Notes on insertion in Distributed Morphology and Nanosyntax

Pavel Caha

0. Abstract

This paper considers two questions that many people ask themselves (or should ask themselves). What is actually the difference between Nanosyntax (NS, Starke, 2009) and Distributed Morphology (DM, Halle and Marantz, 1993)? And which one of them is right? These questions remain as important now as they were some fifteen years ago, when Michal Starke introduced the basics of the NS theory. Despite the fact that several written sources on NS have been available since 2007, there is still a lot of confusion about what NS actually is, and how NS and DM square together. The present paper is an attempt to clear things up.

1. Three differences between NS and DM

The theories of Nanosyntax (cf. Baunaz and Lander this volume) and Distributed Morphology (cf. Harley and Noyer 1999, Embick and Noyer 2003, Bobaljik 2015) provide each a particular theory of how to get from the basic building blocks of language to complex sentences. Any such theory has multiple components, and the two theories diverge on a number of important questions. For instance, an important difference (**dif 1**) concerns **the nature of the basic building blocks**. Distributed Morphology projects its syntactic structures from complex objects that correspond to pre-packaged feature structures (bundles) taken directly out of the pre-syntactic lexicon. On the other hand, the main architectural claim of Nanosyntax is opposed to this; the idea is that the only component of grammar capable of constructing complex feature structures is syntax, which thus has to start from individual features. This leads to a syntax with large trees, increased number of movement steps, and a quite different outlook than the tree

structure representations used within DM. NS also tries to dispense with head movement, which leads to an increase in phrasal movement and remnant movement derivations.

Another difference (**dif 2**) is the use of **phrasal spell out**, which in NS represents one of the core analytical tools, while at the same time being rarely used and often argued against in DM, as in Embick and Marantz (2008), or most recently in Embick (2014).[1] This paper probes deeper into this particular issue, because it seems that in this particular aspect, the theories are close enough for a meaningful comparison to be made. My main point here will be that the insertion procedure used in DM (the so-called Subset Principle) is incapable of governing non-terminal lexical insertion, and this is where the NS conception (the so-called Superset Principle) is better suited to handle the relevant data. The move to non-terminal spell out is theoretically attractive; once adopted, it replaces *some* of the morphological operations proposed in DM, namely Fusion, Fission and Readjustment Rules (Caha, 2009:ch. 2).

The prospect emerging from this conclusion is the possibility of a mapping between syntax and the pronunciation which is direct, mediated only by lexical access and nothing else. This theoretical goal ultimately represents a third difference (**dif 3**) between NS and DM: is there (DM), or is there not (NS), **a separate morphological module/structure**, with rules and operations specific to that module? This issue – independent of the difference 2 – revolves

[1] For clarity, let me mention that phrasal spell out is a type of non-terminal spell out (phrase is a type of a non-terminal node). I mention this here because there seems to be some confusion about this in the literature. Thus, Haugen and Siddiqi (2013) think that phrasal insertion „inserts entire phrases into non-terminal nodes rather than just single Vocabulary Items." This is incorrect, phrasal spell out inserts Vocabulary Items into phrasal nodes.

Please note that what NS calls "phrasal spell out" is often best recast in DM as the spell out of a non-terminal inside a complex head. This is because NS often uses phrasal derivations where DM uses head movement, and so for reasons of cross-theoretical comparison, it is best to understand "phrasal spell out" to actually mean "non-terminal spell out."

around the questions of whether the remaining post-syntactic operations used in DM are needed to derive the surface forms of language, or whether these operations can also be dispensed with.

The specific operations that are covered under **dif 3** (on this way of cutting the pie) are **post-syntactic operations that change constituency and/or linear order**. Specifically, Marantz (1989) as well as Embick and Noyer (2001) propose that in a number of cases, structures of the type [X [Y Z]] can be transformed onto [[Y X] Z] by **Merger**, which (according to them) takes place after syntax. According to my perception of the data, the difference to NS here is not so much about what the constituency is (it is [[Y X] Z], i.e., the one which DM derives by Merger), the question is whether Merger is the only way to derive such structures.

Obviously, the logically independent issues covered under 2 (is there or is there not phrasal spell out) and 3 (what is the constituency of strings and how we can derive it) interact. If lexicalization targets (potentially phrasal) constituents, then it matters what type of constituent structure feeds into lexicalization. To give a concrete example: it matters for spell out of X and Y whether the constituency is [X [Y Z]] (X and Y cannot spell out together) or whether it is [[X Y] Z] (X and Y can be spelled out together).

Most of the discussion of NS from DM positions has focused on issues surrounding **dif 3**. So despite the fact that my main goal is to look at the topic of insertion, it feels wrong to simply ignore the concerns that researchers working in DM have about the constituency required by phrasal spell out. So I devote some remarks to this topic in section 2, before I come back to insertion in sections 3 and 4.

2. Spell out and constituency in a model of syntax that keeps changing

As highlighted under **dif 1,** NS and DM have diverging opinions concerning the general outlook of syntactic structure. In order to illustrate how one's conception of syntax tends to correlate

with one's choice of morphological framework (and vice versa), let me come back to the 1990s when DM itself was taking off as a framework. In one of the papers of that time, Pullum and Zwicky (1990) reflected over DM's conception of morphology from the positions of an "A-morphous" approach (based on the idea that words correspond to unstructured feature matrices which in turn are the atoms of structure building). The authors noted that the structured morphological representations used in DM fare quite well with Pollock's groundbreaking (1989) system, where complex words arise by head-movement, rather than being simply pulled out of the lexicon. To the authors, such a system involved "baroque-style derivations," where "it is not clear why affixes or stems could not be moved apart by syntactic rules to yield (e.g.) Affix Topicalization or Heavy Stem Shift constructions." They went on to claim that [Pollock's system,] "even if it worked…, it would … generate strings of the wrong sort to be input to a morphological module of a theoretically and empirically optimal sort." In effect, Pullum and Zwicky took their paper to "cast serious doubt on whether its basic operation of head-to-head movement was doing any real work." Some twenty five years later, we are where we are. Head movement has become a common stance, and DM a point of reference to compare new alternatives with.

