Język. Komunikacja. Informacja
Ilona Koutny, Ida Stria (red./ed.)
13/2018: 188–211

Probal Dasgupta

Whole Word Morphology Reloaded: The Case for a Semiotic Turn

Abstrakt (Morfologia całego wyrazu – reaktywacja: argument za zwrotem semiotycznym). GP-WWM to program badawczy wykorzystujący morfologię całego wyrazu (WWM, Whole Word Morphology) w morfologii, fonotaktykę generatywną (GP, Generative Phonotactics) w fonologii oraz wyznaczanie domen, które zrównuje moduł fonologiczny z automatycznymi procesami. W niniejszym artykule przekonujemy, że należy pozwolić mechanizmom opartym na semiotyce zmienić sposób, w jaki WWM wdraża strategie tworzenia słów (Word Formation Strategies). Proponujemy pakiety LSSG (Language-Specific Semiotic Guidelines), przedstawiając naszą główną propozycję, tj. czysto semiotyczny projekt koherencji kohortowej (Cohort Coherence Design) dla takiego pakietu wobec szkicowo nakreślonego projektu równań diglosyjnych (Diglossic Equations Design).

Abstract. GP-WWM is a research programme that uses WWM (Whole Word Morphology) in morphology, Generative Phonotactics in phonology, and a domain delineation that equates the phonology module with automatic processes. In this paper, we advocate letting semiotically based mechanisms reshape the way WWM deploys its Word Formation Strategies. We propose LSSG (Language-Specific Semiotic Guidelines) packages, pitting our main proposal, the purely semiotics-driven Cohort Coherence Design for such a package, against a sketchily delineated Diglossic Equations Design.

0. Synopsis

The framework for linguistic description developed by Rajendra Singh and Alan Ford (Ford & Singh 1991; Ford, Singh & Martohardjono 1997; Singh & Starosta 2003), here called GP-WWM, delimits phonology from morphology in the following terms. GP (Generative Phonotactics) – which proposes that the phonology module handles only automatic processes, relegating all other alternations to the morphology – postulates universal principles, and a universal set of repair strategies from which

particular languages choose suitable resources. But the Ford-Singh approach to morphology, WWM (Whole Word Morphology), rests on universal criteria that handle one language at a time, scanning all paradigmatic associations throughout its lexicon. The linguist describing some language L_i uses these criteria to diagnose the connections guiding a speaker-listener of the language as she produces or comprehends (what for her are) 'new' words based on 'old' word-sets. The WWM model's morphological description of L_i formalizes the set of active connections operative in that language as Word Formation Strategies (WFS's). WWM's **language-particular strategies** reflect universal diagnostics – unlike GP, whose operational devices themselves are drawn from a **universal set of repair strategies**.

The present study, which consolidates the substantivist thread of inquiry (Dasgupta 2016; Dasgupta, Ford & Singh 2000), argues in favour of extending the analytical apparatus of the GP-WWM approach by adding Language-Specific Semiotic Guideline packages LSSG(L_1), LSSG(L_2), … at a point located half-way between universal phonological principles and particular lexical data. Every LSSG(L_i) package consists of semiotically motivated guidelines modulating the procedures whereby the universal diagnostics constitutive of WWM address the particular data patterns of any concrete set of L_i words. This paper takes up the morphology of Bangla (a.k.a. Bengali, spoken in Bangladesh and Eastern India) and considers two designs for LSSG(Bangla): a more universalistically framed Cohort Coherence Design (CCD) and a locally anchored Diglossic Equations Design (DED). This study argues in favour of CCD, which is focused on semiotics alone, over against DED, which leans heavily on the sociolinguistics of diglossia.

This is not a mere descriptive issue; CCD is likely to prove superior not only for Bangla, but in general.[1] Once we separate sociolinguistic variability from the architecture of morphology, the universal credentials of semiotics allow us to confidently demarcate the theoretical apparatus (which guides the operations of morphological analysis) from language-particular lexicology (where one must address socio-cultural variability). If CCD prevails, then the LSSG method strengthens morphological theory in a direction that fruitfully extends the GP-WWM approach. The choice of such an intermediate site of intervention – LSSG, half-way between the universal and the particular – is a strategic measure. If true universality proves attainable, this tentative single-language focus will stand superseded.

Section 1 provides empirical motivation for semiotic guidelines in morphology. Section 2 outlines the co-articulation we advocate between semiotically grounded machinery and GP-WWM's formal devices. Sections 3 and 4 elaborate this co-articulation further in the context of formulating specific proposals for Bangla. Section 5 sharpens the methodological issues by pitting this exclusively semiotics-focused description of the Bangla facts, CCD, against a sociolinguistically grounded alternative analysis, DED. Section 6 discusses cognitive load considerations. Concluding remarks appear in section 7.

[1] LSSG type mechanisms are independently required in the syntax (Dasgupta 2016) to describe certain differences between Bangla and English with regard to recoverability phenomena.

1. Broaching the project

To maximize accessibility, we begin by presenting familiar data that show how the semiotic considerations invoked in our LSSG proposal are relevant. Consider the following contrast in English[2]. The philosopher Boruah, at a conference, coined the adjective *Pradhanian* /prə'da:niən/ to characterize positions developed by his interlocutor Pradhan /prə'da:n/. In that register, the philosophers Biswas or Sarkar would have elicited *Biswásian, Sarkárian*. Notice that **Pradhán**an**, **Biswás**an**, **Sarkár**an** are unacceptable. However, the noun *monsóon* – which is etymologically as non-Anglophone as *Pradhan, Biswas, Sarkar* – yields the adjective *monsóon**al***. Here it is **monsoon**ial*** that is ill-formed. The *Pradhan*(i)an* vs *monsoon(*i)al* contrast calls for comment.

Can we attribute this *X*(i)an* vs *X(*i)al* contrast to the preponderance, in the *X*(i)an* class, of country names like *Arabia, Bolivia, India, Slovenia* and their derivatives *Arab**ian**, Boliv**ian**, Ind**ian**, Sloven**ian***? This claim has no firm basis: the pattern of *gerund/ gerund**ial**, baron/ baron**ial**, manager/ manager**ial**, professor/ professor**ial*** might have generalized instead. If it had, **monsoon**ial*** would have been well-formed – despite the absence of nominal counterparts like **gerundia, *manageria* etc. We know that in point of fact the *Xial* schema has not generalized. When we look at canonical English adjectives with penultimate stress (thus setting aside *régional, séasonal*), we consistently find *asteroid(*i)al, rhomboid(*i)al, decad(*i)al, octan(*i)al, dialect(*i)al, acrolect(*i)al, cyclon(*i)al, hormon(*i)al* (and also cases like *neuron(*i)al, axon(*i)al*, which a naïve approach might add to this list, but which a serious WWM account would place under a different rubric; see note 3). On the *X(i)an* side of the picture, also, well-formed *Hamilton**ian**, Smithson**ian*** are in sharp contrast with ill-formed **Anderson**an**, **Johnson**an*** (despite a few entrenched forms like *Elizabethan, Aztecan, Mohammedan*, which do not generalize). Our account needs to handle both halves of this robust pattern[3].

We are inclined to attribute the contrast between the *X(*i)al* facts and the *X*(i)an* facts to one crucial difference. Namely, *Xial* competes with the *Xual* schema instantiated by *sensual, ineffectual, perceptual, conceptual, aspectual, intellectual*. It is neither *Xial* nor the equally specific schema *Xual* that generalizes, but the shorter, minimally specified *Xal*. We conjecture that *Xian* is able to spread in the absence of any *Xuan* competitor (or any other similar competitor) that might have stopped it.

The *X*(i)an/ X(*i)al* contrast is no isolated phenomenon. To see what principle it instantiates, let us schematize it, introducing toy terms. What we shall call the two-member **schema cohort** comprising the minimally specified **mini**-schema *Xan* and the

[2] These examples are presented here as 'toy' material illustrating the issues, with no bearing on serious proposals for English morphology.

[3] This section's naïve exercise ignores the phonic special effects observed in *residential* and *Lilliputian* – only a naive approach would group them under "*X(i)al, X(i)an"*. Readers keyed into Ford and Singh's work can recast these formulations in terms of actual WWM machinery at once (excluding *neuronal* and *axonal*, which unlike the other examples involves a phonic special effect).

maximally specified **mega**-schema *Xian*[4] stands in contrast to the three-member cohort that features the mini-schema *Xal* and two competing intermediate **meso**-schemas *Xial* and *Xual* (these rivals settle for meso status, as it were, keeping each other from the mega crown). The formal highlight is that a **mega**-schema like *Xian* – in a schema cohort that does have a mega member – becomes the productive member. A schema that only pits **meso**-schemas against each other and has no mega member awards diffusion privileges, instead, to the **mini**-schema (if there is one), e.g. *Xal*.

