should model ignorance, not actual behavior.

"...the reification of probabilities as cognitively represented quantities invites a problematic conception of randomness and probability, in which randomness plays a shallow role in the generation of behavior [...] The states of physical systems are not governed by probabilities. Probabilities are merely statistical descriptions that are useful when we are mostly ignorant about the detailed dynamics of a system and its surroundings." - Tilsen (2023: p. 265)

(8) Grammars should model knowledge of language, not use of language (for an individual).

"...it is perhaps worth while to reiterate that a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language by a speaker-hearer." - Chomsky (1965: p. 9)

(9) Phonological knowledge is only one of many factors governing behavior.

"...linguistic knowledge is only one of the inputs to language production, language comprehension, and other forms of language performance. What accounts for the facts of performance is a conjunct of a theory of linguistic knowledge ('What is the nature of the representation of linguistic knowledge?') and a theory of language performance ('How is this knowledge put to use?')." - Mohanan (1986: p. 183)

Model-Theoretic Phonology

(10) A **model signature** $\mathcal{M}$ is a collection of symbols for the functions, relations, and constants that describe structures.

a. This provides the ingredients for formally defining phonological representations.

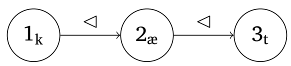
b. For example, the signature $\langle \lhd, \{R_\sigma \mid \sigma \in \Sigma\}\rangle$ gives us two types of operations. The *ordering relation* $\lhd$ lets us linearly order individual elements and the *labeling relations* R_σ allows us to give a label to an element where the labels are drawn from a finite set Σ.

(11) An $\mathcal{M}$**-structure**, A, contains a set called the **domain**, as well as **denotations**

a. The domain is typically an initial sequence of the natural numbers: {1,2,...}.

b. A denotation takes the domain and says which elements of the domain satisfy which properties.

c. Given a set of symbols Σ = {æ,t,k}, the structure for the word *cat* is $\langle \mathcal{D} = \{1, 2, 3\}; \lhd = \{(1, 2), (2, 3)\}; \{R_{\text{æ}} = 2, R_{\text{t}} = 3, R_{\text{k}} = 1\}\rangle$.

d. We can also display this graphically: 

(12) A **logical language** in logic x is defined by combining the symbols of that logic with a specific model signature $\mathcal{S}$. I will focus primarily on **Monadic Second Order** (MSO) logic and its weighted variant.

a. The model signature provides the tools to formalize representations.

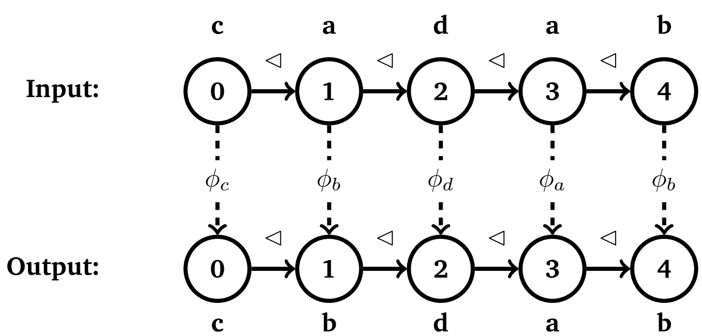
b. The logic provides a way to do inference over the representations that are built.

(13) An **interpretation** of structure A in terms of structure B is a function denoted by a set of n formulas $\{\phi_i, \ldots, \phi_n\}$ where n is equal to the number of functions, relations, and constants in A's model signature (Courcelle 1994, Engelfriet & Hoogeboom 2001).[1]

a. A formula $\phi_P(x) := Q(x)$ denotes that domain element x has property P in the output structure if it has property Q in the input structure.

(14) These types of logical interpretations can be used to define phonological input-output maps such as **a** $\rightarrow$ **b/c_d** (Strother-Garcia 2019, Chandlee & Jardine 2021, Nelson 2024, Heinz Forthcoming).

a.
$$\begin{array}{lcl}
\phi_a(x) & := & a(x) \wedge \neg\exists y, z[y \lhd x \lhd z \wedge c(y) \wedge d(z)] \\
\phi_b(x) & := & b(x) \vee (a(x) \wedge \exists y, z[y \lhd x \lhd z \wedge c(y) \wedge d(z)]) \\
\phi_c(x) & := & c(x) \\
\phi_d(x) & := & d(x)
\end{array}$$

b. 

