sciendo

Linguistic **Frontiers**

On certain consequences of the objectification of languages: a substantivist approach

Probal Dasgupta
Independent Researcher; ORCID number 0000-0001-9168-7642

0. PRELIMINARIES

The practitioners of linguistics (in all its forms) hope to converge on tools suitable for describing all human languages within a shared terminological and conceptual framework, demarcating phenomena that lend themselves to meaningful cross-linguistic comparison from those that do not. To this end, linguists are obliged to treat languages, and speech communities, as **objects** of analysis. In this respect, the characteristic posture of linguistics authors vis-à-vis their readers contrasts with the non-objectifying attitude associated with Traditional Lexicography and Grammar (here called TLG). Those who write dictionaries and (normative and pedagogical) grammars that count as authoritative for various points on the literacy scale, ranging from schoolchildren to the most proficient users of the written language, address their readers as potential writers (and, crucially, as potential editors) of the language.

The practices and attitudes characteristic of TLG reference a single editorial-normative community. As such, they are particularistic, but may occasionally involve more than one nation-state. Country A's TLG workers negotiate with their counterparts in country B, to calibrate orthographic or other norms of a shared language like Dutch or German. Bilingual dictionaries operate with the TLG equipment of both the societies. As an enterprise, TLG crosses national boundaries only on this limited, transactional scale. It does not aspire to a universal scientific standpoint, and thus has no reason to objectify its language or its speech community. TLG represents, and intersubjectively addresses, only a circumscribed editorial-normative collectivity, the "we" to which its authors and readers belong.

But linguistics references "us scientists of language," a global professional network. Linguists hope to converge on a universal theoretical and descriptive framework applicable to all languages. Its scientific gaze theoretically places every language and every speech community under objective, descriptive scrutiny. The practical application of these principles has led to difficulties. We argue in this paper that these difficulties have to do with certain unresolved aspects of the relation between the 'science' of linguistics and the 'cultural practice' of TLG.

Linguistics claims to deal primarily with spoken language (for linguistics to focus on written language would have made it non-universal; only a proper subset of spoken languages is wedded to writing systems). But every literate society's TLG manages the pedagogy and the editorial-normative functioning of its written language, treating the spoken language as one implementation of the written. The task of optimizing the linguistics-TLG equation, then, is closely related to that of adequately articulating the relation between speech and writing. It is at this level that this paper hopes to contribute to the field of linguistics. We set out by adhering to the received wisdom that linguistics is an enterprise that has been improving and superseding TLG's practices and operational machinery. As our argument develops, the limitations of this view will become evident. The argumentation in this paper represents a viewpoint anchored in formal linguistics but focused on the speech-writing equation. Given the talk of spoken and written 'substances,' referenced in the dictum that 'language is form, not substance,' we call our viewpoint **substantivist** linguistics, and describe those of our colleagues who focus on form alone as **formalists**.

On certain consequences of the objectification of languages: a substantivist approach

This paper further explores the notion that the pursuit of a certain **precarious exactitude** has been the driving force behind the study of formal linguistics, i.e. phonology, morphology, syntax, and semantics. As such, we first explore the traditional practices of grammar and lexicography, disciplines that have always tacitly pursued the same goal, the unusual character of which never came to the fore. We propose that examining the pursuit of this goal in all its precariousness — heightened by particular factors operative today — will help illuminate the practice of linguistics and enable fruitful conversations with other disciplines.

In section 1, we argue that the practice of formal linguistics today is in an anomalous position; the wording of the section title is an allusion to Smolin (2006). Section 2 assembles some elements of a productive response to the predicament diagnosed. Sections 3 and 4 gather these ingredients into operational proposals and conceptual responsibilities, respectively.

1. THE TROUBLE WITH LINGUISTICS

What prompts us to claim that traditional grammar pursues a **precarious exactitude** in linguistic phenomena? School geometry highlights circles, triangles, and squares, whose regularity can be exactly specified. Map drawing in geography showcases irregular shapes, resisting exactitude. But grammar occupies a curiously intermediate position. Grammatical irregularity phenomena tempt us to try and devise an account that would couple fully exact generalizations with maximally exact local specifications of the distribution of irregularities. However, if one succumbs to this temptation and follows this path to its logical conclusion, one gets into deep trouble, as we are about to see — trouble that calls for deep remediation.

Regardless of the diversity of schools of linguistics, we have all been exposed to irregular verbs as part of our basic pedagogy. Regular past tenses and perfect participles (*kicked, taped, flitted*) contrast with irregular forms (*spoke, shook, sat*; *spoken, shaken, sat*) — an obstacle every learner of a language like English must negotiate. Our lifelong familiarity with these examples prevents us from seeing what makes grammatical irregularities a unique set of phenomena. Let us look past the familiar tip of the irregularity iceberg, then, and notice that *writers,* who *write; drivers,* who *drive; directors,* who *direct*, constitute a majority pattern contrasting with an outlier pattern exemplifed by *authors,* who do not *auth; gladiators,* who never *gladiate; testators,* who never *testate;* and even *dictators,* who do not *dictate*. Such outliers, hidden from the ordinary view, make the picture look awkward.

How does one deal with this disorder if one is pursuing exactitude? One extracts, even from outliers, as much tribute as possible to the principle of order. One notes *dictator*'s potential metaphoric connection with *dictating*. One acknowledges *auth**or*** and *gladi**ator*** as agent nouns with these bold-font flags, despite the non-existence of *authing*/*gladiating*.

Variability is an additional factor for the pursuit of lexico-grammatical exactitude to contend with. Alongside the irregular past/perfect *learnt* and *burnt*, English allows the regular *learned, burned*. *Creep* harbours an amusing duality: ordinary creeping sticks to the irregular *crept*, but the baroque verb in *That guy creeps me out* chooses the regular inflection, *creeped*. Such eccentricities keep traditional grammarians and lexicographers on their toes.

But surely modern linguistics, with its scientific mandate, has sent prescriptive traditional grammar packing, changed the agenda completely, and placed variability squarely on its pluralistic-descriptive map; right? Wrong, actually. Normative dictionaries and grammars still control educated speech and writing in literate societies. The methods of that enterprise call the shots. Descriptive linguistics, a scientific **supplement** to the **TLG** (traditional lexicography and grammar) **apparatus**, has relatively recently persuaded it to become more inclusive and accommodate marginal usage. To see just how this affects the pursuit of exactitude, consider some examples.

As explained in the prefatory sections of the Merriam-Webster (2003: 12a, 34a), alongside the normal standard pronunciations of the words *cupola* and *nuclear*, this dictionary also registers the fact that some Americans pronounce them as *cupalo* and *nucular*. To be sure, every such outlier is flagged as 'a pronunciation that occurs in educated speech but is considered by some to be unacceptable' (12a). By recording them, however, the Webster is making an inclusionary cultural statement.

Dictionaries still serve as authoritative resources for teachers and editors prescriptively enforcing standards. But this authority has been moving towards descriptive inclusivism. Traditional education instructed the public to **catch up** with textbook models. But, given that the entire public will never say *cupola, nuclear, February, library*, the TLG machinery today tries to **catch down** with the stragglers, letting *cup**ola**, nu**cular**, Fe**buary**, li**bary** also count as 'educated speech'. How far 'down' does the process go? The Oxford English Dictionary Team (2020) has added to its agenda the goal of covering 'world English' — varieties spoken in New Zealand and Nigeria, in India and Ireland (cf. Nelson et al. 2019, Schreier et al. 2020). Similar moves may well be under way for other major languages used across the North-South divide. It is time formal linguists began to look at this divide as a factor relevant to their disciplinary cartography and more.

Is TLG biting off more than its pursuit of exactitude can chew? Its newfound concern to include what the "old" authorities frowned upon compels it to aim for exact characterizations of these déshabilles and at the same time also of norm-setting speakers whose impeccability is admired and emulated. Pursuing such a precariously heterogeneous species of exactitude amounts to conflating the Olympics with the Paralympics, surely?

Readers who find our comments unfair may attribute the apparent scrambling of priorities to TLG's dual role: keeping a tradition-bound pedagogic system satisfied and responding to pressures from linguistics. Modern

linguistics is seen as having successfully pursued rigorously delineated goals; and, within its limits, it has. But in order to continue its advance, linguistics — in our view — needs to renegotiate its equation with TLG and set aside its focus on the quest for a homeland in the natural sciences, even as it keeps pursuing its own species of exactitude.

The point becomes clearer when one looks at the tacit equations now in place. TLG keeps languages codified by running an administrative apparatus that serves the public authorial **agora** (the iconically creative performances staged in the agora are standardly known as *literature*; terms like *the public space*, *civil society*, *free speech* point to the agora itself; using the neutral term agora helps us to prescind from the diversity of societies with regard to the valorization of free speech priorities). The TLG apparatus is assisted by its **scientific supplement**, *linguistics*. The public and its formal institutions hold literature in high esteem, respect the literature-backed editorial-normative authority of their society's TLG apparatus, but see no societal role for linguistics.

Linguists, in turn, ignore their triangular relation with the agora and TLG. As they conduct their formal business in phonology, morphology, syntax, and semantics, linguists project their work as just an extension of TLG. This projection makes the precarious exactitude challenge even harder to meet. To see why, let us take a closer look at the way some current syntactic writings characterize what does and does not fall within standard languages.

When Massam (2017) and Bayer (2020) add a few more bricks to the formal edifice of generative syntax on the basis of data drawn from markedly casual English and German;[i] when Cinque (2010) emphasizes the near-uniformity of key aspects of adjectival sequencing in 'the national Romance languages' in contrast to the 'greater variation' seen in 'dialects, or older stages, of these languages;'[ii] or when Barbiers (2010, 2014, cited by Bayer 2020) presents some 'colloquial Dutch' facts resembling Bayer's German data — we find TLG-normative terminology guiding the delimitation of their phenomena. They come across as providing a rigorous and elaborate, but 'catching-down'-driven, account of **sentence** structure supplementing TLG's lexicographic catching down vis-à-vis the **words** of these languages.

Bayer (2020: 53–54) explicitly argues that sentences with two occurrences of a discourse particle like *denn*, such as *Vor was denn ist er denn geflüchtet?* 'What on earth did he flee from?', are not 'speech errors which by some accident made it onto a web page' or cases of 'blending in running speech,' but count as acceptable in standard German. That he needs to argue for the point leads us to wonder how widely his perceptions are shared. I showed some native speakers and some German-proficient Indians Bayer's examples and requested acceptability judgments. The Indians all said that the normative German taught to foreigners prohibits discourse particle doubling. Among the native speakers, those concerned with connotations characterized the sentences as

instances of highly nuanced modulation. Native speakers favouring robust delimitation rejected the sentences as careless speech, unacceptable in standard German.