But syntax has of course not stopped in the 90s. New approaches appeared where head movement has been replaced by phrasal movement. Consider, for instance, the development from Cinque (1994) to Cinque (2005). Both papers try to derive the order of nouns, adjectives, numerals and demonstratives. In 1994, Cinque proposed that when the noun preceded any of these, it moved from its base position at the bottom of the tree to the left of those modifiers by head-movement. In 2005, Cinque included many more languages into his sample and noted that the original system encounters problems, because a noun may move across two modifiers without inverting their order, something that the traditional head movement analysis could not do. So Cinque proposed that the noun actually crossed its modifiers by phrasal movement, and

provided a system which restricts movements in such a way that it can generate all and only the attested orders using exclusively phrasal movement. This – among other things – meant that the complement of the noun (which often remains last in the string even when the noun moves) had to obligatorily undergo extraposition (Cinque 2005:327).[2]

With such changes in syntax arriving, it became possible to contemplate new models of morphology which make use of such derivations. To see where such derivations fit in, consider one of the examples discussed in Baunaz' and Lander's introduction. In their ex. 4, they introduce the form *puell-as*, which is an accusative plural (*-as*) of the Latin noun 'girl' (*puell*). As far as the syntactic structure is concerned, there are reasons to think that number and case occupy distinct slots in the cross-linguistically fixed sequence of projections. But in the morphology, it is impossible to separate the accusative meaning from the plural meaning; *-as* is a portmanteau for both. To remove tensions such as these (a non-divisible form corresponding to two syntactic positions), NS proposes (as Baunaz and Lander point out) that these two projections are spelled out together as a phrase after the noun moves to the left of them, see (1):

(1) N **[K [Num ~~N~~]]**

puell ----- *as* -----

Suppose now that the noun had a complement. What would happen to it? If the complement did not move, it would remain in between the root and the affix, which never happens. So the conclusion is that the complement undergoes extraposition to an even higher position

[2] Extraposition in this context means movement of the noun's complement to a high position in the extended NP such that it c-commands and precedes all the noun's modifiers. After that, in languages where complements follow the noun, the whole extended NP undergoes remnant movement, so that the complement ends up to the right of the whole noun phrase (again).

(minimally above the landing site of the noun), followed by a remnant movement of the noun with its affix. There is nothing new here compared to Cinque's (2005) work: complements undergo obligatory extraposition.

For many researchers, such derivations are simply too baroque to be of any interest. For instance, Embick (2014) correctly points out that "the main predictions of [phrasal spell out] derive from constituency." But keeping the 1990s ideas about syntax as a reference point, he points out that for instance, "synthetic forms should never be found with adjectives that take complements," because that would lead to the problem of the intervening complement in (1). In general terms, Embick concludes that the representations required by phrasal spell out are „in the crucial cases … incorrect." He notes, though, that „in all of these, it is possible to incorporate additional assumptions [extraposition of complements] to neutralize incorrect predictions", but for these „there is little evidence" (p. 23). However, keeping Cinque's conclusions in mind, one cannot help feeling that we are taking part in a similar debate that took place some 25 years ago: we simply believe in different syntaxes (one standard by now, the other still too baroque), and so we are talking cross purposes.

Embick's paper is one of the few critiques of phrasal spell out that at least acknowledges these issues. Merchant (2015) fails to acknowledge even the possibility of extraposition when he writes that „whenever [phrasal spell involving the root] must occur, the root should allow for no internal arguments, a prediction that is clearly false."

Merchant's unwillingness to contemplate extraposition in these cases contrasts with his own work on ellipsis. Consider, for instance, gapping examples like: *Fred likes to pet the cat, and Sally ~~likes to pet~~ the dog*, concerning which van Craenenbroek and Merchant (2013:743) say that "it has become fairly standard to analyze gapping as involving movement of the remnants [*the dog*] to the left followed by deletion … of the rest of the clause." Clearly, the authors are relying on extraposition to create constituents that are later targeted by ellipsis. The

reason why Merchant denies this option for phrasal spell out (in effect subjecting phrasal spell out to a different theoretical standard than ellipsis) is a mystery to me.

Another construction where Embick (2014) as well as Haugen and Siddiqui (2013) think that phrasal spell out does not work are cases where a preposition spells out as one piece with the determiner (they mention French and Spanish respectively). They start from the idea that because of the scope relations, the preposition and the determiner cannot form a constituent to the exclusion of the noun. But since they spell out together, they conclude that phrasal spell out cannot deal with this fact.

What is slightly puzzling is that Embick (2014) ultimately does adopt the structure [[P D] N] for his own analysis, i.e., a structure which makes his point about wrong predictions of phrasal spell out irrelevant if evaluated against that structure. The reason why he denies the phrasal-spell-out model access to the correct structure is because he says that as far as syntax is concerned, the structure must be [P [D N]], and the structure which correctly describes the P-D interaction is derived by Merger. The idea is that phrasal spell out cannot make use of the post-syntactic Merger which, according to Embick, is the only way to derive the structure.

