Multiple Multiple Spellout*

Meaghan Fowlie

1 Introduction

In the Minimalist Program (Chomsky, 1995), Spellout is the syntactic operation that delivers to the phonological component of the language faculty everything it needs in order to form an utterable sentence. This includes all of the phonological features of the terminal nodes and the order in which they are to appear. Spellout must also be assumed to deliver to the semantic component everything it needs in order to form a comprehensible sentence. Whether this LF-Spellout is the same operation as PF-Spellout or a separate operation is yet unclear. However, it is common to use the term *Spellout* to refer primarily to PF-Spellout – consider for example Nunes and Uriagereka (2000) – and as such I will be using the term *Spellout* mainly for PF-Spellout.

Bresnan (1971) introduced the notion of Multiple Spellout: that Spellout can occur several times throughout the derivation. Working within Minimalism, researchers ask what exactly Spellout is, and how a Spelled-out constituent behaves. Uriagereka (1999) in particular explores Multiple Spellout in the context of Minimalism. Given that phonological features and linear order are quite different sorts of entities, I suggest that Spellout is in fact two operations: one that assigns linear order to the terminal nodes of a constituent, and one that sends phonological features to the phonological component, rendering the constituent inaccessible to the derivation.

The first of these operations, LINEARISE, I base closely on Fox and Pesetsky (2005)'s operation *Linearise* in combination with Nunes and Uriagereka (2000)'s revised Linear Correspondence Algorithm, and the second, ATOMISE, is based on Nunes and Uriagereka (2000)'s Spellout. Of particular interest is the free word order typology predicted by the interface of these two operations.

*This article is based on my undergraduate thesis written at McGill University in 2007. I would like to thank the many students and professors there who took time to advise me on this project, in particular Jon Nissenbaum, Tobin Skinner, Raphael Mercado, and especially my undergraduate thesis supervisor Lisa deMena Travis, who was also responsible for pointing out the possibilities this approach presents for free word order. I extend many thanks as well to the Principles of Linearisation workshop at GLOW 31 and the UCLA syntax-semantics seminar for advice and questions and support, as well as to Ed Stabler for opening my eyes to the importance of computational considerations.

Published in: Theresa Biberauer and Ian Roberts, editors, *Challenges to Linearization*, volume 114 of Studies in generative grammar, pp. 129-69. Mouton de Gruyter, Berlin, 2013.

This paper will begin with four seemingly unrelated puzzles, to which I will suggest solutions in section 7. I then give some necessary background in linearisation and Spellout, including Kayne (1994), Fox and Pesetsky (2005) and Nunes and Uriagereka (2000). In section 4, I will present the details of my proposal and we will look closely at an example sentence. Finally, sections 5 and 7 explore the gains made by separating Spellout into two operations; in particular, section 7.4 presents the free word order predictions.

2 Some Mysteries

Here I will present four seemingly unrelated mysteries: CED and PIC effects, exceptions to CED effects, the lack of CED effects in objects in OV languages, and free word order. Section 7 will show how my proposal can account for all four of them.

2.1 CED and PIC Effects

The Phase Impenetrability Condition (PIC) and Conditions on Extraction Domains (CED) are very similar. They both govern movement out of a constituent. On the face of it, it seems like they might be able to be collapsed as an effect of Spellout, but it is not immediately clear how.

> **Phase Impenetrability Condition (PIC)** If H is the head of a phase, HP is not accessible to operations outside HP; only H and its edge are accessible (Chomsky, 2001).
>
> **Condition on Extraction Domains (CED)** (Huang, 1982): Verbal complements can be extracted from but adjuncts and specifiers cannot.

If phases are spellout domains, multiple spellout gives a nice solution to phase impenetrability: once something is spelled out, it is no longer accessible to syntax. Phases do not seem to help with CED effects: embedded CPs – classic phases – are often *object* CPs. The CED says objects are *easier*, not *harder*, to escape from. Moreover, CED means *all* extraction out of specs & adjuncts is impossible, not just non-edge extraction.

2.2 Exceptions to CED

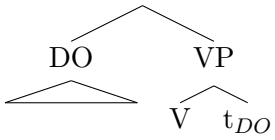
Sometimes specifiers and adjuncts can be extracted from. For example, in some free word order languages like Dyirbal, noun markers, adjectives, and nouns can be separated. There are also languages like Hindi whose word order is not so free, but which allow possessors to be extracted from DPs other than direct objects:

(1) Kiskii tum socte ho ki [$_{SUB}$ t$_{wh}$ kitaab] corii ho gayii
whose you think that [$_{SUB}$ t$_{wh}$ book] stolen was
'Whose book do you think was stolen?'

2.3 Objects in OV languages

Under Antisymmetry, direct objects in OV languages start as rightward complements of V, and then move to a position to the left of V.

(2)



All movement is necessarily to a specifier or adjunct position, so we expect direct objects in OV languages to show CED effects. They do not, as this example from Dutch shows:[1]:

(3) [Van wie] heb je [$_{DO}$ een foto t] gezien?
[Of whom] have you [$_{DO}$ a picture t] seen?
'Who did you see a picture of?'

2.4 Free Word Order

Free Word Order: Optional and semantically vacuous word-order variation

Some "free word order" is clearly A-movement as it has binding effects (Saito, 1992). However, some free word order is very likely not movement at all. An example is Dyirbal (Dixon, 1972).[2]

(4) a. **bayi** **wangal** *bangul* *yarangu* *bulganu* banggun dugumbiru
the-NOM boomerang-NOM the-GEN man-GEN big-GEN the-ERG woman-ERG
buran
see-PRES/PST
'The woman saw the big man's boomerang'

b. **bayi** *yarangu* dugumbiru buran **wangal** banggun
the-NOM man-GEN woman-ERG see-PRES/PST boomerang-NOM the-ERG
bangul *bulganu*
the-GEN big-GEN

The **bold** (nominative), *italics* (ergative) and plainface type (genitive) are to clarify which elements have the same case. Notice that in this second order, noun markers are separated from their nouns and adjectives are separated from the nouns they modify.

All word-orders are claimed to be grammatical. Indeed, example (4-b) was chosen here because this order was "made up" by Dixon to illustrate dramatic free word order. He

[1] Dutch data from Floris van Vugt p.c. 2010

[2] r = semi-retroflex continuant, d = lamino-palatal/alveolar. *the* in the glosses are noun-markers. Abbreviations are as follows: NOM = nominative case GEN = genitive case ERG = ergative case PRES/PAST = present or past tense

writes:

> A well-known linguist took exception to this, categorically denied that freedom of word-order of this magnitude was possible in any language, and accused the writer of exaggerating. [(4-b)] was put to informants at the next opportunity, and they castigated the writer for asking a trivial and unnecessary question – "you know that's alright!" (Dixon 1972 p. 107-8)

There are eight words in the sentence so there are $8! = 40\ 320$ possible word-orders. If free word-order were optional movment, there would have to be 40 320 optional movement combinations to account for these sentences alone. Moreover, Minimalism requires motivation – generally in the form of features – for every Move. Thus each word order must have unique features in its numeration. There must be 40 320 separate numerations if the orders are created by movement. Finally, it must be semantically vacuous movement. A better solution would derive all orders from one structure.

3 Background

The model of Spellout I propose is a combination of the Linear Correspondence Axiom (Kayne, 1994; Nunes and Uriagereka, 2000) and Fox and Pesetsky (2005)'s *Linearise* operation. Here I will outline each of these proposals.

3.1 Linear Correspondence Axiom (LCA)

X-bar theory offers no algorithm to map hierarchical structure to the surface linear form of language. Any pair of sisters can be stipulated to appear in either order.

In his 1994 monograph *The Antisymmetry of Syntax*, Richard Kayne proposed that linear order is in fact derivable from hierarchical structure. He showed that it was possible to derive a restrictive version of X-bar theory from c-command relations. In particular, he proposed the Linear Correspondence Axiom (LCA):

> **Linear Correspondence Axiom** (Kayne, 1994): For any pair of non-terminal nodes $\langle X, Y \rangle$, if X asymmetrically c-commands Y then each terminal node dominated by X precedes each terminal node dominated by Y. Moreover, the set of all such correspondences constitutes a total ordering on the terminal nodes.

Kayne assumes irreflexive dominance and that the terminal nodes (e.g. lexical items) project up to a syntactic head without branching. At least one of these two assumptions is necessary to derive a total ordering on the terminal nodes.

3.2 Fox and Pesetsky

Fox and Pesetsky (2003, 2005) propose that Spellout fixes the relative order of the lexical items in a Spelled-out domain. At the end of the construction of each Spellout Domain (SD) D_i,

1. The elements of D_i are linearised by some linearisation algorithm
2. Resultant ordered pairs are stored on an *Ordering Table*

They call this operation *Linearise*. The rule of *Order Preservation* states that no information is ever lost from the Ordering Table.

When the next SD D_{i+1} is Linearised, it is Linearised with respect to the first element of the previously spelled-out SD D_i. "First" is defined by the previous application of Linearise.[1] After an SD is Linearised, constituents thereof can still move out of the SD, but they cannot change their order. This elegantly derives successive cyclicity effects such as Quantifier Movement in Scandinavian langauges and the Holmberg Generalisation, including aspects that are rarely accounted for. Fox and Pesetsky also derive other types of successive cyclicity. Consider, for example, English object *wh*-movement. The object *wh*-word in its theta-position is after the verb and subject. For example:

(5) $[_{CP}$ what did $[_{vP}$ ~~what~~ he read ~~what~~$]]$

Ordering Table (vP)
what < he
he < read
what < read

Ordering Table (CP)
what < he
he < read
what < read
what < did
what < vP

In the final form of the sentence, *what* is before the verb. If vP is an SD, *what* must be before the verb when the vP is Linearised. Otherwise, the Ordering Table will store the information that *read* precedes *what*, and then when the next SD is Linearised, the Ordering Table will receive the information that *what* precedes *read*. This contradiction will crash the derivation.

What exactly constitutes a Spellout Domain is the subject of debate, and neither Fox and Pesetsky, Nunes and Uriagereka, nor I fully agree on this point. My own proposal will be tentative, and based on a combination of Cinque (1999)'s adverbial domains and free word order data.

In their 2003 handout, Fox and Pesetsky propose a possible linearisation algorithm that relies on stipulated ordering of heads, specifiers, and complements. A major advantage

over Nunes and Uriagereka's version of the LCA is that simple sisters can be linearised (Cf. section 3.3 below). Nevertheless, Fox and Pesetsky comment: "we suspect that all or most of our proposal could be reformulated if c-command, rather than sisterhood, were the central notion for the Laws of Precedence (as in Kayne 1995)" (Fox and Pesetsky (2003) p. 15). We will not examine their proposal for laws of precedence in any more detail, as ultimately I will take them up on their suggestion that the central idea of their theory is not dependent on any particular linearisation algorithm. In fact, I will be using the LCA as the ordering algorithm in my proposal.

