

Head Movement Before and After Spellout

Abstract. This paper argues that head movement must exist in the syntactic derivation, contra current opinion in Minimalist thinking. I show that in Bangla, there is movement of the minimal classifier element to the left edge of the nominal that carries semantic implications, resulting in the Quantificational Approximateness construction (QA). Moreover, I show that this operation enters into differential (counter)feeding relationships with a noun-to-classifier movement operation, argued for classifier languages generally in Simpson (2005). Specifically, I show that there is noun-to-classifier movement of unit terms which subsequently feeds the construction of QA, but is counterfed for non-unit terms. I take this to suggest to mean that there are displacement operations targeting minimal elements that occur in the narrow, overt syntax, as well as after SPELLOUT.

1. Introduction

Chomsky (2001) claims that head movement may be distinct from the operations available in the overt, narrow syntax. This is motivated primarily by theory-internal concerns. For one, on standard conceptualizations of head movement, a head adjoins to another, creating a complex head in the process. Within a Bare Phrase Structure approach, there is no coherent way of stating that a head is *complex*, since all projection levels are derived, relational notions. That is, if a head is defined as the smallest projection of a lexical item, then there is no way to distinguish an “intermediate” X^0 from an X' . Additionally, head movement by its nature seems to violate most notions of cyclicity in the Minimalist Program, most notably the Extension Condition (Kitahara, 1995; Watanabe, 1995; Boeckx & Stjepanovic, 2001; Chomsky, 1995, 2001; Hornstein, 2009). Thus, being able to attribute head movement effects to other components of the grammar would be a welcome outcome, and one consistent with the general Minimalist approach. In fact, there

I would like to thank Aniqua Hasan, Samira Choudhury, Nandita Rahman, Satarupa Das, Priyanka Biswas and others for being patient informants. Also thanks to Hooi Ling Soh, Jean-Philippe Marcotte, Brian Reese, Jeanette Gundel, Howard Lasnik and Colin Phillips for comments on previous versions of this paper. All other errors are mine.

have been a number of attempts at reducing head movement effects entirely to the phonological component (Boeckx & Stjepanovic 2001 is representative) or to an interaction between the syntax and an interpretive morphological component (Matushansky, 2006). An alternative approach has been to suggest that head movement is actually compatible with constraints on the overt syntax after all, and that we have mischaracterized head movement, such as Bobaljik & Brown's (1997) formulation of head movement as a subspecies of Nunes's (2001) sideways movement. Most of the discussion has been focussing on how head movement could *in principle* be possible given theoretical technology, or whether it needs to be delegated to a post-syntactic interface. However, there has been little discussion as to whether there is empirical evidence that firmly places head movement effects at a particular point in the derivation. It has been shown that there is indeed head movement (or head movement-like effects) *somewhere*. It has been argued that we need head movement *at least* on the PF-branch of the derivation (Boeckx & Stjepanovic, 2001; Merchant, 2002).

However, there has been very little investigation into whether the theoretically-motivated abolishment of head movement from pre-SPELLOUT syntax is well motivated empirically, or whether there are any phenomena that are best explained by making reference to head movement pre-SPELLOUT. Indeed, much of the discussion has relied on showing that head movement has semantic effects (Pollock, 1989; Höhle, 1992; Hale & Keyser, 1993; Baker, 2009; Benedicto, 1997; Lechner, 2007). However, I do not believe that this is enough. It is possible that there could be some feature on a lexical item that is interpretable to the semantics, but then induces head movement after SPELLOUT, or that the relevant constructions are confluences of multiple operations – some morphological, and some in the narrow syntax. A true defense of head movement as an operation available to the pre-SPELLOUT syntax should rely specifically on ordering relations between operations that are (at first blush) syntactic. This is the goal of this paper.

First, I will show evidence given originally Chacón (2009) that the quantificational approximateness (QA) construction in Bangla must occur in the overt syntax, because it has obvious se-

semantic consequences and because of its interactions with other elements in the syntax.¹ Secondly, I will further show that a certain, semantically-defined class of nouns substitute into the classifier position in the Bangla DP, and that this head-substitution feeds the QA construction. I will refer to this class as self-individuating nouns, for semantic reasons that will become evident shortly. I take this to indicate that this substitution must also occur in overt syntax, given this feeding relationship. Finally, I will suggest that there is another class of nouns that show similar diagnostics, but that do not seem to incorporate in the narrow syntax as indicated by counterfeeding the QA movement. Using Embick & Noyer's (2001) framework, I will suggest that these facts are due to word formation post-Linearization which, by hypothesis, is in the post-SPELLOUT branch of the derivation. This will suggest that there are indeed head movement(-like) operations in the post-syntactic morphological components, per Chomsky's (2001) suggestion. The feeding and bleeding relations will suggest that, contra current trends in Minimalist syntax, there are head-movement/word-formation operations in the narrow, overt syntax, in addition to the PF-branch, argued for on syntactic terms.

2. Classifier-incompatible nouns and the QA Construction in Bangla

In this section, I briefly describe the properties of the QA construction and the proposed derivation. I also describe the patterns of substitution, and demonstrate the feeding and counterfeeding relationships alluded to earlier.

2.1 The syntax of the QA construction

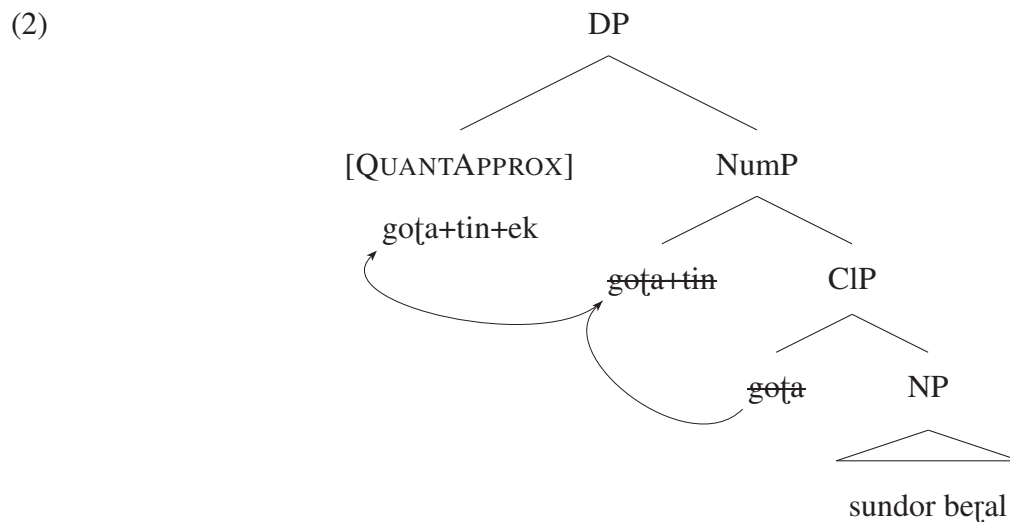
Bangla is a language that uses numeral classifiers, unlike most of the languages in the South Asian Sprachbund. One of the constructions that has puzzled a number of syntacticians is what I call the quantificational approximateness (QA) construction, demonstrated in (1).

- (1) a. tin-ʈe beʈal
 three-CL cat
 'three cats'

¹Although Chacón (2009) give an account for this using a head movement operation, it has been argued to be derived via phrasal movement in Tanmoy Bhattacharya's work Bhattacharya (1999b,a, 2001). Either way, due to its semantic conditioning, we can be certain whatever operations result in this construction must be before SPELLOUT. Bhattacharya's view does not obviously account for any of the feeding or bleeding relationships I will focus on here.

