Poverty of the Stimulus Arguments in Phonology

William J. Idsardi, University of Delaware

Abstract: Recent discussions of Chomsky's poverty of the stimulus argument have focused on the empirical side of the argument to the exclusion of its conceptual and philosophical foundations in the general problem of induction (Hume, Goodman, etc.). Moreover, it has been claimed that there are no poverty of the stimulus arguments in phonology. Against these claims this article offers an example of a phonological poverty of the stimulus argument which is made stronger by its relation to Berwick's Subset Principle for language acquisition.

Keywords: phonology, poverty of the stimulus, Subset Principle, induction, acquisition

1. Introduction

Blevins 2004, (1), and Carr 2005, (2), assert that there are no poverty of stimulus arguments in phonology.

(1) "The evidence from language acquisition and the patterns of word-based syllabification presented in this section argue strongly for data-driven learning in phonology. Within the domain of sounds, there is no poverty of the stimulus." Blevins 2004: 235

(2) "Given this, I have tried to sustain an argument to the effect that, if UG exists, then it provides a set of abstract semantico-syntactic objects and relations which are distinct in kind from the sorts of information internalisable from the signal. Phonological objects and relations are internalisable: there is no poverty of the stimulus argument in phonology. No phonological knowledge is given by UG.

> The relations contracted by phonological objects are distinct in kind from the relations given by UG, and alleged syntax/phonology parallelisms, said to be given by UG, are questionable." Carr 2005: 21 (in print, page number subject to change in published version)

As a historiographical comment, it may indeed be true that there are a paucity of cases of poverty of stimulus arguments from phonology (though see Halle 1961, 1978). However, Blevins's and Carr's claims go much further – that phonological poverty of stimulus arguments are either empirically or conceptually impossible. Blevins's work represents a trend in poverty of stimulus arguments that focuses narrowly on the empirical portion of the argument. A particularly relevant example of this restricted debate is given by the exchange of views in the special issue in volume 19 of this journal where Pullum and Scholz explicitly limit their analysis to *empirical* issues: "we were concerned only with the empirical premise in one kind of stimulus-poverty argument" (Scholz and Pullum 2002: 187). However, it is clear from Chomsky's writings, especially Chomsky and Fodor 1980, that the poverty of stimulus argument, at its core, is a conceptual and philosophical argument about the problem of induction. Chomsky and Fodor explicitly and repeatedly cite the philosophical literature on induction and curve-fitting: Hume, Hempel, Goodman and others. As Chomsky puts it, "Once you understand the paradox, it is obvious that *you have to have a set of prejudices in advance for induction to take place;"* (Chomsky and Fodor 1980: 259, emphasis in the original text). Sober 1994 has similar, if slightly less trenchant, comments:

(3) "For me, the fundamental lesson of the grue problem is that empirical

> assumptions that go beyond the content of past observations are needed to establish an epistemic asymmetry between GREEN and GRUE. Whereas philosophers often formulate this point by appealing to the need for "auxiliary assumptions," scientists of a statistical bent often stress the importance of specifying a "model" of the relation of data to the various hypotheses under test. Without assumptions of this sort, the data cannot be interpreted. The slogan for scientists is: *NO MODEL, NO INFERENCE.* This entirely familiar point from the practice of science should not be forgotten when we investigate the theory of that practice." Sober 1994: 237 (emphasis in the original text)

In contrast, none of the contributions in volume 19 (not even Thomas 2002) discuss the modern problem of induction, nor does Vallauri 2004. Chomsky, however, has seen the problem as central from the beginning of generative grammar; Goodman 1943 is one of only five citations in Chomsky's MA thesis (published as Chomsky 1979). And poverty of the stimulus arguments still do merit philosophical discussion (Laurence and Margolis 2001).