There are two observations to be made. The first one is that phrasal spell out is in principle independent of whether one does or does not have Merger, and so these things should not be automatically lumped together. Even more so that they introduce a double standard for comparison where a terminal-spell-out model is evaluated against the correct structure, and the phrasal-spell-out model is claimed to make wrong predictions because it is evaluated against a different (and incorrect) structure. In other words, even if one adopts a theory with Merger, it is still better (as I argue in section 3 and 4) to spell out the non-terminal derived by Merger than let insertion target terminals.

The second observation is related to the issue of whether syntax can or cannot build structures of the sort [[P D] N]. Here it is worth noting that apart from the French/Spanish facts,

there is evidence for such structures independent of phrasal spell out. For instance, Baker and Kramer (2014) discuss the fact that in Amharic, a prepositional marker which scopes over the whole extended NP is located inside (i.e., in the middle of) a constituent which is located on the left branch of that NP. Now given such converging evidence for the existence of [[P D] N] structures, the question to ask is: can we design our theory of syntax in a way that it can accommodate the evidence without the involvement of a post-syntactic component?

Proposals addressing such issues are already out there. For instance, Leu (2008) investigates structures analogous to the unexpected [[P D] N], but for a slightly different case. He looks at examples where the evidence suggests that definiteness morphemes are present in the projection of the adjective, yielding surface structures of the sort [[D A] N], against the semantically expected [D [A N]]. Leu then proposes a theory which base-generates the definiteness marker inside the projection of the adjective. The fact that definiteness scopes over the whole NP is attributed to a silent D sitting in the main projection line, so we actually have the structure [[D A] [**D** [N]]], where the boldfaced D is silent due to a generalized Doubly Filled Comp (Koopman 1996; Starke 2004). For a related way to generate such structures in NS, see the description of the operation CONSTRUCT in Baunaz and Lander's introduction.

The point of the narrative is that the reflections of NS in the DM literature resemble the welcome that DM got some 25 years ago. The reason for the resemblance is that people back then as much as now have different ideas about what syntax looks like. The syntactic structures of NS are "too baroque, and even if they worked, there is little evidence for them." I have tried to make some justice to that debate here; my main point is that the constituent structures needed for phrasal spell out (and claimed to be problematic) are needed also for reasons that have nothing to do with phrasal spell out, a point that I think should not be controversial.

The only thing that is potentially controversial is whether we should generate such structures in the syntax or not. I think that the null hypothesis is that we should minimally try

to do so. Ultimately, any empirical debate about the structures needed for phrasal spell out cannot take the 1990s syntax for granted. Rather, we have to focus on questions such as these: is there any evidence for or against the claim that the constituent structure needed for phrasal spell out is accessed also by other processes? At least for the case of the disputed extraposition, the answer seems to be yes; the same constituent structure that is required by phrasal spell out is also required by ellipsis.

As for the [[P D] N] structures, it is worth considering the fact discussed in Miller (1992), namely that a coordination of the type [P [D N] and [D N]] is available for non-contracting determiners, but unavailable for exactly those P-D combinations that undergo contraction, i.e., those that require the [[P D] N] structure. For those who accept the conclusion that the *syntactic* structure in such cases is [[P D] N] indeed, the coordination facts are as expected. Embick (2007) (who accounts for the contraction by Lowering/Merger), has to rely on the proposal that Lowering has to apply across the board just like regular syntactic operations. The fact that post-syntactic operations are subject to the same constraint as syntactic operations, should minimally raise some interest.

Whatever the ultimate answer to such questions will be, in what follows, I sidestep as many of these issues as possible. My goal will be to discuss the differences that exist in the insertion procedure independently of the diverging opinions about what syntactic structures look like, and how they are constructed. The way I do this is that I assume here only structures independently adopted in DM, and show how the theory can be improved if non-terminal spell out is adopted.

3. Insertion in DM

This section illustrates the paradoxes that arise in classical DM, where spell out targets terminal nodes only. I also show why it is impossible to extend the insertion procedure to non-terminals,

presenting what I think are unsurmountable puzzles. The solution to the puzzles will come in section 4, when we turn our attention to the Superset Principle.

3.1 The Subset Principle

Vocabulary insertion in DM is governed by the so-called Subset Principle. Its canonical formulation comes from Halle (1997), see (2). The principle has two parts, which can be called The Subset Clause and The Elsewhere Clause. The Subset Clause determines the conditions for the applicability of a Vocabulary Item. The Elsewhere Clause says what happens when several Vocabulary Items are applicable in a given environment.

(2) The Subset Principle

> [*The Subset Clause*:] The phonological exponent of a Vocabulary Item is inserted into a morpheme of the terminal string if the item matches all or only a subset of the grammatical features specified in the terminal morpheme. Insertion does not take place if the Vocabulary Item contains features not present in the morpheme. [*The Elsewhere Clause*:] Where several Vocabulary Items meet the conditions for insertion, the item matching the greatest number of features in the terminal morpheme must apply.

The main empirical bite of the principle is that it can nicely model syncretism. To see that on an example, consider the following fragment of the singular declension in Latin. In the table, I also indicate one possibility of attributing morpho-syntactic features to the relevant cells borrowing some of the features from Halle (1997).