3.3 Nunes and Uriagereka

In contrast to Kayne, Nunes and Uriagereka (2000) propose that terminal nodes do not project up to a syntactic head before branching, i.e. they maintain the Bare Phrase Structure assumption that there is no head-terminal distinction. They also propose an LCA which is simpler than Kayne's 1994 version. In particular, they remove the notion of dominance from the definition of the LCA.

> **Linear Correspondence Axiom (Nunes and Uriagereka 2000)**: A Lexical Item α precedes a Lexical Item β iff α asymmetrically c-commands β.

The removal of dominance from the definition of the LCA means that Nunes and Uriagereka are no longer interested in anything but terminal nodes. This is in contrast to Kayne, who uses mathematical relations among non-terminals to determine linear order of terminals. Nunes and Uriagereka determine linear order of terminals directly from the mathematical relations among the terminals themselves.

Clearly, this simplified version of the LCA fails to yield a total ordering on the Lexical Items in many sentences. For example, in a phrase with a complex specifier, there is no asymmetric c-command relation between the elements of the specifier and those of the sister of the specifier.[2] Nunes and Uriagereka seek to derive such an asymmetric c-command relation using Bresnan's notion of Multiple Spellout.

> **Multiple Spellout** (Bresnan, 1971): Spellout may occur more than once in the course of a derivation.

Nunes and Uriagereka propose that Spellout linearises according to their LCA (above). After being spelled out, the structure of the constituent is removed from syntax. Only the label, including the syntactic features – which encode everything the syntax needs to know to manipulate the spelled-out constituent as a single unit – remains in the derivation. This label, being a terminal node, behaves as a single lexical item. It acts as a "bookmark" to note the location of the spelled-out constituent in the main structure. When PF is generated, the phonological component uses the label to determine where the constituent belongs in the order.

Under this theory, specifiers must be spelled out before merging with the main structure. What is merged is only a label, which appears to the derivation as a terminal node. This simple specifier is clearly in an asymmetric c-command relation with the subconstituents of its sister.

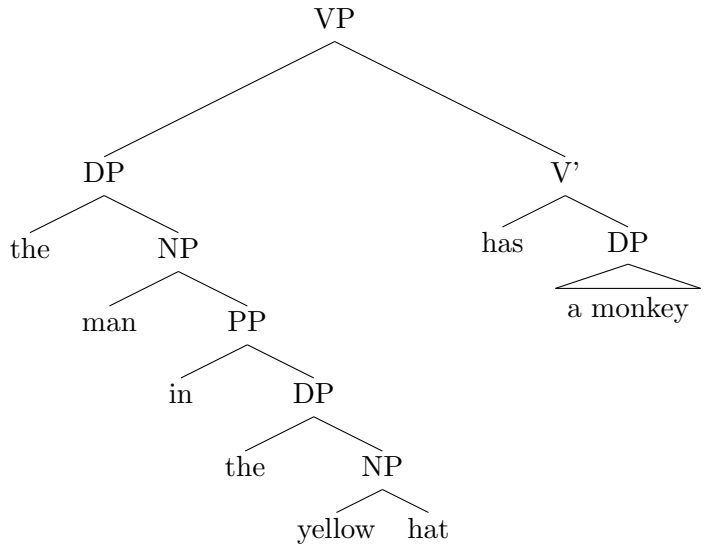


Figure 1: Lexical Items of a complex specifier have no c-command relationship with Lexical Items of V'

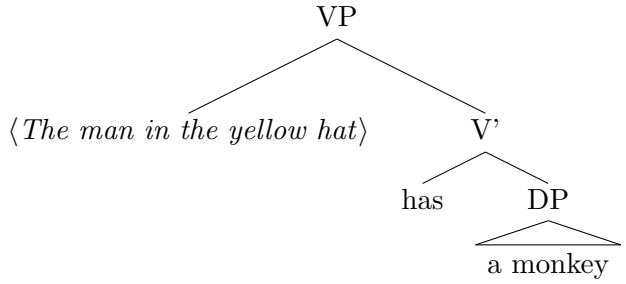


Figure 2: A Spelled-out specifier is a terminal node and is in an asymmetric c-command relation with Lexical Items of V'.

4 Proposal

Both Fox and Pesetsky and Nunes and Uriagereka propose compelling accounts of the mechanisms of the Spellout operation. Fox and Pesetsky are able to account for cyclicity

effects and order preservation effects. Their tentative linearisation algorithm (Fox and Pesetsky, 2003) is effective, but stipulative. Nunes and Uriagereka use the LCA, which derives order rather than relying on stipulation. They also account for extraction domain effects such as *wh*-islands and subject islands: since any Spelled-out constituent is no longer part of the derivation, it is clearly inaccessible for extraction.

It is tempting to claim they are both correct. The problem with doing so is that Fox and Pesetsky need to be able to access spelled-out constituents to account for object shift, *wh*-movement, and so on, while Nunes and Uriagereka stipulate that this is impossible. A closer look at the problem reveals that Fox and Pesetsky and Nunes and Uriagereka are looking at different domains. Fox and Pesetsky are interested in the "spine" of the tree and its sisters, while Nunes and Uriagereka are looking at the internal structure of the "satellites".

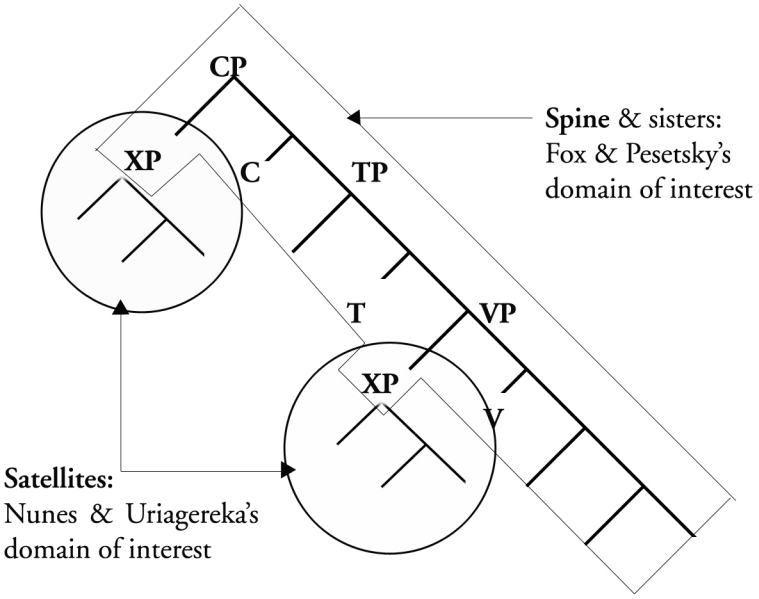


Figure 3: Spine and Satellites

In technical terms, *spine* refers to the extended projection line of the semantic head (i.e. the verb) of the tree; informally, this is the main trunk of the tree and the heads thereof. The *satellites* are essentially the specifiers and adjuncts: anything but the spine. (see Figure 3.)

I claim that both conceptions of Spellout are correct but incomplete. Two separate operations apply: LINEARISE and ATOMISE.

4.1 Atomise

Atomise is essentially Nunes and Uriagereka's Spellout. It sends the phonological features of the constituent to the phonological component of the grammar and in so doing renders the constituent inaccessible to the derivation. Atomise applies only when the derivation is "done" with the constituent. The label left behind acts as an atom in the derivation. However, unlike Nunes and Uriagereka's Spellout, atomise does not linearise the constituent.

4.1.1 LF Spellout

Nunes and Uriagereka claim that after Spellout, "there is literally no syntactic object within [the Spelled-out domain]" (Nunes and Uriagereka (2000) p. 25). Taken literally, this means that LF receives no structure whatsoever by the end of the derivation. This is clearly impossible: semantics relies heavily on structure. We are left with two possibilities:

1. The syntactic structure is still there, but simply inaccessible to the syntactic derivation: a closed "suitcase" that can be moved around but which only LF can open.

2. When a constituent is atomised it is indeed removed from the derivation, but it goes both to PF and to LF. In other words, there is also LF Spellout, as per numerous theories (eg. Chomsky (2001)).

It would be desirable to preserve Nunes and Uriagereka's intuition that the spelled-out constituent is literally gone from the derivation.[3] This leads us to option two: LF Spellout exists. Suppose this is true. Then we must ask, does LF Spellout occur simultaneously with PF Spellout? Within this framework, LF Spellout must occur whenever PF spellout occurs; otherwise LF would not receive structure from the derivation. LF-Spellout may also occur separately from atomise, but this can only be true if there are more applications of LF Spellout than atomise.

4.2 Linearise

Linearise is Fox and Pesetsky's Spellout/Linearise, with the ordering algorithm specified as Nunes and Uriagereka's LCA. Linearise takes a "snapshot"[4] of the derivation so far and applies Nunes and Uriagereka's LCA, storing the resultant ordered pairs on the Ordering Table. Linearise treats an atomised constituent as a terminal node. This allows satellites to linearise with their sisters.

A linearised constituent is still accessible to the derivation, unlike under Nunes and Uriagereka. Linearise simply marks the order of Lexical Items. As noted in section 4 above, Fox and Pesetsky require accessibility to the spine of the tree in order to derive cyclicity effects. As long as linearise but not atomise applies to the phase, accessiblity is retained, but Spellout has still occured.

4.3 A note on economy

Phases are meant to reduce computational load. Whether any of the proposals presented here (Nunes & Uriagereka's, Fox & Pesetsky's, and mine) would truly do so hangs on what is costly and what is cheap. If parallel processing has a very large effect on overall time or energy required, all three proposals have a strong advantage there. However, Fox & Pesetsky's proposal, and therefore mine as well, requires building and remembering a partial order, and then turning it into a linear order in the end. This may be quite costly.

If removing structure from the derivation significantly reduces work in the derivation, both Nunes & Uriagereka's and my proposals have an advantage here as well. However, any proposal that takes sentences apart for parallel processing must also put them back together at least in part. For example, in these two proposals, the structures do not need to be put back together but the linear order does: the "bookmarks" need to be replaced by the terminals that they stand for. This too could be quite costly.

4.4 Spellout Domains

The precise nature and location of Spellout Domains is much debated, and, as discussed earlier, is not agreed on between Fox and Pesetsky (2005) and Nunes and Uriagereka (2000). My claims here regarding the nature and junctures of SDs are tentative, and should not be taken as more than examples of how a split Spellout might function.

Like Fox and Pesetsky (2005), I take Spellout Domains to roughly correspond to phases (Chomsky, 2001). I remain agnostic as to whether there are separate numerations for each phase, or if phases are *defined* by Spellout Domains.

4.4.1 When spellouts occur

I have no reason to suppose that ATOMISE and LINEARISE have different domains. However, I do have evidence that LINEARISE occurs (more or less) as soon as a phase is completed, while ATOMISE is delayed until the next phase is completed.

The evidence that LINEARISE is not delayed comes directly from Fox and Pesetsky (2005). Their theory of linearisation requires that ordering information be recorded as soon as a phase is completed, hence cyclic patterns such as Holmberg effects.

That ATOMISE is delayed follows directly from basic facts like *wh*-movement: if an embedded *wh*-clause were ATOMISED immediately, the *wh*-word would not be able to escape.