(15) The model-theoretic approach also provides a way for identifying substructures with logic. This is useful when thinking about static phonotactic or morpheme structure constraints.

a. Does structure A contain the substring cad?

b. $\varphi := \exists x[\exists y[y \lhd x \wedge c(y)] \wedge \exists z[x \lhd z \wedge d(z)] \wedge a(x) \wedge \texttt{TRUE}]$

c. The last conjunct isn't strictly necessary, but will make more sense as we progress through the handout.

(16) These can be thought of as boolean functions over strings, but we can generalize these to be functions from strings to some member of a set S.

a. $f : \Sigma^* \rightarrow \{\texttt{TRUE}, \texttt{FALSE}\}$

b. $f : \Sigma^* \rightarrow S$

[1] Interpretations also require a domain formula, copy set, and licensing function which I will ignore for the purpose of this talk.

Weighted Logic

(17) The types of set S we are interested in are called **semirings** (Mohri 1997, Goodman 1999, Droste & Gastin 2009, Roark & Sproat 2007, Chandlee & Heinz 2017, Heinz Forthcoming). The four properties below are crucial for today's talk but not complete.

a. A binary "addition" operator $\oplus$

b. A binary "multiplication" operator $\otimes$

c. A value **0** which is an identity for $\oplus$

d. A value **1** which is an identity for $\otimes$

(18) The boolean semiring contains the values $\mathbb{B} = \{\texttt{TRUE}, \texttt{FALSE}\}$.

a. Disjunction acts as addition with `FALSE` as its identity 0 since $([b \in \mathbb{B}] \vee \texttt{FALSE}) = b$.

b. Conjunction acts as multiplication with `TRUE` as its identity 1 since $([b \in \mathbb{B}] \wedge \texttt{TRUE}) = b$.

(19) Both the probability semiring ($\mathbb{R}_{\geq 0}$) and the natural number ($\mathbb{N}$) semirings have standard addition and multiplication as their operators and 0 and 1 as their identities.

(20) **Weighted MSO Logic** adds the following conditions to standard MSO logic.

a. MSO logic (roughly): quantification over individuals and sets of individuals plus all standard logical connectives and definitions.

b. $s \in$ S is an atomic formula

c. Negation is only allowed in atomic formulas

d. $\phi \wedge \psi$ is interpreted as $\phi \otimes \psi$

e. $\phi \vee \psi$ is interpreted as $\phi \oplus \psi$

f. $\forall x \phi$ is interpreted as $\phi(x_1) \otimes \phi(x_2) \otimes \cdots \otimes \phi(x_n) \;\; \forall x \in \mathcal{D}$

g. $\exists x \phi$ is interpreted as $\phi(x_1) \oplus \phi(x_2) \oplus \cdots \oplus \phi(x_n) \;\; \forall x \in \mathcal{D}$

(21) Weighted logic allows us to use what are essentially the same formula, but make different types of claims about a structure (*cf.* Mayer 2025). Suppose we had a string $abab$.

a. $\exists x, y[a(x) \wedge x \lhd y \wedge b(y) \wedge \texttt{TRUE}]]$

b. $[a(1) \otimes (1 \lhd 1) \otimes b(1) \wedge s] \oplus [a(1) \otimes (1 \lhd 2) \otimes b(2) \wedge s] \oplus \ldots \oplus [a(4) \otimes (4 \lhd 4) \otimes b(4) \wedge s]$

c. If we use the boolean semiring we turn all the $\oplus$ symbols back to disjunction and all the $\otimes$ symbols back to conjunction. Assuming s once again is `TRUE`, this returns a truth value which tells us whether or not the string contains the sequence ab *(binary wellformedness)*.

d. If we use the natural numbers semiring we turn all the $\oplus$ symbols into regular addition and all the $\otimes$ symbols into regular multiplication. Assuming s equals 1, this returns a count of how many times a string contains the sequence ab *(counting violations)*.

e. If we use the natural numbers semiring we turn all the $\oplus$ symbols into regular addition and all the $\otimes$ symbols into regular multiplication. Assuming s equals an arbitrary value, this returns a weight based on how many times a string contains the sequence ab *(weighted violation/continuous wellformedness)*.