Considered formally, the pattern comes close to instantiating the Elsewhere Condition, 'specific overrides general', a principle that lies at the heart of the semiotic support system that serves grammatical operations. What emerges in this domain, and needs broader contextualization, is a concrete semiotic principle, the **mini-mega generalization**: Where one particular relatively information-rich member of a schema cohort, such as *Xian*, is paradigmatically adjacent to a less richly specified mini-member (to *Xan*, in our concrete example) and counts as its mega-partner, it gets diffusion privileges. Where competition for mega status jeopardizes this, the unrivaled mini-member of the cohort (if there is one, like our *Xal*) receives these privileges instead.

Allowing the mini-mega generalization to rest on such a slender basis would have been absurd. There is, fortunately, additional evidence. The few entrenched instances of the X_{Adj}/Xty_{N} pattern (*loyalty, royalty, certainty, sovereignty, suzerainty, specialty*) do not jeopardize the mega-schema (associated with the $X_{Adj}/Xity_{N}$ pattern), which generalizes. An unfamiliar adjective triggers $Xity_{N}$, not Xty_{N}; thus, **coronalty, *anteriorty* are not frequent beginner's errors in linguistics pedagogy. These facts confirm our generalization. The *Xity* schema counts as the mega-member of its cohort, and spreads: the nouns for *perspicuous, continuous* etc. end in *uity*, not *uty*.

We pause to consider apparent counterexamples. Consider *interpretive*. Why does its mega-rival, *interpretative*, not block it? Conceivably *interpretative* sounds like stuttering and encourages (or, variably, triggers) haplology. Controlling for this factor, one finds no preference for the off-beat alternatives *instrumentive* or *instrumentative* over the frequent *instrumental*. *Qualitative, quantitative, authoritative* have no mini-rivals: **qualitive* and **authoritive* are ill-formed. Even though coiners of technical terms have some discretion, forms like *capabilitative*, *possibilitative* decisively prevail over **capabilitive, *possibilitive.* Thus, we can attribute *interpretive*, an isolated exception to the mini-mega pattern, to contingent factors converging.

A second potential counterexample to mini-mega, *precarity*, though falling short of the high frequency of *interpretive*, outcompetes both **precariety* (putatively modeled on *notoriety, variety, impiety*[5]) and **precariosity* (modeled on *curiosity, obliviosity,*

[4] If examples are not enough, here are some definitions: a 'schema' is an expression appearing on either side of a WFS (Word Formation Strategy). A 'strategy cohort' is a set of WFS's wedded to the same grammatical and semantic function and exhibiting maximal phonic similarity. A 'schema cohort' is the set of schemas featured in a WFS cohort. Our ad hoc intra-cohort comparison tools 'mini, meso, mega' pertain to information richness and can refer either to schemas or to WFS's.

[5] Note that *Xiety* really instantiates the schema /X'σiti/; it stresses the syllable immediately preceding /iti/.

abstemiosity, religiosity, melodiosity, contagiosity). But there are good reasons for not regarding *precarity* as a real counterexample.

First of all, a mini-mega issue would have formally arisen only if *Xiety* were really available as an affixation model, but it is not. Not enough *Xiety* words exist to push that pattern into spreading. Secondly, not a single adjective in the model set cited for **precariosity* shares as much phonic material with *precarious* as *hilarious, tenacious, pertinacious, perspicacious, capacious, rapacious* do; these lead to the *precarity* outcome, making **precariosity* a non-starter. Thirdly, consider *Xary/Xarity* pairs (*solidary/solidarity, sedentary/sedentarity, exemplary/exemplarity, capillary/capillarity*) and *Xar/Xarity* pairs (*singular/singularity, regular/regularity, granular/granularity*): clearly several word formation strategies converge onto *Xarity*. That convergence makes the *Xarity* schema salient, pushing *precarity* ahead of potential rivals.

As an expository toy, the mini-mega generalization – even if other, decisive counterexamples do deflate it eventually – serves to exemplify semiotics-anchored moves of the sort that this paper advocates. Section 2 begins to explore serious semiotic considerations in relation to morphological theory.

2. The WWM Model and Semiotic Principles

We begin by rehearsing the core assumptions of Whole Word Morphology (WWM) and then show where semiotic considerations become relevant.

WWM rests on the following universal set of criteria for diagnosing morphological relationships and of conventions for their formalization. A Word Formation Strategy or WFS is supported for some language L_i when at least two word-pairs in L_i instantiate identical sound and meaning differences, and is formalized by means of strategy template (1), subject to specifications (1a-f):[6]

(1) $/X/_{\alpha} \longleftrightarrow /X'/_{\beta}$, where

a. $/X/_{\alpha}$ and $/X'/_{\beta}$ schematize words; X and X′ are schemas representing classes of words belonging to categories represented by the feature (bundle)s α and β;
b. the prime (′) represents the form-related differences between /X/ and /X′/ that fall outside of automatic phonology as specified by Generative Phonotactics (GP);
c. the symbol $\longleftrightarrow$ represents a bidirectional implication;
d. the interpretation of $/X/_{\alpha}$ is a semantic function of that of $/X'/_{\beta}$, and vice versa;
e. the content of the prime (′) can be null iff $\alpha \neq \beta$.

[6] This formulation is based on Singh (2006: 578) and a pedagogic outline written with input from him, Bhattacharja (ms: 9-10).

Writings such as Ford & Singh 1991, Ford, Singh & Martohardjono 1997, Singh & Agnihotri 1997, Singh & Starosta 2003 expand this set of formulaic tenets, and the associated claim that (1a-f) is a necessary and sufficient basis for morphology, into concrete proposals amounting to a 'radically amorphous' model of morphology, i.e. one in which the word is the lower bound of grammatical segmentation. Like other frameworks proposed along similar lines,[7] WWM drew its initial inspiration from a seminal paper in which Jackendoff (1975) argued that even a system postulating morphemes requires a 'measure of independent information'. The presence of such an evaluation procedure, Jackendoff showed, rendered the 'morpheme' concept redundant and unsustainable.

Understandably, those WWM writings focused on refuting putative arguments for segmenting a word into syntagmatic constituents – on demonstrating that the restrictive universal postulates (1a-f) suffice. However, both Singh and Ford (p.c. in 1997) recognized that WWM's **resources** (though not its **core postulates**) would require augmentation to deal with such unresolved problems as Walsh & Parker's (1983) finding[8] that affixal *s* in English is phonetically longer than the *s* of *lapse* or *corpse*. The question was not whether further resources would be needed, but exactly which ones would be chosen.

From 1997 onwards, Ford and Singh (p.c.) hoped that adequate answers could be built around Wurzel's (1989: 8-17) articulation (based on Mayerthaler 1981) of 'transparency' and 'iconicity' within his approach to 'naturalness'. They acknowledged that the WWM-GP model, while preserving its formal foundation, would need to use these resources to guide its operations. They co-authored a book (Dasgupta et al. 2000) featuring proposals that invoke 'semiotics' and appeal to Wurzel-type principles[9]. After Ford's passing, Singh (p.c.) endorsed papers (Dasgupta 2009, 2010a, b, c) that gave semiotically anchored apparatus specific tasks to perform in tandem with WWM. A recent demonstration that semiotic tools can address apparent violations of syntactic deletion recoverability (Dasgupta 2016) encourages us to co-articulate semiotics with core devices of morphological and phonological description; hence the present study.

We shall now begin the work of co-articulating semiotic principles with WWM-GP devices. On this basis, sections 3 and 4 offer concrete proposals for Bangla. We then

[7] Readers may wonder if Bochner's (1993) 'Lexical Relatedness Morphology' is a similar framework. While Ford and Singh (p.c.) endorsed Bochner as a broadly like-minded thinker, their approach co-articulated morphology with Generative Phonotactics (GP), whereas Bochner's work lacked any co-articulation. 'WWM-GP' uniquely characterizes the Ford and Singh approach.

[8] This and related results are surveyed in the context of a fresh set of studies by Plag et al. (2017), who argue that few current theories of phonology or morphology have the wherewithal to meet this challenge.

[9] Some readers may recall that Ford and Singh consistently rejected Dressler's semiotically argued case for a 'morphonology' intermodule (Dressler 1994 and related writings). That rejection was indeed common knowledge. Their endorsement of the Mayerthaler-Wurzel approach never reached that stage – they kept working on a fuller articulation that they hoped would eventually niche that approach within WWM.

sharpen the methodological issues by pitting that description against a sociolinguistically grounded alternative.

We take from Wurzel the proposition that *Other things being equal, transparent or iconicity-maximizing word formation strategies prevail* (wording ours, content Wurzel's). This *Mayerthaler-Wurzel Proposition*, the **transparency**-focused special case of what we regard as a 'shared opacity' principle, maximizes word shape recognizability across inflectional and derivational variation. It makes sense of the rarity of suppletion. In contrast, the symmetrically **opacity**-focused special case (our *anti-Mayerthaler-Wurzel Proposition: Other things are not equal where saliently shared opacity is entrenched and counts as significant*) renders a paradigmatic subsystem easily recognizable: e.g., by umlaut-marking words like *Äpfel* 'apples', German increases their recognizability qua plurals. Both these propositions embody the idea that specific, richly informative cues maximize identifiability.