4.4.2 Where spellouts occur

I claim that the junctures at which these operations may occur is universal, but that whether they occur or not is parameterised. Ideally, the Spellout Domains should be motivated both theory-internally and by natural language data. The precise locations of

SDs is less important than the basic idea of universally parameterised domains over which one or more spellout operation may occur.[5]

I have taken as a starting point the phases suggested by Chomsky (2001): CP and vP, and follow many authors including Fox and Pesetsky in adding DP. From there I suggest that Cinque (1999)'s hierarchy adds TP as a phase, as the TP domain of the sentence is characterised by sentential functional heads, while vP is characterised by the argument structure of the sentence and lexical adverbs, and CP is characterised by propositional attitudes.[6]

Empirical evidence for TP phasehood comes from free word order data in section 8: if verb-initial languages are generally taken to have V move to C, the phasehood of TP will explain the free-word-order effects in Tagalog and Warlpiri.

It has also been proposed (Bittner and Hale, 1996) that DP can be separated in a parallel fashion: [KP [DP [NP]]] (where KP = case phrase).[7] We will see evidence in sections 8.2 and 8.3 below that there exists at least one phase edge within satellites.

For simplicity's sake, the model I will present in section 6 below has only five spellout domains, and the worked example in section 4.6 below has only three.

4.4.3 When a Spellout Domain is finished

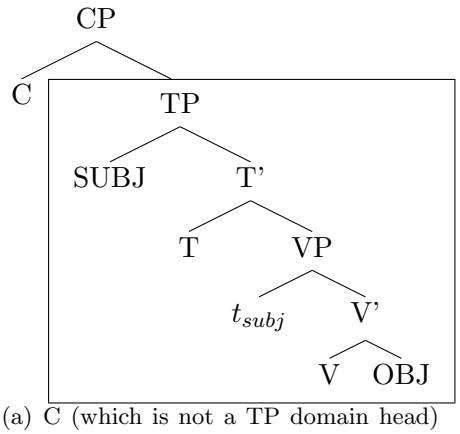
A phase is finished when a Spellout Domain is finished being built. What signals this completion depends on what defines a phase. If each phase is marked by a separate numeration, it is simplest to assume that the SD is finished when the numeration is spent.

If there is no numeration, and if the upper end of a phase is not always present (e.g. if not all CPs have a SpeechActP), the last merging node must somehow "know" it is the last. I suggest that the merger of a new head that does not belong to the same domain (in the sense of Cinque (1999)) as the previous projections triggers Spellout of its sister – the phase. This is illustrated in figure 4(a).

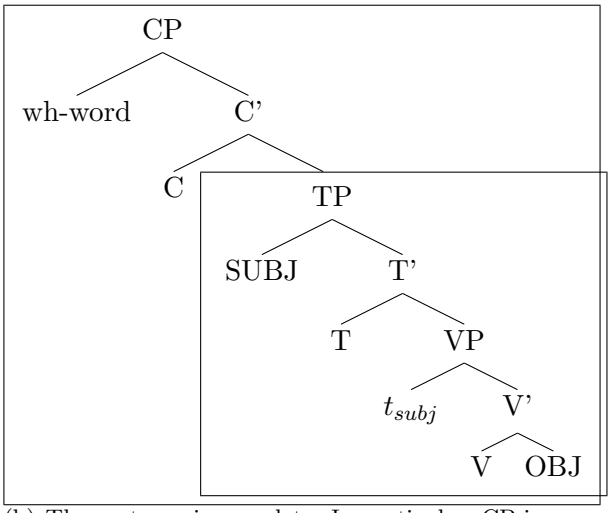
It is necessary to stipulate that a shift in head type is not the only motivation for spellout operations. When the work in a derivational workspace is completed, there is no head merging above it to signal the SD is complete. However, we still want spellout operations to be available at this juncture.

Whether or not satellites are ATOMISED as soon as they are built – i.e. before they are merged to the main tree – must also be parameterised. As noted in section 2, some languages like Hindi allow movement out of a left branch. Additionally, languages like Dyirbal which allow discontinuous expressions even with satellites cannot, under this proposal, have ATOMISED satellites; otherwise the constituent will stick together even if it is freely-ordered internally.

Finally, when the sentence is complete, ATOMISE in particular must be available, or else parts of the sentence will not be pronounced.[8] Please see figure 4(b).



(a) C (which is not a TP domain head) spells out its sister, TP



(b) The sentence is complete. In particular, CP is complete, but there is nothing Merged above CP, and therefore no sister to spell CP out. CP must be spelled out on its own.

Figure 4: TP and CP spellout

4.5 Parameters

The basic model is this: each Spellout Domain D is specified for whether each operation occurs, thus:

$$\mathbf{D}[\pm\text{L}\pm\text{A}]$$

For example, $\mathbf{D}$[+L–A] is LINEARISED but not ATOMISED .
There are four possibilities for each SD:

	+L	–L
+A	[+L+A]	[–L+A]
–A	[+L–A]	[–L–A]

Table 1: Four possibilities for each Spellout Domain

1. [+L+A]: Linear order is strict; nothing can escape. e.g. English Subjects
2. [+L–A]: Linear order is strict; escape is possible. E.g. Scandinavian VP
3. [–L–A]: Linear order is not (yet) determined; escape is possible. i.e. acts like non-phase

4. [–L+A]: No linear order determined; nothing can escape. e.g. Dyirbal CPs

In addition to the parameters for each phase, there is also a Workspace parameter **W**[±A]. If a language is **W**[+A] then the constituent built in a workspace is ATOMISED as soon as it is complete. What is merged to the main tree is then an "atom", a single node. If the language is **W**[–A], satellites are merged as complex trees.

These parameters combine to make Spellout Types. I will write them just as a list of SDs and their types. For example, supposing there are three SDs A, B, C, we have:

W[±A] **A**[±L±A] **B**[±L±A] **C**[±L±A]

4.5.1 The Procedure for Spellout

When a Spellout Domain *D* is complete:

1. Check for SDs embedded in *D*
2. For each embedded SD *E*, if *E* is [+A], ATOMISE *E*
3. If *D* is [+L] LINEARISE *D*

When you are finished with a workspace:

1. If the phrase built is a [+L] SD, LINEARISE it
2. If the language is [+W], ATOMISE the phrase

When a sentence is finished:

1. If the SD of the finished sentence is [+L], LINEARISE the finished sentence
2. Always ATOMISE the finished sentence

4.6 An illustration

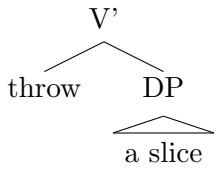
Let us now look at a worked example: building the English sentence *The monkey with pizza will throw a slice*. In the interests of sticking to bare phrase structure as much as possible, this declarative English sentence will be assumed to be only a TP. This model is simplified, with only three SDs: DP (**D**), VP (**V**) and CP (**C**), and with a VP-internal subject.

I will assume here that English LINEARISES all three SDs, and that it ATOMISES DPs. I will also assume, as suggested in section 4.4.2, that the merger of a new type of head signals the completion of a phase. Finally, I assume that English spells out the constituent built in a workspace when it is finished. That is, I will assume English is of type

W[+A] **C**[+L+A] **V**[+L–A] **D**[+L+A]

Instead of an Ordering Table I will notate the order relation being built with a Hasse diagram, which is easier to read. In a Hasse diagram, a precedes b if there is an arrow from a to b. The notations are equivalent.

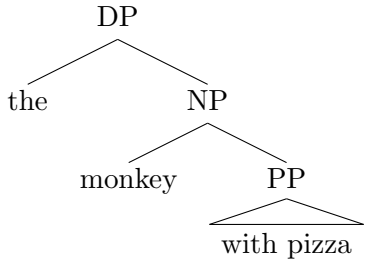
Step 1. Merge V' = *throw a slice*[9]



Step 2. Transition from DP domain to VP domain is signaled by Merger of V. LINEARISE *a slice*.

Figure 5: Hasse diagram of word order: in the final sentence, *a* must precede *slice*.

Step 3. Meanwhile, Merge DP = *the monkey with pizza* in a separate workspace



Step 4. Workspace work complete: LINEARISE *the monkey with pizza*

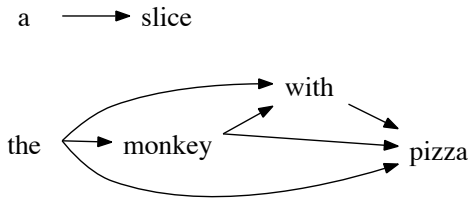
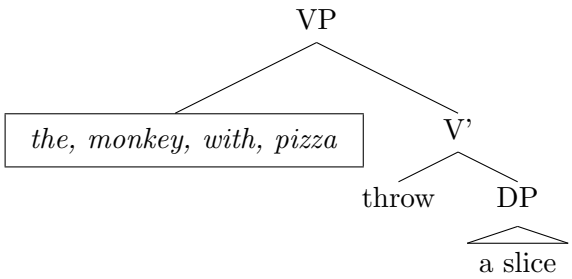


Figure 6: Hasse diagram including LINEARISED subject

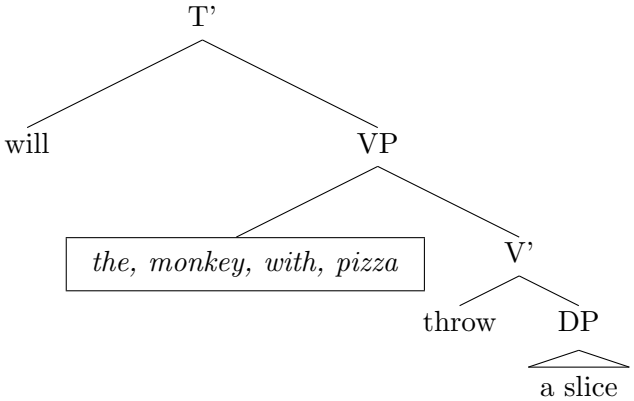
Step 5. ATOMISE *the monkey with pizza*

The phonological features of *the monkey with pizza* are sent to Phonological Component. Let $\boxed{\textit{the, monkey, with, pizza}}$ represent the ATOMISED constituent.

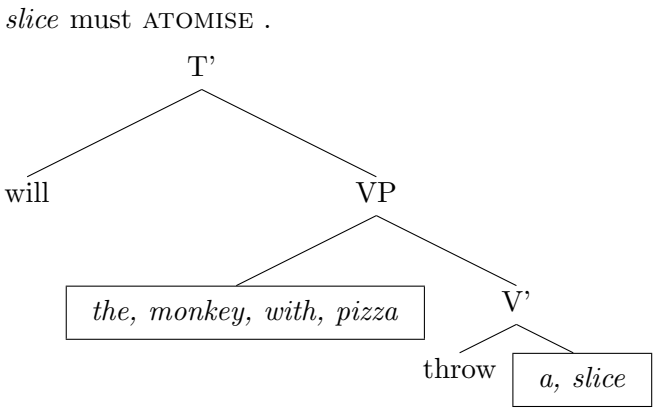
Step 6. Merge $\boxed{\textit{the, monkey, with, pizza}}$ with *throw a slice*



Step 7. Merge *will*



Step 8. VP now being complete, spellout operations may take place. First, we look inside VP for any constituents that are parameterised to ATOMISE in English. The DP *a*



Step 9. LINEARISE may now take place. Recall we are assuming that English is parameterised for LINEARISE of VP. The ATOMISED (boxed) constituents are now terminal nodes, and I'll just use the boxed words as their labels here.