(3) A fragment of the Latin declension

[+SINGULAR]	neuter [-MASC]	masculine [+MASC]
nom [+SUPERIOR]	*-um*	*-us*
acc [-SUPERIOR]	*-um*	*-um*

What we see here is that one of the exponents (*-us*) is tailor made for a particular case and gender. There is no intricacy involved in saying that this marker is the pronunciation of the features [+masc, +sup]. It is much harder to say what the other marker is the pronunciation of. It can appear both with masculines and neuters, and it can appear both in the nominative and in the accusative.

The Subset Principle allows us to generate such paradigms easily. It is enough to say that the marker *–um* simply corresponds to an underspecified singular marker, as in (4a). The Subset principle then predicts that it can occur in any case and any gender, as long as the [+sg] feature (for which *–um* is specified) is present in a given feature bundle/paradigm cell. So in principle, this marker can occur in all the cells of the paradigm (3).

(4) a. *-um* = [+sg]

b. *-us* = [+sg, +sup, +masc]

The number of cells where such an 'underspecified' marker actually surfaces depends on what other markers there are. So when we add the nominative (+sup) masculine (+masc) *–us* into the set of Vocabulary Items, see (4b), this will lead to a clash between *–um* and *–us* in the nominative of the masculine paradigm (in all other cells *-us* does not qualify for insertion on the basis of the subset clause). In this particular environment, *–us* wins, because it is more specific, and *–um* fills the rest of the cells. So the possibility to have underspecified markers

coming in after syntax is very attractive; syntacticians can do their work on fully specified trees/feature structures, over which syntactic generalizations are stated. The Subset Principle takes care of the morphological detail.

3.2 The paradox

Notice now that the insertion principle (2) explicitly states that phonological exponents are "inserted into a morpheme of the terminal string." A morpheme in DM is a term for a bundle of features (a terminal), so in effect this means that insertion only applies at terminal nodes.

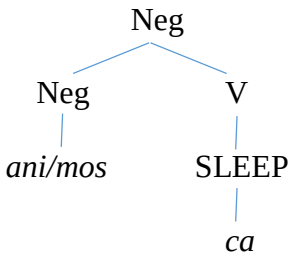
Are there any reasons to doubt this conclusion? Suggestions to this effect were rare in the early days of DM (I think that the first non-terminal spell out analysis in DM is Radkevich 2010). However, some issues did start appearing. An extremely interesting insight into the dilemmas that the theory faced at that point (and still faces today) is provided by a careful and detailed study of negation in Korean by Chung (2007).

The starting point of his discussion is the fact that sentences in Korean can be negated by attaching one of the two negative prefixes *ani* or *mos* to the verb (5a,b).[3] Chung (2007) shows that each of the negations is a separate head in the clausal spine, and the verb combines with it by syntactic movement. The negation sits lower than T, so that the structure of the string *mos/an(i)-ca* is shown in (6).

(5)	a.	*ca-n-ta*	b.	*mos/an(i) ca-n-ta*
		sleep-PRES-DECL		NEG sleep-PRES -DECL
		'is sleeping'		'can't sleep / isn't sleeping' (Chung 2007:ex.1,2,4)

[3] The meaning of the negations differs slightly, *ani* is a simple negation, *mos* has a modal component, and means 'cannot, is not allowed to.' The modal negation does not have an epistemic reading.

(6)



The second relevant fact is that the verb *al-* 'know,' see (7a), does not combine with any of these markers (7b), but shows a suppletive form *molu-* instead, as in (7c).

(7)

a.	b.	c.
al-n-ta	**mos/an(i) al-n-ta*	*molu-n-ta*
know-PRES-DECL	NEG know-PRES-DECL	NEG.know-PRES-DECL
'knows'	'cannot / does not know'	'cannot / does not know'

(Chung 2007:ex.45)

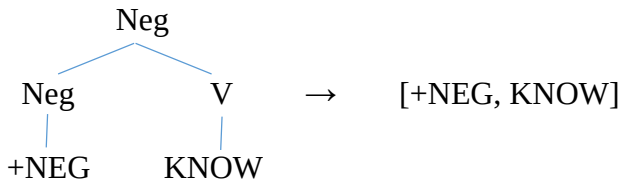
What we see here in abstract terms is that where one meaning ('not sleep') has two markers (Neg and V), another meaning ('not know') has a single non-divisible marker (Neg+V). Chung takes care to show that *molu* is an actual negative form; it does not correspond to a lexical verb like 'ignore, be unaware of.' So it seems that *molu* actually expresses both Neg and V.

But given that insertion only targets terminal nodes, it cannot be the case that *molu* is inserted into the two terminals at the same time. So the question is whether *molu* is inserted under Neg, or under V. The choice feels forced and leads to some obvious issues. If it was inserted under V, we would expect it to combine with negation (which it doesn't). If it was inserted under Neg, we would expect it to combine with a verb (which it doesn't either).

The intuition that the form *molu* in fact conveys the meaning of both 'know' and the negation seems hard to implement. Chung (2007) concludes that within the confines of the DM system, there is only one possible solution. What one has to say is that the structure (6) is

targeted by a special operation, Fusion, which turns the layered representation into a flat node. The procedure is given in (8), taken from Chung (2007:ex.82). The lexical entry (9) is then allowed to apply, since Fusion has collapsed both features under a single terminal node.