These responses to Bayer's data, as well as Barbiers and Massam's labels (cited above) for analogous Dutch and English material, suggest that TLG's scientific supplement, formal linguistics, is 'catching down' with casual déshabille. It is thereby bringing the **sentence** level also under the inclusionary regime (like TLG's toleration of substandard pronunciations of **words** like *cupola* and *nuclear*) — expanding downwards the range of what counts as normatively acceptable. The Cinque quotes exemplify the fact that linguistics has always referenced normatively administered national standard languages (ranging over ordinary language use and the 'formal styles' mentioned in note 2) and their earlier stages and dialects.

But Bayer, Barbiers, Cinque, and Massam's syntax writings are academically impeccable. How can they be methodologically troubling? When we suggest that extending inclusivism from words to sentences makes the pursuit of our precarious exactitude even harder — when we claim that linguistics, to keep advancing, must reset equations with TLG and the agora — just where do we see a problem?

The formal syntax mainstream, here called Mainstream Generative Grammar (MGG), has consistently taken what we may call Standard Average Prose as its only object of study, a default baseline from which stylistic adventures take off. Morphology, semantics, and phonology follow suit. This first approximation has overstayed its utility. MGG workers fail to see the historically developed nature of 'ordinary language.'

Virginia Woolf (1932/2003: 9), notes, for example, that English prose in Shakespeare's time, 'for all its beauty and bounty, was a very imperfect medium. It was almost incapable of fulfilling one of the offices of prose which is to make people talk, simply and naturally, about ordinary things.' For example, 'Mercy the milkmaid writes a natural and noble style, which is incapable of vulgarity, and equally incapable of intimacy' (p. 14). When we find that Mercy really wrote like this — 'The thing you wot of, Milord, were a great trespass towards God, a great offence to the world, a great grief to my friends, a great shame to myself, and, as I think, a great dishonour to your lordship' — we see Woolf's point. 'Ordinary prose', a product of modernity, was unavailable in Elizabethan England. Practitioners of generative syntax have unreflectively deployed a contingent historical product as a *scientific baseline* for a *theory* of language. Formal linguistics, led by MGG, cannot afford to stay trapped in this quaint belief, which the TLG enterprise has managed to avoid.

The TLG apparatus, a conglomeration of practical responses to needs, treats doctrines eclectically. Obliged to handle the agora's demands, rhetorical conventions present and past, living and frozen figures of speech, the inclusionary agenda promoted by descriptive linguists, the poetic license of literary authors, the mandatory

inclusion of usages of Shakespeare-level icons, the editorial practices of publishers, and much else, TLG functions pragmatically. The very practicality of its apparatus, however, dramatizes a significant difference regarding codification. It is here that our 'precarious exactitude' issue arises. We use **bold italics** below to flag this difference between TLG and MGG.

TLG **codifies** *only* words and idiomatic **collocations**. It handles the **sentence** both as a **structure** (under a grammatical normativity regime) and as a **composition** (editorially modulating this regime on the genre axis) — at the frontier where **codification** negotiates with the agora's **free** discourse. In doing so, TLG serves both the speaker-writer (attentive to codified norms as well as to editorial practices connected to the exercise of free speech), and the listener-reader (who must comprehend many locutions she herself would never use).

In contrast, MGG (i) builds a model (***with certain scientific idealizations***) offering a 'particular grammar' of every language, **codifying** (i.e. formally associating pronunciations with meanings for) all the **words** *and* the in principle infinitely many **sentences** available in the language, seen only as **structures**; (ii) shows how key points in these language-particular accounts stem from 'universal grammar,' i.e. from MGG's model of *C(HL)*, the **C**omputational calculus that is responsible for the unique properties of **H**uman **L**anguage and is a biological characteristic of our species; and (iii) leaves all biology-independent properties of language to be handled by socio-cultural scholars, as the natural science of linguistics shall have nothing to say about them.

The 'scientific idealizations' include unifying words and sentences under C(HL). This uniform computation throws the pursuit of exactitude into jeopardy; let us see how. Consider the word level. The TLG treatment of exceptions like *slept* instead of *sleeped* is absorbed in MGG's formal machinery under the Lexical Particularization Principle: *The lexicon's particularizations upstage the grammar's generalizations* — my wording, taken from Dasgupta (2009), for a consensual principle sometimes called 'the elsewhere condition'. Thus, the 'lexicon' endows the word *sleep* with a **particular** past form, *slept*, forestalling the **general** form *sleeped* that the grammar's rules would otherwise supply. TLG confines its codification to words — acknowledging that a sentence, qua structure and qua composition, involves negotiation between grammatical pressures and the author's preferences. But MGG brings words and sentences under a unified codification. How does particularization fare under such a regime?

Let us focus on the empirical worry first: do sentences and words really handle exceptional patterns identically? When a quirky sentential pattern competes with a general pattern instantiating grammatical rules, does the exception indeed upstage the rule?

To find out, we compare the quirky syntactic patterns seen at (1i)–(3i) with their regular competitors (1ii)–(3ii):

(1) i. What say you?
    ii. What do you say?

(2) i. What have we here?
    ii. What do we have here?

(3) i. Suffice it to say that this is satisfactory
    ii. It should suffice to say that this is satisfactory

If some variant of the Particularization principle were operative here, (1i)–(3i) should have driven (1ii)–(3ii) out of business. But this is not so; (1ii)–(3ii) are well-formed.

Not only in English do quirky sentential patterns systematically coexist with quirkless rivals. Irregular patterns like (4i) and (5i) in Bangla also leave their regular counterparts (4ii) and (5ii) alone (we are deliberately using a TLG-referencing transliteration rather than an IPA transcription, to flag the specific concerns of this paper):[iii]

(4) i. *Dilip nā pereche śarkārke hāte rākhte, nā pereche biplabi bole nām kinte*
    'Dilip has neither managed to retain the government's goodwill nor managed to become well-known as a revolutionary'
    ii. *Dilip śarkārkeo hāte rākhte pāre ni, biplabi bole nām kinteo pāre ni*
    'idem'

(5) i. *Rinā Lipike śoukhin churiṭā pherot debei debe*
    'Rina will certainly give Lipi the fancy knife back'
    ii. *Rinā Lipike śoukhin churiṭā niścoyi pherot debe*
    'idem'

Quite generally, quirky sentential patterns do not upstage their grammar-obedient, regular counterparts, but coexist with them. Particularization is indeed Lexical, applying to words alone; there is no Syntactic Particularization.

We are now able to specify how MGG's choices throw lexico-grammar's pursuit of a precarious exactitude into deeper jeopardy.

Empirical jeopardy: Had MGG sought to describe **only** Standard Average Prose, leaving all else to the humanities, one could have tried to systematize a formal linguistics so circumscribed. But MGG in practice begins at TLG's normative codification (in societies that have one) of each language X (including extra styles/dialects/periods of history), and builds a formal description whose verdicts roughly echo TLG's codification for X.[iv] Thus, if TLG delimitations prevalent for language Y happen to be more permissive or more broadly polysystemic than those applicable in Z, then formal linguists working on Y and Z end up pursuing different exactitudes for science-external reasons.

Conceptual jeopardy: Not only does the descriptive burden become challenging and intricate. Even technical resources for meeting the challenge are thrown into conceptual jeopardy once the codification of words, inherited from TLG, is formally unified with MGG's innovative codification of sentences (the heart of the generative

revolution). The coexistence of quirky and regular syntactic patterns, versus exceptions (*slept*) upstaging regular morphology products (*sleeped*), is a systematic asymmetry. Making sense of this asymmetry and yet maintaining MGG's unitary codification of words and sentences is a conceptual challenge for every variant of the theory.[v]

The two issues meet at a surprising point, at which remediation also becomes conceivable.

2. FACING THE PREDICAMENT

At this juncture, readers may ask where we stand on familiar contestations of MGG by sociolinguistics (Blommaert 2018), Cognitive Grammar (Langacker 1987, 1991), or MiGG (minority generative grammar, e.g. Bresnan et al. 2015); a valid question. Since we want a productive discussion of grammatical methodology, we find it convenient to address colleagues interested in methodological matters, to focus on the regularities/exceptions issue — an issue every normative reader/writer is acquainted with — and to keep in view TLG and our literature-iconizing agora. The known contestations mentioned above also lock horns with MGG on other issues. They do not help **this** conversation, although their insights are vital elsewhere.

It turns out both our 'jeopardies' have to do with the agora. For the empirical jeopardy this is obvious. That TLG is forced to codify more than the Standard Average Prose of language X under the X-norm reflects the way the agora keeps the X-norm porously open to styles above or below the average; to the past; to dialects; to archaisms; to whimsical formulas that catch the public's fancy; to piecemeal formulas from Shakespeare or others whom the agora iconizes as 'literature.' For the conceptual jeopardy, more needs to be said. We begin by looking at the morphology-syntax unification itself.

The conduct of words is known to be far less systematic than that of phrases/sentences. Morphology is not just riddled with exceptions. Even its regular processes are constrained. Typical of syntax is the grammar of interrogation, which systematically ferries between statements and questions. A statement *Mary asked Dilip five questions* licenses a corresponding question *Did Mary ask Dilip five questions?* But even the far-reaching morphological mapping of adjectives (*happy, sad, brisk*), into adverbs (*happily, sadly, briskly*), is systematic only within limits. Adjectives already ending in *ly* cannot adverbialize (**friendlily, *likelily*). The adverb *fast* lacks the *ly* marking. The compact format for adjectival comparison (*sadder, saddest*), competes with adverbialization — **sadderly, *sadlier*. Being *roundly/squarely condemned* has no adjectival counterpart (**the round/square condemnation*).

Traditional grammar took this contrast seriously, and never unified morphology with syntax. Why, then, did most schools of formal linguistics, including MGG, choose to unify them? Several linguistic concerns converged on discrete 'logical' (or 'cognitive')

meaning-segments. To regard *Sheela and Ahmed remembered the fact that they visited Egypt in 2010* as 'logically' equivalent to *Sheela and Ahmed remembered visiting Egypt in 2010* justifies placing them in the same grammatical computation. Thereupon **de**bugging, featuring the 'morpheme' (the minimal carrier of 'logical' meaning), **de**, aligns with **eliminat**ing bugs.

Holding this notion of segmentable cognitive meanings constant, we find a systemic contrast between morphology and syntax: word-level irregularity (*slept*) upstages regularity (**sleeped*), but syntax allows for the coexistence of exceptional (*What say you?*) and regular sentence patterns (*What do you say?*). Despite some syntactically quaint literary quotes (*Lend me your ears*), most examples of irregular syntax do not come from classics. We are flagging here some material circulating in the agora that syntactically departs from Standard **Average Prose** but extracts recognition from TLG as falling within Standard **Prose** and is tenuously connected to literature. We propose to call ambient material such as *What say you?* **obliterature**, suggesting an imaginary "pulverizer," obliterating textual recognizability but retaining some literary flavor.