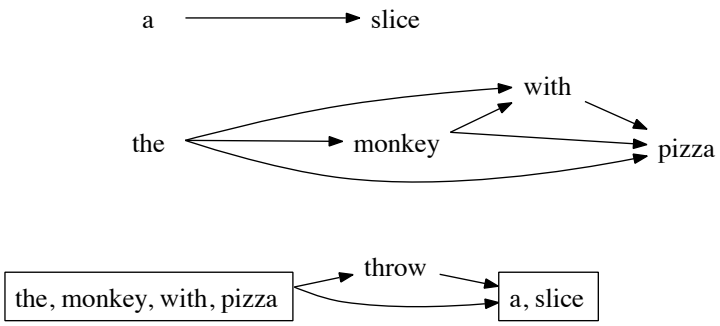
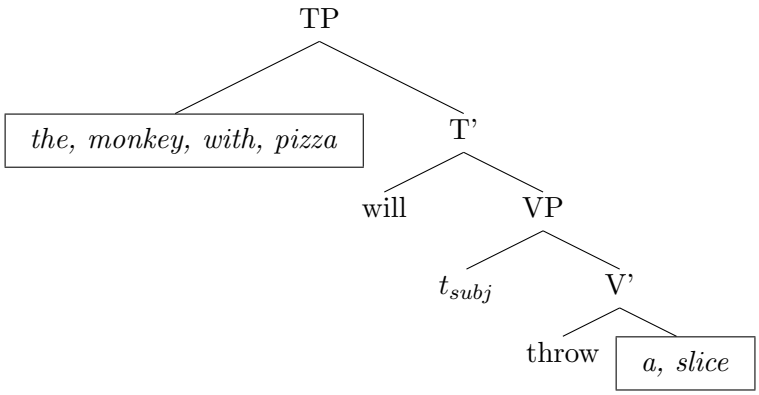


Figure 7: Linerised VP

Step 10. Move the subject to spec-TP



Step 11. LINEARISE TP

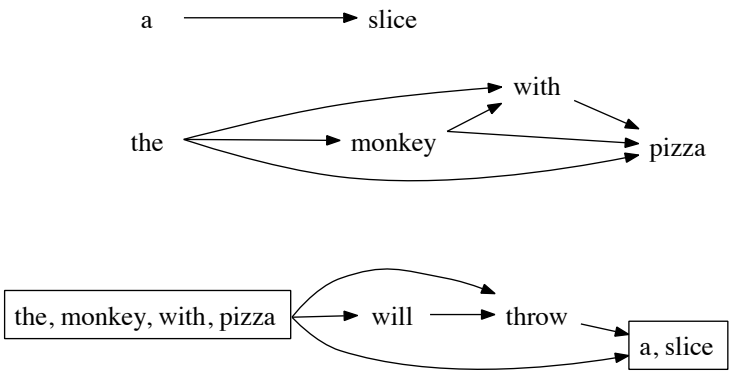


Figure 8: Linearised TP

Step 12. ATOMISE whole sentence

The Hasse diagram contains everything that was ever a terminal node in the course of the derivation. This includes non-lexical items such as [a, slice]. Recall that these are meant to be "bookmarks": we have kept track of which terminal nodes each bookmark refers to. In this example I represented it by these boxes with the terminal in them.

Any model of Spellout that takes a sentence apart for parallel processing needs an algorithm for putting it back together. Since this is English and the word-order will be strict, we can derive the total ordering on the terminals by taking out each box and replacing it with the orders containing the box's members. For example, in place of the arrow from *throw* to [a,slice], we can draw arrows from *throw* to *a* and *slice*. I will revise this algorithm in section 8.4.

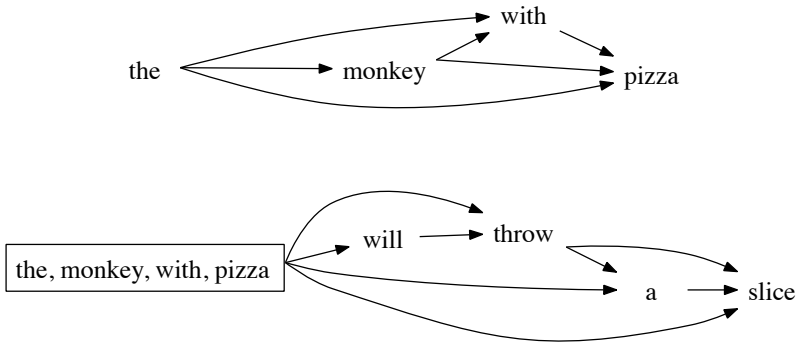


Figure 9: a,slice removed, replaced by arrows to *a, slice.*

We replace the,monkey,with,pizza with an arrow from *pizza* (since *pizza* is in the,monkey,with,pizza). To reduce clutter I will leave out the arrows from *the, monkey,* and *with* since they follow from transitivity.

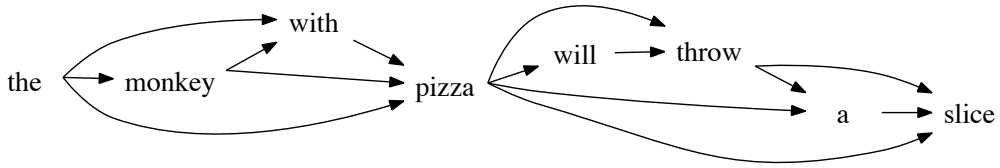


Figure 10: the, monkey, with, pizza removed, replaced by arrows from *pizza.*

Now we have total ordering on the LIs in the sentence:[10]

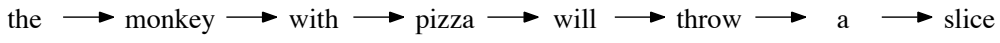


Figure 11: Total ordering on the lexical items

All phonological features are sent to Phonological Component. Notice also that because

the monkey with pizza moved before LINEARISE was applied to TP, the fact that it was once below *will* is irrelevant. The phonological component was never told that *the monkey with pizza* was below *will*.

4.7 Interim summary

I propose that what we normally think of a Spellout in fact comprises two separate operations that can apply independently. LINEARISE assigns an inviolable order to the spellout domain, but does not render the domain inaccessible to syntax. ATOMISE removes the domain from the syntax, leaving behind only a label, which acts as a bookmark noting where the ATOMISED constituent fits into the main tree. That bookmark is used to derive a final order on the lexical items.

At any given spellout domain boundary, each operation may or may not occur. The domain boundaries are proposed to be universal, but whether or not the operations occur is parameterised.

5 What is gained by dividing Spellout

We have seen that separation of LINEARISE and ATOMISE is possible. Now we will see why it is advantageous. Let us consider what is predicted when each operation acts without the other.

5.1 LINEARISE without ATOMISE

A Spellout operation that allows for accessibility after application makes it possible for such phenomena as Object Shift, Quantifier Movement and *wh*-movement to occur across linearisation domains. These are the primary applications in Fox and Pesetsky (2005). LINEARISE constrains movement across linearisation domains by requiring that any moving elements not change their relative order.

5.2 ATOMISE without LINEARISE

Everything in section 5.1 above is really just a consequence of Fox and Pesetsky's conception of Spellout. The following, however, is predicted only by combining Fox and Pesetsky with Nunes and Uriagereka, not by either approach alone.

Normally, we would expect that ATOMISE can only apply to fully LINEARISED constituents. Suppose, though, that ATOMISE applies to an un-LINEARISED domain. Then no information about the order of those constituents will be given to the phonological component. I predict then that true free word order occurs when objects that have not been ordered by LINEARISE are instead assigned a (random) order by the phonological component.[11]

6 Typology

Because it is possible to separately LINEARISE and ATOMISE different parts of the sentence, combinations of these two operations in various domains should derive many types of free word order.

6.1 Model

Suppose there are five basic Spellout Domains, plus Workspace:

W – Completed workspace
C – CP
T – TP
V – vP
D – DP
N – NP

Spellout Types:

$$\mathbf{W}[\pm\text{A}]\ \mathbf{C}[\pm\text{L}\pm\text{A}]\ \mathbf{T}[\pm\text{L}\pm\text{A}]\ \mathbf{V}[\pm\text{L}\pm\text{A}]\ \mathbf{D}[\pm\text{L}\pm\text{A}]\ \mathbf{N}[\pm\text{L}\pm\text{A}]$$

For example, Spellout Type

$$\mathbf{W}[+\text{A}]\ \mathbf{C}[-\text{L}+\text{A}]\ \mathbf{T}[+\text{L}-\text{A}]\ \mathbf{V}[+\text{L}-\text{A}]\ \mathbf{D}[-\text{L}+\text{A}]\ \mathbf{N}[-\text{L}-\text{A}]$$

has un-LINEARISED but ATOMISED DPs and satellites, LINEARISED but un-ATOMISED vP and TP and un-LINEARISED but ATOMISED CP.

There are 2048 logically possible Spellout Types in this model:

$$(2^2)^5 \times (2^1)^1 = 2^{11} = 2048 \tag{1}$$

However, a Type is impossible if Linearisation problems occur when satellites (**D**) are un-ATOMISED but the spine is LINEARISED. LINEARISE requires that the LCA be used, but un- ATOMISED complex satellites make that impossible.

7 Solutions to our Mysteries

7.1 CED vs PIC

The question to be answered here is whether CED and PIC effects can both be explained by phases. Since I have two operations, there is one for each effect.

This approach explains CED effects as follows: subjects and adjuncts are islands because they are built in separate Workspaces. If the language is **W**[+A], they are ATOMISED before they even merge with the rest of the tree, so nothing can escape them. We see this for example in English wh-movement:

(6) a. $[_{\text{CP}}\text{What}_j$ did $[_{\text{TP}}\langle$the monkey with pizza$\rangle_i[_{v\text{P}}\ t_j\ t_i$ throw $t_j]$
b. $*[_{\text{CP}}\text{What}_j$ did $[_{\text{TP}}\langle$the monkey with $t_j\rangle_i[_{v\text{P}}\ t_i$ throw a slice]

Example (6-b) is ungrammatical because it is not possible to extract anything from an ATOMISED domain, while (6-a) is acceptable since VP was not ATOMISED.

The PIC on the other hand is weaker than CED effects because it is the result of LINEARISE. In many cases, only the edge can move because moving anything below the edge requires moving the edge as well to retain linear order (Fox and Pesetsky, 2003).

7.2 CED violation:

Recall that there is evidence some languages violate CED. This is easily explained if **W**[±A] is a parameter. If a language is **W**[–A] **D**[+A], escape from DPs is possible after the DP is MERGED to the main tree but before the next phase up is complete (when embedded DP will be ATOMISED). This explains the Hindi data.