(8)



→ [+NEG, KNOW]

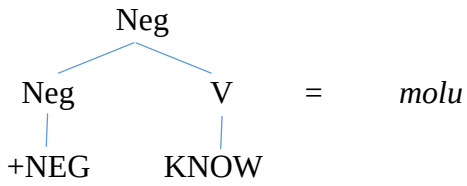
(9) [+NEG, KNOW] = *molu*

The solution in terms of Fusion makes justice to the observation that *molu* conveys both Neg and V, but it leads to a paradox (identified already in Chung 2007:ftn.22). On the one hand, Fusion must precede lexical insertion, because lexicalization targets the structures which Fusion creates. On the other hand, Fusion happens only when the lexicon contains a portmanteau for the fused heads. (Thus, for instance, Fusion cannot happen when the verb is 'sleep,' as in (5)). Thus, an operation which precedes lexicalization is conditioned by lexicalization.

3.3 Non-terminal spell out as a solution to the paradox

Chung then goes on to note that a natural solution for the paradox would be to actually say that the lexical item *molu* spells out the whole structure in (8). Its lexical entry would then look as in (10):

(10)



= *molu*

Once (10) is adopted as a possible format for a lexical item, the paradox disappears. There is no need for Fusion to apply before insertion, because the non-terminal in (8) can directly be mapped onto its pronunciation.[4]

So the natural thing to do at this point would be to simply drop the restriction that Vocabulary Items may only be inserted at terminals. If one could do that, one would get a more general theory of insertion that applies to all syntactic nodes (without a restriction to terminals which is anyway stipulated in (2)). Moreover, the ugly paradox would disappear; we would in fact derive the fact that Fusion is driven by the existence of a particular Vocabulary Item, because insertion at a non-terminal would only be available for such items. Or wouldn't it?

3.4 The communication breakdown

Let us check. Consider for example the Vocabulary Item for SLEEP, given in (11).

(11) *ca-* = [$_V$ SLEEP]

Looking at the item, let us ask the question whether this lexical item can spell out the structure (6). After we have dropped the restriction to terminals, The Subset Principle now says that the entry can spell out a node "if the item matches all or only a subset of the grammatical features" specified in that node. And since (11) corresponds to a subset of the features contained in (6), it seems that *ca-* should be able to spell out that node, in effect meaning both 'to sleep' and 'not to sleep,' a consequence that Chung notes in his ftn. 22. Generalizing this observation, Bye and Svenonius (2010:ftn12) point out that this solution would in fact lead to the expectation that

[4] For completeness, let me also mention that I assume Baunaz and Landers's Principle of Cyclic Override in the discussion (i.e., the spell out of phrasal nodes has preference over terminals) so that (6b) is excluded.

every sentence is pronounced by a single morpheme. I will call this consequence 'the communication breakdown.'

Once Chung realized the grim prospects of simply extending The Subset Principle to non-terminal spell out, he made no attempts at developing the idea further. Compared to the problem of communication breakdown, the Fusion rule in (8) (which he ultimately adopts) looks quite innocent. But is there really no way to avoid this problem?

3.5 Trying to fix things by adding principles: the VIP

One possibility to avoid communication breakdown has been investigated in the work by Radkevich (2010) and Bobaljik (2012). They suggest that adopting in addition to the Subset Principle the 'Vocabulary Insertion Principle' will do the job.

(12) The Vocabulary Insertion Principle (Radkevich 2010:8)

The phonological exponent of a vocabulary item is inserted at the minimal node dominating all the features for which the exponent is specified.

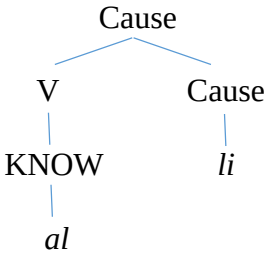
With the VIP in place, some of the most pressing problems of communication breakdown disappear. For instance, (11) can no longer spell out the whole structure in (6), because the minimal node containing all the features of (11) is the V node. Consequently, Neg must be spelled out separately (by one of the negative markers) and we get the right result.

However, (12) ultimately fails in a slightly more complex set of cases. To see that, consider an additional fact noted by Chung (2007), namely that suppletion for negation interacts with causativization. The first piece of the relevant data is in (13a), which is a causative form of the verb 'to know,' meaning 'let know, inform.' The causative component corresponds to the affix *–li*. The structure of the causative as proposed in Chung is shown in (13b). The VIP

based insertion procedure correctly inserts *al-* 'know' at the V node, and *–li* spells out the causative.

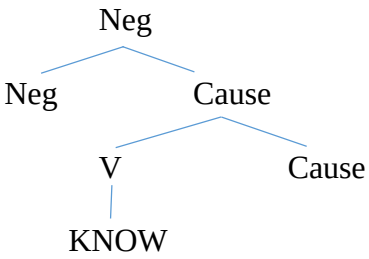
(13) a. *al-li-*
know-CAUS
'let know, inform'

b.



The question to ask is what happens when we negate the causative form, yielding the meaning of 'not to inform,' where the negation scopes over the causative, see (14), reproduced from Chung (2007:ex.86).

(14)



The prediction of The Subset Principle together with the VIP is that the lexical item (10) can be inserted on the top node of the tree (14). That is because the features of the lexical entry (Neg and V) correspond to a Subset of the features in (14), and the top node in (14) is the minimal node containing those features.

But that is a wrong result. As Chung (2007) points out, the structure (14) is actually pronounced as (15a). *Molu* is not possible here, as shown in (15b). If we dropped the *–li*, we would get a non-causative reading.