Demarcating obliterature as distinct from the default genre of a language might give MGG a tidy agenda to implement; does this gambit work? If obliterature is cordoned off, regular *What do you say?* technically no longer coexists with quirky *What have you?* in ordinary English. This should help. The principle that exceptions upstage regular counterparts prevails in morphology **and** syntax (which, now exception-free, 'trivially satisfies' the criterion by technically not violating it). The anomaly mentioned earlier stands resolved, then, in the Standard Average Prose genre of English. Does this mean our troubles are over?

Alas, addressing one difficulty changes the scene, bringing other challenges into view. Having found MGG's morphology-syntax unification feasible on the syntax side, we revisit morphology — primarily to check whether Lexical Particularization is indeed in full working order — only to encounter fresh challenges there.

Fresh Challenge #1: MGG's syntax-morphology unification rests on the belief that morphology chops words up into discrete segments (roots and affixes), exactly as syntax analyzes sentences into words, apart from morphology's exception-proneness. Does this belief stand up to scrutiny, though? Words are **discrete** units, indeed; but far too many 'morphemes' (MGG's exoteric[vi] cover term for roots and affixes), are **discretionary** segments. Articulating 'discretionary' phenomena is Fresh Challenge #1.

Most of us have been brought up on a diet of comfortable examples like *anti-dis-establish-ment-arian-ism*, where the hyphen-marked 'morphemes' seem to match words in the syntax-assembled phrase *doctrine of movement against movement against established order*. You equate *ism* with *doctrine* and *anti* with *against*, but you allow for morphology's addiction to irregularities — like liber**tar**ian/commun**itar**ian instead of the expected liber**ty**ian/commun**ity**ian; you accept this eccentricity.

On certain consequences of the objectification of languages: a substantivist approach

To grasp the magnitude of Fresh Challenge #1, try scrutinizing some *ism* words and their kindred. *Buddhism, Jainism, Judaism, Hinduism, Sikhism* — but *Christianity*; suffixless *Islam*; obsolete term *Mohammedanism*. What about followers? *Jains, Jews, Hindus, Sikhs* (*Juda* vs *Jew* is morphology being whimsical as usual). For religions named after founders, followers get called *Christians, Buddhists*, and (obsolete) *Mohammedans*. So far, so good.

Follower adjectives, as in *Christian* culture, *Jewish/Judaic* culture? *Hindu/Sikh/Jain* double as adjectives, then *Buddhist*, and (obsolete) *Mohammedan*. *Buddhistic* is just an obsolete variant in English. French, however, distinguishes *la culture bouddhique* from *la religion bouddhiste*. In English, *Buddhic* is a technical theosophical term unrelated to Buddha, and *Christic* is a fringe word too: *christique* is more normal in French. Now try German. *Christentum* does not designate Christendom (false friends) but Christianity; *Judentum* is Judaism. But the follower nouns and adjectives display a spectacularly different system: *Buddhisten, buddhistisch; Hinduisten, hinduistisch; Jainisten, jainistisch; Sikhs, sikhistisch;* and *Juden* springs a surprise: adjective *jüdisch*; the pattern-adherent *judaistisch* conveys a specialized meaning 'pertaining to *Judaistik*, the study of Judaism,' Why do the specialized disciplines *Buddhistik/Jainistik* not snatch the words *buddhistisch/jainistisch* away from pedestrian adjectivehood?

Similar questions arise when *science* in English takes the pedestrian adjective *scientific*, with *scientistic* an outlier wedded to *scientism*, but pedestrian *juristic/artistic* prevent the *scientism*-parallel *jurism* from being coined and *artism* from flourishing.

It is not a matter of detail, then, that English morphology's *X-ism, X-ist*$_{Noun}$, *X-ist*$_{Adj}$, *X-an, X-istic* cannot be syntax-glossed, respectively, as 'doctrine associated with X,' 'follower of X,' 'associated with X-doctrine,' 'follower of X,' 'associated with X-doctrine,' leaving a few wrinkles for some irregularity broom to magically sweep up. 'Irregularity' is a suitable term for *went* answering the question **goed* was expected to answer; but here we are clueless. Just what questions are seeking answers in vain? If we devise rough and ready glosses for English *ism, ist, an, istic*, they fail in French or German, and the differences are too intricate to allow systematization. But gloss-cheques, if they have no cash value at language boundaries, become useless as semantic interpretations.

What tentative conclusions follow? It certainly matters that the discipline of *Judaistik* claims ownership of *judaistische*; that the negligible strength of *Jainistik* and *Buddhistik* leaves *jainistisch* and *buddhistisch* free to be pedestrian adjectives; that, outside the proper noun domain, ownership effects recur in *scientistic, artistic, juristic*. These are indeed salient discretionary phenomena. However, 'ownership' hardly helps make sense of the way the limited *X-wālā* pattern of Bangla (*śobjiwālā* 'vegetable.*wala*, vegetable seller;' *phalwālā* 'fruit.*wala*, fruit seller;' *bāṛiwālā* 'house.*wala*, landlord'), contrasts with the broader *X-vālā* pattern in Hindi-Urdu

(which features similar examples plus, unlike Bangla, *āsmānvālā* 'heaven.*vala*, the one in heaven, god;' *denevālā* 'giving.*vala*, (a) one who gives, (b) about to give;' *yehvālā* 'this.*vala*, this particular one;' *kaunsāvālā* 'which.*vala*, which one in particular' and many other hand-waving uses). We use quasi-glosses like 'pertaining to' for Hindi-Urdu *X-vālā* and 'person systematically occupied with' for Bangla *X-wālā*, inadequately[vii], having no exactitude-style purchase on morphology's discretionary ways. When we examine morphology closely, the notion of segmentable word-internal units of cognitive meaning stops being systematically operable.

In response to this challenge, we propose that words like *diner, miner* (see note 7) develop **tessellated meaning-relations** with their morphological associates (e.g. *dining, mining*). What **kind** of move this is becomes clear in connection with the second challenge.

Fresh Challenge #2 shakes core assumptions guiding TLG and MGG: in certain domains, Lexical Particularization collapses, giving way to **superregularity.** The following examples feature proper noun plurals and sarcastic causatives. The plurals of *life* (English), and *ciel* 'sky' (French), as common nouns harbor irregularity: *lives/cieux~ciels*; but magazines called *Life* and *Ciel* have **superregular** (unexpectedly and exclusively regular) plurals: *Lifes/Ciels* (vs **Lives/*Cieux*). The plural of *child* is *children*; but a *Child* family comprises *Childs*, not *Children*. Moving to irregular verbs, in a special sarcastic context (see Dasgupta 2007 for details), Bangla disables standard irregular causatives like *mārā* 'to kill' or *cālāno* 'to drive' (base verbs *marā* 'to die', *calā* 'to run'), and deploys, instead, the superregular *marāno* 'to make (someone) die,' *calāno* 'to make (a vehicle) run.' Even English, with no causative system, overrules other regularities in this context. If I say *Shall we proceed?* or *We just might succumb to the temptation*, your sarcastic retort *Just you wait, **I'll** proceed **you**!/**I'll** succumb **you**!* exemplifies some special factor authorizing transitivization even from otherwise robustly derivation-resistant intransitives, another superregularity phenomenon.

We have two choices here. The choice preferred by MGG's **formalists** (who seek to encode the language in an omniscient lexico-grammar) is to build an even more **intricate** code, installing Ptolemaic epicycle-type counter-exceptions within the grammar. But the parsimonious choice, which we **substantivist linguists** prefer, is to devise an **extricating** strategy (we explained the reference to 'substance' in the Preliminaries section). Our selectively extrovert code opens appropriate interfaces with **discourse** — with the articulate face the agora wears when it talks to language.

Both these moves — the **extrication move** and the **tessellation move**, responding to Fresh Challenges #1 and #2, and incorporated into operational proposals in section 3 — draw on **semiotics,** the study of all signs and sign-combinations. Lexical Particularization itself is a semiotic principle; unsurprisingly, characterizing the limits of its applicability (Challenge #2) also calls for semiotic measures — a hallmark of substantivist inquiry.

As we consider how to face the predicament, we pause to situate our little journey on the map of semiotics, using a metaphor. On your visit to a park, having lunch at a restaurant, located in the park in the *broad* sense, you look at the green reality outside, the actual park in the *narrow* sense. In the park of broad semiotics, linguistics is the restaurant. Semiotics proper is the greenery outside. Describing nature as 'agile' and buildings as 'sedate' helps classify the signs involved. Linguistics handles *sedate* signs, known as **symbols**. The currency of (narrow) semiotics is *agile* signs, subcategorized as **iconic** (resembling something), versus **indexical** (pointing to something).

There is a twist: when the grammar assembles symbols, these assemblages obviously resemble their members. The phrase *heavy desk*'s shape-sharing bond with the words *heavy* and *desk* endows sedate symbols with a touch of agile iconicity. In a further twist, grammatical irregularity obscures certain shape-similarities, reducing the agility of those items. Thus, regular *halfwit* and *two--timing* directly resemble *half* and *two*, while irregular *ha'penny* and *tuppence* only half share their shape. This makes *halfwit/two-timing* more agile than *ha'penny/ tuppence*. The grammar's rule-governed assembling of symbols happens in the linguistics restaurant, then; but even within this assembling process, the semiotics park makes its agile presence felt.

Having assembled the words you need (eaten your restaurant lunch), you intone your sentence as a statement/question/exclamation (you step into the park). TLG's punctuation marks <. ? !> recognize these **intonations** as distinct. A classic paper by Karcevskij (1931) shows how semiotics governs intonation. His account notes that, while uttering a sentence, we exhaust the pulmonic air supply; thus, our pitch keeps falling as we speak. Unmanipulated, intonation would therefore always show a downward contour. But human language deliberately raises the pitch at sentence-end to say 'the speaker hasn't finished.' Raised pitch may serve as a *semi-colon*, ushering in *my* next sentence, or as a *question mark,* inviting *you* to continue my thought by answering. The normal *full-stop* intonation mimics nature by letting the pitch fall naturally. Here, letting nature take its course is part of language design. Semiotics counts this design feature as **iconic**, mimicking the natural phenomenon. Its opposite, rising pitch, counts as a **symbolic** device, saliently anti-iconic, *cultural*. Notice, however, that this is nonetheless the park's *horti*culture, not the restaurant's cuisine. Lexico-grammar does not encode intonation the way it encodes our sedate signs. Intonation is agile — on the [+iconic] axis.

We can now say that the semiotics of exclamative intonation overrules sedate grammar in our sarcastic constructions. This is the park authorizing superregular transitives (English) and causatives (Bangla), pushing agility beyond the restaurant's limits.

What about superregularity in proper nouns? Those also involve semiotics, but of an indexical kind, a matter addressed in section 4. For immediate purposes we merely note that names obviously invoke discourse; it is thus natural that their superregular behaviour also responds well to the substantivist restaurant's anti-hermetic strategy of letting the park's air and sunshine come in through selected windows.