- b. goṭa-tin-ek beṭal
 CL-three-EK cat
 ‘three or so cats’

This construction is apparently derived by “swapping” the positions of the classifier element (CI⁰) and the numeral (Num⁰), with an affixed *ek* morpheme at the end of the morpheme cluster. This construction expresses an uncertainty or vagueness on the quantity of the noun. Chacón (2009) previously suggested that these constructions are derived via a successive head movement operation that adjoins a CI⁰ to a Num⁰, and then subsequently adjoins Num⁰ to D⁰. On this account, the distinct phonological shape of the classifier in the QA construction and the addition of the *ek* morpheme are indicative of the features of the drive the movement on this analysis. This is schematicized in (2).²



I hypothesize that these constructions are head movement operations in the *narrow, overt* syntax, since there is a semantic reflex of the movement, as indicated in the gloss.

How do we know that this construction is syntactically derived, and that *goṭa-tin-ek* is not an individual word in the syntax? It is not incoherent to claim that there could be a lexical rule that

²Similar claims have been made about the literature describing a similar phenomenon in Russian, called the Approximate Inversion. See Yadroff & Billings (1998) for an account similar to the analysis proposed here. Also, note that this account requires that D⁰, Num⁰ and CI⁰ be distinct heads, and that that Bangla has a D⁰ position. A similar syntactic spine for classifier languages has been proposed by Tang (1990).

takes a numeral and outputs an “approximate” numeral. One reason why this kind of approach is unsatisfying is that, despite the apparent phonological distinction between *goṭa* and the *-ṭe* clitic CI^0 in (1), there is a consistent mapping of the form of the post-numeral classifier and the left-most morpheme in the morpheme cluster in the QA construction, when the appropriate context for that classifier obtains. This is shown in (4). Here, when the human classifier *jɔn* is licensed, *jɔna* appears in the QA construction, and similarly, the flat, square object classifier *khana* appears in both positions when licensed.

- (3) a. *tin-jɔn sromik*
 three-CL laborer
 ‘three laborers’
- b. *jɔna-tin-ek sromik*
 CL-three-CL laborer
 ‘three or so laborers’
- (4) a. *tin-khana boi*
 three-CL book
 ‘three books’
- b. *khan-tin-ek boi*
 CL-three-EK book
 ‘three or so books’ (Dasgupta, 1983)

Encoding this information both in a lexical rule that generates the hypothetical approximate numerals and in a syntacticosemantic constraint that ensures their co-occurrence would be redundant, whereas an approach that forms the QA construction via a transformational rule would need no additional mechanisms. Note that the fact that the phonological form varies based on whether it cliticizes to the right or left side of a numeral would need to be encoded in either case.

Additionally, I claim that the CI^0 incorporates into the D^0 position. This is supported by the fact that the morpheme *ek* in the QA constructions seen above is phonologically similar (if not homophonous) with the indefinite article/numeral *æk* ‘one’. In this proposal, this correspondence

is not an accident – these are both the same *Ek* morpheme abstractly.³ That is, I would predict that if another language had a similar construction, we should see a homophony between the final morpheme in the QA cluster formed at the left edge of the DP and that language’s indefinite article. Since, all things being equal, we should accept a more brittle and falsifiable theory, I take this as reason to suppose a syntactic approach to these constructions over one in which this correpondence between the indefinite article and the QA morpheme is accidental.

Note that the analysis given so far does not necessarily entail any kind of head movement *per se*. As suggested in the Introduction, this could be done via a system of phrasal movements and covert movements. In the next section I will illustrate that nouns that do not co-occur with classifiers in the general case feed the QA construction differentially, and that the analysis proposed here can capture this asymmetry in a principled way. It is not clear how an analysis that rejects pre-SPELLOUT movement could capture these systematicities without stipulation.

2.2 *Classifier-Incompatible Nouns*

Simpson (2005) notes that many classifier languages have a class of nouns that do not normally co-occur with a classifier. This is exemplified by Thai in (5). This is especially illustrative, given that the canonical position for nouns precedes the numeral-classifier complex in Thai.

- (5) a. soon pii
two year
‘two years’
- b. saam khrang
three time
‘three times’
- c. sii khon
four person
‘four people’
- (Thai)

³Though, something else needs to be said to account for the apparent singular interpretation that is otherwise concomittant with the indefinite article, but absent here.

In order to account for these facts, Simpson proposes that there is N^0 -to- Cl^0 movement in these cases, which I will call *Simpson's Conjecture*. Simpson suggests that the data in (7) are consistent with this generalization if we accept a copy theory of movement. That is, if there is indeed N^0 -to- Cl^0 movement, then we might expect that, in some languages, the lower copy of the N^0 might be left over. If this were to occur, we should expect to see two copies of the noun root – one *in situ*, and one adjacent to the numeral, by hypothesis in the Cl^0 position. This prediction is borne out in some Thai and Burmese constructions, shown in (7).⁴

(6) **Simpson's Conjecture:** A noun that does not co-occur with an overt classifier has undergone movement to the Cl^0 position.

- (7) a. hoong [_{NumP} saam hoong]
 room three room
 'three rooms' (Thai)
- b. cun [_{NumP} ta cun]
 island one island
 'one islands' (Burmese)

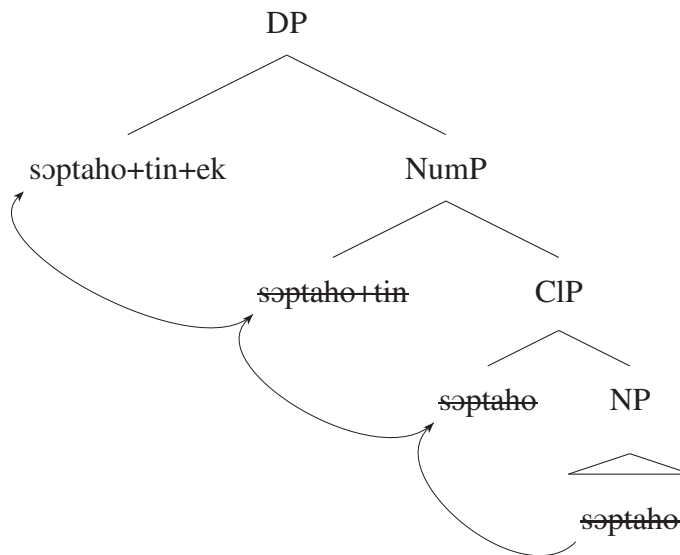
Suppose that, in classifier languages, this is true for all bare nouns that do not co-occur with a classifier. If this were the case, and if the analysis of the QA construction discussed above is on the right track, we predict that the noun root should appear in the left edge of the DP, in the Cl^0 position. This prediction is borne out, shown in (8). Here, *sɔptaho* does not – under normal circumstances – co-occur with a classifier. Supposing that the noun (or the root) first moves to the Cl^0 position, and then there is a later movement of the Cl^0 element to the D^0 position to adjoin to *ek*, we predict that noun should not remain *in-situ*. The derivation is given in (9).

- (8) a. tin(-[?]tɛ) sɔptaho
 three-CL week
 'three weeks'

⁴Note that this assumes that heads leave traces at all, suggested to not be the case in Travis (1991); Lasnik (1998); Omaki (2009), but see Koopman (1984) for empirical evidence that implies the contrary.