(15) a. ani / mos al-li-ess-ta
NEG know-CAUS-PAST-DECL
'did not /could not inform'

b. *molu-li-ess-ta
NEG.know-CAUS-PAST-DECL
(Korean, Chung 2007:ex.58)

What is the general lesson we learn? The Subset Principle – which has been widely adopted in DM as a principle governing insertion – fails badly when used as a principle that regulates insertion at non-terminal nodes. The literature that (for good reasons) tries to extend DM for non-terminal spell out adopted a general strategy of preventing problems by *adding* principles. Their goal is to somehow block the unbounded extension of a particular VI up to a point where this single item spells out the whole sentence. The VIP is one possible way to do this, but it fails in a particular set of cases – namely when a VI contains a feature at the bottom (V) and at the top of a tree (Neg), with one or more features intervening in between (Cause). I will refer to this as a "problem with interveners."

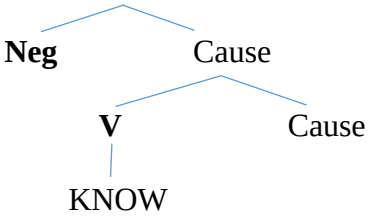
Other principles of a similar sort have been proposed, but they all ultimately fail. I discuss one of these below.

3.6 Global comparisons

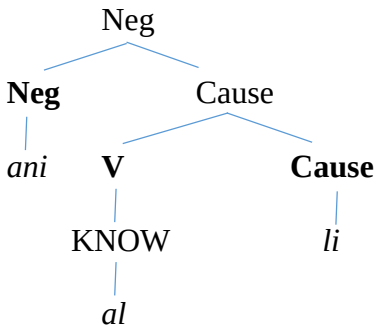
One way to avoid the problem with interveners is to say that there is actually nothing wrong with the fact that *molu* matches the structure in (14). It's just that there is a *better* way of expressing the content of (14), which expresses more features than *molu* alone.

The idea is informally depicted in (16). What we see in (16a) is that when the item *molu* spells out the whole constituent (16a), it leaves the causative feature unexpressed (expressed features are in bold). Compare to this the (correct) spell out in (16b), which – even though more complex in terms of the number of markers – expresses more features. (16c) makes explicit the logic which would lead to the result that (16b) is chosen over (16a).

(16) a. Neg = *molu*



b.



c. Compare possible lexicalizations and choose the one with fewest features unexpressed. (Compare Siddiqui (2006): The most economical derivation will be the one that maximally realizes all the formal features of the derivation with the fewest morphemes.)

This proposal feels quite complex. How many possible lexicalizations are there to consider for more complicated structures? A theory along these lines must have been on Embick and Marantz's mind when they wrote that "when the approaches to blocking involving comparison between otherwise grammatical expressions are made explicit [in our case the comparison of *molu* and *ani-al-li*], they all necessarily involve global competitions …" Embick and Marantz (2008:4).

We will see later that this conclusion is not completely correct, and that The Superset Principle provides a solution that avoids global competitions.[5] Mentioning the quote here serves as an illustration that by weighing the two scenarios in (16) one against the other, we are not pursuing a useless intellectual exercise. It seems that when Embick and Marantz tried to imagine a model of spell out where *molu* blocks *ali-al* (recall (7)), they considered the "global

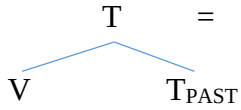
[5] Julien (2015) is an attempt to achieve this goal as well with a modified version of the Subset Principle. I have hard times understanding her proposal, and I think it does not account even for the facts it is claimed to account for in Julien's paper, but there is no space to go through this here. In any case, her account (like any global competition based account) runs against the AbAB problem, which I get to shortly.

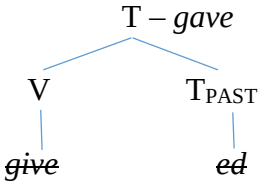
comparison" scenario in (16) an unavoidable consequence. This finding is important when we want to evaluate the actual contribution of NS to the development of a usable phrasal spell out model. Clearly, minimally Embick and Marantz did not see it coming as late as 2008 (which is about the same time when the first Nanosyntax papers started appearing in press, see Caha 2007 and Fábregas 2007, who both give credit to Starke's unpublished work going back to the early 2000s).

Now even admitting that (16) is computationally complex, is it going to work? There is again a set of cases where it won't. In classical instances of phrasal spell out, the expected and regular combination of A and B is blocked by an opaque form C. But there are also cases where the combination of A and B is blocked by A, i.e., a form that is identical to one of the two expected pieces. I will call the the AbAB pattern (**A b**locks **AB**).

One example of such a pattern can be found in English plural formation. As Baunaz and Lander (this volume:ex.31) illustrate, English has some irregular plurals like *mice* which (apparently) block regular plurals such as **mouse-s*. If we extend this treatment to *sheep* blocking **sheep-s*, we get AbAB.

Another case of AbAB is attested for irregular past tense formation. In the classical case, **giv-ed* is blocked by *gave*. One way to encode this is to say that *gave* has the lexical entry in (17a), and so by Cyclic Override, it blocks the analytical spell out *giv-ed*, as shown in (17b).

(17) a. 

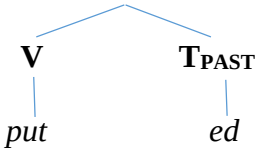
b. 

With the background in place, consider now the verb *put*. Here, the regular form **put-ed* is not blocked by a third form, say **pat*. Instead, the past tense is actually the same as the root (*put*),

so we have a case of AbAB. How can we implement this? According to the Subset Principle, *put* cannot have an entry like (17a), because it can also be inserted under V only. Therefore, it must be lexically specified as V, just like all regular verbs. But with such an entry, it is a mystery why *put* can block **put-ed*, whereas regular verbs cannot. In technical terms, when we run the global comparison model on the two lexicalization patterns in (18), we incorrectly predict that (18b) (where no features are left unexpressed) will be chosen over (18a) (where PAST is unexpressed).