Limited phenomena like superregularity call for windows. The ubiquitous agility that our semiotic gaze notices in the restaurant itself, however, prompts us to open a door, one that has stayed locked for decades. We propose to invoke MGG's core belief that, given that the grammar generates infinitely many sentences, every particular sentence that is uttered counts as new. This view treats the sentence as essentially agile and is not far removed from the approach we are suggesting. Assuming that every sentence uttered makes a fresh contribution to the discourse, semiotics may help us articulate this contribution process.

Familiarity makes it hard to notice freshness in stereotypical ordinary prose. Let us try something exotic, like poetry (we choose classical examples, to minimize the detour). Alexander Pope's poem 'An essay on man: epistle II' opens with: 'Know then thyself, presume not God to scan; /The proper study of mankind is man[viii].' No single line gives access to the work done by Pope's rhyme/metre scaffolding. In this poem, we must inspect at least two lines to see the rhyme (*scan — man*) and the consistent use of the iambic pentameter. In Shakespeare's sonnet XII, the scaffolding emerges over a four-verse span: 'When I do count the clock that tells the time, /And see the brave day sunk in hideous night; /When I behold the violet past prime, /And sable curls, all silvered o'er with white;'[ix] — this is the shortest passage that exhibits the ABAB rhyme scheme, apart from confirming that the iambic pentameter's writ runs here as well.

Readers are aware of the role played by rhyme in shaping our attention patterns. Pope's verses use rhyme to emphasize *man* — the point of the poem. In Shakespeare's sonnet, the ABAB words *time — night — prime — white* ring the changes on the theme of aging. Assuming that these are the salient words, semiotics urges us to notice that the rhyme-carrying phrases in the Shakespeare passage — *tells the time /sunk in hideous night /past prime /silvered o'er with white* — convey decline, greying, aging; and that these 'after' photographs immediately follow *the clock /the brave day /the violet /sable curls*, which show us either neutral or young 'before' photographs.

We call the rhyme-bearing part of each line of verse its **kite**, and the earlier part the **anchor** of the line (in Dasgupta 2010, we note that the anchor-kite division is akin to topic-comment, and that one misses the point if one treats the study of poetry as if it was extraneous to the prose-only enterprise of formal linguistics). The devices used in classical poetry to highlight contrasts and connections *between* sentences have only pale, elusive counterparts in prose (average or otherwise), or in speech; efforts to catch and tame those beasts have failed. A semiotic gaze has a lot more to spot *within* a sentence.

What does it find? In prototypical instances, grammatical agreement connects two major ingredients in a clause; even inflectionally austere English features subject-predicate agreement: *John **loves** Mary*, singular; *John and Bill **love** Mozart*, plural. Perhaps grammar cannot demonstrate the universality of syntactic agreement distinct from non-universal morphological agreement. But semiotics registers this indexical bond as one key factor in the sentence's sentencehood. The iconicity axis comes into the picture too. Any sentence starts with what is 'given' and contributes something 'new,' thereby offering, at its minimal level, some contrast (some **countericonicity**). At its maximal level, this contrast between the given and the new stages a spectacle. The indexical bond in every sentence forms an intimate entente with its (however minimal) countericonicity.

But all this is merely the syntagmatic tip of the semiotic iceberg. On the paradigmatic plane, semiotics diligently juxtaposes the declaratives considered so far with exclamations, questions and other distinct sentence formats (the operation 'juxtapose' is formalized in Dasgupta 2010). It associates intonation with semiotic consequences of lexical and structural choices; it presides over these distinct formats handling the intimate entente differently; it does sundry other footwork we have not yet learnt how to watch.

Some readers are bound to wonder if we have said enough to justify the claim that relying so heavily on semiotics, regarded by many as an untested set of instrumentation proposals, is going to help us to face our predicament. We have indeed not said enough to justify this leap of faith, but here is an example of what we are staring at when we face our predicament — the fact that no known formal linguistic take on language in general or on English in particular is able to say[x] why the following sentence from the first page of *Mrs Dalloway* (Woolf 1925) strikes any literate reader as a normal **sentence**, though somewhat long: *How fresh, how calm, stiller than this of course, the air was in the early morning; like the flap of a wave; the kiss of a wave; chill and sharp and yet (for a girl of eighteen as she then was) solemn, feeling as she did, standing there at the open window, that something awful was about to happen; looking at the flowers, at the trees with the smoke winding off them and the rooks rising, falling; standing and looking until Peter Walsh said, 'Musing among the vegetables?' — was that it? — 'I prefer men to cauliflowers' — was that it?*

And yet, letting semiotics 'take over' will not lead to salvation.

3. OPERATIONAL CONSEQUENCES

To see clearly what is at stake, we visit a domain where semiotics legitimately demands grammatical attention: operational iconicity, as in reduplication performing the grammatical function of plural marking. In Bangla, 'reduplicating' (repeating) an adjective often pluralizes the head noun of the construction: *lal lal phul* (red red

flower), 'red flowers' (a bare noun in Bangla can take a plural value if the context so warrants). Why 'often?' Sometimes it conveys approximation rather than plurality: *bõdā bõdā gandho* (musty musty smell), 'a sort of musty smell.' In a context where both readings work, such as (6), the listener chooses the contextually plausible reading:

(6) Rākār rā̃dhā titke titke tarkāri kārur bhālo lāge nā
　　Raka.Gen cooked pungent pungent vegetable nobody like Aux Aux
　　'Nobody likes the pungent vegetable dishes/the sort-of-pungent vegetable dish cooked by Raka'

Neither the grammar nor the semiotics forces one particular interpretation.

We tentatively conclude that, where the grammar gives reduplication an operational job to do, the semiotics and the grammar scrutinize word choice and contextual plausibility while negotiating over admissible readings that will prevail. When the phenomena themselves suggest that the semiotics-grammar interaction is negotiation-laden, we should neither fear nor welcome a 'takeover' by semiotics, but should visit domains where this negotiation becomes visible.

One domain to visit is relative clauses, which feature **relative pronouns**, e.g. *the place **where** I met them was Milan*. In one corner of English, the relative pronoun turns poltergeist: now you see it (*the woman **whom** you noticed/the pictures **which** you looked at*), now you don't (*the woman you noticed/the pictures you looked at*). Bangla places its poltergeist corner elsewhere. A Bangla **relative** preceding a **demonstrative —** such as (7) — allows relative pronoun omission, as seen at (8):

(7) Kumbhokarnero **jā** ghum bhāngiye debe **èmon** āwāj
　　Kumbhakarna's.even **which** wake up will **such** noise
　　'**Such** a noise **which** will rouse even Kumbhakarna from his slumber'

(8) Kumbhokarnero **jā** ghum bhāngiye debe **èmon** āwāj
　　Kumbhakarna's.even ~~which~~ wake up will **such** noise
　　'idem'

For these cases, MGG resources manage the traffic efficiently, using methods that need not detain us.

But one key contrast between English and Bangla poses a challenge for MGG. An English poltergeist vanishes alone, as we see at (9) below, where the preposition *at* survives the deletion of the *which*, whereas Bangla poltergeists disappear with their content-bearing companions, in apparent violation of the principle of the recoverability of deletion, as we observe at (10)–(11):

(9) the pictures ~~which~~ you looked at

(10) ***jā diye*** *bastā kāṭā jāy èmon kāci*
~~which with~~ sack cut can.be such scissors
'such scissors ~~which~~ we can cut sacks~~with~~'

(11) ***jā dekhe*** *cokh dhādhiye jāy èmon ujjal ālo*
~~which seeing~~ eye dazzled are such bright light
'such a bright light ~~on seeing which~~ one's eyes are dazzled'

When as fundamental a conservation principle as the recoverability of deletion is violated, MGG's syntactic apparatus cannot cope. Semiotics steps in, authorizing a recasting of left-of-demonstrative relatives as complement clauses: 'such a bright light *on seeing which* one's eyes are dazzled' becomes 'such a bright light *that* one's eyes are dazzled.' Certain typological factors enabling such semiotics-guided syntactic tessellation in Bangla but not in English are flagged in Dasgupta (2016). In older work (Dasgupta 1980), on which MGG descriptions of relative clauses in Bangla and sister languages are based, the intractable left-of-demonstrative cases were explicitly flagged and set aside. Analysis became possible only after the substantivist turn endowed formal linguistics with semiotic resources.

Our second domain concerns syntax-harnessed reduplication, this time involving not adjectives but verbs. In Bangla, a single verb.*te* form is an infinitive *(bolte* 'to say/talk/tell,' *dite* 'to give/allow'), but a verb.*te* verb.*te* sequence is an adverbial progressive participle *(bolte bolte* 'while saying/talking/telling,' *dite dite* 'while giving/allowing'). Now, consider sentences like (12) and (13), in which one infinitive is embedded within an exact twin:

(12) Dilipke [[bonnār kathā **bolte**] **bolte**] keu cāibe nā
Dilip.Dat [[flood.Gen about **talk.Inf**] **tell.Inf**] anybody want.Fut Neg
'Nobody will want to **tell** Dilip to **talk** about the flood'

(13) Rākāke [[khārāp khaborṭā Nikhilke tokkhuni **dite**] **dite**] ke rāji habe?
Raka.Dat [[bad news.the Nikhil.Dat at.once **give.Inf**] **let.Inf**] who agree will
'Who will agree to **let** Raka *give* Nikhil the bad news at once?'

When said out loud, these sentences carry a tell-tale intonation helping the listener to parse them. But readers seeing a written sentence of this type out of the blue initially read verb.*te* verb.*te* as a participle, pause when this reading fails, and then do a double take.

MGG syntax handles these simple cases of twin infinitives with ease. But phrasal verbs like *bāron karā* (lit. prohibition to.do) 'to prohibit' give pause. Embedding one infinitive-inflected phrasal verb within a twin, even in writing, proves easy to parse, even for a context-unaided reader; consider (14):

(14) Tridib baḍḍo śabāike bāron kore bèṛāy, tāi bole āmi oke [[**bāron korte**]
bāron korte] pārbo nā
Tridip a.lot everyone.Dat prohibition does.and goes.around, and yet I him
[[**prohibition to.do**] **prohibition to.do**] can.Fut Neg
'Tridib goes around issuing prohibitions, but I can't **prohibit** him from *prohibiting*'

Why is (14) readily parsable? Because here the twin infinitives *bāron korte bāron korte* cannot be confused with the participle *bāron korte korte*.

These observations per se may not compel an invocation of semiotics,[xi] but Jana (2021: 113–126), argues that the cluster of phenomena they come from do require it. Verb.*te*, she notes, is the only non-finite form exhibiting mandatory reduplication, and that too only in its participial use. Jana shows how this reduplication is nested in other iconicities; thus, the facts become irreducibly semiotic, while remaining within the grammar.