These anti/transparency propositions leave the form of grammatical operations like Word Formation Strategies intact, but do guide their functioning: that one WFS is more general than another attracts some version of the (semiotic) Elsewhere Condition, as is found in Dasgupta's (2010a: 28) account of the relative opacity of *cómparable* vs the relative transparency of *compárable*. This is not the whole story, though. The anti/transparency propositions point to a larger set of patterns of (dis)similarity with tangible lexico-grammatical consequences. To clarify this, we shall also highlight semiotic phenomena interacting with the lexicon without affecting the grammar. We start, however, at the grammar-semiotics interface.

How do the anti/transparency propositions interact with the functioning of WFS's? Section 1's 'mini-mega generalization' manifestly instantiates anti/transparency. When a WFS cohort has a mega-member, choosing to apply it maximizes paradigmatic subsystem recognizability for the cohort; hence this choice. When a WFS cohort has only meso-members, more richly specified than its mini-member, only this mini-WFS can maximize the recognizability of the cohort, and is therefore preferred. The anti-transparency proposition entails these corollaries. It is surely clear that the semiotic factors motivating the choice of one WFS over its peers in its cohort are akin to the Elsewhere Condition.

Now consider semiotic factors concerned mainly with the lexicon. The way certain word classes united by shape similarity perform a 'sublexical signalling' function is one salient factor of this sort. The following illustration unpacks the idea.

Dasgupta (2001)'s semiotically grounded analysis rests on the observation that the /CeCa/ template instantiated in Bangla causative verbs like *phera* 'send back' and *bheɟa* 'soak' directly contrasts with /CeCo/ exemplified in denominals like *bero* 'leave', *pero* 'cross'.[10] Due to this paradigmatic fact, the /CeCa/ template 'signals' causativity. Consequently, the lexicon resists the entry of any denominals into the /CeC(C)a/ shape class: their presence would jeopardize the causativity-'signalling' function.

[10] When discussing Bangla verbs, the gerund is the usual citation form. To avoid confusing non-local readers, however, this paper cites verbs in a 'bare' form, 2p intimate imperative. Transcriptions are in IPA.

In contrast, the behaviour of /CV_1CV_2/ verbs with a **back** V_1 differs sharply from that of the **front** V_1 examples considered above. /CoCa/ template causatives like *boɟha* 'explain', *phoʈa* 'boil [tr]' have no denominal paradigmatic neighbour instantiating the template */CoCo/ (whose existence would have made the /CoCa/ template signal causativity). This front V_1 vs back V_1 contrast, Dasgupta (2001) argues, makes possible the existence of /CoCa/-template[11] denominal verbs like *goŋa* 'groan', *khõɽa* 'limp', *bola* 'stroke', *gocha* 'arrange', unimpeded by any lexical resistance.

This front-back asymmetry shows how semiotic signalling affects the composition of the lexicon. Section 3 adduces other front-back asymmetries, emphasizing that verbs in Bangla allow *Vŭ* and *Võ* diphthongs but resist *Vĭ, Vĕ*[12], while nouns and adjectives allow all four diphthong templates. To express such generalizations, we introduce constraints on templates and bring them to bear on grammatical mechanisms.

Summarizing so far, all theories of grammar use the Elsewhere Condition, a semiotic principle that brings the transparency-opacity spectrum to bear on grammatical mechanisms at the point of application. Building on the Elsewhere Condition, we advocate an explicit, differentiated approach to that spectrum's interaction with the lexico-grammatical machinery. Principles of anti/transparency interact with grammatical mechanisms directly. The sublexical signalling properties associated with certain word subclasses (i.e. with canonical templates characterizing these subclasses) serve to template-constrain words – diagnosing them as admissible/inadmissible into the lexicon of the language.

Extending the methodological discussion, section 3 illustrates certain points at which sublexical signalling interacts with the grammar as well, not only with the lexicon. We postulate language-specific semiotic machinery for Bangla – and propose that Bangla is not alone. We assume that many such descriptions need to accumulate before single-language postulates generalize into principles for which universality claims may prove defensible (or serious correlations with syndromes characteristic of specifiable language types). Our stand is a response to the steady accumulation of such evidence. Consider Nandini Bhattacharya's (2017a, b) finding that the semiotically salient reduplication phenomenon interacts with the grammar of plural quantification in Bangla, or Dasgupta's (2016) study, which addresses apparent violations of syntactic deletion recoverability in Bangla by co-articulating semiotically grounded machinery with the grammar. We are formally introducing LSSG (Language-Specific Semiotic Guideline) packages in order to flag the general significance of such studies. One expects the LSSG route to lead to principled solutions, once significant cross-linguistic generalizations emerge.

---

[11] Dasgupta (2001) considered only /CVCV/ examples. The strength of the effect becomes clearer if /CVCCV/ examples are also examined. The availability of *tobɾa* 'bend', *totla* 'stammer', *domɾa* 'twist', *mocɾa* 'twist' (instantiating the /CoCCa/ template for denominals) stands in contrast to the absence of /CeCCa/ denominals in standard Bangla (discounting an outlier or two like the nonstandard variant *neŋɾa* for *niŋɾo* 'wring').

[12] Strict IPA would require [i̯ e̯ u̯ o̯]; our breve-laden substitutes are less unwieldy.

3. From Signalling to Sublexical Signatures

'Signalling' prepares us for 'sublexical signatures'. Given the systematic asymmetry shown in (2), we flag the absence of VG_f diphthongs in the environment __CV (chart (2), cell (ii-A)) as one of the traits in the **sublexical signature** characterizing Bangla verbs[13].

(2) Narrow asymmetry within verbs:

	(A) [Verbs	(B) Non-Verbs
(i) Vowel + G_b (back glide) preceding CV	**aŏ**ɽa 'recite (poetry), parrot (clichés)' t**æŏ**ɽa 'twist'	p**aŏ**na 'dues' ʃ**æŏ**la 'moss' n**æŏ**ɽa 'fond (of)'
(ii) Vowel + G_f (front glide) preceding CV	Not instantiated in [Verbs	g**ɔĕ**na 'ornament' t**oĭ**ri 'ready' m**ɔĕ**da 'flour'

Section 2 mentioned sublexical 'signalling' in connection with the causativity-signalling template /CeCa/. Several traits signalling categorial status (e.g. verb, noun, adjective), clustering together, count as that category's sublexical signature. Thus, Bangla verbs exhibit another trait shown in (3), column (3)A – all and only verbs prohibit $VV_{hm}C\#$ and $VG_{hm}C\#$ (*hm* for 'high or mid'). This trait too forms part of the sublexical signature of Bangla verbs:

(3) Broad asymmetry between verbs[14] and non-verbs:

	(A) Verbs	(B) Non-Verbs
(i) $VV_{hm}C\#$	Not instantiated in verbs	d**eul** 'temple', b**aul** 'mystical sect singer', ɖh**auʃ** 'huge', b**aiʃ** 'twenty-two', d**aer** 'start (a lawsuit)', kh**ɔer** 'catechu'

[13] In (2) we use the ad hoc *[Verb*, with an unpaired left bracket, to pick out only 'bare verbs', second person intimate imperative forms. The later broadening of the discussion of (2) to cover all verbs is indicated by removing '['. We write just *Verb* in (3), which covers all verbs.

[14] (3) concerns all verbs, not just bare *[Verbs*. There is 'free' variation (within and across idiolects) as to which forms go into cell (i)B or (ii)B. The $V\$V_{hm}$ (often $[VGV_{hm}]$) vs VG_{hm} contrast carries a functional load only word-finally, not in pre-consonantal position.