- b. sɔptaho-tin-ek
 week-three-EK
 ‘three or so weeks’

(9)



Thus, if Simpson’s Generalization is accurate, we have evidence that N^0 incorporates into Cl^0 . This demonstrates a number of things. First, it suggests that if the QA construction is derived in the overt syntax, then so is N^0 -to- Cl^0 movement, since the N^0 -to- Cl^0 movement *feeds* the QA construction. This is exactly what we would predict if both of these operations occurred in the narrow, overt syntax, since syntactic derivation is assumed here to be built bottom-up and cyclically, whereas there is no need for any aspect of the phonological component to have these properties. The apparent feeding of the N^0 -to- Cl^0 movement to the QA construction is also to be expected. If, say, there were a lexical feature that would cue the D^0 to attract the Cl^0 in a post-syntactic component, we might not expect there to be any kind of feeding of the nominal root. Secondly, as I will argue in section 3, the N^0 -to- Cl^0 movement may actually be motivated by semantic factors, again suggesting that attributing this effect purely to a morphophonological structure may not capture all the generalizations, or, at least, not give us the most restrictive theory. Additionally, this data shows that the QA construction is unlikely to be derived via a lexical rule, since doing so would require the derived hypothetical “approximate numeral” to have some method of seeing what nominal root it would co-occur with in order to have that root appear in the left edge of the morpheme

cluster. In other words, how would the lexicon “know” to generate N^0 adjoined to Cl^0 before they have merged in the syntactic derivation? Thus, deriving the QA via a lexical rule would cue a look-ahead problem. Moreover, its successive cyclic nature and morphological output suggest that it is *head* movement, specifically. Furthermore, the N^0 -to- Cl^0 movement that feeds it must also be a movement in the overt syntax, since it feeds the QA construction. By hypothesis, this must be head movement as well, in order to avoid a Structure Preservation Condition violation (Emonds, 1976). I address two other analyses – merging no Cl^0 at all, and treating these nominal roots as “intransitive” Cl^0 s – in section .

By these diagnostics, not all nouns that occur bare obviously are compatible with Simpson’s Generalization. For instance, consider the noun *caka* ‘wheel’. In (10), we see that it is marked for *caka* to occur with a classifier, similar to *səptaho* ‘week’ in (8). However, in the QA construction, *caka* does not piggy-back with Cl^0 to the left edge of the morpheme cluster in D^0 , and we get an overt classifier form *goʔa*, as in the case of unmarked nouns. I will suggest that this is because there is a post-SPELLOUT operation that moves the noun to the classifier position if they are adjacent in linear order after Linearization – that is, after SPELLOUT. This is shown in (11), to be discussed in more detail later.

- (10) a. tin-^ʔte caka
 three-CL wheel
 ‘three wheels’
- b. *caka-tin-ek
 wheel-three-EK
- c. goʔa-tin-ek caka
 CL-three-EK wheel
 ‘three or so wheels’

- (11) [Cl [N]] → [Cl ⊕ N] when N is $\sqrt{\text{NATION}}$, $\sqrt{\text{WHEEL}}$, $\sqrt{\text{DIRECTION}}$, $\sqrt{\text{SIDE}}$, $\sqrt{\text{BROTHER}}$,
 or ...

When CI^0 moves before SPELLOUT, and *a fortiori* before Linearization, the post-SPELLOUT N^0 -to- CI^0 raising rule is bled. In this case, the N^0 -to- CI^0 movement occurs after the QA construction is formed, and thereby cannot feed the CI^0 -to- D^0 movement, unless the nouns just presented.

Next, I will attempt to motivate *why* nouns like *sɔptaho* ‘week’ move to the CI^0 position in overt syntax. Specifically, I will propose that this class of nouns is a principled natural class – namely, they are nouns that, by their lexical meaning, do not need to be individuated by an external classifier. I will propose some reasons why the grammar might treat these nouns the way they do, all by way of motivating that these nouns are moving for some formal syntacticosemantic reason, consistent with the approach that these nouns are moving to the CI^0 position in the pre-SPELLOUT syntax. Then, I will turn my attention to the other CI^0 -resisting nouns, like *caka* ‘wheel’, which do not meet the semantic criteria for the former set of nouns. I will suggest also that these nouns do not form any kind of natural class, suggesting any operation that targets these roots specifically will have to be formulated in terms of the root identities. If this is the case, we expect – following [Embick & Noyer \(2001\)](#) – that this operation must follow SPELLOUT.

3. An Excursus into Semantics – Self-Individuating Nouns

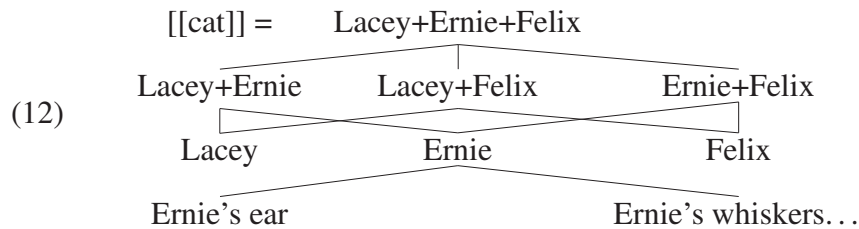
First, I would like to reflect on why the nouns that incorporate into the CI^0 in the pre-SPELLOUT syntax do so, and why the nouns that do so are the nouns that do so.⁵ Essentially, I intend to motivate Simpson’s Conjecture, at least for nouns that incorporate before SPELLOUT. I do this by cashing out [Dasgupta’s \(1983\)](#) intuition that – in Bengali, at least – these nouns are by and large “measuring words”. I will concentrate specifically on time span denoting nouns, such as *din* ‘day’ and *sɔptaho* ‘week’, but these facts also seem true for other nouns, such as *inc* ‘inch’ and *fuṭ* ‘foot’. This should also make predictions about other nouns that might have the individuation conditions as part of the meaning.

First, I will follow much work in the literature on classifier languages in supposing that, abstractly, classifiers serve some computational function that enables an individuated or count refer-

⁵In the process of doing so, I’ll be inadvertently modifying the theory of plurality and the mass/count distinction. A lot of what I say here is a faint echo of [Bale & Barner \(2009\)](#) and [Bale & Khanjian \(2009\)](#).

ence (Chierchia 1998a,b; Doetjes 1996; Simpson 2005; Borer 2005, *inter alia*). A proposal along these lines captures the apparent lack of any language that is able to use a numeral classifier and count morphology in the same construction⁶, if classifiers and plural morphology are different realizations of the same syntactic head. The function that it serves is something like individuating an otherwise unstructured reference into discrete, countable units. One could make the analogy that nominal roots are like a mass of cookie dough, and the classifier and number morphologies are a cookie cutter for stamping out appropriately shaped and individuated units.

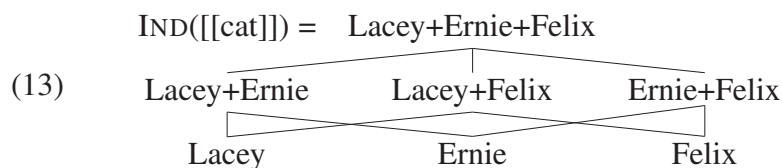
One way to formalize this intuition would be to suppose that nouns, in the general case, denote structured join semi-lattices, in line with the tradition of Link (1983). Similar to Link and Chierchia (1998a; 1998b), I suppose that the lattice consists of a set of individuals at some rung in the lattice, and that there are composed plurals constituted by the sums of these individuals. I suppose that the set of individuals is not the bottom rung of the join lattice in the general case. Instead, each individual is itself composed of (perhaps non-discrete) subparts, with no commitment to contiguity or consistency of the individual subparts. Thus, the “rung of individuals” might actually be a rung of sets which are labeled as individual by virtue of non-linguistic world knowledge. Furthermore, suppose that – again, in the general case – which rung is the rung of individuals is available to computation. That is, there might be some function that intensionally references the individuals of a predicate, regardless of what exactly the composition of the relevant entities are. Thus, an example of a denotation, in this approach, is given below.



With this kind of a denotation, we can trivially define a classifier as containing a function IND,

⁶Though see Nakanishi & Tomioka (2004) for examples that suggest exactly this is possible in Japanese, in which classifiers and the plural marker *tati* may co-occur.

a function from predicates to the set of individuals and the sums composed of those individuals.⁷ This essentially rules out the subindividual parts, forcing a non-mass reading. This is demonstrated in (13).