Readers belonging to the global North tend to feel overwhelmed when they see too much 'exotic' data, their term for the global South. Setting aside some interesting recent findings from Santali (Badenoch et al. 2019, Williams 2019), and the sequel to our own poltergeist story[xii], we therefore turn now to an unexpected and pervasive irruption of the exotic within English itself, first observed by Edward Sapir (1921: 35): 'We are compelled to leave [the word] *unthinkable* as an integral whole, a miniature bit of art' — Sapir's point being that one says 'unthinkable' only of what one *can* **think** but finds un**accept**able.

In a classic article, Tirumalesh (1991) links Sapir's observation to Aronoff's generalization: 'When a word does have a base, it is legitimate to ask about the semantic relationship between the two. Since morphology is not syntax, this relationship will seldom be one of neat compositionality. There will usually be some sort of divergence. Interestingly, this divergence is not between the derivative and the base, but between the actual meaning of the derivative and the meaning we expect it to have, given the independently occurring meaning of the base' (Aronoff 1976: 32).

Tirumalesh (1991: 259–60), uses Aronoff's 'compositionality plus bonus' generalization to make sense of the fact that we can say (15) but not (16):

(15) $repunch_1$ the holes in the paper

(16) $repunch_2$ Bill

He argues that in (15) the word's expected meaning, combining *re* with $punch_1$ 'perforate' contributes to but does not exhaust the actual meaning of $repunch_1$ — the bonus is the element that the second person (who does the $repunching_1$), is completing a predecessor's shoddy work; but that (16)'s derivation from $punch_2$ 'hit' offers

no bonus whatsoever, leaving us with no rationale for forming such a word. *Repunch₂ Bill* fails, Tirumalesh suggests, because its exact synonymy with *punch₂ Bill again* is crushingly pedantic.

The reasoning here rests on 'every derived word is implicitly oriented to some bonus it is expected to carry,' a semiotic idea. Sentences violating it sound absurd — but remain intelligible, as humor: 'This further supports our analysis since the humorous effect [of any actual use of *repunch₂*] is produced because of the "cost" involved in its use: so much is spent to achieve so little!' (Tirumalesh, p. 260).

In order to see clearly how this 1991 analysis inflects the pursuit of exactitude, we apply its ideas to an example that emerged well after that date: the word *resend₇* (imagine that 'send by e-mail' is sense #7 of *send*). When I *resend₇* an email message, I certainly do '*send₇* it *again*,' The 'bonus' — that my doing so resumes and completes an earlier unsuccessful attempt to *send₇* — is only one ingredient that goes into the texture of *resend₇*. Even if you later tell me my first *sending₇* was in fact successful, we cannot say you *rereceived₇* my message. In his rich study (cited fragmentarily here), Tirumalesh highlights the concrete texture of words, saying, in conclusion (p. 266): 'The pragmatic **landscape** of *re*-words has only certain signposts to guide us and these — we hope to have demonstrated — resist all attempts at semanticization' (emphasis mine). If 'semanticization' (an exact mapping between sounds and meanings — including precise differentiation of options where one 'morpheme' juggles two or more senses) had indeed been in charge of language, as formalist linguists claim it is, then *rereceive₇* would echo *resend₇*, we would *reask* and *reanswer* endless questions, correction, we would avoid ever calling a series of questions *endless* — bowing to the verifiable fact that they do end — and exactitude would be equally attainable in morphology and in syntax.

In a **landscape**, the word we emphasize above, no single factor dominates. Linguists must address the fact that every word₁ meeting word₂ in a sentence activates discourse, a multi-factor game. To Tirumalesh's insights we add that *thank**less*** but not *thank**ful*** collocates with (appears in the same phrases as) *task*, scrambling the arrows of gratitude/ingratitude. Actions are *lawful*: places/persons are *lawless*. The *helpful* provide help: the *help**less*** fail to receive it. Untypical *less/ful* word-pairs that neatly obey semanticization (*fearful/fearless, thoughtful/thoughtless, joyful/joyless*) look tidy on account of accidental collocation partner 'choices.' The question is not what semantics can contribute to a derived word; the question is what the word, with its phrasal companions, can offer to the cause of tessellation.

Tessellation functions not just for meanings, but for forms too. One cannot delimit the adjectivalizing 'affix' in *melodic, electrical/electric, philosophical/*(rare) *philosophic, theoretical/*(game-, set-)*theoretic, periodical ≠ periodic, historical ≠ historic*. Some of these words expand into *Xicity*: *electricity, periodicity* (and *historicity*, which is

happy to (semantically) match *historical* and not *historic*), whimsically avoiding *Xicality*. Adverbs choose the *Xically* look instead: *melodically, philosophically*, etc. No principled delimitation of the 'affix(es)' *ic, ic-al, ical* turns out to be defensible.

'Morpheme'-believing formalists may agonize over rational cuts in such words. But substantivist methods, based on the seamless integrity of the word, attain two goals simultaneously by stating word-to-word regularities in the 'Word Formation Strategy' format [Xic(al)]$_{\text{Adjective}}$ ←→ [Xically]$_{\text{Adverb}}$. Substantivism handles the formal details optimally and, by taking its stand at the word, aligns with the social fact that the play of discourse affects words (and phrases), freezing *histor**ic*** and *period**ical*** into particular senses, without the public ever noticing what happens to morphological units like *ic* and *ical*; for this game, they're the wrong chips (Singh, Starosta and Neuvel 2003; Ford, Singh and Martohardjono 1997).

How about syntactic units? If you revisit sarcastic causatives and transitives (the *I'll succumb you!* material mentioned earlier), you find that the play of discourse contextualization affects even the delimitation of those units. Only specialists will notice syntactic unit boundaries in that instance. We therefore focus on more obvious data: in (17) and (18), the verbs bold-faced in the glosses are silent in the German originals:

(17) Er will nach Berlin (~~gehen~~)
　　　'He wants (**to go**) to Berlin'

(18) Sie muß nach Köln (~~gehen~~)
　　　'She must (**go**) to Cologne'

A silent 'go' can piggyback ride the modal verbs *will/muß* 'wants to/must,' Discourse tweaks syntactic unit boundaries in German to authorize these special rides.

Kayne, who has worked for decades (with results partly reported in Kayne 2005) to systematize the MGG understanding of such silent lexical items — and whose system would easily handle those German cases or the Hindi example *vo merī$_{fem}$ (bāt$_{fem}$) nahī̃ suntā* 'he doesn't listen to my (words)' — informs me that one problem that Gouet (1976), and Morin (1977), discussed decades ago still eludes his net. *Bourgogne* 'Burgundy,' the district, is feminine in French, but the wine is masculine. Ordering wine at a restaurant, one says *Un$_{masc}$ bourgogne blanc$_{masc}$, s'il vous plaît* 'white Burgundy, please,' the masculine article and adjective possibly agreeing with a silent *vin$_{masc}$* 'wine,' whereupon the waiter shouts *Une$_{fem}$ bourgogne blanc$_{masc}$ pour la 2!* 'one white Burgundy for table 2!' — with a feminine article possibly agreeing with a silent *bouteille$_{fem}$* 'bottle', but leaving the adjective masculine. Gouet proposed processes deleting contextually chosen words like *bouteille*. Morin advocated aligning visible agreement features with absent words identifiable from the context. Kayne has been trying to place these examples on a serious map of silent word phenomena.

15

How is this Gouet-Morin problem, or the Sapir-Aronoff-Tirumalesh material, crucially different from section 2's 'obliterature'? The word-to-word issues highlighted here arise in **conversational** contexts, whereas obliterature invokes (pulverized) **writing**. Linguistic terminology distinguishes the **form** of language from the written, spoken and other[xiii] **substances** in which it is embodied. Taking the naive conception focused on physical manifestations as a starting point, **substantivism** rearticulates the speech/writing binary in ways elaborated in section 4.

In this 'operational' section 3, reconfiguring our disciplinary equations with TLG and the agora, we must acknowledge a clash: TLG treats language as primarily written ('How do you pronounce *Gwladys*?'); MGG inherits the structuralist view that language is primarily spoken ('How do they write εksãprovãs in French?'). Section 3 copes by focusing on the practice of parsing sentences, shared by both sides. The **spoken/written sentence** is the consensual upper limit of grammar. We substantivists assume, with MGG, that the syntax of the spoken sentence is an exact business; and, with TLG, that whenever you finalize a written sentence, you aspire to enjoy editorial autonomy as part of normal democratic citizenship. TLG's normativity and MGG's exactitude converge on this amphibian site where we must reset our equations. Unscrambling the speech/writing binary stands postponed to section 4.

The 'contextual' factors one invokes are typically located just outside the sentence. By choosing a domain where they are at a minimal *paradigmatic* distance from the target *re*-word, Tirumalesh was able to comment rigorously on the grammar-semiotics negotiation. We gain comparable leverage by choosing the following case, where the affector-affectee *syntagmatic* distance[xiv] is minimal.

Expanding the older term 'cranberry morphemes' (*cran* and *boysen* occur only in *cranberry/boysenberry* — contrasting with the 'versatile' *black/goose* of *blackberry/gooseberry*), linguists today speak of **cranberry words**, or **bound words**, e.g. *betwixt*, confined to the locution *betwixt and between*. Their distribution in Bangla verb collocations is patterned as follows (Dasgupta 2012a). Phrases of the 'compound verb' type (e.g. *phele deoā*, drop.and give, 'to throw away'), or the 'conjunct verb' type (e.g. *āloconā karā*, discussion do, 'to discuss'), may harbor bound words (**emphasized** in *hediye jāoā* 'to get bored', ***kẽce** jāoā* 'to come to naught', ***thoke** jāoā* 'to get tired'; ***jigeś** karā* 'to ask', *ṭer pāoā* 'to notice'). If two words are separated by a predication boundary, like V_1 and V_2 in (19) or (20), neither of them is ever bound:

(19) āmrā [cā khete$_1$] cāi$_2$ ni
     'we didn't want$_2$ [to have$_1$ tea]'

(20) Dilip tomāke [ḍimgulo śeddho korte$_1$] bale$_2$ ni
     'Dilip didn't ask$_2$ you [to boil$_1$ the eggs]'

Bound word phenomena involve semiotics quite centrally. After all, semiotics underwrites the fundamental colligation associating the shape of any word with its meaning. When a bound word depends on a sponsor word for interpretability, semiotics must make special arrangements. The Bangla case study suggests that a phrase may harbor bound word sponsorship, but that the bond cannot cross a syntactic boundary. Substantivism's investment in semiotics leads it to interpose a Phraseology module between the word-managing Lexicon and the sentence-assembling Syntax module. More needs to be said on 'modules' for this to become clear.