(ii) $VG_{hm}C\#$	Not instantiated in verbs	d**oŭɽ** ‘dash, run (N)’, **aŭʃ** ‘a type of paddy’, **aĭn** ‘law’, l**aĭn** ‘line’, k**ɔĕn** ‘coin’

Trait (3)A is a good place to start scrutinizing WFS/sublexical signature trait interactions. Recall that Singh and Ford’s delineation of Generative Phonotactics (GP) and Whole Word Morphology (WWM) eliminates cross-modular redundancies by requiring a WFS to implement only those modifications that GP’s Repair Strategies would not perform. Recognizing sublexical signature traits puts redundancy elimination on our agenda vis-à-vis category-wide phonic phenomena. Bearing in mind the signature trait highlighted in (3), we note at (4) and (5) – closely related sets of verb-to-verb WFS’s – that the strategies add *iʃ, en, uk* to C-ending bare verbs but just *ʃ, n, k* to V-ending ones:

(4)i. $[XC]_{V\ 2p\ Intim\ Imp} \longleftrightarrow [XCiʃ]_{V\ 2p\ Intim\ Pres}$ (5)i. $[Xa]_{V\ 2p\ Intim\ Imp} \longleftrightarrow [Xaʃ]_{V\ 2p\ Intim\ Pres}$
rakh/rakhiʃ ‘keep’, kac/kaciʃ ‘wash’ kha/khaʃ ‘eat’, ba/baʃ ‘row’

ii. $[XC]_{V\ 2p\ Intim\ Imp} \longleftrightarrow [XCen]_{V\ 2/3p\ Hon\ Pres}$ ii. $[Xa]_{V\ 2p\ Intim\ Imp} \longleftrightarrow [Xan]_{V\ 2/3p\ Hon\ Pres}$
rakh/rakhen ‘keep’, kac/kacen ‘wash’ kha/khan ‘eat’, ba/ban ‘row’

iii. $[XC]_{V\ 2p\ Intim\ Imp} \longleftrightarrow [XCuk]_{V\ 3p\ Neu\ Imp}$ iii. $[Xa]_{V\ 2p\ Intim\ Imp} \longleftrightarrow [Xak]_{V\ 3p\ Neu\ Imp}$
rakh/rakhuk ‘keep’, kac/kacuk ‘wash’ kha/khak ‘eat’, ba/bak ‘row’

Note that *baʃ* and *khak* in (5)i, iii narrowly miss coming out as *baiʃ* and *khauk*. Had the WFS’s been *X* ⟷ *Xiʃ/Xen/Xuk* rather than (4)i-iii, they would have attached *iʃ/en/uk* to vowel-ending bare verbs as well. Juxtaposing (5) with the non-verbs *baiʃ* ‘twenty-two’ and *ɖhauʃ* ‘huge’ shown at (3), we find that GP’s Repair Strategies would leave the hypothetical *baiʃ* and *khauk* intact in such a counterfactual scenario. The GP-WWM framework, which confines GP to automatic processes, must of course regard the verbal subsystem’s choice of *baʃ*/*khak* over *baiʃ*/*khauk* as **morphological** rather than phonological. That point stands. Our question is how to acknowledge that this is a **systematic** morphological phenomenon.

The choice of *baʃ*/*khak* over *baiʃ*/*khauk* reflects a signature trait pervading the verb system. The alternation observed at *ba**ʃ**/kha**k*** (in (4)) vis-à-vis *rakh**iʃ**/kac**uk*** (in (5)) is neither automatic (i.e. **phonological**) nor regular-but-nonautomatic (i.e. an **arbitrary** morphological generalization). It is a **systematic** morphological phenomenon, which counts as **phonology-sponsored**; it reflects a natural process, not an arbitrary ‘crazy rule’, although it falls outside the Bangla-wide automatic phonology filters. Thus, GP-WWM must express this generalization – with some help from the semiotics of signatures.

For this purpose, sublexical signatures must learn how to repair morphology-induced signature violations by applying analogues to GP’s Repair Strategies. We propose **Signature Audit** procedures with this mandate. To show how they work,

we first replace (4) and (5) with the less richly specified WFS's (6)i-iii, to obtain bad output requiring repair:

(6)i. $[X]_{\text{V 2p Intim Imp}} \longleftrightarrow [\text{Xi}\int]_{\text{V 2p Intim Pres}}$
ii. $[X]_{\text{V 2p Intim Imp}} \longleftrightarrow [\text{Xen}]_{\text{V 2/3p Hon Pres}}$
iii. $[X]_{\text{V 2p Intim Imp}} \longleftrightarrow [\text{Xuk}]_{\text{V 3p Neu Imp}}$

The discrepant output of the vowel-ending cases (which (5) was designed to address) is now handled by Signature Audit. For 'eat' and 'row', (6) yields output violating the Sublexical Signature for verbs – *khaiʃ, khaen, khauk; baiʃ, baen, bauk*. Signature Audit repairs the forms, deleting V_{hm} in the environment $V\_\_C\#]_V$ – yielding *khaʃ, khan*, etc., as required.

Caution is called for; GP-WWM's core insights must be retained. How do we stop these proposals from reversing the decision to recognize Velar Softening in English (the *k/s* alternation in *electric/electricity*) as morphological, not phonological? What prevents such backsliding is the fact that a 'crazy', phonologically unsponsored process like *k > s* never pervades a lexical category. It has some semiotic significance (local opacity enhances word identifiability); but signature traits register only configurations, not alternations like *k/s*. The semiotic supplementation proposed here serves only to extend GP-WWM's core insights, not to undermine them.

As we keep unpacking our proposals at the logistic level, we also begin to motivate them. Section 6, which revisits Jackendoff's evaluation metric in the light of the present study, provides further justification.

We have seen that word-level semiotic properties (including signature traits) paradigmatically mark a word of lexical category LC_1 as falling outside LC_2, LC_3 etc. But certain properties of this sort fortify a word syntagmatically, highlighting its integrity. Thus, consider the *i*-ending reduplicated action nouns in (7); compare the penultimate syllables of (7)i-v with (8)i-v, (9)i-v:

(7) **Reduplicated action nouns**	(8) **Bare verbs**	(9) **3p.Neu.Pres verbs**
i. meʃamiʃi 'mingling'	i. meʃ 'mingle'	i. meʃe 'mingles'
ii. khõɟakhũɟi 'searching'	ii. khõɟ 'search'	ii. khõɟe 'searches'
iii. ʈhælaʈheli 'jostling'	iii. ʈhæl 'jostle'	iii. ʈhæle 'jostles'
iv. bɔlaboli 'talking'	iv. bɔl 'talk'	iv. bɔle 'talks'
v. maramari 'fighting'	v. mar 'hit'	v. mare 'hits'

The final *i* in (7)i-iv might be said to trigger 'vowel raising' to its left.[15] In process terms, [+High, +ATR] features 'spread' in (7)i-ii, and [+ATR] 'spreads' in (7)iii-iv (linguists resort to informal process talk, even in models that deal with configurations and not processes). Hence – considering the penult of (7)i-iv – the alternations /miʃ~meʃ, khũɟ~khõɟ, ʈhel~ʈhæl, bol~bɔl/.

[15] (7)v illustrates the fact that /a/ is not targeted; we should strictly call it 'non-low vowel raising'.

This effect of word-internal final *i* on the adjacent syllable contrasts with the absence of such effects in (10), where the particle *i* is cliticized:

(10) Examples with enclitic focus particle /i/
i. deʃ-i 'country-Foc: the country, indeed'
ii. roɟ-i 'every.day-Foc: absolutely every day'
iii. æk-i 'same-Foc: the very same'
iv. ɟɔl-i 'water-Foc: only water'
v. bhat-i 'rice-Foc: only rice'

Finding that the 'vowel raising' syndrome in (7) is word-bound, we infer that delimiting the word is one of its semiotic functions.

How does this syndrome fit into the signature trait story? While vowel raising in nouns and adjectives is very frequent, it stops short of taking over. But verbs uniformly undergo it; the syndrome is a verbal signature trait. Comparison of the raised vowels at (11)- (12) with the non-raised vowels at (8)-(9) makes this obvious (the low /a/ at (11v), (12v), as in (7)v, eludes raising):

(11) **Vowel Raising with /i/, in 2P Intimate Present**
i. miʃiʃ 'you.mingle'
ii. khũɟiʃ 'you.search'
iii. ʈheliʃ 'you.jostle'
iv. boliʃ 'you.talk'
v. mariʃ 'you.hit'

(12) **Vowel Raising with /u/, in 3P Neutral Imperative**
i. miʃuk 'let.them.mingle'
ii. khũɟuk 'let.them.search'
iii. ʈheluk 'let.them.jostle'
iv. boluk 'let.them.talk'
v. maruk 'let.them.hit'

GP ignores vowel raising, which is non-automatic in Bangla. However, languages like Turkish do harbour vowel harmony as an automatic process. The raising syndrome, part of the Bangla verb's sublexical signature, is universal-phonology-sponsored, then; it counts as a cluster of natural processes rather than crazy rules.

These statements work perfectly for (11) and (12), where the [+High, +ATR] triggers *i* and *u* attract obvious phonological support for the 'raised vowel plus trigger' configuration. However, in (14) and (15) – where Signature Audit (which, one will recall, deletes V_{hm} in V__C#]$_V$) removes these triggers – how do we respond to the fact that, even without overt triggers, (13)'s unraised *e, o, ɔ* alternate with (14)-(15)'s raised *i, u, o*? Does universal phonology withdraw support, forcing vowel raising to revert to 'crazy' status in such examples?

(13) **Bare Verbs**
i. de 'give'
ii. ʃo 'lie.down'
iii. hɔ 'be'

(14) **2p.Intim.Pres Forms**
i. diʃ 'you.give'
ii. ʃuʃ 'you.lie.down'
iii. hoʃ 'you.are'

(15) **3p.Neu.Imp Forms**
i. dik 'let.them.give'
ii. ʃuk 'let.them.lie.down'
iii. hok 'let.them.be'

We must now choose either Path One, to continue our quest and claim naturalness even for (14) and (15), or Path Two, a retreat to unrevised GP-WWM, diagnosing all non-automatic alternations as 'crazy' and invoking diachronic and socio-cultural factors to address the Bangla facts. We devote section 4 to Path One, section 5 to Path Two, and section 6 to articulating 'phonology-supported signature traits' into an explicit methodological proposal that helps choose between One and Two.