2. A Phonological Example

Let us consider now one of the most common phonological processes found in languages of the world, word-final obstruent devoicing. This rule neutralizes the difference between voiced (i.e. slack vocal folds) obstruents such as /b/ and unvoiced obstruents such as /p/ by changing the voiced ones to their unvoiced cognates, in this case [p]. Underlying /p/ is also pronounced as [p]. A rule-based analysis of this process using SPE (Chomsky and Halle 1968) notation and the feature [voice] would be the rule in (4).

(4) [-sonorant +voice] → [-voice] / _ #

Now consider a hypothetical rule of last obstruent devoicing, which devoices the last obstruent in the word, regardless of whether it occurs finally, given in (5).

(5) [-sonorant +voice] → [-voice] / _ $[+sonorant]_0$ #

Rule (5) will not only devoice all word-final obstruents, it would also devoice word-internal obstruents as long as no obstruent follows in the word. For example, rule (5) will change /navar/ to [nafar]. While rule (4) is attested in a wide variety of languages around the world, (5) is completely unattested. The theory in SPE gives a principled explanation for this observation: the evaluation metric for phonology favors (4) over (5). That is, rule (4) is simpler than rule (5). But, as we will see in the next section, the simplicity metric in SPE is not the only logically possible or practicable one; thus the cross-linguistic preference for (4) over (5) is evidence for the innateness of the particular simplicity criterion, a phonological poverty of the stimulus argument.

The above reasoning is not limited to rule-based phonological analyses, as analogous arguments can be constructed for Optimality Theory (OT) grammars (Prince and Smolensky 2004). For example, Grijzenhout and Krämer 2000:66 formulate a word-final devoicing constraint and constraint ranking essentially as in (6).

(6) *[-sonorant +voice]# >> Ident[voice]

The relation of (6) to (4) is clear – (6) contains the structural description of (4), and obviously the constraint ranking in (7) would correspond to (5).

(7) *[-sonorant +voice]$[+sonorant]_0$# >> Ident[voice]

The constraints in (6) and (7) meet the computability requirements proposed in Frank and

Satta 1998 (they fall within the class of constraints which can implemented by finite-state automata, see also below). Thus, there is no general computational reason to exclude (6) or (7) from OT grammars. To exclude (7) by fiat from the starting state for the language learner (i.e. Universal Grammar, UG), then, is to offer a poverty of the stimulus argument. Otherwise, the learning theory must somehow value (6) in preference to (7) if we are to arrive at the correct grammar for word-final devoicing. However, current learning algorithms for OT (Prince and Tesar 1999) if they allow both (6) and (7) as part of UG would rank them both in the top initial stratum because they are both markedness constraints, obfuscating the contribution of each constraint to the evaluation of the candidates; see Halle and Idsardi 2000: 207-211 for extended discussion of this point. That is, (7) would be active in the initial state. We will observe a similar result for learning via the Subset Principle, below.

3. Simplicity

Chomsky and Halle 1968 define the evaluation metric for the enumeration of phonological grammar hypotheses in terms of a set of abbreviation conventions and a symbol-counting procedure (SPE: 334). They specifically limit the application of this procedure to the language acquisition problem, as quoted in (8).

(8) "It should be observed in this connection that although definition (9) has commonly been referred to as the "simplicity" or "economy criterion," it has never been proposed or intended that the condition defines "simplicity" or "economy" in the very general (and still very poorly understood) sense in which these terms usually appear in writing in the philosophy of science." SPE: 334-5

In contrast, Sober 1975 finds the SPE evaluation metric of general interest in the use of simplicity and parsimony criteria in science generally, as shown in (9).

(9) "Their work has a unique interest for our theory in that it is probably the most detailed attempt by scientists to articulate a simplicity criterion in a form explicit enough to yield unequivocal applications to particular cases." Sober 1975: 87

Sober 1975 recognizes Chomsky and Halle's explicit restrictions on the use of the evaluation metric, (10a) but argues for a non-domain specific understanding of simplicity, (10b).