Discontinuous expressions are possible in languages that are **C/T/V**[–L+A] **W**[–A] **D**[–A]
since nothing keeps the DPs together and order is free in CP/TP/VP. I will explain this further in section 7.4.

7.3 OV Languages

The problem with OV languages was this: under antisymmetry, the Direct Object moves to become a satellite, so it should be an island, but it isn't.

The solution is simple: the direct object doesn't start as a satellite, but rather as a complement, so there is time for things to move out before it is ATOMISED.

7.4 Free Word Order

Claim: If no linear order is determined on lexical items, order is free.

This approach derives multiple word orders from a single structure. LF gets the same information no matter the order in which the Lexical Items (LIs) are pronounced. It is not necessary to come up with multiple optional move operations to account for the optionality of word order.

7.4.1 A Sampling of Types

1. **W**[+A] **C**[+L+A] **T**[+L±A] **V**[+L±A] **D**[−L+A] **N**[−L−A]

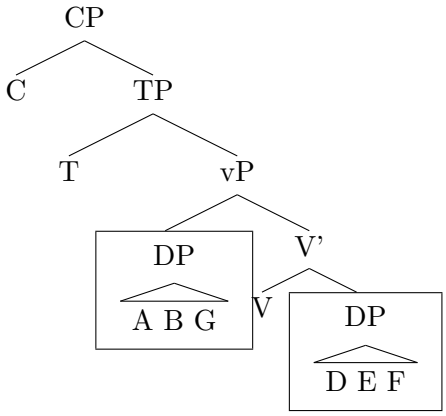


Figure 12: Type 1 (Framed constituents are ATOMISED.)

→ Free Word Order within satellites only
→ Example word orders:

- C T ABG V DEF
- C T AGB V EFD
- C T GBA V FED
- C T ABG V DFE

2. **W**[+A] **C**[+L+A] **T**[−L+A] **V**[−L−A] **D**[+L+A] **N**[±L±A]

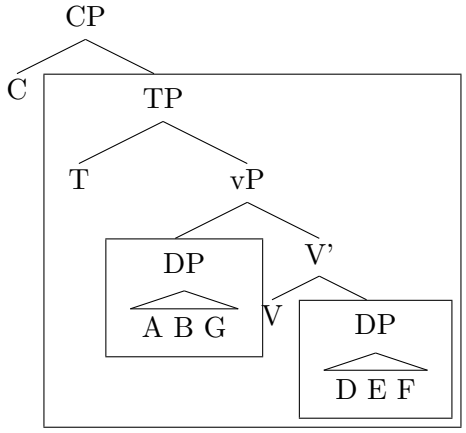


Figure 13: Type 2 (Framed constituents are ATOMISED.)

→ Freely-Ordered LINEARISED, ATOMISED satellites within TP; C-initial
→ Example word orders:

- C [V ABG T DEF]
- C [ABG DEF T V]
- C [DEF V ABG T]
- C [T V DEF ABG]

3. **W**[–A] **C**[−L+A] **T**[−L−A] **V**[−L−A] **D**[−L−A] **N**[−L−A]

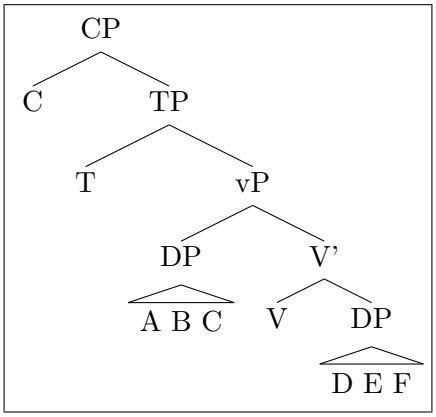


Figure 14: Type 3 (Framed constituents are ATOMISED.)

→ Totally free word order
→ Example word orders:

- A G C V T D F C E
- V F G E T A C B D
- A B C D E F G V T

4. **W**[+A] **C**[+L+A] **T**[−L–A] **V**[+L−A] **D**[+L+A] **N**[±L±A]
   No Free Word Order, but vP elements cannot change their relative order.

8 Free Word Order data – some examples

As discussed in Baker (2001), *non-configurational* can refer to a number of scrambling phenomena. Languages like Japanese, Korean, and Hindi exhibit a certain amount of free

word order, which has been shown to be the result of movement. These languages also have unmarked basic word orders (Saito (1992); Hoji (1985); Webelhuth (1992); Mahajan (1990) etc.). These "non-configurational" languages are accounted for by movement, not by a failure to linearise.

However, languages like Dyirbal, Tagalog, Warlpiri and Mohawk, which do not show movement effects or default word order (Baker, 2001), are not as easily accounted for. It is languages whose freedom of word-order is not easily explained as movement that will be of interest here.[12]

Many of these languages show peculiar apparent c-command properties. Mohawk, for example, appears to lack c-command relations between subject and object, while Warlpiri appears to show mutual c-command between subject and object (Baker, 2001). Because I claim that languages with free word order nonetheless have structure, whatever c-command relationships that exist must describe the LF structure. In these languages it is not word-order that reflects the c-command relationships, but only referetiality. For example, in Mohawk, weak crossover does not hold in some environments. (7-a) is ungrammatical *in any order*, if the arguments corefer.[13]

(7) a. Úhka wa' -akó -[a]ti-' ne akaúha ako -núhkwa?
who fact -FSgO -lose-Punc PRT her FSgP -medicine
*'Who$_i$ lost her$_i$ medicine?'/ 'Who$_i$ lost her$_{j\neq i}$ medicine?'

b. Úhka yako -ya'takéhnha -s ne akaúha ako -núhkwa?
who NSgS/FSgO -help -Hab PRT her FSgP -medicine
*'Who$_i$ did her$_i$ medicine help?'/'Who$_i$ did her$_{j\neq i}$ medicine help?'

Jelinek (1984) argues that the NPs in non-configurational languages such as Warlpiri are not arguments but adjuncts. The theta-roles are fulfilled by clitics or perhaps *pro*s. Although this remains an oft-cited theory, there are arguments against it. It is not clear that simply being adjuncts is enough to allow for free word order. Languages are full of adjuncts whose order is fairly strict. For example, if English adverbs are considered adjuncts, their order is freer than arguments, but still constrained:

(8) a. Apparently, Sir Robin bravely ran away.
b. Sir Robin apparently bravely ran away.
c. *Bravely, Sir Robin ran apparently away.
d. *Sir Robin bravely apparently ran away.

My proposal does not preclude flat structure or adjunct arguments for these langauges. An advantage of this approach is that multiple word orders are possible from a single numeration and/or a single syntactic structure, even within the deterministic confines of Minimalism.

Let us now consider how this approach would treat three non-configurational languages, Dyirbal, Tagalog, and Warlpiri.

8.1 Dyirbal

Recall that Dyirbal seems to be entirely free, with determiners separable from their NPs, at least within CP:

(9) a. **bayi** **wangal** *bangul* *yarangu* *bulganu* banggun dugumbiru
the-NOM boomerang-NOM the-GEN man-GEN big-GEN the-ERG woman-ERG
buran
see-PRES/PST
'The woman saw the big man's boomerang'

b. **bayi** *yarangu* dugumbiru buran **wangal** banggun
the-NOM man-GEN woman-ERG see-PRES/PST boomerang-NOM the-ERG
bangul *bulganu*
the-GEN big-GEN

I propose that these discontinuous DPs occur when satellites are not ATOMISED before being merged to the tree. I propose that Dyirbal is of Type 3 above:

$$\mathbf{W}[\text{–A}]\ \mathbf{C}[-\text{L}+\text{A}]\ \mathbf{T}[-\text{L}-\text{A}]\ \mathbf{V}[-\text{L}-\text{A}]\ \mathbf{D}[-\text{L}-\text{A}]\ \mathbf{N}[-\text{L}-\text{A}]$$

Recall that there are 8! = 40 320 possible word-orders. If these word-orders were produced by optional movement, there would have to be 40 320 different optional movement combinations. If instead they are created by a lack of linearisation, only one derivation is required.

8.2 Warlpiri

I propose that Warlpiri is of type $\mathbf{S}[+\text{L}+\text{A}]\ \mathbf{T}[\text{–L}+\text{A}]\ \mathbf{V}[\text{–L–A}]\ \mathbf{D}[+\text{L}+\text{A}]\ \mathbf{N}[+\text{L}+\text{A}]$. It is well known that Warlpiri has discontinuous expressions. In the examples below, adjectives are separated from the nouns they modify.

(10) a. **Maliki** **wiri-ngki** ∅-ji yalku-rnu
dog big-ERG PERF-1-OBJ bite-PAST
'The/a big dog bit me'[14]

b. **Maliki-rli** ∅-ji yarlku-rnu **wiri-ngki**
dog-ERG PERF-1-OBJ bite-PAST big-ERG

(11) Nyangu jana **rdaku** walyangka **wita-wita** **manu** **wiri-wiri**
saw 3-PL-DAT hole ground-LOC small-small and big-big
'He saw their big and little holes in the ground.' [15]

Despite the apparent discontinuity of DPs, I claim for two reasons that satellites are ATOMISED.

First, in some cases, Warlpiri DPs must appear continuously. If there is only one case marker, the adjective cannnot be separated from the noun. Moreover, the modifier always follows the modified expression and the case-marker must appear at the end of the expression, cliticised to the modifier.

(12) a. warna maru-ngku
snake black-ERG
'The/a black snake'

b. *warna-ngku maru-ngku*
snake-ERG black-ERG

c. *warna-ERG maru
snake-ERG black

d. *maru warna-ngku
black snake-ERG

Second, both Hale and Laughren (1989) frequently translate discontinuous expressions as secondary predication. An example is the second translation offered for (11) above: *'He saw their holes in the ground, small ones and big ones.'*. Further examples can be seen in (13) and (14) below:

(13) a. *Maliki-rli ka marlu wajilipi-nyi mata-ngku*
dog-ERG PRES kangaroo chase-NPST tired-ERG
'The dog, tired, is chasing the kangaroo'

b. *Mata-ngku ka marlu wajilipi-nyi maliki-rli*
tired-ERG PRES kangaroo chase-NPST dog-ERG

(14) a. *Maliki-rli ka marlu wajilipi-nyi mata*
dog-ERG PRES kangaroo chase-NPST tired-ERG
'The dog is chasing the kangaroo (and the latter is) tired'

b. *Mata ka wajilipi-nyi maliki-rli marlu*
tired PRES chase-NPST dog-ERG kangaroo

Let us call expressions with case marking only on the noun such as (12-a) *singly case-marked* and those with case marking on both the noun and the adjective such as (12-b) *doubly-case-marked*. Putting the above two observations together, one can conclude that singly-case-marked and doubly-case-marked expressions are distinct entities. I suggest that doubly-case-marked expressions such as *warna-ngku maru-ngku* 'black snake' are in fact secondary predicates. and the noun and adjective are merged separately. This explains their ability to appear separately. Singly-case-marked expressions such as *warna maru-ngku* 'black snake', on the other hand, are LINEARISED, ATOMISED DPs. This explains their requirement to remain together and in a set order. The phrase marker below is an example of how this might work, within the Kaynian framework. The boxed NP is LINEARISED and ATOMISED before being moved to spec-DP. This is evidence of NP as a Spellout Domain.