4. The Semiotics-Focused Account

The foregoing remarks are compatible with letting universal conventions govern all grammatical operations. Sections 4 and 5 make certain language-particular proposals; we develop here some **Language-Specific** Semiotic Guidelines for Bangla, as promised. Section 4 proposes a semiotics-only version of this LSSG(L_i) package. Section 5's counterproposal combines sociolinguistic resources with semiotic principles. Even section 3's innovations –amenable in principle to universalistic construal – are provisionally placed in the LSSG (Bangla) package.

Returning to the recalcitrant data of section 3, we now take up 'Path One'. We maintain that even in (14)-(15), despite Signature Audit deleting the triggers *i* and *u*, universal phonology sponsors the *raised vowel plus trigger* configurations, which therefore count as 'natural' rather than 'crazy'. In defence of this position, we argue that Path One comprehensively handles most of the raising phenomena throughout the Bangla verb system.

We begin by introducing 'quirky' verbs, which feature the semivowel *ĭ* as a vowel raising trigger. Compare (16)'s unraised *ɔ, o* with the raised *o, u* in (17)-(18):

(16) Bare Verbs	**(17) Infinitive Forms**	**(18) 3p Neutral Past Forms**
i. bɔ 'carry'	i. boĭte 'to.carry'	i. boĭlo 'they.carried'
ii. ʃɔ 'tolerate'	ii. ʃoĭte 'to.tolerate'	ii. ʃoĭlo 'they.tolerated'
iii. no 'bend (intr)'	iii. nuĭte 'to.bend'	iii. nuĭlo 'they.bent'
iv. ro 'sow'	iv. ruĭte 'to.sow'	iv. ruĭlo 'they.sowed'

Now consider ordinary, non-quirky verbs with near-identical bare forms. In the infinitive and the third person neutral past, they instantiate vowel raising despite the surface absence of the *ĭ* trigger:

(19) **Bare Verbs**	**(20) Infinitive Forms**	**(21) 3p Neutral Past Forms**
i. hɔ 'become'	i. hote 'to.become'	i. holo 'they.became'
ii. ʃo 'lie.down'	ii. ʃute 'to.lie.down'	ii. ʃulo 'they.lay.down'
iii. dho 'wash'	iii. dhute 'to.wash'	iii. dhulo 'they.washed'

Although the specific Audit process (deleting V_{hm} in the environment V__C#]$_V$) that eliminates the triggers in (14)-(15) is evidently inapplicable here, Signature Audit is active in (20)-(21) as well; we provide details forthwith. For now, note that yet another part of the verb system displays vowel raising despite the surface absence of a high trigger; and that, again, independent evidence pinpoints the trigger involved. If we state the Infinitive and 3p Neutral Past WFS's with maximal generality, as (22)-(23)[16], then the right results for (24)'s C-ending bare forms follow immediately – i.e. (25)-(26) also exhibit the raising syndrome despite the surface absence of the trigger *ĭ*:

(22) $[(C)V(C)]_{\text{V 2p Intim Imp}} \longleftrightarrow [(C)V_{\text{raised}}(C)\check{\imath}te]_{\text{V Inf}}$
(23) $[(C)V(C)]_{\text{V 2p Intim Imp}} \longleftrightarrow [(C)V_{\text{raised}}(C)\check{\imath}lo]_{\text{V 3p Neu Past}}$

(24) **Bare Forms**	(25) **Infinitive Forms**	(26) **3p Neutral Past Forms**
i. dhɔr 'catch'	i. dhorte 'to.catch'	i. dhorlo 'they.caught'
ii. mɔr 'die'	ii. morte 'to.die'	ii. morlo 'they.died'
iii. ʃon 'hear, listen'	iii. ʃunte 'to.hear'	iii. ʃunlo 'they.heard'
iv. tol 'lift'	iv. tulte 'to.lift'	iv. tullo 'they.lifted'

The reason we need not invoke any special Signature Audit process here is that the ordinary provisions of Generative Phonotactics do the job without help. In (25)-(26), *ĭ* occurs sandwiched between consonants and, being unpronounceable there, gets deleted by a Repair Strategy. Before it does, the 'vowel raising' template in such words elicits due recognition as a 'natural' bit of systematic morphology forming part of the Bangla verb's Sublexical Signature. This recognition – relevant to the 'naturalness evaluation' discussed in section 6 below – survives the sandwiched glide's deletion.

However, the deletion of postvocalic *ĭ* in (20)-(21) cannot be a matter of Bangla phonotactics as a whole[17], but must reflect the Sublexical Signature of verbs. The Signature Audit process required is presumably an expanded version of the prohibition of postvocalic high *ĭ*/mid *ĕ* in bare verbs, the front glide constraint highlighted in (2). The prohibition barring VG_fCV is indeed absolute for **bare** verbs. In the larger class of verbs, the only forms eluding it are the 'quirky' infinitive and past forms at (17)-(18) that kickstarted our discussion. Quirky forms allow *ĭ* in the designated position. All verbs prohibit *ĕ* there.

Bangla verbs are basically allergic, then, to VG_fCV – with the exception of 'quirky' verbs. If this allergy translates into a category-wide *VG_fCV constraint within the Sublexical Signature of verbs, a Signature Audit operation deleting *ĭ* in (20)-(21) follows without stipulation. If we tweak nothing else, however, that prohibition ends up making the 'quirky' (17)-(18) collapse into the well-behaved (20)-(21), failing even to attain observational adequacy.

[16] We save space here, writing V_{raised} to abbreviate 'turn non-high non-ATR into non-high ATR, and turn non-high ATR into high ATR'.

[17] The phonotactics must allow VG_{high} at (17)-(18) and thus cannot be responsible for the deletion in (20)-(21).

Mechanical fixes are easy to devise; but dismissing the problem as an insignificant conundrum means abandoning descriptive adequacy. We therefore consider the option of making the Signature Audit of verbs delete only *ĕ* in V__CV (the surface facts allow us to use a surface-absent mid glide to make the raised vowels look 'natural') and bending this operation to our ends. Call this idea Plan B.

Motivated by the availability of a Signature Audit operation of *ĕ* deletion alone, Plan B receives serendipitous support from 'strong' raising, a configuration characteristic of the conjunctive participle and the second person neutral future imperative, in which the pivotal *ĕ* appears on the surface (such raising is 'strong' because it targets even the low vowel *a*). Implementing Plan B would involve replacing every high *ĭ* trigger in our earlier proposals with a mid *ĕ* trigger.

Despite these advantages, we reject Plan B, for it would force quirky forms like *ʃoĭte* 'to.tolerate' and *nuĭte* 'to.bend' to choose one of two bad options: either to insert *ĕ* and then get it deleted (this involves a WFS that says either …*Vĭĕte* or …*Vĕĭte*, both of them anomalous), or to revert to classical GP-WWM and postulate WFS (22) exclusively for quirky verbs. The latter option would render Plan B conceptually incoherent: if we postulate abstract **mid** glides for the sake of phonological sponsorship, and our naïve analysis of outliers featuring **high** glides keeps them disconnected from the abstract postulation, our account becomes descriptively inadequate. Hence Plan B's demise.

What we propose is a Plan C, under which all WFS's follow strict GP-WWM norms. Consonant-ending bare verbs attract WFS's that tweak the vowel and add whatever the surface form requires – postulating no abstract glides. Plan C makes only one semiotic move to elicit the blessings of universal phonology for raising (both 'weak' and 'strong'): we invoke the 'mega', 'mini' and 'cohort' notions from section 1, also introducing the auxiliary notion of 'propria' (which we define by example: the mega *Xial* of section 1 exhibits an extra *i* not shared with its mini cohort-mate *Xal*; this *i* shall count as *Xial's* **propria**). So equipped, we empower a mega schema to abstractly attach its propria (for naturalness evaluation purposes) to the corresponding niche in the mini schema in its cohort. We place this proposal within the *LSSG (Bangla)* package, leaving the grammars of other languages unaffected.