(22) What will be key for the proposed analysis is the **finite language semiring** (*cf.* Beros & de la Higuera 2014).

a. The set S is the set of all finite languages (stringsets).

b. $\oplus$ is union ($\cup$), $\otimes$ is language concatenation ($\cdot$), **0** is the empty set ($\emptyset$), and **1** is the set containing the empty string ($\{\lambda\}$). Importantly, $\emptyset \cdot A = A \cdot \emptyset = \emptyset$.

c. When this is used, we can define functions from strings to *sets of strings* ($f : \Sigma^* \rightarrow \mathcal{P}(\Sigma^*)$)

d. We can now write logical formula that compute optional **b→c/a_a**.

e. $\forall x \exists y, z[(a(x) \wedge \{a\}) \vee (b(x) \wedge \{b\}) \vee ((b(x) \wedge a(y) \wedge a(z) \wedge y \lhd x \lhd z) \wedge \{c\})]$

Phonological Knowledge and Phonological Usage

(23) The general idea for the system is to split the computation into two functions: one for phonological **k**nowledge (competence) which determines possible outputs and one for phonological **u**sage which determines a specific output at a given moment (performance).

a. Knowledge/Competence generates possible outputs by mapping input strings to output sets of strings using the finite language semiring ($k : \Sigma^* \rightarrow \mathcal{P}(\Sigma^*)$).

b. Usage/Performance generates weights for the set of possible outputs using the probability semiring ($u : \mathcal{P}(\Sigma^*) \rightarrow \mathbb{R}^n_{\geq 0}$) and chooses an output with something like `argmax` ($x \in \mathcal{P}(\Sigma) \mid \texttt{argmax}(u) = x$).

c. Crucially, this only works if the factors that determine the weights include at least some dynamic properties. This seems to be the exact desired result if we honor the spirit of competence/performance. Knowledge is static(ish) but performance is dependent on environmental conditions.

Nasal Place Assimilation

(24) In English, coronal nasals assimilate in place to a following obstruent across word boundaries. The status of non-coronal nasal assimilation is contested, but it has been argued to be absent in certain (Borowsky 1986, Avery & Rice 1989, Coleman et al. 2016). This gives us pairs like i[m] [p^h]ort Jefferson $\sim$ fro[m] [p^h]ort Jefferson, i[ŋ] [k^h]anada $\sim$ fro[m] [k^h]anada, and i[n] [t^h]acoma $\sim$ fro[m] [t^h]acoma

(25) Production and perception rates vary.

a. Production (Buckeye Corpus; Dilley & Pitt 2007): pre-labial (18.5%) and pre-velar (21.8%)

b. Production (Audio British National Corpus; Coleman et al. 2016): overall (20%)

c. Perception (Coetzee 2016): pre-labial (45%) and pre-velar (35%)

(26) Coetzee (2016) shows that factors like *speech rate*, and *experiment type* are influence the rate of assimilation. We can turn these into functions that give weights to use in our logic.

a. `speechRate` : {slow,faster,fastest} $\rightarrow \mathbb{R}_{\geq 0}$ b. `expType` : {exp1,exp2} $\rightarrow \mathbb{R}_{\geq 0}$

(27) We can first implement the **k** function using the finite language semiring.

a. $\varphi := \forall x[(\text{ŋ}(x) \wedge \{\text{ŋ}\}) \vee (\text{n}(x) \wedge \{\text{n}\}) \vee (\text{m}(x) \wedge \{\text{m}\}) \vee (\text{n}(x) \wedge \texttt{preLabStop}(x) \wedge \{\text{m}\}) \vee (\text{n}(x) \wedge \texttt{preVelStop}(x) \wedge \{\text{ŋ}\}) \vee (?(x) \wedge \{?\})$

b. Extension: {(nt,{nt}), (np,{np,mp}), (nk,{nk,ŋk}), (mp,{mp})}

(28) We can implement the **u** function using the probability semiring.

a. $\varphi := (\texttt{speechRate} \vee \texttt{expType}) \wedge \forall x[(n(x) \wedge \texttt{preLabStop}(x) \wedge w_1) \vee (n(x) \wedge \texttt{preVelStop}(x) \wedge w_2)]$

b. This gives a weight to each output of the k function which can be turned into a probability distribution by using `softmax`. A hypothetical example: $\{np, mp\} \rightarrow [4.3, 17.11] \leftrightarrow [0.21, 0.79]$.