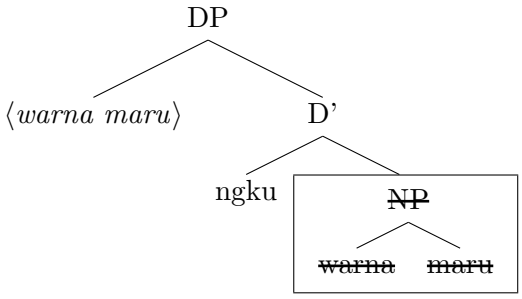


Figure 15: Singly-case-marked DP

This DP is then LINEARISED and ATOMISED before being Merged to the main tree, yielding both the set order of elements within the DP and the continuousness of the expression.

As for the spine of a Warlpiri sentence, the evidence indicates that elements are freely ordered in TP but the elements in CP are strictly ordered. Consider the following data: [16]

(15) ***Karinganta wangumarnanypa-lku waja***-*npa nyuntu-ju*
I-ASSERT orphan-now I.say-2-SG you
'It's that you are now an orphan' (In reply to the question, "Why are you two crying?")

(16) ***Pangurnu-ju nyarrpara-wiyi*** *ka-nkulu marda-ni*?
shovel-KN where-before PRES-you-S hold-NPST
'Where have you got a shovel?' (In reply to 'Let's dig with a shovel.')

The elements in **bold** are pre-auxiliary elements. The auxiliary is underlined. *Karinganta* is an evidential Speech Act particle. Contrary to traditional belief, more than one element may appear before AUX. (15) shows that evidentials are possible before AUX, and (16) shows that both a focus and topic can appear in pre-AUX position; in other words, elements of a complex CP domain (Simpson, ress). Moreover, these elements are strictly ordered.

While the evidential *karinganta* is presumably merged in the CP domain, and the pre-AUX verb *waja* 'say' in (15) may in fact be part of the auxiliary rather than before it, it is probable that the *wh*-phrase *nyarrpara-wiyi* 'where-before' in (16) and the noun *wangumarnanypa-lku* 'orphan-now' in (15) come from the theta structure of the VP. Warlpiri word order is free below AUX, and AUX appears to be on the right periphery of the CP domain. It should therefore be assumed that VP is neither LINEARISED nor ATOMISED, but that TP is ATOMISED.

Figure 16 illustrates a schematic of the derivation of (15). First, notice that the auxiliary is represented here as the result of head movement through V and T to C.

Wangumarnanypa-lku 'orphan-now' moves out of VP, all the way up to the specifier of TopicP (it may stop along the way, of course, but that is not important here). *Nyuntu-ju* 'you' moves out of VP into spec-TP, though this is not necessary.

TP is not ATOMISED until CP is completed, so *wangumarnanypa* 'orphan' is able to move out of VP and into TopicP in the CP domain.

My conclusion is that Warlpiri ATOMISES TP without LINEARISATION, and both LINEARISES and ATOMISES CP. When we compare this result with the following analysis of Tagalog, the surprising conclusion is that Warlpiri word order is stricter than Tagalog, as the only difference is that Tagalog is **N**[–L+A] and Warlpiri is **N**[+L+A]. That is, Warlpiri LINEARISES its NPs but Tagalog does not.

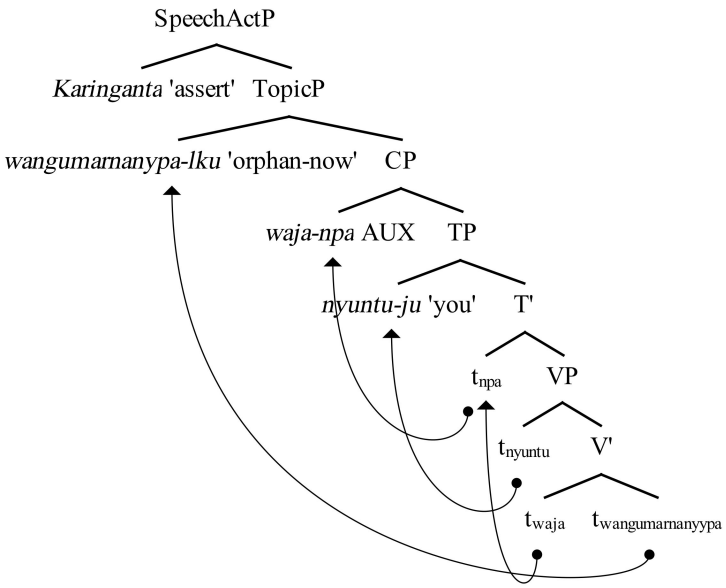


Figure 16: Two pre-AUX elements in Warlpiri

8.3 Tagalog

I propose that Tagalog is of type Type

$$\mathbf{W}[+A]\ \mathbf{C}[+L+A]\ \mathbf{T}[-L+A]\ \mathbf{V}[-L-A]\ \mathbf{D}[+L+A]\ \mathbf{N}[-L+A]$$

Its ATOMISED satellites scramble within TP and word order is free inside NP. To see that Tagalog is of type **C**[+L+A] **T**[–L+A] **V**[–L–A], consider the following data:[17][18]

(17) *Nagbigay* [***ng-libro*** *sa-babae* ang-lalaki]
gave [GEN-book DAT-woman NOM-man]
'The man gave the woman a book'
*Nagbigay [**ng-libro** ang-lalaki sa-babae]*
*Nagbigay [sa-babae **ng-libro** ang-lalaki]*
*Nagbigay [sa-babae ang-lalaki **ng-libro**]*
*Nagbigay [ang-lalaki sa-babae **ng-libro**]*
*Nagbigay [ang-lalaki **ng-libro** sa-babae]*

We can see from these word-orders that satellites are freely ordered within TP, but the verb is in the initial position. This can be explained if the sentence is built thus (The next section gives a detailed derivation.):

1. Satellites are LINEARISED and ATOMISED before being Merged → DPs are internally ordered and are contiguous
2. V moves from its merged position to the first head in the CP domain → V-initial sentence
3. TP is ATOMISED without being LINEARISED → ATOMISED satellites scramble
4. CP is LINEARISED and then ATOMISED → V stays in first position

When an adjective is added, more is revealed. The case marker is consistently in first position within the DP, but the adjective and noun can be in either order with the same meaning.[19]

(18) a. ng libro-ng **malaki**
GEN book-LK **big**
'the big book'
b. *ng* **malaki**-ng libro
GEN **big**-LK book

These data are explained if NP, in parallel with vP, is a phase. In Tagalog, NP is ATOMISED without being LINEARISED, resulting in free word order within NP. DP, though, is LINEARISED and ATOMISED, deriving a consistent linear location for the case marker (which I am calling D, head of DP).[20]

Interestingly, clauses and DPs acts very similarly in Tagalog. In both, the highest SD (i.e. CP and D/KP) is LINEARISED, and contains a subtree with free word-order (i.e. TP and NP). Further research into more free word order languages will show whether this is a common pattern, or mere coincidence.

8.4 Procedure

In the English example in section 4.6 I mentioned that the procedure for deriving the word order from the Hasse diagram and bookmarks would be revised. In the English example word order was strict: there was only one possible order to derive from the information given. I approximated that part of the procedure by simply allowing arrows to be added to all of the words that were contained in that box (set of words from an ATOMISED constituent). Unfortunately nothing in that procedure forces freely-ordered terminals of an ATOMISED constituent to stick together.

The procedure I will outline here is very similar to a Bare Grammar model of Tagalog given in Fowlie (2010). The boxes of terminal nodes are multisets (like sets, but an element can occur more than once). Sets are inherently unordered. If an order on the terminals is given in the Hasse diagram, that order will be enforced by the requirement that the uttered sentence be a total order.

The basic idea is that the mechanism for determining linear order has access to the Hasse diagram and the bookmarks – the atomised constituents. In choosing an order for the sentence, if a bookmark is encountered, say $\boxed{\text{A,B,C}}$, it is taken as instructions to choose an order on A,B, and C, and to add it to the Hasse diagram right where the bookmark was.[21] For example:

(19) $\text{X} \rightarrow \boxed{\text{A,B,C}} \rightarrow \text{Y} =$

a. $\text{X} \rightarrow \text{A} \rightarrow \text{B} \rightarrow \text{C} \rightarrow \text{Y}$, *OR*

b. $\text{X} \rightarrow \text{A} \rightarrow \text{C} \rightarrow \text{B} \rightarrow \text{Y}$, *OR*

c. $\text{X} \rightarrow \text{B} \rightarrow \text{A} \rightarrow \text{C} \rightarrow \text{Y}$, *OR*

d. $\text{X} \rightarrow \text{B} \rightarrow \text{C} \rightarrow \text{A} \rightarrow \text{Y}$, *OR*

e. $\text{X} \rightarrow \text{C} \rightarrow \text{A} \rightarrow \text{B} \rightarrow \text{Y}$, *OR*

f. $\text{X} \rightarrow \text{C} \rightarrow \text{B} \rightarrow \text{A} \rightarrow \text{Y}$

If the choice made does not result in a total order, the derivation crashes. This can happen because the sentence cannot yield any total order – i.e. there is a contradiction in the original Hasse diagram, so that A precedes B and B precedes A, or because the choice contradicts an existing pair (arrow) already determined. In the latter case, a different order may work.[22]

Let us look at an example from Tagalog, for which twelve orders are possible. This is just (17) from above but with a modifier *makali* 'big' on *libro* 'book'. I will ignore the linker *ng*.

(20) Nagbigay [[ng libro-ng malaki] sa-babae ang-lalaki]
gave [[GEN book-LK big] DAT-woman NOM-man]
'The man gave the woman a big book'

First, *libro* and *malaki* are merged.

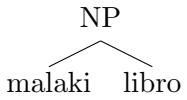


Figure 17: Workspace 1: modifier and noun form an NP

Meanwhile, build the satellites. Case markers in all three DPs are merged with NPs, signaling the end of the NP domain. Tagalog is **N**[–L] so NPs are not LINEARISED.

In Workspace 1, V is merged, signalling the end of DP. Workspaces 2 and 3 are finished with, signalling the end of DP there too. We look inside DP for embedded [+A] SDs. Tagalog is **N**[+A] so we ATOMISE all the NPs.