In the founding formal moves of substantivist linguistics, Singh and Ford (Ford et al. 1997, Singh et al. 2003), drew a semiotic line demarcating the grammar's Phonology module (Generative Phonotactics) from its Morphology module (Whole Word Morphology), placing all arbitrariness-laden word formation processes in the Morphology. The Lexicon, the repository of all *arbitrary* or 'sedate' information, is the stage on which modules run their operations.[xv]

To see how to parse this, consider some examples. English Morphology posits a wrinkle-free WFS (Word Formation Strategy) $[X]_V \leftrightarrow [Xing]_{V, Gerund}$ for word-pairs like *go/going, come/coming*. A second WFS, $[X]_{Adj} \leftrightarrow [Xly]_{Adv}$, easily handles *rapid/rapidly*, but coordinates with the Phonology module for *simple/simply*. Why? The unsupplemented WFS maps the pronunciation *simpl* onto *simplli, unpronounceable in English. Phonology, using processes not elaborated here, arrives at *simpli*. Now take another WFS, $[Xic]_{Adj} \leftrightarrow [Xicity]_N$, handling words pronounced as *ilektrik/ilektrisiti*.[xvi] Why does Phonology not turn *ik* into *isiti* here? Why force Morphology to handle it? Because *ilektrikiti* is **pronounceable** —that *ilektrisiti* is in fact pronounced with an **s** is an arbitrary fact about English, which belongs to the Morphology module. The module boundary, as you now see, is drawn semiotically.[xvii]

When substantivism first applied semiotic resources to syntactic phenomena (Dasgupta, Ford and Singh 2000), we noted that syntax juxtaposes discourse-adjacent with morphology-adjacent core properties. Our module demarcation strategy therefore precluded establishing a unitary syntax module. We posited Phrase Formation Rules in the syntax-morphology frontier region. They later grew into the Phraseology module invoked above. Other aspects of syntax have been modularized differently (see Dasgupta (2011) for details). We are trying to highlight certain compelling points that substantivism's semiotic toolkit enables it to articulate, and to argue on this basis for a reconfiguration of the TLG/linguistics/agora triangle.

Contemporary substantivism treats semiotics as a place like a national repository bank. It runs a Microsemiotics module, one operational bank among others, managing relevant patterns in coordination with other modules, underwriting a level of formal structure empirically motivated elsewhere (Dasgupta, forthcoming). But semiotics also works at the repository bank level as a pervasive supramodule, overseeing the traffic (through principles like Particularization, which counter-principles sometimes overrule), and micromanaging a domain or two.

Perhaps the most tangible cases of such micromanagement are at the morphology-phonology interface; the reader is referred to Dasgupta 2018 for an explicit, step-by-step exposition.

MGG linguists and similar formalists also rely on Particularization and a semiotic armature they usually sideline as 'mere lexical arbitrariness'. They have simply refrained from articulating the consequences of their manifestly semiotic premises. Focusing on these issues, substantivism develops packages of remedies for such omissions in formalist work. Our architecture is an elaborate case for a semiotics-driven course correction; validating it as a remediation package should be enough. Linguists using our package will only take specific remedies they need.

The current question: How might this body of work improve the health of the linguistics/TLG/agora triangle? Our answer rests, not on the **agile** keys to our semiotic turn, but on the **sedate** tools added in section 4. With that upgrade, semiotics-aided linguistics — because semiotics runs the bloodlike circulation system of language, as our remarks above have shown — can at last face and half-agree with TLG's writing-centered view of language. To half-agree is to begin to articulately question that center from a standpoint informed by exactitude and literate access to the agora.

The division of labor is surely clear to readers who have been following our chain of argumentation. Semiotics bears the burden of the **inexact** flesh and blood of a **particular** language, in dialogue with the agora where it is spoken. The modes of formal pattern **exactitude** — niched in the modules of the formal architecture (reflecting analyses of the hundreds of languages linguists have been studying carefully), and micro-interacting with the semiotics of the particular language in ways exemplified in our discussion — offer a **universalistic** (if not necessarily natural-scientific[xviii]) basis for credibly questioning the forces in the agora of language L that promote center-worship and systematically erode or undermine cognitive activity. This questioning needs to be led, *in every possible instance*, by L-speaking linguists who, by writing *in* L, anchor themselves and the linguistics discipline in the discourse of the analysand speech community, adding vertical strength to the horizontal-universalistic validation of the discipline.

So augmented, semiotics-aided linguistics can launch a natural sequel to the virtuous but relatively stress-free documenting of 'endangered' languages. We can no longer postpone facing contradictory definitions of literacy, rationality, autonomy, in our various milieux, and their relevance to seriously reciprocal cross-cultural cooperation.

But the seriously interested reader is waiting for the sedate upgrade that this story rests on, which involves the Semantics module.[xix]

4. CONCEPTUAL RESPONSIBILITIES

Sedate signs, called symbols, have the 'arbitrariness' property. The naive approach to arbitrariness homogenizes them, thereby missing the **deutero-arbitrariness twists**. Section 4 articulates two such Twists, which are interconnected, and which give rise to certain responsibilities for linguists: Twist W — **writing**, which encodes speech, is deutero-arbitrary; Twist N — **names** are deutero-arbitrary as well, in contrast to common nouns. In this section, we also discuss perpetrate Tweak H — **hypo-indexicality** (hypocoristic indexicality), which relabels a Sarah Ann Timmons by the **nickname** *Sally*.

We have been using classical 'iconic/indexical/symbolic' labels for expository accessibility; this was an expository oversimplification. It is true that, by pointing at a table, *this one* does indexical work. But at the lexical encoding level the English expression *this one*, in contrast to its French equivalent *celle-ci* and its German equivalent *dieser*, is constitutively symbolic, arbitrary. To ask 'Is *this one* mainly symbolic or mainly indexical?' is inappropriate. Upgrading our toolkit involves more than just terminological enrichment: we must raise deeper questions.

Most grammatical work, most saliently MGG, rests on the perfection idealization: *Imagine an infinite-memory-and-attention-laden fictional person P instantaneously learning language L perfectly and completely, characterize PKL (P's knowledge of L), and you have described L. The PKL idealization sets aside individual deficits irrelevant to linguistics*.

Normal MGG work, however, is done by imperfect individuals I_1, I_2, I_3 presenting their judgments (as to which sentences are well-formed/ill-formed) as an approximation to PKL and expecting other L-speakers to concur. (Untypical MGG work by I_is on Foreign Language FL, relaying FL-speaking J_1, J_2, J_3's judgments, toes the line set by normal MGG and leaves the epistemic scene unaltered). Does an I_i wear her grammarian hat while judging well-formedness? Given the discipline's anti-prescriptivist posture, linguists seek 'naturalness' judgments, leaning away from their schooling. Nobody has inquired to what degree unconscious attitudes to TLG, and to written discourse, affect intuition-mining practices; we are not about to embark on such a fruitless course of investigation here.

We prefer a different strategy, that of noting that semantics stands to benefit from a more dynamic idealization than the standard idealization in current use. An individual speaker-hearer I_i's active knowledge of L, at any age, at any educational level, is far more limited than her passive knowledge, which is fed by scraps of words (and more), intimately known to speakers familiar with other sectors of L's division of lexical and generic labour. The linguist's characterization of I_i's semantic knowledge, we hereby suggest, needs to track a child's idealized growing comprehension (see Dasgupta 2000 for a full articulation of this point). The child assimilates all the adult expressions she deciphers. She recognizes as infantile certain words she earlier accepted as valid (*this bow-wow, that meow*), perhaps retaining them as toys. She critically scrutinizes her adults, seeking to out-adult them.

Now, an *empirical* older child typically seeks to out--adult her adult reference figures from the standpoint of *an ideal citizen of her society*. But the substantivist semanticist's *idealized* older child seeks to expose, from *a rational panhuman standpoint*, the indefensibility of that society's priorities (embodied in its semantic defaults and other indicators).

It is here that linguistics and TLG part company, intellectually. TLG implicitly represents the writ (of variable scope and definability) that runs in the L-speaking society. Linguists must explicitly resist this writ, and insist: 'If *we* cannot fully comprehend everything in the agora (since the task is unending), *your* alleged key to the order of things, whatever it is, must be invalid.' You will never fully **under-stand** your language, for reasons of principle. The same principles enjoin you to **with-stand** claims implicit in the written language to the effect that somebody owns a key to it.

A serious linguistics adopting this approach ironically agrees with TLG that language is **writing**-**centered** — and then proceeds to denounce this **centering** as illegitimate. To be able to do this, linguistics has to intervene in L-speaking society as a counterwriting which, notionally on behalf of panhuman suprawriting (or, equivalently, on behalf of the Older Child Idealization OCI), constantly questions the power of the written in L, keeping this power on its toes. On the ground, such intervention[xx] involves leveraging multilingual speakers whose individual cognitive repertoire references the TLG codifications of several languages, inevitably opening up contradictions and creating opportunities for critical scrutiny, playing one TLG off against another (call this exercise the macroscopic tessellation).

The above is offered as an update of anti-prescriptivist linguistics, as a prolegomenon to our contribution to democratic discourse — a contribution motivated not by incidental beliefs of individual practitioners but by constitutive properties of our science. MGG, by moving away from the agora, has in our view thrown the democratic responsibilities of science out with the ideological bathwater.

We turn now to these 'constitutive properties,' in the context of upgrading our semiotics to handle Twist W (Writing), Twist N (Names), and Tweak H (Hypoindexicality). Most of semiotics detects and analyses implicit signalling. But the Twists are matters of Explicit Coding. Both Writing and Names constitutively refer to recorded, culturally stored codes and thereby add an extra layer of arbitrariness, making them deutero-arbitrary. Let us unpack this.

When Sheehan and Hinzen (2011) propose to align categories representing 'the basic referential dimensions of human language' with 'the three major "Phases" of recent [MGG syntax]' and to this end adapt Longobardi's (2005) model 'for the modes of nominal reference to the clausal domain,' they exemplify MGG's continued acceptance of his syntactic and semantic account of names.

Substantivism has used Longobardi's analysis since the nineties. Oversimplifying drastically for exoteric

purposes, his account in our recasting characterizes a **name** as an **indexical** noun pointing at an entry in a culturally stored code. When we wrote (Dasgupta 2012b) a comprehensive substantivist account arguing that his MGG story meets Derrida's approach to names half-way, Longobardi generously sent updated versions of his story and (without explicitly endorsing it) did not object to our retelling, which invoked one strand of Derrida's work to upscale linguistic semiotics, and which now reinvokes that strand to niche deutero-arbitrariness.

Derrida (1967) has maintained that the received 'literate/preliterate' dichotomy reflects a naive misunderstanding of the phenomenon of writing. Noting that every known society present or past gives each person a name, invoking the community's implicit register — a notional record in the culture's archive — Derrida argues that all societies are fundamentally literate. 'If one stops understanding writing in its narrow sense of linear and phonetic notation,' he writes, 'one should be able to say that all societies capable of producing, that is to say of obliterating, their proper names, and of playing classificatory difference, practise writing in general. No reality or concept would therefore correspond to the expression "society without writing". This expression is redolent of ethnocentric oneirism, abusing the vulgar, that is to say ethnocentric, concept of writing' (Derrida 2016: 118). Linguists rejecting this view will want to develop their own take on orature, cultural archives, and whether Twists W and N intersect (perhaps on the basis of the approaches to writing articulated by Coulmas 1989, 2016).