Conclusion

(29) a. Composing the functions suggests they may be viewed as a single function, but only the latter is affected by extra-grammatical factors. This plus the choice of semiring echo the competence/performance distinction.

b. The law of frequency matching becomes a second order effect. Variation is rarely "free" and is regularly explained by internal and external factors. If these create probability distributions, then behavior would match the observed probabilities if language users correctly learn which factors affect choice.

c. Formal logic provides a way to be explicit about the systems one is analyzing and how certain commitments such as discrete phonological knowledge and modularity can coexist with gradient usage data.

References cited. Anderson, S.R. 1985. *Phonology in the twentieth century*. • Avery, P. & K. Rice. 1989. Segment structure and coronal underspecification. *Phonology* 6(2). 179–200. • Beros, A. & C. de la Higuera. 2014. A canonical semi-deterministic transducer. In *International conference on grammatical inference*, 33–48. • Borowsky, T. 1986. *Topics of lexical phonology of english*: University of Massachusetts dissertation dissertation. • Chandlee, J. & J. Heinz. 2017. Computational phonology. In *Oxford research encyclopedia of linguistics*, . • Chandlee, J. & A. Jardine. 2021. Computational universals in linguistic theory: Using recursive programs for phonological analysis. *Language* 97(3). 485–519. • Chomsky, N. 1965. *Aspects of the theory of syntax*. • Coetzee, A. W. 2016. A comprehensive model of phonological variation: Grammatical and non-grammatical factors in variable nasal place assimilation. *Phonology* 33(2). 211–246. • Coleman, J., M. E. L. Renwick & R. A.M. Temple. 2016. Probabilistic underspecification in nasal place assimilation. *Phonology* 33(3). 425–458. • Courcelle, B. 1994. Monadic second-order definable graph transductions: a survey. *Theoretical Computer Science* 126(1). 53–75. • Cummins, R. 1983. *The nature of psychological explanation*. • Dilley, L. C. & M. A. Pitt. 2007. A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America* 122(4). 2340–2353. • Droste, M. & P. Gastin. 2009. Weighted automata and weighted logics. In *Handbook of weighted automata*, 175–211. • Engelfriet, J. & H. J. Hoogeboom. 2001. MSO definable string transductions and two-way finite-state transducers. *ACM Transactions on Computational Logic (TOCL)* 2(2). 216–254. • Goodman, J. 1999. Semiring parsing. *Computational Linguistics* 25(4). 573–606. • Guest, O. & A. E. Martin. 2021. How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science* 16(4). 789–802. • Hayes, B. 2022. Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics* 8(1). 473–494. • Hayes, B., P. Siptár, K. Zuraw & Z. Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 822–863. • Heinz, J. 2018. The computational nature of phonological generalizations. In *Phonological typology*, 126–195. • Heinz, J. Forthcoming. *Doing computational phonology*. • Magri, G. 2023. Advanced phonology. Unpublished Course Notes. • Marr, D. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. • Mayer, C. 2025. Reconciling categorical and gradient models of phonotactics. *Proceedings of the Society for Computation in Linguistics* 8(1). • Mohanan, K.P. 1986. *The theory of lexical phonology*, vol. 6. • Mohri, M. 1997. Finite-state transducers in language and speech processing. *Computational linguistics* 23(2). 269–311. • Nelson, S. 2024. *The computational structure of phonological and phonetic knowledge*: Stony Brook University dissertation. • Roark, B. & R. Sproat. 2007. *Computational approaches to morphology and syntax*, vol. 4. • van Rooij, I. & G. Baggio. 2021. Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science* . • Strother-Garcia, K. 2019. *Using model theory in phonology: A novel characterization of syllable structure and syllabification*: dissertation. • Tilsen, S. 2023. Probability and randomness in phonology: Deep vs. shallow stochasticity. *Studies in Phonetics, Phonology and Morphology* 29(2). 247–269.