Operationally, our proposal gives Bangla quirky verbs a semiotically interesting role to play. To see the main point, consider quirky (22) – repeated for convenience as (27) – vis-à-vis regular (28)-(29); to reduce ambiguity, we flag the types of raising involved:

(27) $[(C)V(C)]_{\text{V 2p Intim Imp}} \longleftrightarrow [(C)V_{\text{weak-raised}}(C)\breve{\text{i}}te]_{\text{V Inf}}$
(28) $[(C)VC]_{\text{V 2p Intim Imp}} \longleftrightarrow [(C)V_{\text{weak-raised}}Cte]_{\text{V Inf}}$
(29) $[(C)V]_{\text{V 2p Intim Imp}} \longleftrightarrow [(C)V_{\text{strong-raised}}te]_{\text{V Inf}}$

The few forms like *ʃoĭte* 'to.tolerate', *nuĭte* 'to.bend' that fall under (27) count as quirky because the majority of Bangla infinitives conform either to (28), e.g. *dhorte* 'to.catch', *ʃunte* 'to.listen', *ʧhelte* 'to.push', *miʃte* 'to.mingle', *rakhte* 'to.keep', or to (29), e.g. *hote* 'to.become', *ʃute* 'to.lie.down', *khete* 'to.eat'. Our Plan C identifies (27) as

the mega member of the infinitival cohort and notes that its propria relative to its mini cohort-mate (28) is /ĭ/. Plan C proposes that, exercising a privilege that we call **Cohort Peer Pressure**, (27) virtually (abstractly, as an evaluation-motivated exercise) inserts this propria in (28). As a result, (28) looks like (28′) when the naturalness evaluation algorithm inspects it; we are using the strike-through notation to show that the inserted propria is not pronounced, for it is visible only to that special evaluation algorithm, not to the regular grammatical mapping that uses phonological mechanisms to yield phonetic output:

(28′) $[(C)VC]_{\text{V 2p Intim Imp}} \longleftrightarrow [(C)V_{\text{weak-raised}}C\check{\text{i}}te]_{\text{V Inf}}$

In other words, the **morphological output** of WFS (28′) – as the evaluation algorithm sees it – looks like *dhoršte* ‘to.catch’, *ʃunšte* ‘to.listen’, etc. The semiotic buffer that arranges for naturalness checking of harmonically motivated vocoid patterns is thus able to submit the *...o...š...* and *…u…š…* patterns of *dhoršte* and *ʃunšte* respectively (in conjunction, obviously, with their non-raised paradigm-mates) to universal phonology for inspection and licensing as ‘natural’ configurations.

A closer look at (27)-(29) reveals that (29) is the odd man out: it specifies strong raising, while (27)-(28) choose weak raising. Consequently, we propose, (29) does not belong to the same cohort, formally, as (27)-(28), and stays away from this application of Cohort Peer Pressure. This leaves infinitives like *hote* ‘to.become’, *ʃute* ‘to.lie.down’, *khete* ‘to.eat’ unavailable for naturalness licensing. If that is how the chips fall, Plan C formally accepts this outcome without hesitation: those particular infinitives then count as relatively opaque. Beyond formal issues, perhaps more delicate means of assessing naturalness will eventually emerge from such a research program. Semiotics needs to combine exact and inexact considerations in ways that are seldom explored (Tirumalesh 1991 is a rare exception).

Our approach gains serendipitous support from the behaviour of Future forms and Future-Laden Gerund forms. In the Future, quirky verbs show *ĭ* alongside weakly raised vowels, and CVC verbs show only the latter; however, regular CV verbs show neither. (30) saves space by presenting only examples featuring *ɔ* (raised to *o*), but the rest of the paradigm is easy to infer. Now, alongside ordinary Gerunds, shown in (31), Bangla also harbours a second type, which, because it shares a characteristic *b* with the Future, is best called the Future-Laden Gerund. (32) shows that it exhibits selective free variation – for all and only CVC verbs. This free variation requires explanation.

(30) Future Forms

	Quirky Verbs	CVC Verbs	Regular CV Verbs
Bare Form	ʃɔ ‘tolerate’	dhɔr ‘catch’	hɔ ‘become’
3p Neu Fut	ʃoĭbe ‘they.will. tolerate’	dhorbe ‘they.will. catch’	hɔbe ‘they.will. become’

(31) Ordinary Gerund Forms

	Quirky Verbs	CVC Verbs	Regular CV Verbs
Bare Form	ʃɔ ‘tolerate’	dhɔr ‘catch’	hɔ ‘become’
Gerund	ʃɔ̆a ‘tolerating’	dhɔra ‘catching’	hɔ̆a ‘becoming’

(32) Future-Laden Gerund Forms

	Quirky Verbs	CVC Verbs	Regular CV Verbs
Bare Form	ʃɔ ‘tolerate’	dhɔr ‘catch’	hɔ ‘become’
F-L Gerund	ʃoĭba ‘tolerate.FLG’	dhorba ~ dhɔrba ‘catch.FLG’	hɔba ‘become.FLG’

In the semiotics-focused account proposed here – at this point, we drop the ‘Plan C’ label and officially designate it as CCD, the **Cohort Coherence Design** for LSSG(Bangla) – we notice that the raised-vowel form of the Future-Laden Gerund, *dhorba*, shares this raised vowel with, and thus reminds us of, Future forms like *dhorbe* ‘they.will.catch’. By the same token, the intact-vowel form *dhɔrba* reminds us of the regular Gerund *dhɔra* ‘catching’, whose vocalism it echoes. The significance of the Future-Laden Gerund’s manifestation of dual allegiance becomes clearer once we outline the sociolinguistic/ diachronic alternative account in section 5 – a task to which we now turn.

5. The Sociolinguistic Alternative

Even if we had not decided that our proposal should be compared with a diachronically and sociolinguistically grounded alternative, the full disclosure imperative would in any case have obliged us to inform the reader that Bangla as a natural language exists in the unusual state of diglossia – often regarded as a topic for exclusively sociolinguistic study. A diglossic language harbours a special archaic ‘High’ code coexisting with the ordinary or ‘Low’ norm; in Bangla, these codes are called <Sādhu Bāṅlā> / ʃadhu baŋla/ and <Calit Bāṅlā> /colit baŋla/ respectively. The dominance of the High code that keeps a diglossia vibrant is a delicate constellation of alignments. As far as the diglossia of Bangla is concerned, that constellation weakened a hundred years ago; the vitality of its diglossia has steadily diminished. The use of High Bangla – which had been the default choice in discursive writing and ceremonial speeches – is now confined to newspaper editorials. Thus, linguists today tend to describe only Low Bangla and to call it just Bangla. However, a literate user of the language must inherit a legacy that includes classics in High Bangla. Thus, despite the moribund state of

the High code, schooling continues to impart proficiency in High Bangla; educated speakers are aware that the High code of their language preserves many features of older varieties that were spoken five or six centuries ago. One must assume that the specific bits of lexical and morphological knowledge pertaining to the High and Low codes are aligned within their knowledge of Bangla. This alignment has never become an object of psycholinguistic study.

The reason these facts are pertinent is that the most of the differences between the High and Low codes in Bangla have to do with the form of verbs. Consider the infinitive, the future, the past, where Low Bangla has a semivowel *ĭ* that appears in intervocalic position in quirky forms but disappears in interconsonantal position, as we see in the examples repeated in the Low code cells of chart (33); High Bangla, in all these cases, has a corresponding medial vowel *i*, invariant for all verb classes (apart from its postvocalic allophonic form *ĭ*), as shown in the High code cells of (33):

(33) **Comparing High Code and Low Code Forms of Bangla Verbs**

	High Code Quirky Verb *no* 'bend'	High Code CVC Verb *dhɔr* 'catch'	Low Code Quirky Verb *no* 'bend'	Low Code CVC Verb *dhɔr* 'catch'
Infinitive	nuĭte 'to.bend'	dhorite 'to. catch'	nuĭte 'to.bend'	dhorte 'to. catch'
Future 3p Neu	nuĭbe 'they.will. bend'	dhoribe 'they. will.catch'	nuĭbe 'they.will. bend'	dhorbe 'they. will.catch'
Past 3p Neu	nuĭlo 'they.bent'	dhorilo 'they. caught'	nuĭlo 'they.bent'	dhorlo 'they. caught'

Likewise, the Low code forms cited at (5)i-iii – where the plain *f, k* of *khaf* 'you. eat', *khak* 'let.them.eat' alternate with the high vowel laden *if, uk* of *rakhif* 'you.keep', *rakhuk* 'let.them.keep' – correspond to High code forms *kha**if*** 'you.eat' and *kha**uk*** 'let. them.eat', which show the vowels in their full glory. There are many such instances of morphological invariance in the High code. One consequence is that vowel harmony based constraints apply transparently throughout the conjugation in High Bangla, and can be reckoned as part of the Sublexical Signature of the verb category.

Given these salient facts, questions arise about the pressure, if any, that those relatively transparent patterns from High Bangla morphology exert on their obfuscated counterparts in the Low code as far as proficient bicodal speakers are concerned. In the absence of case studies, one is of course bound to speculate, ostensibly to establish valid terms of reference for case studies that can be undertaken. In this context, one may imagine a critic (willing to use semiotics in principle but skeptical about our account in particular) constructing a sociolinguistically grounded alternative to the purely semiotic proposals made in sections 3 and 4 labelled as the Cohort Coherence Design (CCD) for LSSG(Bangla).