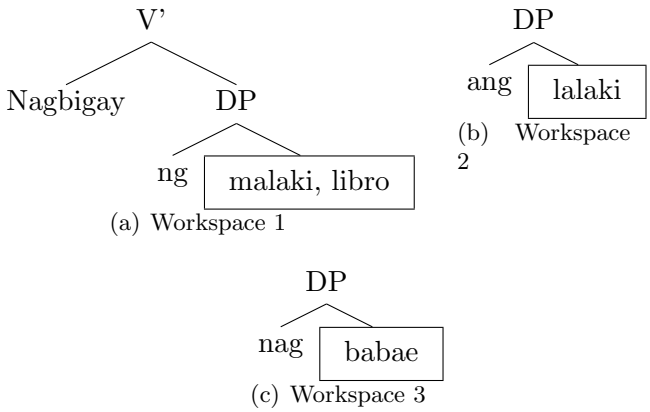


Figure 18: Case markers merge. NPs are ATOMISED and replaced by multisets of their terminals.

Workspaces 2 and 3 are finished with. Tagalog is **D**[+L] so LINEARISE DPs. LINEARISE gives us the beginnings of the order, which I will again represent with a Hasse diagram:

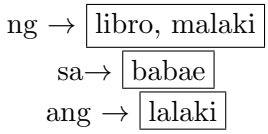
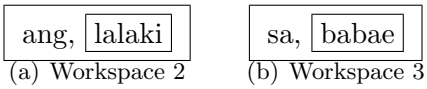


Figure 19: Order so far: all the DPs

It is also **W**[+A] so ATOMISE the satellites.



(a) Workspace 2 (b) Workspace 3

Figure 20: Whole satellites LINEARISED and ATOMISED

Note that terminals include the ATOMISED constituent, represented here by multisets, so a member of a multiset can also be a multiset.

The rest of the tree is built without spelling anything out since Tagalog is **V**[–L–A]. V moves up to C. The merger of C signals the end of the TP domain. Tagalog is **T**[–L+A] so TP is *not* LINEARISED .

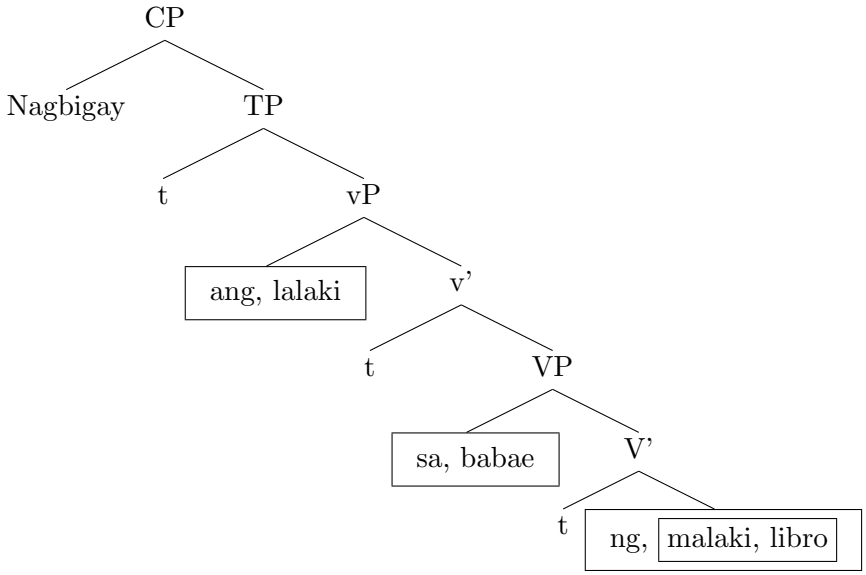


Figure 21: Workspace 1

The sentence is now fully built. According to our procedure in section 4.5.1, we look inside CP to see if there are any embedded phases that are [+A]. TP is, so it is ATOMISED .

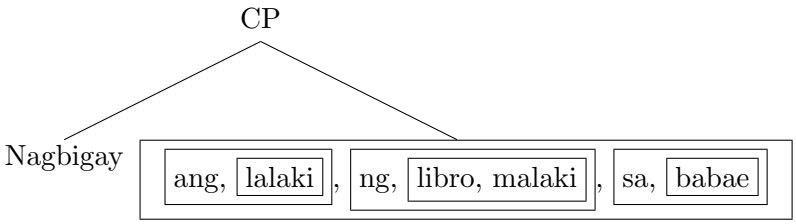


Figure 22: Workspace 1: TP ATOMISED

Finally, CP is LINEARISED and ATOMISED .

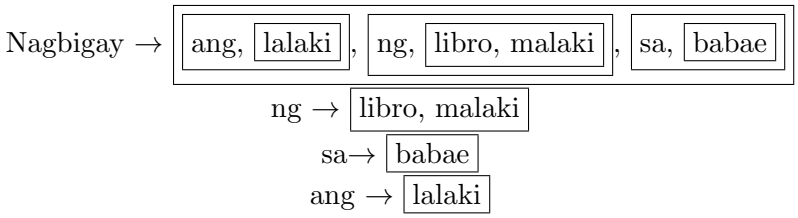


Figure 23: Complete Hasse diagram

Now we need a linear order. There is no one total order derived from this diagram. Each box in a choice, and that choice is constrained only by a few parts of the diagram. Nothing constrains the choice between *libro* 'book' and *malaki* 'big'. However, the choice between *sa* 'DAT' and *babae* 'woman' is constrained by the arrow from *sa* to *babae*.

1. Start (arbitrarily) with *Nagbigay*.
2. *Nagbigay* must precede everything in TP, the big box. Choose an order on the three DPs.

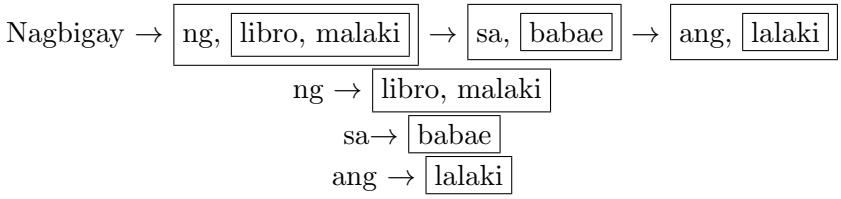


Figure 24: Order on arguments chosen arbitrarily

3. Now choose an order between *ng* 'DAT' and libro, malaki. We could theoretically choose libro, malaki before *ng*, but we would have a problem in the Hasse diagram: there is already an arrow from *ng* to libro, malaki. This could not yield a linear

order. So we choose *ng* to be first. This means we have replaced the arrow from *Nagbigay* to [ng, [libro, malaki]] with one from *Nagbigay* to *ng*. We have also replaced the arrow from [ng, [libro, malaki]] to [sa, [babae]] to one from [libro, malaki] to [sa, [babae]].

Nagbigay → ng → [libro, malaki] → [sa, [babae]] → [ang, [lalaki]]
sa→ [babae]
ang → [lalaki]

Figure 25: Order between case marker *ng* and NP chosen to avoid contradiction

4. Similarly, we choose the D-NP order for the other two DPs.

Nagbigay → ng → [libro, malaki] → sa → [babae] → ang → [lalaki]

Figure 26: Order in DPs was already determined by LINEARISE

5. Now we choose an order for *libro* and *malaki*. Let's choose *libro* → *malaki*.

Nagbigay → ng → libro → malaki → sa → [babae] → ang → [lalaki]

Figure 27: Order between adjective and noun chosen arbitrarily

6. Finally, we vacuously "choose" the only possible orders for each of *babae* 'woman' and *lalaki* 'man'.

Nagbigay → ng → libro → malaki → sa → babae → ang → lalaki

Figure 28: Final order

Now we have a linear order, one of twelve possible ones for this tree.

9 Theoretical consequences

If my proposal is correct, the originally simply-shaped Y-model of syntax is yet less simple. Not only are there multiple branches from the derivation to PF, but there are (at least) two different types of branches to PF: one for LINEARISE (depicted as boxes) and one for ATOMISE (depicted as arrows). There are also branches going the other direction: to the conceptual-intentional component of language.

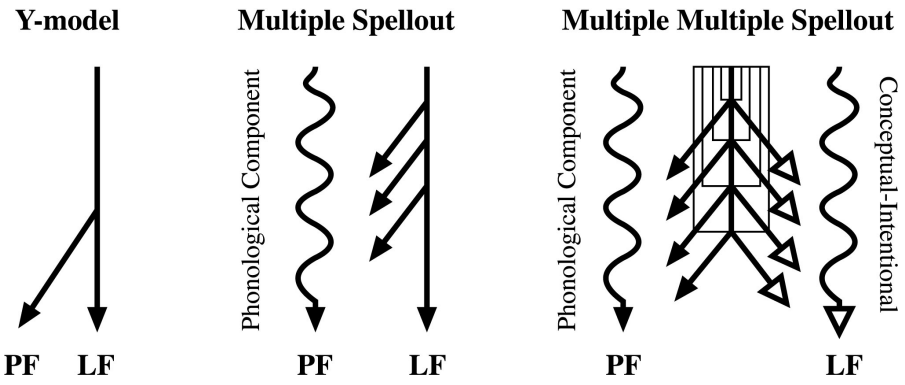


Figure 29: Three models of Spellout

Different types of information are given to the phonological component at different times. It seems to me that the recipient of the information that LINEARISE and ATOMISE generate is better thought of as the Chonological Component rather than the Phonological Form (PF), given how much processing the Phonological Component is assumed to be doing in order to generate the Phonological Form. If I am correct, the Phonological Component is processing not only phonological features, but also reinsertion of previously ATOMISED constituents in the linear order, interpretation of the ordering information, and order-assignment to un-LINEARISED domains. The distinction between the syntax as a process and PF as a form may continue to weaken, especially with further research into how morphology might fit into this model. Whether this will prove beneficial or detrimental to the pursuit of a minimalist, computational system of language is only one of many questions to explore.

Another point made explicit here is the work required for any model of Spellout that literally removes constituents from the derivation: parts need to be put back together at least in the linear order. Depending on what turns out to be computationally expensive and what turns out to be computationally cheap, this could be a big problem, or negligible. The computational advantages are the availability of parallel processing and simplification of the main tree in the derivation. The extra computational work is in the need to put the order together from the multisets.

10 Conclusions

10.1 Summary

My proposal is as follows. Syntactic Spellout is at least two separate operations: LINEARISE defines the linear order of terminal nodes; ATOMISE sends phonological features to the phonological component, removing the ATOMISED domain from the syntactic deriva-

tion. Spellout Domains are assumed to be universal, but whether or not each Spellout operation occurs is proposed to be parameterised. If LINEARISE applies without ATOMISE, phenomena such as Holmberg's Generalisation, *wh*-movement, and other such cyclic movements are able to occur. LINEARISE allows movement out of a LINEARISED constituent, while constraining final order. If ATOMISE applies without LINEARISE, free word order occurs within the ATOMISED constituent. The ATOMISATION without LINEARISATION of different constituents accounts for and predicts a free word order typology.

We saw too how Nunes and Uriagereka (2000) focus their Spellout operation on the satellites (DPs etc.) while Fox and Pesetsky focus on the spine. Nunes and Uriagereka claim that Spellout removes the constituent from the derivation. This works well for satellites, but less well for subparts of the spine such as *v*P. Here Fox and Pesetsky shine, with a Spellout operation that linearises but does not remove the constituent, thus allowing movement out of the spelled out constituent. A split spellout operation takes the best of both worlds: ATOMISE removes constituents, and LINEARISE allows further movement.