(18) a. T – *put*

b. T



The conclusion is that even admitting global competition does not help, because the signals we are getting from the data are contradictory. For (18), it seems that it is better to use a portmanteau rather than to express all the features. But in (17), it seems that it is better to express all the features, rather than use a portmanteau. I can see no clear way out.

It could of course be proposed that while *gave* blocks **giv-ed*, *put* does not block **put-ed* in the same way; instead, a Fusion rule applies, and yields a single terminal where *put* is inserted. But that brings us back to where we started from: Fusion is paradoxical, so we tried to do better only to find out that this is impossible.

3.7 Giving up the Subset Principle for non-terminal spell out

Technical details aside, the main point is clear: there are attempts in the literature that try to reach two goals at the same time: (i) use DM style insertion and (ii) spell out non-terminals. What we have seen is that they have hard times doing so: the subset logic in (2) which works

so beautifully for paradigms such as (3) breaks down when one tries to extend it to non-terminal spell out.

The strategy has been to avoid these problems by adding principles that somehow restrict the logic inherent in the Subset Principle; we have seen Radkevich's VIP, or (something resembling) a global comparison method to maximize matching of features (Julien 2015). None of these really work, and so we are still waiting for a homegrown DM framework to achieve (i) and (ii). This may well be impossible, as Embick and Marantz seemed to think back in 2008.

But in a sense, all these strategies equal to giving up The Subset Principle for non-terminal spell out, because their ultimate goal is to allow underspecification only at terminal nodes, and neutralize its effects at non-terminals. This is explicitly hinted at in Svenonius and Bye (2010), who try to stay neutral concerning the issue whether spell out is driven by the Subset or the Superset Principle. In their ftn. 12, they write "The Subset Principle […] would lead to the result that every sentence would be at most one morpheme. To avoid this empirically false result, [… w]e might for example assume that [an additional principle of] Exhaustive Lexicalization holds of projecting categories […]". Exhaustive Lexicalization Principle (Fábregas 2007) says that every feature must be lexicalized; in the passage quoted above, Bye and Svenonius note that if the principle only held for heads (each head must be lexicalized), the communication breakdown would be avoided. Ultimately, this approach equals to admitting two insertion principles: The Subset Principle for terminals, and the "Spell Out Every Head" Principle for non-terminals. (But even that will still not explain the AbAB pattern, because here one head goes unexpressed.)

4. The Superset Principle

There is a single reason underlying all these puzzles: the insertion procedure is not governed by the Subset Principle. In order to make things work, we have to adopt the Superset Principle

instead (Starke 2009). I give one possible formulation below, which is quite different from Starke's wording, but it is close to the original Subset Principle (so that we can see what the difference actually is). It is also not too remote from Starke's idea.

(21) The Superset Principle

[*The Superset Clause*:] The phonological exponent of a Vocabulary Item is inserted into a node if the item contains all (or a superset of) the grammatical features contained in the node. Insertion does not take place if the vocabulary item is not specified for all features contained in the node. [*The Elsewhere Clause*:] Where several items meet the conditions for insertion, the item containing fewer features unspecified in the node must be chosen.

Potentially, this principle may apply to terminal, non-terminal and phrasal nodes. Let me now show how adopting it solves all the puzzles we have seen.

4.1 AbAB

I start from the AbAB. With the Superset Principle in place, we can specify both *put* and *gave* in the same way, namely as irregular past tense forms, see (22a). With this lexical entry, they can spell out the non-terminal (17b), because they contain all its features. This is not the case for regular verbs, which have a lexical entry like the one in (22b). They do not contain the T node, and so they cannot be inserted at the non-terminal in (17b).

(22) a. *put, gave, sang, …* = [V T_{PAST}]

b. *kiss, locate, …* = V

c. *give, sing, …* = V

Notice now that all the verbs in (22a) can also spell out the V node only, because V is contained in them. So this explains why *put* can have the same form for the past tense and the root, while regular verbs can't.

Finally, the reason why *gave* and *sang* do not show up as the root form of the relevant verb is because of competition. The VIs *sing* and *give* also spell out the V node, and in this capacity, they contain fewer unused features, so they outcompete the past tense form in this environment.

4.2 Korean Negation

Recall now the problem we had with *molu.* It was specified as [Neg KNOW] (recall (10)) and the question was why it can't spell out the causative form [Neg [KNOW Cause]], recall (14). The answer is now clear: it is because (10) does not contain all the features of the top node in (14); Cause is missing.

4.3 How to do a simple paradigm

Recall finally the paradigm (3), repeated below in (22) for convenience. How does the Superset Principle deal with this example?

(22) A fragment of the Latin declension

[+SINGULAR]	neuter [-MASC]	masculine [+MASC]
nom [+SUPERIOR]	*-um*	*-us*
acc [-SUPERIOR]	*-um*	*-um*

Let me first say that because of independent assumptions (recall **dif 1**), there are no complex terminals in NS, so the following discussion is artificial with respect to the actual working of the Superset Principle in that framework. Still, for completeness of the argument, let me show that the Superset Principle could also be used to handle these cases.