By choosing 'Diglossic Equations Design (DED) for LSSG(Bangla)' as our label for the alternative account that our critic may construct, we allude to the possibility that the alignment between a bicodal Bangla speaker's two codes may in part take the form of lexical cluster-to-cluster 'equations'. Constructing an actual DED conceptually independent of (and thus amenable to meaningful comparison with) the CCD proposed in sections 3 and 4 would be relatively straightforward if we or our imaginary critic had at our disposal an account with a robust characterization of the cognitive alignment driving an educated speaker's knowledge of her two codes. Such an account could, on the basis of the facts itemized above, articulately claim that speakers of the language schooled in both codes at an early age acquire such deep-rooted bicodal settings in their knowledge of Bangla that their Low code alternations (as in *kha**ʃ**/rakh**if***, *kha**k**/rakh**uk***) are never psychologically opaque. An account of the type we are imagining would argue that their invariant, single-WFS counterparts in the High code (like *kha**if**/rakh**if**, kha**uk**/rakh**uk***) are aligned with them, illuminating the alternations.

Of course, the absence of such a completely independent account does not make it impossible to produce a DED for the LSSG(Bangla) package, and to compare that alternative with our CCD. Someone designing a DED would have to choose between a diachronically grounded approach – one that keeps separate tabs on the history of Bangla over the last six centuries and on its weakly diglossic state today – and an approach focused on the diglossic alignments alone. Either way, the proponent of a DED might want to adopt some of CCD's relatively 'neutral' semiotic propositions in order to make meaningful theory comparison feasible. The point would be to avoid adopting those characteristic proposals that distinguish CCD from widely held positions – for instance, the 'Cohort Peer Pressure' idea.

It serves no purpose for the author proposing CCD himself to play the devil's advocate with full sincerity and design an entire counterproposal in the DED format. The remarks in this section are only an attempt to outline with some clarity the context for such a task. An actual proponent of DED will want to rigorously compare certain lexical and morphological phenomena in the High and Low codes of Bangla. If such a proponent does not speak the language and needs to use a (preferably theory-neutral) description, perhaps the most comprehensive sourcebook for this purpose that is readily available is Thompson (2010).

In the spirit of itemizing problems that someone constructing DED will want to address, we flag one architectural issue and two empirical issues. The architectural issue arises if the critic opts for partial acceptance of the semiotic approach. In that case the difficulty is that adding diglossic (and perhaps diachronic) data to the material to be handled runs the risk of placing a far more baroque grammatical/ lexical/ semiotic puzzle on the table. Once the discussion takes up a larger class of sets of elements in the context of inquiring about paradigmatic relations between some of them interacting with syntagmatic and other phenomena, there will inevitably be new semiotic facts to consider, requiring fresh generalizations and manoeuvres. If this happens, all that the DED-proposing critic can hope to show is that specific diglossic (and/or diachronic) factors explain certain phenomena that the rival CCD account handles by semiotic

means, and that these need to be redescribed. This set of minor quibbles will not affect the overall methodological debate regarding the desirability of a semiotic turn. To get that far, DED will need to pare the semiotics down to the Elsewhere Condition and perhaps Mayerthaler-Wurzel, and to show that everything else proposed in the present paper amounts to a series of mere artifacts that dissolve on serious diachronic and sociolinguistic scrutiny.

The architectural issue fails to arise if our critic adopts a standpoint well outside any version of the semiotic turn. Such an author will need to consider matters that we raise in section 7, when we return to this topic.

We turn now to matters empirical. The first issue our critic needs to face pertains to quirky verbs. Certain examples give us the initial impression that Low Bangla's quirky verbs, which exhibit that typical *ĭ* – like the infinitivals *ʃoĭte* 'to.tolerate', *duĭte* 'to.milk', contrasting with the non-quirky *hote* 'to.become', *dhute* 'to.wash' – correspond to High Bangla verbs with the bare verb template CV*h*. These particular examples do have CV*h* counterparts in the High code, *ʃohite* 'to.tolerate', *duhite* 'to.milk', which stand in contrast with the *h*-less *hoĭte* 'to.become', *dhuĭte* 'to.wash'. Such examples invite the inference that – during the historical period when the vowel harmony trigger /i/ was diachronically lost in forms like *hote* or *dhute* – the /h/ in *ʃohite, duhite* protected their /i/ from loss; and that when a subsequent sound change targeted the /h/ (hence the outcomes *ʃoĭte, duĭte* in Low Bangla today), the high-vowel-targeting sound change had run its course and left the quirky *ĭ*'s untouched.

There is an issue here, because in fact several High code counterparts to quirky verbs neither harbour nor have ever harboured such an /h/. We find exactly the same *h*-less forms in the High and Low codes for *nuĭte* 'to.bend', *ruĭte* 'to.sow', *khoĭte* 'to.decay', *dhaĭte* 'to.run'. Loose talk of 'analogy' is all very well, but the actual traffic of just how the chips fell, and when, has never been worked out for quirky verbs and their non-quirky comrades. Without having at least an overall story that says where /h/ fits into the High-Low alignment with regard to quirkiness, a proponent of DED cannot critique the core of CCD.

The second issue pertains to Future forms. The High code's Future consistently exhibits medial /i/ (with the postvocalic allophone [ĭ]). CVC class: *ʃunibe* 'they.will.listen', *dhoribe* 'they.will.catch', *miʃibe* 'they.will.mingle', *ʧhelibe* 'they.will.push'. Quirky CV class: *ʃohibe* 'they.will.tolerate', *nuĭbe* 'they.will.bend'. Regular CV class: *hoĭbe* 'they.will.become', *dhuĭbe* 'they.will.wash'. In the Low code, while the first two classes behave as expected (CVC class: *ʃunbe, dhorbe, miʃbe, ʧhelbe*; quirky CV class: *ʃoĭbe, nuĭbe*), the regular CV class surprises us by deploying not raised vowels, but plain vowels. We would expect High *hoĭbe* 'they.will.become', *dhuĭbe* 'they.will.wash' to have the Low counterparts *hobe, dhube*. Instead, Low Bangla features *hɔbe, dhobe* etc. throughout this verb class.

The problem is not just that our understanding of the diachrony leading to Low Bangla has no clear niche for the combinations of historical processes responsible for these exceptional Futures. We must also parse the fact that the divergent, non-raising behaviour of the regular CV Future is not spreading to the CVC template, coupled with

the puzzling variability facts focused on the very same template in the context of the Future-Laden Gerund form – facts mentioned at the end of section 4 and flagged as due for a revisit at the end of section 5. We now end section 5, leaving to real critics the task of developing an alternative that covers our facts, either with some semiotics in their toolkit, as suggested in this section, or with none, a project discussed in section 7.

6. Revisiting Jackendovian 'Independent Information'

Throughout this paper, we nebulously appeal to some 'evaluation metric' – formulated along the lines of Jackendoff's (1975) 'Measure of Independent Information', but with some phonological insight attached – whose adjudication would distinguish natural from crazy connections in terms of phonological naturalness. Our exposition implies a metric capable of treating all natural and crazy connections as light and heavy cognitive burdens respectively (in terms of acquisition by children, recognition by adult listeners, and production cost for adult speakers). A semiotic turn warrants, not a literal, penny-wise-pound-foolish evaluation metric, but only appropriate signposting. In this section, we provide the indications required, without formalizing in a pointlessly rigorous style. Readers committed to formalization may parse these remarks as constraints on the class of admissible metrics.

In the substantivist tradition, especially its GP-WWM sector, cognitive burden differences do not reflect quantitative properties of surface notational entities. WWM pursues a non-atomizing word-based approach to the analysis of inter-word relations, and must therefore express non-syntagmatically the intuition that the word-set {*aggression, aggressive, aggressor*} is a lighter burden to carry than {*theft, thief, steal*}. This task is handled by what Jackendoff (1975) called a Measure of Independent Information (MII). In a theory devoid of any imaginary verb 'aggress', what counts is the MII burden difference between the heavier set {*theft, thief, steal*}[18] and the lighter set {*aggression, aggressive, aggressor*}. It is this MII that we seek to adapt to our purposes.

A theory formalizing inter-word relations in terms of WFS's must have WFS applications evaluated by MII. MII's broader Jackendovian task (identifying and assessing word-to-word relations within the lexicon) includes such evaluation as a special case. If we are to appreciate the full spectrum of MII's verdicts, we should expand its mandate to include assessing the weight of GP's application of Repair Strategies. MII evaluates these processes as burdenless, as infinitely light. At the other end an unnatural or 'crazy' WFS substituting /s/ for /k/ in *electricity* is very heavy in MII's reckoning; if it earns some redeeming value as a flag that maximizes word-cluster recognition, then

[18] Note that {*theft, thief, steal*} is a significantly heavy burden in the eyes of WWM; this model does not regard *theft* and *thief* as morphologically related, since no WFS (which would have to be based on at least two phonetically/ semantically matching word-pairs) subsumes it (*heft/heave* is unhelpful: there is no **heaf*).

the relevant semiotic subsystem mitigates MII's verdict (without impugning MII), but still leaves the WFS marked as moderately heavy. MII faces interesting questions of relative evaluation at the points on the spectrum that lie between 'infinitely light' and 'very heavy'.