Unpacking this function a bit, we might define a classifier as returning only the “whole” elements in the denotation of its predicate argument, where some x is “whole” iff it is atomic (defined lexically), or any of its subparts are whole. By mathematical induction, then, every “rung” of the join semilattice from the atoms on up are classified as “whole”.

$$(14) \quad [[\text{CI}]] = \lambda P. \lambda x. x \in \{x \mid \text{Whole}(x) \wedge P(x)\}$$

$$(15) \quad \forall x[\text{Whole}(x) \leftrightarrow \text{Atom}(x) \vee \forall y \leq x[\text{Whole}(y)]]$$

Using a massifier or not merging a classifier (or perhaps using [Pelletier’s \(1975\)](#) “Universal Grinder” in a language like English) makes all subindividual entities available for reference, allowing a mass interpretation. Broadly accepting the position that classifier selection (the cookie cutter shape) is non-asserted meaning (such as in [McCready’s \(2009\)](#) approach), I will assume that the classifier restrictions are presuppositions. One way of writing the meanings of the Bangla classifiers, then, is given below.

$$(16) \quad [[\text{jɔn}]] = \lambda P : \forall x[P(x) \rightarrow \text{Person}(x)].\text{Ind}(P)$$

$$(17) \quad [[\text{khana}]] = \lambda P : \forall x[P(x) \rightarrow \text{Flat}(x) \wedge \text{Square}(x)].\text{Ind}(P)$$

$$(18) \quad [[\text{go}[\text{a}]]] = \lambda P : \forall x[P(x) \rightarrow \text{Flat}(x) \wedge \text{Square}(x)].\text{Ind}(P)$$

⁷This is distinct from Chierchia’s PL, since it is meant to include the the atomic units. Thus, we can describe why English plural morphology only works with multiple reference in the general case (**one cats*), whereas having a classifier with the numeral *one* is the norm in classifier languages.

Now, we are equipped to consider why nouns like *sɔptaho* ‘week’ move into the Cl^0 position, and why Simpson’s Conjecture may be accurate for these kinds of nouns crosslinguistically. I propose that nouns like *sɔptaho* ‘week’, by virtue of being nouns that denote *spans of time*, cannot have any rungs below the rungs of individuals. That is to say, there is no subpart of a week (or day, month, year, inch, foot, so on) that can be truthfully referred to as a week. In other words, Dasgupta’s (1983) “measuring words” are nouns such that IND applies vacuously:

(19) An N is a *self individuating noun* iff $[[N]] = \text{IND}([[N]])$.

The intuition that the individuals in the extensions of these nouns form the bottom rung is corroborated by their resistance to the Universal Grinder. This is because, on the proposal defended here, these nouns are by their lexical meaning already “count”, with no possible subparts to support a mass interpretation. Thus, the classifier only serves a vacuous semantic function.

(20) #We spent a lot of day at the Bagel Place.

(21) #I spent a lot of week in Dhaka.

At this point, we might consider that what I have been calling N^0 -to- Cl^0 movement may not necessarily be movement at all. That is, perhaps for some nominal predicate P , $\text{IND}(P)$ and P are different (enough) categorically to be distinguished by syntactic rules. That is, the structure contains a Num^0 directly selecting for an NP, without any movement to N^0 -to- Cl^0 movement. If this is the case, then one could suppose self-individuating nouns share the same type as $\text{IND}(P)$, either by virtue of their meaning or via some lexical rule. If we suppose that the QA construction is only available with individuated predicates, we could propose that self-individuating nouns are classifier-incompatible simply because there is no reason to merge a classifier, given standard Economy considerations. It could be the case that, since QA is modulating aspects of quantification, attempting to form this kind of meaning with a non-individuated predicate might result in semantic anomaly, and thus classifiers are in some respect a Last Resort option to appropriately

unitize a potentially massy predicate for quantification. After all, it is likely no accident that the QA construction is derived via movement of the *classifier* element, and not some other projection, I wager. In this case, we would be saying that the operations underlying QA do not care about the syntactic identity of the element being moved, but rather whether they have the relevant features for expressing a QA-type meaning, and that classifiers only are merged to license an individuated interpretation when no individuated interpretation would otherwise obtain.

I think that, in theoretical terms, this approach is attractive. However, I will continue to follow the proposal that we are observing syntactic head movement from the N^0 to the Cl^0 positions. This is the most plausible move, in my view, since we can maintain Simpson's copy theory-based explanation of the facts in (7). That is, on an analysis in which self-individuating nouns never are merged with a classifier, the fact that we have two copies surfacing in (7) is completely unexpected. This suggests that there actually is *movement*, at least in Thai and Burmese, which would be lost on an analysis in which no Cl^0 projection appears due to semantic incompatibility. If we must maintain at least a movement account of these facts in order to explain the Thai and Burmese data, the maximally simple approach for Bangla – and in general – would be to suggest that these types of structures are always driven by movement. Anything different would place a seemingly arbitrary and difficult burden on the child acquiring the appropriate syntax for these constructions. In other words, the Thai and Burmese facts coupled with general concerns about Occam's Razor and acquisition strongly imply Simpson's Conjecture should be universal, conditioned on the presence of classifiers.

Another option springs to mind. I have ruled out that these constructions are formed by merging Num^0 and NP, and argued that there must be movement of N^0 to an intervening Cl^0 . Alternatively, it might be the case that *sɔptaho* 'week' and other unit terms *are* classifiers, and do not enter the derivation as nouns. This too seems like a plausible move. In fact, crosslinguistically, this may be the case. Masahiko Takahashi (p.c.) points out that in Japanese, unit terms do seem to be classifiers, and not nouns, given they may select for nouns:

- (22) san-shuu no ryokou
 three-week GEN journey
 ‘three week(-long) journey’ (Japanese)

I believe there is a confound with this kind of data, however. In Japanese, the “linker” *no* appears between the numeral-classifier complex and the noun it is construed with. This is also homophonous with the genitive particle. In Bangla, however, no such genitive marker is used with prototypical numeral classifier constructions, nor is there any such linker. To express this kind of meaning, however, Bangla must (exceptionally) use a possessive construction. I take this to suggest that neither *shuu* ‘week’ or *sɔptaho* ‘week’ are classifiers after all, at least on the face of it. Granted, this also does not rule out the possibility that it is a third type of quantificational morpheme. However, as far as I know, these roots otherwise are capable of behaving exactly like nouns outside of these particular contexts. Thus, deriving functional uses from nominal uses remains a viable strategy.

- (23) tin sɔptah*(-er) jatra
 three week-GEN journey
 ‘three week journey’

If there is truly N^0 -to- Cl^0 movement, what drives it? Depending on theoretical commitments, there are a number of ways we can encode how and why this dependency occurs. First, in deciding what drives this movement, we need to take seriously Simpson’s and Dasgupta’s observations that Simpson’s Conjecture often holds of “measure words” crosslinguistically – my “self-individuating nouns”. That is, there may be a missed generalization about classifier languages if we blame it purely on idiosyncratic properties of Bangla. One (but hardly the only) way of driving why self-individuating nouns must move to Cl^0 is to suggest that they provide a more stringent and accurate restrictor on the presupposition in the Cl^0 ’s meaning. That is, since classifiers only restrict the domain to general classes – in Bangla, to people, to flat square things, or not at all – self-individuating nouns can more accurately restrict the domain than inserting a less informative classifier. However, due to the apparent “anti-Economy” flavor of this explanation, and the number of other imaginable

ways of encoding this dependency, I will not attempt to have this property fall out naturally of any aspect of the analysis, instead leaving this concern to future research. Still, I do think there is a plausible intuition here – why would the grammar overtly realize a less informative classifier at Vocabulary Insertion (on the assumption that until Vocabulary Insertion, the syntax sees only a [CL] feature) when it could construct a more representationally economical and pragmatically informative expression by moving N^0 to the Cl^0 position? However, this kind of analysis forces a look-ahead problem, in that blaming VI for the lack of a Cl^0 does not trigger N^0 -to- Cl^0 movement before SPELLOUT. Thus, I will put these particular technical concerns aside.