In sections 8.2 and 8.3 we saw that Tagalog seems to have slightly freer word order than Warlpiri. Adjectives and nouns can appear in either relative order in Tagalog as long as they are together. In Warlpiri, when adjectives and nouns are together, the noun must precede the adjective.

10.2 Further research

The preliminary nature of this research means there are numerous unanswered questions; I will list only a few here.

So far, all the evidence indicates that ATOMISATION and LINEARISATION domains are identical. If this is true, further evidence from free word order and other cyclicity effects will continue to support the hypothesis. If they are in fact different domains, further data should indicate that.

LF-Spellout is a vital part of any theory of Spellout that claims that Spellout literally removes spelled-out constituents from the derivation. Either the constituent is spelled out simultaneously to LF or the constituent is not truly removed from the derivation. There are theories that claim LF- and PF-Spellouts occur separately (e.g. Marušič (2005)) and that they occur together. These matters deserve further research, in particular to see how LF-Spellout domains might interact with applications of ATOMISE. For example, it is possible that a single parameter controls whether ATOMISE occurs and whether LF-Spellout occurs at a particular juncture. Languages that go for several phases without ATOMISING may then also go for several phases without spelling out to LF. Only further data can shed light on this question.

Exactly what the phonological component receives from ATOMISE is an important question. If ATOMISE cannot account for PF behaviour, it must be modified or abandoned. It is clear, for example, that the phonological component needs sytnactic structure in order to account for prosody. ATOMISE cannot therefore send only features. An interesting line of

future research would be how prosody behaves in free word order languages. If LINEARISE and ATOMISE are separate operations, we would expect prosody to be independent of word order.

Finally, my proposal predicts many Spellout Types not explored here. Data especially from additional free word order languages can help test and refine this model. If no more Spellout Types are to be found in natural language, there is certainly something wrong with this theory. If further data conform to predictions by this model, the model may in fact be valuable. Languages of interest include Classical Greek, Latin, Tohono O'odham and Hungarian.

Notes

[1]Although it may seem like a lot of extra stipulation that Linearise can tell when something has been spelled out already, the details of Fox and Pesetsky's proposal are more rigorous that what I have presented here. In particular, the operation Linearise is *always* looking at the "first" and "last" elements of a domain; the difference is that sometimes the domain in question is a single Lexical Item (or possibly morpheme) with a "beginning" and an "end" as expected. See their 2003 handout for details. (A monograph is forthcoming.)

[2]It should be noted that Nunes and Uriagereka are not attempting to solve another problem for bare phrase structure and the LCA: the linearisation of simple sisters. If neither sister is complex, and non-branching nodes are disallowed, no asymmetric c-command relation is possible between them. I have no solution to this problem either, and must put it aside as a matter for further research.

[3]As phases are still the subject of debate, it is perhaps not completely clear what the advantage is. Two particular possible advantages are economy and reduced stipulation. We might expect stronger economy effects from the removal of structure from the workspace than we would from keeping it in the derivation. Removal of the structure allows parallel processing without duplication (a copy in PF and narrow syntax). Spellout removing structures also prevents a stipulative or tautological definition of phases: that a phase is that which is difficult to escape. Removing the phase from syntax provides a reason for its impenetrability.

[4]The evocative term "snapshot" comes courtesy of Heather Newell.

[5]For example, evidence from Tagalog indicates that TP is ATOMISED without being LINEARISED. This is based, however, on the assumption that V moves to C. If this is not the case – if for example, V stops at v – Tagalog is instead evidence that the sister of v (presumably VP) is ATOMISED without being LINEARISED. This approach is also compatable with the idea that every phrase is a phase (Mueller, 2010), though its predictions would in that case be so many as to be perhaps uninteresting and difficult to link to evidence

[6]An alternative way to look at potential SDs is to consider the pairs VP, vP and TP, CP. It could be claimed that the "strong phases" (theta-complete) vP and CP spell out their complements (Cf. Chomsky (2001)). Such a model would predict that vP is never spelled out as a unit, as the sister of vP is not a strong phase. The free-word-order data I have gathered thus far sheds no light on the question of vP as an SD, but Fox and Pesetsky do discuss vP/VP in terms of linearisation. Their evidence is not conclusive, but they do have good evidence for vP as an SD, for example from Korean (Ko, 2003, 2004), where the object can only scramble over the subject if the subject does not in turn scramble over the object again. If vP is an SD, the object must first have moved across the external argument within vP, and then vP was LINEARISED. This establishes O<S, which cannot be violated even by scrambling.

In fact, in their attempt to unify languages with apparent vP linearisation domains with those with apparent VP linearisation domains such as English, Fox and Pesetsky (2003) tend toward vP as the universal domain. Given these facts, I am reluctant to rule out vP as an SD at this time.

Another possiblity is that a strong phase spells out both the sister of its head and the whole phase. This would explain the VP/vP contrasts in linearisation domains noted by Fox and Pesetsky.

[7]The basic idea is that NP and VP are characterised by the central Lexical Item of the larger phrase: the noun in a referential phrase and the verb in the sentence. T and D add functional heads. C and K are suggested to act as a kind of glue, attaching a sentence into another in the case of embedded clauses, or attaching the satellite to the sentence.

[8]I am assuming that this is always necessary. It is, of course, logically possible that some languages do not pronounce anything above TP.

[9]I'm leaving the internal structure unexplained – as a triangle– , since, as noted earlier, I don't have a mechanism for linearising simple sisters.

[10]Represented here is the intransitive retraction of the total order. The total order is the transitive closure of the relation represented by the Hasse diagram in figure 11

[11]Or, given that not all grammatical orders in a free word order language are equal (e.g. Dixon (1972)), pragmatics or even sociolinguistic forces may be involved in the choices between grammatical orders. Such choices would have to apply post-PF. It must also be considered that true randomness is not found outside of quantum mechanics, so this "randomness" must the governed by *something*. What is important, though, is that linear order is not linked to the syntax or semantics.

[12]Klaus Abels points out (p.c. 2008) that Japanese was once thought to be nearly as freely-ordered as Dyirbal, but deeper examination (Saito, 1992) showed the apparent free ordering to be optional movement. While I know of no such study of the data presented in this paper, I certainly acknowledge that this entire aspect of Multiple Multiple Spellout would be entirely uninteresting if all apparent free-ordering were shown to be movement. Further research is required.

[13]Data from Baker (2001). fact = factual F = feminine, O = object, Sg = singular, PUNC = punctual, PRT = particle, P = plural, S = subject, N = neuter, Hab = habitual

[14]Data from Hale 1983. New abbreviations for Warlpiri: PERF = perfective; NPST = non-past; 1,2,3 = 1^{st}, 2^{nd}, 3^{rd} person; SG = singular; PL = plural; S = subject; LOC = locative; KN = known

[15]Data from Mary Laughren 2007 p.c.

[16]Data from Simpson (ress)

[17] Data from Kroeger (1993)

[18]Abbreviations are as follows: NOM = nominative case GEN = genitive case DAT = dative case LK = linker

[19]Data from Raphael Marcado, p.c.

[20]Throughout this paper, it is perhaps more theoretically consistent to include KP in the structure, especially as most of the "determiners" are really case markers. I have chosen to keep the satellites to only two phases because I have no evidence for more than two. For now, DPs and KPs can be thought of as equivalent.

[21]Equivalently, a unique label can be chosen for the ATOMISED constituent, such as DP_1, and the information about what the bookmark DP_1 stands for can be stored separately, such as DP_1={the, man}

[22]Note too that instead of allowing any choice and then letting it crash if it was done wrong, the procedure can also check the rest of the Hasse diagram before making its choice.

References

Baker, M. C. (2001). The natures of non-configurationality. In M. R. Baltin and C. Collins (Eds.), *The handbook of contemporary syntactic theory*, Chapter 13, pp. 407–438. Wiley-Blackwell.

Bittner, M. and K. Hale (1996). The structural determination of case and agreement. *Linguistic Inquiry 27*, 1–68.

Bresnan, J. (1971). Contraction and the transformational cycle in English. Ms., MIT, Cambridge, MA.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language*, pp. 1–52. Cambridge, MA: MIT Press.

Cinque, G. (1999). *Adverbs and functional heads: a cross-linguistic perspective*. Oxford studies in comparative syntax. Oxford: Oxford University Press.

Dixon, R. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Fowlie, M. (2010). Little tagalog: Free word order in bare grammar. In T. Graf (Ed.), *UCLA Working Papers in Linguistics: Papers in Mathematical Linguistics 1*, Volume 15.

Fox, D. and D. Pesetsky (2003, July). Cyclic linearization and the typology of movement. online.

Fox, D. and D. Pesetsky (2005). Cyclic linearization of syntactic structure. *Theoretical Linguistics 31*, 1–45.

Hoji, H. (1985). *Logical Form Constraints and Configurational Structures in Japanese*. Ph. D. thesis, University of Washington.

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Ph. D. thesis, MIT.

Jelinek, E. (1984). Empty categories, case and configurationality. *Natural Language and Linguistic Theory 2*, 39–76.

Kayne, R. (1994). *The Antisymmetry of Syntax*, Volume 25 of *Linguistic Inquiry Monographs*. Cambridge, MA: MIT Press.

Ko, H. (2003). Asymmetry in scrambling and cyclic spell-out. In *Approaching asymmetry at the interface conference*, Université de Québec à Montréal.

Ko, H. (2004). Asymmetries in scrambling and cyclic linearization. Ms. Cambridge, MA.

Kroeger, P. (1993). *Phrase Structure and Grammatical Relations in Tagalog*. Stanford, CA: CSLI Publications.

Laughren, M. (1989). The configurationality parameter and Warlpiri. In L. Maracz and P. Muysken (Eds.), *Configurationality: the typoplogy of asymmetries*, Studies in Generative Grammar, pp. 319–353. Foris Publications.

Mahajan, A. (1990). *The A/A' distinction and movement theory.* Ph. D. thesis, MIT.

Marušič, F. (2005). *On Non-simultaneous phases.* Ph. D. thesis, Stoney Brook University.

Mueller, G. (2010). On deriving CED effects from the PIC. *Linguistic Inquiry 41*(1), 35–82.

Nunes, J. and J. Uriagereka (2000, April). Cyclicity and extraction domains. *Syntax 3*(1), 20–43.

Saito, M. (1992). Long distance scrambling in Japanese. *Journal of East Asian Linguistics 1*, 69–118.

Simpson, J. ("In press"). Expressing pragmatic constraints on word order in Walrlpiri. In J. Grimshaw, J. Maling, C. Manning, J. Simpson, and A. Zaenin (Eds.), *Architectures, rules, and preferences: a festschrift for Joan Bresnan.* Stanford CA: CSLI Publications. To appear.

Uriagereka, J. (1999). Multiple spell-out. In S. Epstein and N. Hornstein (Eds.), *Working Minimalism*, pp. 251–282. Cambridge, MA: MIT Press.

Webelhuth, G. (1992). *Principles and parameters of syntactic saturation.* Oxford University Press.