(10) a. "[their] claim that their notion of simplicity is purely internal to the theory of transformational grammar and need have no significant similarities with the notion of simplicity discussed in the philosophy of science." Sober 1975: 88

b. "In the final analysis, we want an adequate simplicity criterion for phonological theory to be a special case of the simplicity criterion used in cognition in general." Sober 1975: 104

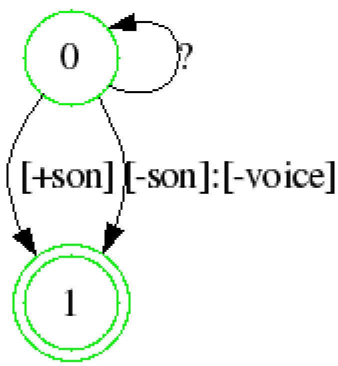
Some 15 years later, Sober 1990 changes his assessment after having examined many parsimony arguments in evolutionary biology – simplicity or parsimony criteria must be domain-specific, as he says in (11).

(11) "These case studies also suggest that there is no such thing as an *a priori* and subject matter invariant principle of parsimony. ... What is perhaps more novel in my proposal is the idea that parsimony be understood locally, not globally." Sober 1990: 77-78

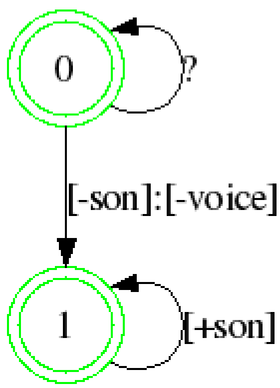
In the end, Sober agrees with Chomsky and Halle: simplicity is domain-specific and in the case of language, innate.

For the rules (4) and (5) we can give a different, non-trivial alternative to the SPE simplicity metric displaying different properties. At least since Johnson 1972 it has been known that phonological rules can be adequately modeled by finite state automata. The automaton for (4) is shown in (12) and that for (5) in (13); see Beasley and Karttunen 2003 for a comprehensive review of finite state automata and their use in phonology.

(12) Final Devoicing FSA



(13) Last Devoicing FSA

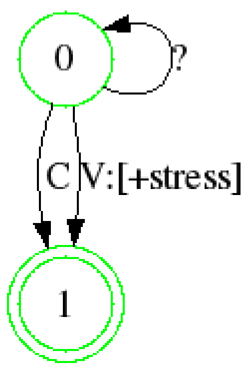


Under the FSA interpretation of the rules it becomes much harder to see what simplicity criteria would favor (12) over (13). Both automata have the same number of states and arcs and the arcs have the same labels. The topology of the automata is different, with the

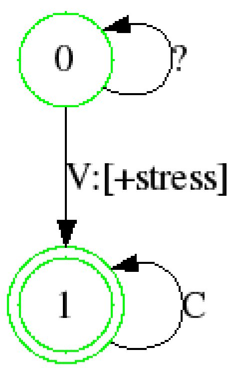
[+sonorant] arc connecting states 1 and 2 in (12), but connecting state 1 to itself in (13). It is difficult to see what simplicity metric over FSA would declare (12) preferable to (13). Minimization techniques for FSA (Skiena 1998:418-419) focus on finding machines with the smallest number of states, not a criterion which will work in this case.

Moreover, criteria for simplicity between the concepts "last" and "final" seem to differ even within phonology. Stress rules stressing word-final vowels, the FSA in (14), are much less common than those stressing the "last" vowel, the FSA in (15).

(14) Word Final Stress FSA



(15) Last Vowel Stress FSA



This is true even for languages which have "default to opposite side" systems. That is, there are languages which stress the last long vowel, otherwise the first vowel (e.g.

Huasteco, Hayes 1995: 296) but none that stress a word-final vowel, else the first vowel. The obvious response is that stress rules are different from devoicing rules; the content of this observation affirms the main point, however, as that difference must be present in the initial state of the language learner. If the language learner can only rely on the formal properties of automata such as (12) – (15) we will not be able to distinguish the stress rule preferences ("last" preferred over "final") from the devoicing rule preferences ("final" preferred over "last").