This excursus was meant as a way of showing that a non-stipulative and strictly grammatical account of N^0 -to- Cl^0 head movement in the narrow syntax is tenable and motivated. However, given that there is good reason to believe that the QA construction is fed by this N^0 -to- Cl^0 movement, if we take the cyclic application of rules characteristic of Minimalist syntax seriously, then we should be surprised if the system worked in any other way. In the next section, I will examine the case of nouns that do not feed the QA construction, although they do not co-occur with classifier elements. First, I will introduce Embick & Noyer’s (2001) theory of the PF-branch, which I will assume, since it makes predictions about what syntactic operations after SPELLOUT should in principle look like. Secondly, I will introduce a rule that targets the Cl^0 -less non-self-individuating nouns and “overwrites” an adjacent classifier. I will then give an explicit formalization of how I conceptualize the relevant rules.

4. Non-Self-Individuating Cl^0 -less Nouns

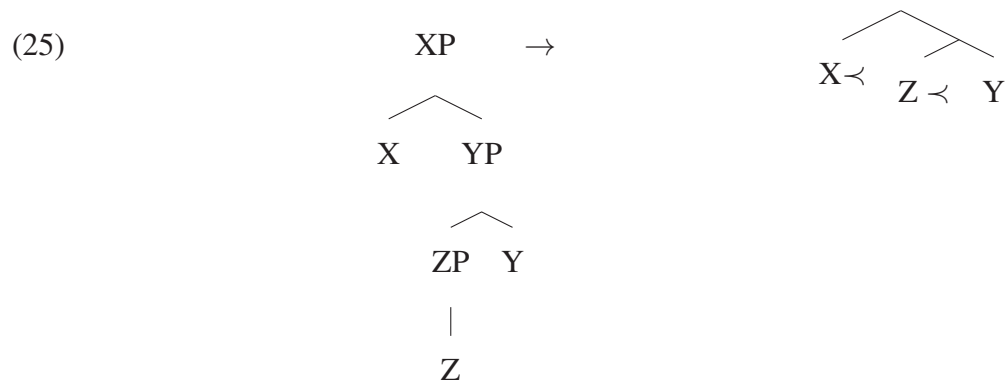
In this section, I introduce a framework proposed by Embick & Noyer (2001), which distinguishes syntactic rules that appear before and after Linearization. I then argue that at least one diagnostic supports the ordering conclusions. Specifically, I show that the idiosyncratic nature of non-self-individuating Cl^0 -less nouns means they *must* occur after Linearization, and *a fortiori* after SPELLOUT. I show that the other diagnostic that Embick & Noyer (2001) propose – strict linear sensitivity – also supports the analysis at hand, with a few outstanding issues.

4.1 Embick and Noyer’s Framework

Embick & Noyer (2001) propose a theory of post-SPELLOUT syntactic operations on the PF-branch within the Distributed Morphology (Halle, 1990; Halle & Marantz, 1993, 1994) framework. For Embick and Noyer, the PF-branch of the derivation is bipartite – there is some point in the derivation at which both Linearization and Vocabulary Insertion occurs, and grammatical operations show different properties depending on whether they occur before or after Linearization and Vocabulary Insertion. This is the *Late Linearization Hypothesis*.

- (24) **The Late Linearization Hypothesis:** The elements of a phrase marker are linearized at Vocabulary Insertion. (Embick & Noyer, 2001, p. 562)

Linearization here refers to making a structure *flatter*, not *flattened*. Critically, the intuition is that in the tree on the right in (25), although it is structural, encodes linear precedence (the original paper used *, here I use \prec to signal precedence), and relational notions such as “specifier of” and “complement of” are no longer visible.



Embick and Noyer propose that Marantz’s (1988) *Morphological Merger* is an operation that is live at the post-SPELLOUT stage of the derivation and has distinct properties depending on whether it occurs pre- or post-Linearization.

- (26) **Morphological Merger:** At any level of the syntactic analysis (D-Structure, S-Structure, phonological structure), a relation between X and Y may be replaced by (expressed by)

the affixation of the lexical head of X to the lexical head of Y.

For Embick and Noyer, Morphological Merger has two distinct flavors on the PF-branch. If an operation occurs before Vocabulary Insertion and Linearization, then it should be sensitive to properties of the syntactic features of the structure that it operates over. Furthermore, a pre-Vocabulary Insertion and pre-Linearization application of Morphological Merger should not be sensitive to surface properties of the morphological elements that are later inserted at Vocabulary Insertion. Alternatively, if the application of Morphological Merger occurs after Vocabulary Insertion and Linearization, then the prediction is that the operation is local in a stricter sense, since the operations only have access to a linearized structure. That is, adjuncts and other kinds of structure that normally do not block movements should block applications of Morphological Merger after Linearization. On the other hand, surface morphophonological features are legible to these kinds of operations, since Vocabulary Insertion would precede. Embick and Noyer consider two particular instantiations of Morphological Merger – Lowering (a pre-Vocabulary Insertion/Linearization rule) and another lowering rule called Local Dislocation (a post-Vocabulary Insertion/Linearization rule). For instance, they analyze T^0 -to- V^0 lowering in English (“affix hopping”) as an instantiation of Lowering. That is, it happens with all verbs, regardless of their phonological shape, and intervening adjuncts do not block it. This latter property is demonstrated in (27). Embick and Noyer submit the attachment of the comparative morpheme *-er* as an example of Local Dislocation. They suggest that since *-er* seems to be sensitive to the phonological shape of the adjective that it attaches to – *-er* seems to only be compatible with monosyllabic words, by and large – that this operation must occur after Vocabulary Insertion. In their theory, then, it should also be sensitive to linear structure – that is, an intervening adjunct should block the Local Dislocation operation. This seems to be true, since it is not possible to attach the comparative morpheme if an adjunct like *amazingly* appears, shown in (28-c).

- (27) a. John ~~s~~ go-es to the store.
b. John ~~s~~ frequently go-es to the store.

- (28) a. John is more intelligent(*-er).
 b. John is ~~more~~ taller.
 c. John is more amazingly tall(*-er).

Embick & Noyer's (2001) rules were, for the most part, *lowering* rules. However, there is nothing that forces us to restrict the set of possible post-SPELLOUT operations to lowering rules or prevents us from using raising or head movement rules at this stage. In the next section, I use Embick and Noyer's diagnostic for rules that apply after Linearization/Vocabulary Insertion to show that N^0 -to- CI^0 is a post-Linearization (and thus, post-SPELLOUT) operation, whenever it applies to non-self-individuating nouns. That is, I will show that whatever ensures the operation is sensitive to the identity of the root when applied to non-self-individuating nouns,⁸ ensures that the operation is bounded by the linear string, and not structural notions like adjunct or head.

4.2 The Rule and Linear Sensitivity

Before explicitly stating the rule that accounts for CI^0 -less nouns, I demonstrate that it is an instance of post-SPELLOUT N^0 -to- CI^0 movement.