In order to attach some terminology to this tweaking of MII's mandate, we call our version of it MI***A***I, a Measure of Independent ***Arbitrary*** Information. In this view, processes reflecting the phonotactics of the language contribute zero weight. Processes recognized as natural by universal phonology but falling outside the phonotactics of the language itself contribute less weight – fine-tuning how little, and why, is where implementations are going to diverge. That level of detail lies beyond the scope of this paper. Simply in order to convey the sort of criteria we have in mind, let us note that the 'strong raising' syndrome (mapping, for example, between /kh***a***/ 'eat' and /kh***e***ĕe/ 'having.eaten') mentioned in the context of the ill-fated Plan B counts as mildly natural, which makes the relevant WFS's moderately light; but that the severe deformation observed in polysyllabic verbs, a topic not discussed in this paper (e.g. the mapping between /pal***a***/ 'escape', /ghænghæn***a***/ 'whine' and /pal***ĩ***e/ 'having.escaped', /ghænghen***ĩ***e/ 'having.whined'), is evaluated by MIAI as moderately heavy. The point is to extend the resources of the GP-WWM tradition, not to undermine it.

What contribution, if any, can MIAI make to the task of choosing between the CCD and DED designs for LSSG(Bangla)? That depends on one's implementation of the MIAI program. It is not yet clear how the 'familiar' vs 'formally learnt' axis interacts with other determinants of ease or difficulty in the linguistic knowledge of a fluent speaker. Thus, the cognitive import of diglossia and related phenomena remains obscure. Authors will, no doubt, argue for or against specific hypotheses as such inquiry makes headway. For the moment, in the absence of any articulation of DED, we cannot comment on the matter.

Some readers may find a way to bring our approach to bear on the Walsh & Parker (1983) or Plag (2017) results mentioned in our introductory remarks. But the concrete proposals made in this paper do not move us any closer to that goal, which we mention only because those results count as 'semiotic' if this term means anything at all.

7. Conclusion

By way of concluding remarks, we would like to make some bibliographic points that skeptics need to take into account. Authors who intend to base their criticism of substantivist proposals on empirical material from the morphology of Bangla (or its sister language Hindi) will need to engage with the existing literature. But the standard morphological works on Hindi – the mother tongue of Rajendra Singh, one of the major architects of GP and WWM, and the reference language for several key proposals in his theorizing – include Singh & Agnihotri (1997) and Agnihotri (2007). These are

WWM descriptions drawing on GP. The only book-length account that examines in detail the processes of word formation in Bangla also uses WWM supplemented by GP (Bhattacharja 2007). Dan (2014) and Mitra (2017a, b) take Bhattacharja's work further. Even authors hostile to GP-WWM or its extensions cannot investigate Hindi and Bangla without engaging with this research tradition, which, needless to say, welcomes empirically grounded criticism.

Bibliography

Agnihotri, Rama Kant 2007: *Hindi: An Essential Grammar.* London / New York: Routledge.

Bhattacharja, Shishir 2007: *Word Formation in Bengali: A Whole Word Morphological Description and Its Theoretical Implications*. München: Lincom Europa.

Bhattacharja, Shishir (Ms): *Introducing Whole Word Morphology*. Unpublished manuscript.

Bhattacharya, Nandini 2017a: Semantics of reduplicated nominal quantification in Bangla. In: R. Ashem & G. Kaur & U. Udaar (eds): *Structures and Syntax in South Asian Languages*. Cambridge, UK: Cambridge Scholars. 73-92.

Bhattacharya, Nandini 2017b: Modified numerals and plurality in Bangla. *Jadavpur Journal of Languages and Linguistics* 1:2.1-7.

Bochner, Harry 1993: *Simplicity in Generative Morphology*. Berlin/ New York: Mouton de Gruyter.

Dan, Mina 2012: Kriyāpade svaradhvanir uccatābhed: ekṭi kāyābādī biśleṣaṇ. [In Bangla: 'Vowel height differences in verbs: a substantivist analysis'.] *Alochonachakra* 32:235-55.

Dasgupta, Probal 2000: The verb stem formative /a/ in Bangla. R.E. Asher, R. Harris (eds) *Linguisticoliterary: A Festschrift for Professor D.S. Dwivedi.* Delhi: Pilgrim. 115-25.

Dasgupta, Probal 2001: On a vowel template asymmetry in Bangla verbs. Anvita Abbi, R.S. Gupta, Ayesha Kidwai (eds) *Linguistic Structure and Language Dynamics in South Asia: Papers from the Proceedings of SALA XVIII Roundtable*. Delhi: Motilal Banarsidass. 164-181.

Dasgupta, Probal 2009: Transparency and arbitrariness in natural language: some empirical issues. Rajendra Singh (ed.) *Annual Review of South Asian Languages and Linguistics 2008.* Berlin/ New York: Mouton de Gruyter. 3-19.

Dasgupta, Probal 2010a: Strategies and their shadows. Rajendra Singh (ed.) *Annual Review of South Asian Languages and Linguistics 2009.* Berlin / New York: Mouton de Gruyter. 3-40.

Dasgupta, Probal 2010b: The Hindi long vowel problem: a substantivist approach. Hasnain, S. Imtiaz; Chaudhary, Shreesh (eds) *Problematizing Language Studies: Cultural, Theoretical and Applied Perspectives: Essays in Honour of Rama Kant Agnihotri.* Delhi: Aakar. 286-291.

Dasgupta, Probal 2016: Pre-demonstrative gaps in Bangla: syntactic and semiotic recoverability. *JKI* (*Język Komunikacja Informacja*) 11:195-212.

Dasgupta, Probal & Ford, Alan & Singh, Rajendra 2000: *After Etymology: Towards a Substantivist Linguistics.* München: Lincom Europa.

Dressler, Wolfgang Ulrich 1994: A functionalist semiotic model of morphonology. Rajendra Singh (ed.): *Trubetzkoy's Orphan: Proceedings of the Montréal Roundtable "Morphonology: Contemporary Responses"*. Amsterdam/ Philadelphia: Benjamins. 67-83.

Ford, Alan & Singh, Rajendra 1991: Propédeutique morphologique. *Folia Linguistica* 25:549-575.

Ford, Alan & Singh, Rajendra & Martohardjono, Gita 1997: *Pace Pāṇini: Towards a Word-Based Theory of Morphology.* New York: Peter Lang.

Jackendoff, Ray 1975: Morphological and semantic regularities in the lexicon. *Language* 51:639-71.
Mayerthaler, Willi. 1981. *Morphologische Natürlichkeit.* Wiesbaden: Athenaion.
Mitra, Nivedita. 2017a. Asymmetries and Whole Word Morphology. *Bulletin of the Department of Linguistics [University of Calcutta]* 20:173-194.
Mitra, Nivedita. 2017b. *Sound Pattern Asymmetries in Bangla Morphology.* Doctoral dissertation, University of Calcutta, Kolkata.
Plag, Ingo; Homann, Julia; Kunter, Gero. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53:181-216.
Singh, Rajendra (2006): Whole Word Morphology, in: *Elsevier Encyclopedia of Linguistics*. 2nd ed. Amsterdam: Elsevier. 578-9.
Singh, Rajendra; Agnihotri, Ramakant. 1997. *Hindi Morphology: A Word-Based Description.* Delhi: Motilal Banarsidass.
Singh, Rajendra; Starosta, Stanley (eds.) 2003. *Explorations in Seamless Morphology*. New Delhi: Sage.
Thompson, Hanne-Ruth. 2010. *Bengali: A Comprehensive Grammar*. London/ New York: Routledge.
Tirumalesh, Kilingar Venkappa. 1991. Why you can't repunch Bill: an enquiry into the pragmatics of '*re*-words' in English. *Journal of Pragmatics* 16.249-267.
Walsh, Thomas; Parker, Frank. 1983. The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics* 11:2.201-6.
Wurzel, Wolfgang Ulrich. 1989. *Inflectional Morphology and Naturalness.* Tr. Manfred Schentke. Dordrecht/ Boston/ London: Kluwer.

Abbreviations

1p, 2p, 3p	first, second, third person
ATR	advanced tongue root
C	consonant
CCD	Cohort Coherence Design
DED	Diglossic Equations Design
FL Gerund	Future-Laden Gerund
Fut	Future
GP	Generative Phonotactics
Hon	honorific
Imp	imperative
Intim	intimate
LSSG	Language-Specific Semiotic Guidelines
MII	Measure of Independent Information
MIAI	Measure of Independent Arbitrary Information
Neu	neutral
σ	syllable
tr	transitive
V	verb
V	vowel
WFS	Word Formation Strategy
WWM	Whole Word Morphology