The domain-specificity and "theory-laden" nature of simplicity criteria can also be seen in other abstract logico-mathematical systems. Phrase structure grammars can be in Chomsky normal form, Greibach normal form or other normal forms (Hopcroft and Ullman 1979: 87-99). Logical formulae can be in conjunctive normal form, disjunctive normal form, or can even be expressed solely in terms of *nand* or *nor*. Equivalent expressions will differ in their complexity as measured by symbol counting based on the axiomatization used, as shown for the logical formula "A and B" in (16).

(16) Equivalent formulae for "A and B"

a. ~, ∧ as primitives: $A \wedge B$

b. ~, ∨ as primitives: $\sim(\sim A \vee \sim B)$

c. $\barwedge$ as primitive: $(A \barwedge B) \barwedge (A \barwedge B)$

d. $\veebar$ as primitive: $(A \veebar A) \veebar (B \veebar B)$

The relative complexity of logical formulae (and other such systems) is thus directly dependent on the axiomatization chosen for the system. So (4) is simpler than (5) given the SPE set of primitives, but if we change to another basis for phonology, then the parsimony of (4) relative to (5) is not necessarily true in the new axiom system. The learner must have an axiomatic system with which to construct grammars; at the very least this axiom system must be innate. The poverty of the stimulus relates in this case to the evidence available to the child for the choice of the axiom system.

4. The Subset Principle

The languages allowed by the rules in (4) and (5) display an interesting property relevant to language learning: the language generated by (5) is a proper subset of the language generated by (4). Forms such as [ba] obey rule (4) but do not obey rule (5). The underlying representation /ba/ would undergo rule (5) and change to [pa]. Berwick 1985 has proposed that in such cases, the learner proceeds by choosing the subset language, as there will (eventually) be positive evidence to move to the superset language. This principle, (17), is the Subset Principle.

(17) "In the special case where one target language is properly contained within another, the point of this condition is to ensure that the acquisition procedure always guesses a subset language if possible, that is, the smallest language that is also compatible with the positive evidence so far encountered." Berwick 1985: 236

Thus in this case, the Subset Principle is at odds with the evaluation procedure of SPE. The SPE metric prefers (4) to (5), predicting that the learner will try (4) before trying (5); the Subset Principle suggests the opposite.

There is another interpretation possible in this case: that both SPE and the Subset Principle are correct. In this case, the enumeration procedure gives a series of available grammars ranked according to complexity. Learners will be able to select a more complex grammar only if positive evidence exists. That is, learning of a more complex rule is possible only if the languages are in a subset relation. If this were to be the case, the rule in (5) would be unattainable, as it is more complex than (4), and the language generated by (5) is a subset of that generated by (4). Therefore, the learner will try (4) first (by the enumeration procedure) and there can be no positive evidence for the learner to abandon (4) in favor of (5).

5. A Thought Experiment

Recent work has been done teaching "novel" phonological contrasts to adults (Maye and Gerken 2000) and infants (Maye, Werker and Gerken 2002). Similar experiments could be designed to test (4) and (5). If subjects are presented with stimuli from the subset language consistent with both (4) and (5), e.g. (18a) will they accept forms such as (18b) – which are consistent only with (4)?

(18) a. [nat], [pa], [fas], [bat]

b. [ba]

The general intuition of phonologists would be that the existence of [pa] and [bat] should imply the grammaticality of [ba] (in fact, many might say [bat] itself was sufficient). But the reification of this intuition is simply that (5) is somehow inaccessible inaccessible to the learner, that is, a poverty of the stimulus argument.

6. Conclusion

The paucity of extant phonological arguments for the poverty of the stimulus is a historical accident. Such arguments can be constructed, as were done above for the phonological environments "last" and "final", and are perhaps more telling in some ways than syntactic arguments, as there is no possibility of semantic boot-strapping for phonological rules. The correct abstract axiomatization for phonological rules (an empirical question for science, see SPE) does real work in guiding the learning algorithm, and thus represents "knowledge unlearned and untaught".