In section 3, I argued that there is a well-defined class of nouns that participate in N^0 -to- CI^0 movement in the pre-SPELLOUT syntax. The CI^0 -less nouns that remain, however, do not seem to form any kind of coherent natural class. Dasgupta (1983) and Bhattacharya (1999a,b) have identified seven, and I am not aware of any others: *def* 'nation', *caka* 'wheel', *dik* 'direction', *paf* 'side', *bhai* 'brother', *bon* 'sister'. I would like to suggest that whatever rule governs the particular behavior of these nouns must be a rule that, in its formulation, targets the appropriate roots. This seems the most plausible, because these roots seem to form no coherent phonological or semantic class, suggesting that there is no inherent property of these roots that would be visible to the pre-SPELLOUT syntax. One could always stipulate that these particular lexical entries come annotated with a diacritic that governs their behavior. That is, a counterhypothesis would be that

⁸Depending on one's theory of where roots initially appear, this may suggest a later point of insertion in the derivation.

there is a feature that is live during the pre-SPELLOUT syntax, but its effects are only realized post-SPELLOUT. However, if this is the case, and whatever governs these nouns' distribution is a syntactic entity supplied by the lexicon, it should be surprising that they behave differently from the self-individuating nouns at all. If the property of being classifier-less is due to a feature that is legible to the pre-SPELLOUT syntax in both cases, then it is conceptually odd that there are two features that have the same effect in the grammar, but their morphophonological footprints are visible at different parts of the derivation. I anticipate that the best way to implement this kind of analysis would be to significantly complicate the syntax in a non-obvious way, and the worst way would ultimately be a re-statement of the facts. Thus, I'll take the approach that says the apparent idiosyncrasy and selectivity of these nouns is not a property of the nouns *as such*, but rather, to the rule that later on manipulates the distribution of these nouns at a point in the derivation.

This would suggest that this operation must occur after the point at which the roots are first inserted into the syntax. On classic models of Distributed Morphology, such as Halle & Marantz (1993), the roots are not available until Vocabulary Insertion. If this were the case, and if it is true that the rule itself must target these nouns by virtue of their root identity, then we would conclude that this operation must be occurring after Vocabulary Insertion, and consequently, at some point after SPELLOUT. However, the condition that no roots are available until Vocabulary Insertion has been challenged on facts from speech errors that have been taken to suggest that root identities may be available earlier (Pfaus, 2000), and that an absence of root identities until after VI complicates the theory of root allomorphy (Siddiqi, 2009). Let us suppose that the earlier view is correct, however, and that roots aren't available until after VI.

A stronger argument in favor of the proposal that this operation occurs after Linearization is the strict linear sensitivity on its application. As has already been mentioned, the proposed post-SPELLOUT N^0 -to- Cl^0 operation seems only to occur when there is linear adjacency between the nominal root and the classifier element. Whereas it is not acceptable to have a classifier element occur with a noun like *caka* 'wheel' when the Cl^0 is adjacent to N^0 , moving Cl^0 in the pre-SPELLOUT syntax – and disrupting linear adjacency – allows an overt classifier to appear. That is, the contin-

gent Cl^0 -lessness of *caka* ‘wheel’-type nouns is due to the linear adjacency of Cl^0 and N^0 being blocked by syntactic operations. This is distinct from nominals headed by nouns like *sɔptaho* ‘week’, in which an overt classifier never seems to be available, regardless of what other movements have occurred.

- (29) a. tin-(ʔtɛ) caka
 three-CL wheel
 ‘three wheel’
- b. goʔa-tin-ek caka
 CL-three-EK wheel
 ‘three or so wheels’
- c. *caka-tin-ek
 wheel-three-EK

- (30) a. tin-(ʔtɛ) sɔptaho
 three-CL week
 ‘three weeks’
- b. ʔgoʔa-tin-ek sɔptaho
 CL-three-EK week
- c. sɔptaho-tin-ek
 week-three-EK
 ‘three or so weeks’

This can be observed outside of the QA construction, as well. There is a construction in Bangla that left-dislocates the NP DP-internally marking definiteness, demonstrated in (31). (Bhattacharya, 1999a,b; Chacón, 2009)

- (31) a. tin-tɛ [NP sundor beʔal]
 three-CL beautiful cat
 ‘three beautiful cats’
- b. [NP sundor beʔal] tin-tɛ
 beautiful cat three-CL
 ‘the three cats’

With self-individuating nouns, the NP left-dislocation is not available for marking definiteness.

This is what we would predict if the N^0 -to- Cl^0 movement operation occurs in the narrow syntax and if extracting an X^0 from XP renders XP immobile (Takano, 2000; Lasnik, 1999). That is, if N^0 has been extracted, NP should not be available to move to a higher position for marking definiteness. However, with a noun like *caka* ‘wheel’, whose incorporation I propose occurs after Linearization (and thus, after all other syntactic operations) and when Cl^0 and N^0 are adjacent linearly, it should be no surprise that Cl^0 appears. Thus, this suggests that the rule I propose is (a) sensitive to linear order and notions like “adjacent” and (b) can be bleeded by pre-SPELLOUT syntactic operations like the DP-internal NP left-dislocation or the QA construction. This is consistent with the approach under consideration.⁹

- (32) a. tin sɔptaho
 three week
 ‘three weeks’
 b. *sɔptaho tin
 week three
 intended ‘the three weeks’

- (33) a. tin caka
 three wheel
 ‘three wheels’
 b. caka tin-tɛ
 wheel three-CL
 ‘the three wheels’

For non-self-individuating nouns, I suggest a rule like the one in (34).

- (34) [Cl [N]] → [Cl ⊕ N] when N is $\sqrt{\text{NATION}}$, $\sqrt{\text{WHEEL}}$, $\sqrt{\text{DIRECTION}}$, $\sqrt{\text{SIDE}}$, $\sqrt{\text{BROTHER}}$,
 $\sqrt{\text{SISTER}}$ or ...

This rule uses the notation given in Embick & Noyer (2001), in which a \oplus signifies a post-syntactic

⁹The skeptical reader might retort that all I have shown is that N^0 -to- Cl^0 substitution fails to feed other operations in addition to QA, and not that there is necessarily a restriction on linear adjacency. This is perhaps true, but consider what the data would look like if this N^0 -to- Cl^0 operation *wasn't* bounded linearly. There wouldn't be anything preventing the noun to move into the Cl^0 position, regardless of where Cl^0 is. Thus, the two sets of nouns under discussion should appear the same in *every construction*, contrary to the data.

Morphological Merger. However, in Embick & Noyer’s (2001) framework, what this connective actually means – that is, how is it distinct from and similar to movement and other operations – is not clear. However, it seems to be a concatenation process in some sense. Thus, this rule takes a N^0 after, and morphologically affixes it to Cl^0 .

5. Empirical and Conceptual Difficulties

There are a number of empirical and conceptual difficulties for the proposals given here, and there are a number of loose ends that could be addressed. In this section, I would like to highlight a few interesting predictions and points of tension in the analysis. Specifically, I would like to address variability and a (possible) ambiguity with regard to span time-denoting nouns, and a surprising sensitivity to linear adjacency between self-individuating nouns and adjectives.

First, as I mentioned earlier, Cl^0 -incompatible nouns aren’t necessarily *ungrammatical* when they co-occur with a classifier. Dasgupta (1983) has noticed a particular semantic nuance that is contributed by the classifier in particular constructions. In (35), we see that a classifier *can* be used with a time span-denoting noun if there is an intent to refer to particular individuated units of time. That is, there might be a distinction between the uses of nouns like *day* when there is reference to a 24-hour span of time, distinct from, say, a day of the week.

- (35) a. Du-ʈo bochor kharap gælo – 1966 ar 1976.
two-CL year bad went 1966 and 1976
‘There were two bad years – 1966 and 1976’
- b. Du-ʈo maʃer namer jeʃe “ari” – januari ar phebruari.
two-CL month’s name’s end.LOC “ari” January and February
‘Two months have names ending in *ari* – January and February.’