Our 2012b conceptualization sidestepped the notion of obliteration (Derrida spoke of 'producing, that is to say of *obliterating*, their proper names'), which pertained to the indigenous peoples' practice of refusing to mention one's real name (to prevent enemies from doing harm by black magic based on knowledge of this name that carries the person's essence). Our societies, however, have a very differently motivated practice formally similar to such obliteration: nicknaming. The name *Sepp* figures saliently in the Josef Bayer webschrift (Haider 2015); the fact that every Bavarian called *Josef* has the nickname *Sepp(i)* bears underscoring. A Bavarian *Josef/Sepp* and a middle American *Joseph/Joe* are different Josefs; a French *Elisabeth/Zaza* and a British *Elizabeth/Bessie* are different Elisabeths. Those formally inclined will construct a phonology-encompassing account — explaining why no *Jacob* gets nicknamed *Cobbie*, parallel to *Bessie* — and characterize the culturally differentiated classes of possible nicknames. Linguistic semiotics will confine its interest to the hypo-indexical domestication of macro-social indexicals. Whether it is akin to the aboriginal obliteration of names Derrida refers to is for anthropology to investigate.

Column writing 'merely' discharges the external responsibility of engaging with the agora in each society and connecting the discourse there to our discipline's investigations. For the latter, the upgraded semiotic toolkit empowers us also to engage theoretically with communities'

linguistic self-images, in whatever forms they currently exist. This aspect of the project linguists must now take up will look like, and partly overlap with, serious decolonization. To handle the overlap, we must coordinate with decolonial and orientalism-dismantling forms of inquiry in sister disciplines, and with projects attentive to hierarchies and fissures within the societies in question. Only so may we prevent chauvinistic identitarianisms from hijacking what needs to be a project run by linguists and led, wherever possible, by linguistics-aware literate native speakers of the languages at stake.

Experience shows that shrill modes of engagement elicit backlash and worse. To humanize the space of this negotiation, we suggest nicknaming the global north and south Septentria and Meridia[xxi], respectively, to characterize the terrains of the wide-ranging negotiation that must now articulate itself. Authors unaccustomed to humour may prefer to call them ex-subject and ex-object discourses. A linguistics of discourses, which seriously watches the ways of codes, instead of merely taking the category for granted and describing them, links the linguist's agency to the speaker's, and takes the speaker's literacy on board. Recall our adoption of Derrida's insight that all societies are conceptually literate; we maintain that, if some societies need institutional pedagogies installed, this process will call for delicate negotiation.

The point is to respond to Alisjabanah's (1965: 14–15), charge that our discipline has been irresponsible[xxii]. Where TLGs exist[xxiii], we deal with them critically, through the OCl's eye. Where they are being invented, we participate in this construction equally critically, deploying the same gaze.

The proposed negotiation becomes more, not less, delicate once we realize that the point is not about a white Septentria and a coloured Meridia alone, but that the disenfranchised communities that speak Bavarian or Irish or Catalan in the physical north belong to Meridia in intercultural terms. Thus, a Catalan-Spanish bilingual linguist becomes a major resource in the negotiation. Her divided loyalties are not a handicap, but a site of our collective labor which will have to be carried out by individuals, in part within their own mental landscape.

Readers objecting to this apparent provincialization have forgotten that TLG itself, by multi-codifying, is already provincializing English (see section 1). Our Meridial move is a conceptual sequel, calculated to add value to this fait accompli. This move keeps faith not with practicing substantivists alone, present and past (including the late Rajendra Singh, Alan Ford, and Stanley Starosta), but with early architects of the semiotic turn, including Nicolas Ruwet, Ashok Kelkar, Ray Dougherty, Michael Helke; with Gayatri Chakravorty Spivak, Josef Bayer, Giuseppe Longobardi, Richie Kayne and Noam Chomsky, who have consistently supported the pursuit of these goals across doctrine boundaries; and with inspiring figures today doing cognate work — Veena Das (2020), who reexamines the ordinary through a partly Wittgensteinian lens, thereby revising our conceptualizations; Susan F. Schmerling, whose work (Schmerling, 2018) invokes

Sapir's contribution, flagged in section 3; and Michael McGhee, who in his new book (McGhee 2021) makes obviously semiotic moves to render intelligible those domains where different religious vocabularies overlap but neither side has a vocabulary that can express the concept-sharing. Nobody brought up on logical exactitude will immediately feel comfortable with questions about just how the exactish, straight-looking bones of our armature do business with the always messy-looking flesh, blood, and brains we seem to also need.

ACKNOWLEDGEMENTS

REFERENCES

Alisjabanah, Sutan Takdir. 1965. *The Failure of Modern Linguistics in the Face of Linguistic Problems of the Twentieth Century*. Kuala Lumpur: University of Malaya Press.

Mark Aronoff. 1976. *Word Formation in Generative Grammar*, Cambridge, MA: MIT Press.

Badenoch, Nathan; Purti, Madhu; Choksi, Nishaant. 2019. 'Expressives as moral propositions in Mundari'. *Indian Linguistics* 80:1—17.

Barbiers, Sjef. 2010. 'Focus particle doubling'. Jan-Wouter Zwart and Mark de Vries (eds.) *Structure Preserved: Studies in Syntax for Jan Koster*. Amsterdam: John Benjamins. 21—29.

Barbiers, Sjef. 2014. 'Syntactic doubling and deletion as a source of variation'. M. Carme Picallo (ed.) *Linguistic Variation in the Minimalist Framework*. Oxford: Oxford University Press. 197—223.

Bayer, Josef. 2020. 'Why doubling discourse particles?'. Ludovico Franco and Paolo Lorusso (eds.) *Linguistic Variation: Structure and Interpretation: Contributions in Honor of M. Rita Manzini*. Boston/Berlin: Mouton de Gruyter. 47—72.

Bayer, Josef; Dasgupta, Probal. 2016. 'Emphatic topicalization and structure of the left periphery: evidence from German and Bangla'. *Syntax* 19(4):309—353.

Jan Blommaert. 2018. *Durkheim and the Internet: On Sociolinguistics and the Sociological Imagination*. London: Bloomsbury Academic.

Bresnan, Joan; Asudeh, Ash; Toivonen, Ida; Wechsler, Steven (eds.) 2015. *Lexical Functional Syntax*. 2nd ed. London: Wiley Blackwell.

Cinque, Guglielmo. 2010. *The Syntax of Adjectives: A Comparative Study*. Cambridge, MA: MIT Press.

Dasgupta

Coulmas, Florian. 1989. *The Writing Systems of the World*. Oxford: Basil Blackwell, 1989.

Coulmas, Florian. 2016. *Guardians of Language: Twenty Voices through History*. Oxford: Oxford University Press.

Das, Veena. 2020. *Textures of the Ordinary: Doing Anthropology after Wittgenstein*. New York: Fordham University Press.

Dasgupta, Probal. 1980. *Questions and Relative and Complement Clauses in a Bangla Grammar*. Ph.D. dissertation, New York University. [Distributed by University Microfilms, Ann Arbor; a 2020 revised version, with a preface by Richard S. Kayne, is in private circulation and will be publicly web-available after the pandemic.]

Dasgupta, Probal. 1993. 'The roots of structuralism and Loharam Shiroratna'. Alok Bhalla and Sudhir Chandra (eds.) *Indian Responses to Colonialism in the Nineteenth Century*. New Delhi: Sterling. 48—92.

Dasgupta, Probal. 2000. 'Tesnière indicators and Indian languages'. Rajendra Singh et al. (eds.) *The Yearbook of South Asian Languages and Linguistics 2000*. New Delhi/London/Thousand Oaks: Sage. 109—119.

Dasgupta, Probal. 2007. 'Look across'. Eric Reuland, Tanmoy Bhattacharya, and Giorgos Spathas (eds.) *Argument Structure*. Amsterdam: John Benjamins. 213—37.

Dasgupta, Probal. 2009. 'Characterizations in linguistic science'. Amiya Dev (ed.) *Science, Literature and Aesthetics (History of Science, Philosophy and Culture in Indian Civilization, Volume XV, Part 3)*. Delhi: Centre for Studies in Civilization. 149—165.

Dasgupta, Probal. 2010. 'Compositionality, the prose default, and poetry: a cognitive approach'. *International Journal of Mind, Brain and Cognition* 1(2): 55—76.

Dasgupta, Probal. 2011. 'Agreement and non-finite verbs in Bangla: a biaxial approach'. Rajendra Singh and Ghanshyam Sharma (eds.) *Annual Review of South Asian Languages and Linguistics 2011*. Berlin/Boston: De Gruyter Mouton. 35—48.

Dasgupta, Probal. 2012a. 'Rephrasing the question of complex predicates in Bangla: a biaxial approach'. Rajendra Singh and Shishir Bhattacharja (eds.) *Annual Review of South Asian Languages and Linguistics 2012*. Berlin/New York: Mouton De Gruyter. 3—44.

Dasgupta, Probal. 2012b. *Inhabiting Human Languages: The Substantivist Visualization*. New Delhi: Indian Council of Philosophical Research/Samskriti. [The book is out of print. Full text available at: https://wp.me/p2Aw52-3M.]

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

Dasgupta, Probal. 2018. 'Whole Word Morphology reloaded: the case for a semiotic turn'. *Język Komunikacja Informacja* 13:188—211.

Dasgupta, Probal. 2019. 'Clause particles and cleft sentences in Bangla: some preliminary generalizations'.

Josef Bayer and Yvonne Viesel (eds.) *Proceedings of the Workshop Clause Typing and the Syntax-to-Discourse Relation in Head-Final Languages*. 51—67. https://kops.uni-konstanz.de/bitstream/handle/123456789/47461/Bayer_2-kqn8xj9d41c69.pdf sequence=1&isAllowed=y , accessed on 15 May 2020.

Dasgupta, Probal. 2020. 'Riding the rhythm: some reflections on alternating indexicality'. Paper presented at Jadavpur University, 2020. To appear in Samir Karmakar (ed.) *Proceedings of International Conference cum Workshop on Rhythm in Speech and Music from Neuro-Cognitive Perspectives*. Available at https://wp.me/p2Aw52-3S, accessed on 26 August 2021.

Dasgupta, Probal. (Forthcoming). 'Compositionality and Kamal Majumdar's prose'. To appear in Aditi Ghosh (ed.) *Language, Communication and Conflict in South and Southeast Asia*. Available at https://wp.me/p2Aw52-3W, accessed on 26 August 2021.

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

Derrida, Jacques. 1967. *De la grammatologie*. Paris: Minuit, 1967.