References

Berwick, Robert C. 1985. *The Acquisition of Syntactic Knowledge.* MIT Press.

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns.* Cambridge University Press.

Carr, Philip. 2005. Universal Grammar and Syntax/Phonology Parallelisms. *Lingua.* In press.

Chomsky, Noam. 1979. *Mophophonemics of Modern Hebrew.* Garland.

Chomsky, Noam and Jerry Fodor. 1980. The Inductivist Fallacy. In Piattelli-Palmarini, Massimo (ed.) *Language and Learning: The Debate between Jean Piaget and Noam Chomsky.* Harvard University Press. 255-275.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English.* Harper and Row.

Frank, Robert and G. Satta. 1998. Optimality Theory and the Generative Complexity of Constraint Violability. *Computational Linguistics* 24: 307-315.

Goodman, Nelson. 1943. On the Simplicity of Ideas. *Journal of Symbolic Logic* 8: 107-121.

Goodman, Nelson. 1979. *Fact, Fiction and Forecast.* Harvard University Press.

Grijzenhout, Janet and Martin Krämer. 2000.Final Devoicing and Voicing Assimilation in Dutch Derivation and Cliticization, in Stiebels, Barbara and Dieter Wunderlich (eds), *Lexicon in Focus*, Berlin: Akademie Verlag, 55-82.

Halle, Morris. 1961. On the Role of Simplicity in Linguistic Descriptions. In Jakobson, Roman (ed.) *Structure of Language and its Mathematical Aspects.* American Mathematical Society.

Halle, Morris. 1978. *Knowledge Unlearned and Untaught: What Speakers Know about the Sounds of their Language.* In Halle, Morris, Joan Bresnan and George Miller (eds) Linguistic Theory and Psychological Reality. MIT Press. 294-303.

Halle, Morris and William J. Idsardi. 2000. Stress and Length in Hixkaryana. *The Linguistic Review* 19: 199-218.

Hopcroft, John E. and Jeffrey D. Ullman. 1979. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley.

Johnson, C. Douglas. 1972. *Formal Aspects of Phonological Description.* Mouton.

Laurence, Stephen and Eric Margolis. 2001. The Poverty of the Stimulus Argument. *British Journal for the Philosophy of Science* 52: 217-276.

Legate, Julie Anne and Charles D. Yang. 2002. Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review* 19: 151-162.

Maye, Jessica, and LouAnn Gerken. 2000. Learning phonemes without minimal pairs. *Proceedings of the 24th Annual Boston University Conference on Language Development.* 522-533.

Maye, Jessica, Janet F. Werker and LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82: B101-B111.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell.

Prince, Alan and Bruce Tesar. 1999. Learning Phonotactic Distributions. Rutgers Optimality Archive. http://roa.rutgers.edu/view.php3?roa=353

Pullum, Geoffrey K. and Barbara C. Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19: 9-50.

Scholz, Barbara C. and Geoffrey K. Pullum. 2002. Searching for arguments to support linguistic nativism. *The Linguistic Review* 19: 185-223.

Skiena, Steven S. 1998. *The Algorithm Design Manual.* Springer.

Sober, Elliott. 1975. *Simplicity*. Oxford University Press.

Sober, Elliott. 1990. Let's Razor Ockham's Razor. In Knowles, Dudley (ed.) *Explanation and its Limits.* Cambridge University Press. 73-93.

Sober, Elliott. 1994. A Bayesian Primer on the Grue Problem. In Stalker, Douglas. 1994. *Grue! The New Riddle of Induction.* Open Court. 225-240.

Thomas, Margaret. 2002. Development of the concept of “the poverty of the stimulus”. *The Linguistic Review* 19: 51-71.

Vallauri, Edoardo Lombardi. 2004. The relation between mind and language: The Innateness Hypothesis and the Poverty of the Stimulus. *The Linguistic Review* 21: 345-387.