One possibility may be that there are two distinct roots realized as *din* ‘day’ – $\sqrt{24 \text{ HOUR PERIOD}}$ and $\sqrt{\text{DAY OF THE WEEK}}$, and that only the former is a self-individuating noun. This seems to capture Dasgupta’s intuition, since in using the $\sqrt{\text{DAY OF THE WEEK}}$ *din*, which by hypothesis does not incorporate into the Cl^0 position, an overt Cl^0 is needed.¹⁰ This makes an interesting

¹⁰Justin Nuger (p.c.) tells me that French has a similar lexical distinction, between *jour* and *journée* ‘day’.

prediction, however. Namely, we should expect there to be no feeding or bleeding relations with operations like the QA construction or the DP-internal NP left-dislocation as long as there is a “day of the week” interpretation, instead of the span of time interpretation as such. That is, the “day of the week” interpretation should be compatible with a classifier, and should pattern like any other unmarked noun.

Secondly, there is a surprise sensitivity to linear order for N^0 's that incorporate into Cl^0 in the narrow syntax. Since we find linear order being relevant for defining locality in the post-SPELLOUT N^0 -to- Cl^0 movement, and not with the pre-SPELLOUT N^0 -to- Cl^0 movement, this sensitivity is not predicted. This is shown in (36).

- (36) a. tin din [_{NP} ~~din~~]
 three day
- b. *tin din [_{NP} lɔmba ~~din~~]
 three day long
- c. tin-ʈe [_{NP} lɔmba din]
 three-CL long day

On the proposal that I offer in this paper, we should not expect the intervening adjective *lɔmba* ‘long’ to block incorporation into the Cl^0 position. This is because adjectives are generally considered adjuncts, and as such, should not be relevant for movements that are defined over hierarchical structure. There is no clear way to account for this, though there are a few ways that we might derive this fact, with some stipulation. The first thing that comes to mind is to suggest that adjectives are *not* adjuncts, at least in Bangla. A similar notion is found in Abney (1987), who posited that there is an AP layer between DP and NP, in order to account for the facts in (37).

- (37) a. The mother is proud of her daughter.
- b. *The proud of her daughter mother

On Abney’s analysis, the adjective *proud* can select for a PP, as in its predicative use above. However, (37-b) is out by simple structural reasons – if adjectives are functional heads that select for

NP in their attributive use, there should be no way of building a structure in which the A⁰ *proud* can select for both an NP and a PP, on a strict interpretation of a binary branching X'-theory.

A different account of these facts was given in Chacón (2009). There, Chacón (2009) hypothesized that, if Merger is completely unrestricted, and all functional items are inherently optional, there should be situations in which the nominalizer n⁰ might not merge. I proposed that incorporation is a reflex of a failure to merge n⁰. Supposing that lexical roots must be affixed onto *some* functional element in order to be targeted by Vocabulary Insertion, the root must adjoin to the nearest functional head – in this case, the Cl⁰ element. I intended to derive two facts about incorporation from this proposal – namely, incorporated nouns do not tolerate a specific interpretation, and they do not, in the general case, tolerate external modification from their base generation site. This is demonstrated in English noun incorporation below.

(38) We went bear-hunting [_{NP} ~~bear~~], # and it was huge.

(39) We went bear-hunting [_{NP} (*big) ~~bear~~].

The intuition here is that failure to merge n⁰ derives the fact that these lexical roots were, in some sense, not “real” nouns. With the auxiliary assumption that adjectives need to modify *nouns*, and that *nouns* alone are capable of referentiality, this particular clustering of data obtains. A similar approach to DP-internal incorporation also captures the fact that adjectives are not available with self-individuating nouns, if we take the mechanism driving N⁰-to-Cl⁰ movement to be the absence of n⁰. Either of these approaches seems to capture the data, and both are motivated by cross-linguistic concerns, as I see it. In either case, something would need to be said in order to ensure that adjectives block N⁰-to-Cl⁰ movement for self-individuating nouns.

Additionally, there is a very serious problem regarding the notion of head substitution. In a Bare Phrase Structure theory, stating adjunction of a head to another head should be doable by whatever mechanisms we use for adjunction, meaning head adjunction is no more problematic than phrasal adjunction. However, since Merge operates over elements of a lexical array, it isn't clear how to

Merge a head X^0 , and then merge another element Y^0 “in X^0 ’s position”. At any rate, this is not an issue only applicable to the analysis at hand, but extends to the immense literature addressing verb movement, and any head movement analysis in general in which there is substitution. Thus, I will not attempt to solve it here.

One might imagine that the apparent head substitutions in this proposal are actually adjunction operations that only apply when a null element appears. This does not seem satisfying, due to appealing to a null element that only appears in the constructions under discussion. This encodes an oddly circular dependency – a null Cl^0 that selects only for N^0 s that must move. Jeff Lidz (p.c.) points out that taking such an approach actually may simplify the data. If we suppose that non-self-individuating Cl^0 -less nouns are selected by a null classifier, we do not need to postulate movement. Additionally, we can say that the “mysterious” classifier that appears when N^0 and Cl^0 aren’t adjacent may be due to some rule that allows for some emergence of the unmarked, given that all the data presented so far uses the “default” classifier *goʔa*. This does not seem to be the case, since it is (somewhat) acceptable to use other classifiers, shown below. Additionally, this would mean that the null classifier must have the nouns that are compatible with it idiosyncratically listed, given the fact that these nouns seem to make no natural class. Although such idiosyncrasies are well attested for noun classes, classifiers generally are defined in terms that make reference to natural categorical aspects of the noun being counted, and do not admit “exceptions”. Thus, this null classifier would have a number of peculiarities.

- (40) (?)khan-tin-ek caka
CL-three-EK wheel
‘three or so wheels’

Alternatively, one might imagine that there is some principle of Vocabulary Insertion that, when targeting a complex X^0 that consists of an adjunction of a lexical element and a functional element, that Vocabulary Insertion preferentially realizes the lexical element at the cost of the functional element. That is, adjoining an N^0 to Cl^0 , only the N^0 survives VI. However, in articulating such an idea, we must be careful to not rule out V^0 -to- T^0 adjunction. I will leave this for further

investigation.

6. Conclusion

In this paper, I accounted for a series of feeding and bleeding relationships between operations that create a number of structures in the Bangla DP. Specifically, I argued that the N^0 -to- Cl^0 movement, what previously was considered one operation, is best thought of as two different word formation operations – before and after SPELLOUT. I intended to show that the former cases are formed by *head movement*, bearing on the issue of whether there is such an operation, and if so, where it can appear in the grammar. Secondly, I showed that, in line with Embick & Noyer's (2001) general framework, there are post-SPELLOUT word formation rules that operate over a linearized structure, as opposed to a purely hierarchical structure. Additionally, I gave a semantic motivation for Simpson's Conjecture, when it applies in the pre-SPELLOUT syntax.

I did not attempt to address the larger technical concerns that motivate the questioning of head movement as a legitimate operation in the syntax, however. A number of these kinds of concerns still exist – how do we account for its apparent anticyclicity and its particular locality constraints? What is a “complex head”, and how could such a notion be stated without doing violence to the tenets of Bare Phrase Structure? These are questions that must remain unanswered for now. However, I hope to have demonstrated the need for something like head movement in multiple components of the grammar, distinct from a generative lexicon and in the narrow syntax. Hopefully more empirical work of this kind can elucidate the nature and the distinct flavor of head movement and other kinds of extremely local syntactic operations in the syntactic and morphosyntactic domain.

References

- Abney, S. P. 1987. The English noun phrase in its sentential aspect. Ph.D. thesis, MIT.
- Baker, M. 1985. The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16:373–415.
- Baker, M. 1996. *The Polysynthesis Parameter*. Oxford University Press.