Derrida, Jacques. 2016. *Of Grammatology.* Tr. by Gayatri Chakraborty Spivak. 40th anniversary edition. Baltimore: Johns Hopkins University Press.

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

Gouet, Michel. 1976. 'On a class of circumstantial deletion rules'. *Linguistic Inquiry* 7(4):693—697.

Haider, Hubert. 2015. 'Wià effdà daß mà nochdenggd — Koàn bessàn weàds need findn!' [in Bavarian; 'The more one comes to think of it — you won't find a better guy!'] Ellen Brandner, Anna Czypionka, Constantin Freitag, and Andreas Trotzke (eds.) *Webschrift for Josef Bayer*, online, at https://ling.sprachwiss.uni-konstanz.de/pages/WebschriftBayer/2015/contents_files/Haider.pdf. 72—73. (Accessed on 9 April 2020.)

Jana, Devarati. 2021. *The Syntactic Derivation of Bangla Infinitivals*. Calcutta University Ph.D. dissertation.

Karcevskij, Sergej. 1931. 'Sur la phonologie de la phrase'. *Travaux du Cercle Linguistique de Prague* 4:188—227.

Kayne, Richard S. 2005. *Movement and Silence*. Oxford: Oxford University Press.

Langacker, Ronald W. 1987—1991. *Foundations of Cognitive Grammar, Vol. I: Conceptual Prerequisites* (1987); *Vol. II: Descriptive Application* (1991). Stanford: Stanford University Press.

Longobardi, Giuseppe. 2005. 'Toward a unified grammar of reference'. *Zeitschrift für Sprachwissenschaft* 24: 5—44.

McGhee, Michael. 2021. *Spirituality for the Godless: Buddhism, Humanism, and Religion*. Cambridge: Cambridge University Press.

Massam, Diane. 2017. 'Extra *be*: The syntax of shared shell-noun constructions in English'. *Language* 93(1): 121—152.

Merriam-Webster. 2003. Merriam-Webster's Collegiate Dictionary. 11th edition. Springfield, MA: Merriam-Webster.

Morin, Yves-Charles. 1977. 'Re: On a class of circumstantial deletion rules'. *Linguistic Inquiry* 8(4):747-51.

Nelson, Cecil L.; Proshina, Zoya G.; Davis, Daniel R. (eds.) 2019. *The Handbook of World Englishes*. 2nd edition. London: Wiley.

Oxford English Dictionary Team. 2020. 'Would you like to see World English recorded in the OED?' https://linguistlist.org/issues/31/31-1285.html, accessed on 8 April 2020.

Pal, Palash Baran. 2020. *Kathār Kathā* [in Bangla; 'words on words']. Kolkata: Parampara.

Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace and Company.

Schmerling, Susan F. 2018. *Sound and Grammar: A Neo-Sapirian Theory of Language*. Leiden: Brill.

Schreier, Daniel; Hundt, Marianne; Schneider, Edgar W. (eds.) 2020. *The Cambridge Handbook of World Englishes*. Cambridge: Cambridge University Press.

Sheehan, Michelle; Hinzen, Wolfram. 2011. 'Moving towards the edge'. *Linguistic Analysis* 37(3-4):405-58.

Singh, Rajendra; Starosta, Stanley; Neuvel, Sylvain. 2003. *Explorations in Seamless Morphology*. New Delhi/London/Thousand Oaks: Sage.

Smolin, Lee. 2006. *The Trouble with Physics.* Boston: Houghton Mifflin.

Tirumalesh, K.V. 1991. 'Why you can't repunch Bill: An inquiry into the pragmatics of '*re*-words' in English'. *Journal of Pragmatics* 16:249—267.

Weydt, Harald. 2006. 'What are particles good for?'. Kerstin Fischer (ed.) *Approaches to Discourse Particles*. New York: Elsevier. 205—211.

Williams, Jeffrey P. 2019. *The Aesthetics of Grammar: Sound and Meaning in the Languages of Mainland Southeast Asia*. Cambridge: Cambridge University Press.

Woolf, Virginia. 1925. *Mrs Dalloway*. London: Hogarth.

Woolf, Virginia. 2003. 'The strange Elizabethans'. Virginia Woolf. *The Common Reader: Volume 2*. Ed. Andrew Macneillie. [1st ed. London: Hogarth, 1932]. 9—23.

ENDNOTES

i The abstract of Massam's paper says that it 'examines the syntax of Extra *Be* constructions, common in nonprescriptive English and often considered a curiosity, such as: *The problem is, is that she hates apples.*' Bayer's paper flags utterances with two occurrences of the same discourse particle — here is an example where the particle *denn* is doubled: *Bitte erleuchten Sie mich, was denn sind denn die Pyramiden und auf welchen Forschungen beziehen Sie ihr Wissen?* (p. 56, 'Please enlighten me, what are pyramids after all, and from which research do you derive your knowledge?'). Readers unaware of the nature and functioning of discourse particles may consult Weydt 2006 for an exoteric account.

ii Cinque (2010: 74), writes: 'The national Romance languages (Italian, French, Catalan, Spanish, Portuguese, Romanian), despite certain limited differences,[4] are remarkably similar in this respect. Greater variation is found when one takes dialects, or older stages, of these languages into consideration' (his note 4, on p. 136, unpacks the 'limited differences'). Furthermore, Cinque states (p. 135, n 1) that certain characteristics of Romance occur only in a '*very formal style*' (emphasis ours).

iii Our version of the Geneva system, a compromise between transliteration and transcription, uses short *a* for a sound pronounced as [ɔ] in Bangla, the digraph *ng* for [ŋ], and, as will be seen from example (7) onwards, *è* for [æ].

iv To be sure, MGG also explores delicate grammaticality issues that involve native speaker judgments regarding moderately vs egregiously ill-formed sentences, thus going far beyond TLG's domain on **that** front. In the text, we slightly overstate the TLG-hugging agenda to make our point.

v An anonymous reviewer wonders if this statement does hold for all forms of generative grammar. The term 'MGG' is not intended here as a cover term for everything articulated by all soi-disant generative grammarians in every land and clime. MGG is the coalition that speaks for formal linguistics as a whole in the agora. Practitioners of particular variants of generative grammar who wish to claim a share of the substantivist pie are welcome to do so.

vi Its *esoteric* designation, in the Distributed Morphology framework, is "Vocabulary Item" and is approximately equivalent to "morpheme". In substantivist linguistics, we do not chop the word up into grammatical segments, but only into phonological sounds, syllables and other prosodic units. Of this, more anon.

vii Readers unconvinced about Hindi-Urdu *vālā* and Bangla *wālā* are welcome to ponder *X-er* words in English and try to summate the precise semantic contributions made by *er*: *dishwashers, can-openers; pageturners, thrillers; farmers, miners* (who get paid)*; boarders, lodgers* (who pay)*; diners* (where paying happens)*; fortyniners, women's-libbers; fivers, tenners; astrologers, geographers; photographers; biographers*.

viii https://www.poetryfoundation.org/poems/44900/an-essay-on-man-epistle-ii, accessed on 14 May 2020.

ix https://www.gutenberg.org/files/1041/1041-h/1041-h.htm, accessed on 14 May 2020.

x An anonymous reviewer notes that formal linguists (especially Platonists) who distance themselves from MGG's psychologism should not be expected to comment, qua linguists, on what individual readers perceive or believe. I agree; but my remarks in note 5 carry over to this objection as well. Furthermore, the substantivism-Platonism engagement, which goes back to the seventies, led to substantivist insights whose latest form appears in Dasgupta (2009); for earlier articulations, see Dasgupta (1993), and references therein. A Platonist who thinks her

doctrine through should be attracted to some form of semiotics and may wish to engage with current substantivist proposals.

xi Surely the unique salience of the 'word' unit — a key element even in MGG accounts — is a semiotic matter; but we choose not to press the point.

xii The poltergeist story is narrated in Dasgupta (2019). That paper presupposes the analysis of clause particles in Bayer and Dasgupta (2016).

xiii The gestural medium used in Sign Language is a salient example.

xiv The phrase *the tall girl* associates *girl* with its potential substitute *boy* on the **paradigmatic** axis (of selection), and links *girl* to its fellow constituent *tall* on the **syntagmatic** axis (of combination).

xv An anonymous reviewer requests us to comment on lexicon-free architectures. MGG in its default settings does not work with such architectures, but tolerates them as variants. If their users ever find them outperforming substantivist machinery, then reports of such discoveries, with empirical details, will be helpful for our plans for self-improvement.

xvi Normative accents require *è*l/*əktrísiti* with initial (and secondary-stressed), *è*, neutral ə, primary-stressed *í*; we are cutting corners here.

xvii Within morphology, some phenomena are more capricious than others, and the phonology does get a word in edgewise after all. For discussion in recent, fully semiotics-equipped substantivist terms, see Dasgupta (2018, 2020).

xviii The status of methods in linguistics should not depend on where the chips fall in certain debates about the proportions of biological and cultural factors determining core properties of human languages. The point is to delimit exact patterns wherever they occur, to use relevant types of mathematical and other tools to characterize them, and to remember that not every exact discipline is a natural science: music is just as mathematical.

xix An anonymous reviewer has requested a philosophy-minimizing summary of section 4. Readers interested in parsing section 4 who know the pertinent issues in semantics but are put off by some expository characteristics of this section may find it useful to consult Dasgupta (2012b); it provides guidance, at some length, to readers not familiar with the philosophical style.

xx When linguists are too busy to even write a language column, a gifted physicist familiar with basic linguistics steps into the breach (Pal 2020).

xxi These pickwickian region names are back-formed from *septentrial/meridial*, which in turn are based on the French (and archaic English) words *septentrional* 'northern' and *meridional* 'southern.'

xxii Alisjahbana (1965: 14–15) writes: 'What they [= the leaders of new nations] need is not descriptive, but prescriptive linguistics. It is thus very regrettable that precisely in these processes and problems that are crucial for the languages of the developing societies, processes and problems which can be formulated succinctly in the terms standardization and modernization, modern linguistics, through its static, formal and micro approach, is least able to contribute.'

xxiii An anonymous reviewer has requested references specifying what we mean by TLG. The reviewer seems to believe that some readers may imagine that TLG may include pre-generative forms of linguistics. That would be a serious misconstrual. TLG is our term for the society-specific apparatus which, in every literate society, produces the authoritative dictionaries as well as those normative (including the pedagogic) grammars that lead the learner up the ladder to the point of being able to make optimal use of unabridged dictionaries for her writing and editorial labor at the highest level of proficiency. The TLG enterprise does not reflect the views of any school of linguistics, present or past. To be sure, particular individuals have occasionally played a dual role, working both as practitioners of linguistics and of TLG. This empirical fact, however, need not distract us from the principles that drive TLG and MGG as institutions. In many Meridial societies, TLG has yet to be built, and we argue that linguists should help in the labor of this construction.