If we want to encode the same intuition as in (4), namely that *–um* is a marker that can appear in any cell of the paradigm, it has to contain all the potential feature combinations. Its specification is thus as shown in (23a).[6]

(23) a. *-um* = [+MASC, - MASC, + SUP, - SUP, + SG]

b. *-us* = [+ MASC, + SUP, + SG]

If we now specify *–us* as shown in (23b), it will only qualify for insertion in the nominative singular of the masculine gender, because it does not contain the features of any other cell. The lexical entries in (23) thus derive the paradigm (22), which shows that we haven't lost the cases that were working fine under The Subset Principle driven insertion.

5. Conclusions

By looking at the failures of classical DM, we come to realize that NS is not just a particular constellation of the bits and pieces that were independently available before (like Cartography and phrasal spell out). Without changing the principle of insertion, the combination of these proposals simply does not take off. In retrospect, employing the Superset Principle and thereby creating a well working theory seems trivial. At the same time, when Embick and Marantz were contemplating these issues back in 2008, the prospects of any breakthrough in this domain

[6] This entry is quite ugly, a side effect of the terminal-spell-out scenario. See Caha (2009) for a system of case decomposition which makes the lexical items look neater.

seemed impossible (without entering the shaky grounds of global comparisons). From that perspective, the Superset Principle and NS is general have opened a whole new field of competition in grammar to look at. Unlike the half-working alternatives, it has made it possible to abandon Fusion and the paradoxes that came with it. By doing so, it creates the prospect for a simplified architecture of grammar where syntactic structures are mapped on their pronunciation by lexical access only.

6. References

Baker, M. C., & Kramer, R. (2014). Rethinking Amharic prepositions as case markers inserted at PF. *Lingua*, 145, 141-172.

Baunaz, Lena and Lander, Eric. this volume. *Nanosyntax: The basics*.

Bobaljik, J. (2007). *On comparative suppletion*. Ms., University of Connecticut.

Bobaljik, J. (2012). *Universals in Comparative Morphology. Suppletion, Superlatives, and the Structure of Words.* The MIT Press.

Bye, P., & Svenonius, P. (2010). *Exponence, phonology, and non-concatenative morphology*. Ms. CASTL, University of Tromsø.

Caha, P. (2007). Case movement in PPs. *Nordlyd*, 34(2), 239-299.

Caha, P. (2009). The nanosyntax of case. Doctoral dissertation, CASTL, Tromso.

Chung, I. (2007). Suppletive negation in Korean and distributed morphology. *Lingua*, 117(1), 95-148.

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In: *Paths towards Universal Grammar: Studies in honor of Richard S. Kayne*, eds. G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi & R. Zanuttini, 85–110. Georgetown University Press.

Cinque, G. (2005). Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry*, 36(3), 315-332.

Embick, D. (2007). Linearization and local dislocation: Derivational mechanics and interactions. *Linguistic Analysis*, 33(3-4), 2-35.

Embick, D. (2013). *On the targets of phonological realization*. Ms., UPenn.

Embick, D., & Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry*, 39(1), 1-53.

Embick, D., & Noyer, R. (2001). Movement operations after syntax. *Linguistic inquiry*, 32(4), 555-595.

Fábregas, A. (2007). The exhaustive lexicalisation principle. *Nordlyd*, 34(2), 165-199.

Halle, M. (1997). Distributed Morphology: Impoverishment and fission. In: PF: *Papers at the interface*, ed. by Benjamin Bruening, Yoonjung Kang, and Martha McGinnis, 425–450. Cambridge, MA: MIT.

Harley, H. & Noyer, R. (1999). Distributed Morphology. *Glot International* 4(4), 3-9.

Haugen, J. D., & Siddiqi, D. (2013). Roots and the derivation. *Linguistic Inquiry* 44, 493–517.

Julien, M. (2015). Theme vowels in North Sámi: Spanning and maximal expression. *Lingua*, 164, 1-24.

Koopman, H. (1996). The spec head configuration. In: *Syntax at Sunset: UCLA Working Papers in Syntax and Semantics*, edited by Edward Garrett and Felicia Lee.

Leu, T. (2008). *The internal syntax of determiners*. Doctoral dissertation, NYU.

Marantz, A. (1989). Clitics and Phrase Structure. In: *Alternative Conceptions of Phrase Structure*, ed. by M. Baltin and A. Kroch, 99-116. University of Chicago Press, Chicago.

Merchant, J. (2015). How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry* 46, 273–303.

Miller, P. (1992). *Clitics and constituents in Phrase Structure Grammar*. London: Garland.

Pollock, J. Y. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry*, 365-424.

Pullum, G. K., & Zwicky, A. M. (1991). A misconceived approach to morphology. In: *Proceedings of the West Coast Conference on Formal Linguistics*, Vol. 10, 387-398.

Radkevich, N. V. (2008). *On location: The structure of case and adpositions*. Doctoral dissertation, University of Connecticut.

Siddiqi, D. (2006). *Minimize exponence: Economy effects on a model of the morphosyntactic component of the grammar*. Doctoral dissertation, University of Arizona.

Starke, M. (2004). On the inexistence of specifiers and the nature of heads. In: *Structures and Beyond: The Cartography of Syntactic Structures, vol. 3*, edited by Adriana Belletti, 252–268. Oxford University Press, New York.

Starke, M. (2009). Nanosyntax: A short primer to a new approach to language. *Nordlyd*, 36(1), 1-6.

Siddiqi, D. (2006). Minimize exponence: Economy effects on a model of the morphosyntactic component of the grammar (Doctoral dissertation, University of Arizona).

van Craenenbroeck, J., & Merchant, J. (2013). Ellipsis phenomena. In: *The Cambridge Handbook of Generative Syntax*, ed. by M. den Dikken, 701–745. Cambridge: Cambridge University Press.