- Baker, M. 2009. Is head movement still needed for noun incorporation? *Lingua* 119:148–165.
- Bale, A. C. & D. Barber. 2009. The interpretation of functional heads: Using comparatives to explore the mass/count distinction. *The Journal of Semantics* 26:217–252.
- Bale, A. C. & H. Khanjian. 2009. Classifiers and number marking. In *Proceedings of SALT XVIII*, eds. T. Friedman & S. Ito, 55–72. Ithaca, NY: Cornell University.
- Benedicto, E. E. 1997. V-movement and its interpretational effects. Talk presented at GLOW 20, Rabat, Morocco, March.
- Bhattacharya, T. 1999a. Specificity in the Bangla DP. In *Yearbook of South Asian Languages and Linguistics* 2, ed. R. Singh, 71–99. New Delhi/London: Sage Publications.
- Bhattacharya, T. 1999b. The structure of the Bangla DP. Ph.D. thesis, University College London.
- Bhattacharya, T. 2001. Numeral/quantifier-classifier as a complex head. In *Semi-Lexical Heads*, eds. H. C. van Riemsdijk & N. Corver, 191–222. The Hague: Mouton de Gruyter.
- Bobaljik, J. D. & S. Brown. 1997. Interarboreal operations: Head movement and extension requirement. *Linguistic Inquiry* 28:345–356.
- Boeckx, C. & S. Stjepanovic. 2001. Head-ing towards PF. *Linguistic Inquiry* 32:345–355.
- Borer, H. 2005. *In Name Only*, volume 1 of *Structuring Sense*. London: Oxford University Press.
- Chacón, D. A. 2009. Head movement in the Bangla DP. Ms., University of Maryland, College Park, MD.
- Chierchia, G. 1998a. Plurality of mass nouns and the notion of semantic parameter. In *Events and Grammar*, ed. S. Rothstein, 53–104. Dordrecht: Kluwer.
- Chierchia, G. 1998b. Reference to kinds across languages. *Natural Language Semantics* 6:339–405.

- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. 2001. Derivation by phase. In *Ken Hale: A Life in Language*, ed. M. Kenstowicz, 1–52. Cambridge, MA: MIT Press.
- Dasgupta, P. 1983. On the Bangla classifier *ta*, its penumbra, and definiteness. *Indian Linguistics* 44:10–26.
- Doetjes, J. 1996. Mass and count: syntax or semantics? In *Proceedings Meaning on the HIL*, eds. A. Arregui & C. Cremers, 34–52. Leiden: HIL/Leiden University.
- Embick, D. & R. Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32:555–595.
- Emonds, J. 1976. *A Transformational Approach to English Syntax: Root, Structure-Preserving, and Local Transformations*. New York, NY: Academic Press Inc.
- Hale, K. & S. J. Keyser. 1993. On argument structure and the lexical expression of grammatical relations. In *The view from Building 20*, eds. K. Hale & S. J. Keyser, 53–110. Cambridge, MA: MIT Press.
- Halle, M. 1990. An approach to morphology. In *Proceedings of NELS 20*, ed. J. Carter, 150–184. Amherst, MA: University of Massachusetts.
- Halle, M. & A. Marantz. 1993. Distributed morphology and the pieces of inflection. In *The View from Building 20*, eds. K. Hale & S. J. Keyser, 20, 111–176. MIT Press.
- Halle, M. & A. Marantz. 1994. Some key features of distributed morphology. In *MIT Working Papers in Linguistics*, eds. A. Carnie & H. Harley, volume 21, 275–288. Cambridge, MA: MIT Press.
- Höhle, T. 1992. Über verum-fokus im deutschen. In *Informationsstruktur und Grammatik*, ed. J. Jacobs, 112–141. Opladen: Westdeutscher Verlag.

- Hornstein, N. 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.
- Kitahara, H. 1995. Target α : Deducing strict cyclicity from derivational economy. *Linguistic Inquiry* 26:47–75.
- Koopman, H. 1984. *The Syntax of Verbs: From Verb Movement Rules in the Kru Languages to Universal Grammar*. Dordrecht: Foris Publications.
- Lasnik, H. 1998. Some reconstruction riddles. In *University of Pennsylvania Working Papers in Linguistics 5*, eds. A. Dimitriadis, H. Lee, C. Moisset & A. Williams, 83–98. Philadelphia: University of Pennsylvania.
- Lasnik, H. 1999. Pseudogapping puzzles. In *Proceedings of the North East Linguistic Society 31*, eds. M. Kim & U. Strauss, 301–320. Washington, DC: Georgetown University.
- Lechner, W. 2007. Interpretive effects of head movement. LingBuzz/000178.
- Link, G. 1983. The logical analysis of plural and mass terms: A lattice-theoretical approach. In *Meaning, use, and interpretation of language*, eds. C. S. Rainer Bäuerle & A. von Stechow, 302–323. Berlin: Mouton de Gruyter.
- Marantz, A. 1988. Clitics, morphological merger, and the mapping to phonological structure. In *Theoretical Morphology*, eds. M. Hammond & M. Noonan, 235–270. San Diego, CA: Academic Press.
- Matushansky, O. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:69–109.
- McCready, E. 2009. Classifiers and conventional implicature. *The Journal of Cognitive Science* 10:195–208.
- Merchant, J. 2002. Swiping in north Germanic. In *Studies in comparative Germanic syntax*, eds. C. J.-W. Zwart & W. Abraham, 295–321. Amsterdam: John Benjamins.

- Nakanishi, K. & S. Tomioka. 2004. Japanese plurals are exceptional. *Journal of East Asian Linguistics* 13:113–140.
- Nunes, J. 2001. Sideward movement. *Linguistic Inquiry* 32:303–344.
- Omaki, A. 2009. Verbal morphology: Return of the affix hopping approach. In *Proceedings of NELS 38*, eds. A. Schmidt, M. Walkow & M. Abdurrahman, volume 2, 193–204. Amherst, MA: GLSA.
- Pelletier, F. J. 1975. Non-singular reference: some preliminaries. *Philosophia* 5:451–465.
- Pfau, R. 2000. Features and categories in language production. Ph.D. thesis, University of Frankfurt.
- Pollock, J.-Y. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20:365–424.
- Roberts, I. 1991. Excorporation and minimality. *Linguistic Inquiry* 22:209–218.
- Siddiqi, D. 2009. *Syntax within the Word: Economy, allomorphy, and argument selection in Distributed Morphology*. Number 138 in *Linguistics Today*. Amsterdam: John Benjamins.
- Simpson, A. 2005. Classifiers and DP structure in southeast Asia. In *The Oxford Handbook of Comparative Syntax*, eds. G. Cinque & R. S. Kayne, 1–22. London: Oxford University Press.
- Takano, Y. 2000. Illicit remnant movement: an argument for feature-driven movement. *Linguistic Inquiry* 31:141–156.
- Tang, C.-C. J. 1990. Chinese phrase structure and the extended X-bar theory. Ph.D. thesis, Cornell University.
- Travis, L. 1991. Parameters of phrase structure and verb-second phenomena. In *Principles and Parameters in Comparative Syntax*, ed. R. Friedin, 339–364. Cambridge, MA: MIT Press.

Watanabe, A. 1995. Conceptual basis of cyclicity. In *MIT Working Papers in Linguistics 27: Papers on Minimalist Syntax*, eds. R. Pensalfini & H. Ura, 269–291. Cambridge, MA: MIT Press.

Yadroff, M. & L. Billings. 1998. The syntax of approximate inversion in Russian (and the general architecture of nominal expressions). In *Proceedings of 6th Annual Workshop on Formal Approaches to Slavic Languages*, eds. Ž. Bošković, S. Franks & W. Snyder, 319–38. Ann Arbor, MI: Michigan Slavic Publications.