Á. Gallego & D. Ott (eds.). 2015. *50 Years Later: Reflections on Chomsky's* Aspects. Cambridge, MA: MITWPL.
Available for purchase at http://mitwpl.mit.edu

FEATURES FROM *ASPECTS* VIA THE MINIMALIST PROGRAM TO COMBINATORY CATEGORIAL GRAMMAR

NEIL SMITH
ANNABEL CORMACK
University College London

1 Background

One major contribution of *Aspects* (Chomsky 1965)[1] was to initiate the development of a theory of syntactic features. There is no use of features, either syntactic or morphophonemic, in Chomsky's earliest work *The Morphophonemics of Modern Hebrew* (1951/79); they do not appear in his monumental *The Logical Structure of Linguistic Theory* (1955/75); nor in *Syntactic Structures* (1957) (except for category labels); nor, except for phonological distinctive features, in *Current Issues in Linguistic Theory* (1964). In *Aspects* features play a crucial role. Chomsky suggests (p.214) that "We might, then, take a lexical entry to be simply a set of features, some syntactic, some phonological, some semantic". But even here the use of features was somewhat haphazard, apparently unconstrained, and a terminological mess. In this contribution we want to trace – in broad outline – the development of a theory of morphosyntactic features and propose a fully Merge-based and parsimonious version of such a theory. Our work is intended as a contribution to Chomsky's Minimalist Program, pursuing the same ends on the basis of some different background assumptions. These assumptions and the specific analyses derived from them have been developed using an austere version of Combinatorial Categorial Grammar, but are of wider applicability.

All current theories use phonological features and 'semantic' features which we will largely ignore except to point out that the term 'semantic' is used in radically different ways to cover truth-theoretic, conceptual or encyclopaedic distinctions. We will rather concentrate on Syntactic, Morphological and Morphosyntactic features, each of which, depending on the definition of the last category, also occurs in all theories. Our pre-theoretic intuition is that the syntactic features of a lexical item are things like category V, and the encoding of transitivity, which have no immediate morphological effects; morphological features are things like [3rd

[1] Hereafter references to Chomsky's work will be cited only by date.

Declension] which have morphological but no syntactic repercussions (the last being referred to by Chomsky as a morphosyntactic feature); and morphosyntactic features are things like [±Common, ±Count] which have morphological effects with syntactic implications e.g. for agreement.

2 *Aspects* and On

In the 1975 Introduction to *The Logical Structure of Linguistic Theory* Chomsky says (1955/75:17) "Another modification in the ATS (*Aspects*) theory was the development of a system of syntactic features … permitting a sharp separation of the lexicon from the remainder of the phrase-structure grammar". By reconceptualizing the nature of lexical insertion he was restricting the 'categorial component' to a context-free grammar and making the scope of the new features much wider. There were now Syntactic and Morphological features and their interaction in a form which presaged the later development of Morphosyntactic features. One implication of this relation can be seen in the remark: (1965:86-87) "many of the grammatical properties of formatives can now be specified directly in the lexicon, by association of syntactic features with lexical formatives … In particular, morphological properties of various kinds can be treated in this way". He further observes (1965:171) that "we can restate the paradigmatic description directly in terms of syntactic features". That is, a lexical item such as the German *Brüder* can be associated directly with the set of features (*masculine, genitive, plural* in his example) which characterize a cell in a nominal paradigm. To capture the relation to syntax, rules were simplified by replacing [+TRANS] with [+ ___ NP], where the latter is a new construct, a 'contextual feature', subcategorizing the verb in terms of the linear context it can occur in.[2] Such features entered into Phrase structure rules, and appropriate combinations were bundled together as 'syntactic' features on a lexical head (1965:107), with the two being connected by Conventions for Complex Symbols (1965:102).

Syntactic features were subdivided into contextual features, and 'inherent' or 'intrinsic' features, (such as [+Common] or [–Count]) whose cross-classificatory property motivated the development of 'complex symbols' (a set of specified features) such as the underlined part of: N → <u>[+N, ±Common]</u>). Contextual features were in turn divided into Subcategorization features such as [+ — NP] or [+ Det —], and Selection features, determining the semantically felicitous environments in which items could occur. For example, to account for the acceptability of a sentence like "Sincerity may frighten the boy" and the anomalous nature of "The boy may frighten sincerity", the verb *frighten* was assigned appropriate features to require an [abstract] subject and an [animate] object. There was subsequently prolonged discussion as to whether the 'syntactic' features involved were really syntactic or rather semantic. The consensus gradually emerged that they were semantic, to be replaced later still by the view that because such constraints could be over-ridden in pragmatic exploitation of language, they were not part of the grammar at all.

The next major development in feature theory came with "Remarks on Nominalization" (1970). In this paper Chomsky postulated (1970:215) "a feature [+ cause] which can be assigned

[2] At this period, Chomsky talks of "strict subcategorisation of Verbs in terms of the set of syntactic frames in which V appears", noting that "Verbs are not strictly subcategorized in terms of Subject NPs or type of Auxiliary, apparently." Current usage rather speaks of a Verb as being subcategorized for an NP object.

to certain verbs as a lexical property", and made the radical suggestion (1970:208) that "We might … eliminate the distinction of feature and category, and regard all symbols of the grammar as sets of features" … leading to the possibility of "complex symbols of the form [+def, +NP]", to describe a phrase. Concomitantly, features could now be associated with non-lexical categories such as an article. X-bar theory for lexical heads was formulated (1970:210), introducing the notion of 'specifier' of an X′ phrase. The feature system was used in determining well-formed Deep Structures, and to specify the correct input for Transformations.

Neither *Lectures on Government and Binding* (1981) nor *Barriers* (1986) developed syntactic feature systems in any depth, and the next advance came in *The Minimalist Program*, especially Chapter 4 of (1995).[3] However, before this landmark there was significant development in his (1993: 172 ff.) [4] in the form of Checking theory. As part of the drastic re-thinking of linguistic theory at this time, the notion 'Government' was eliminated in favor of the independently motivated relations of X-bar theory ('Spec−head', 'head−complement', with 'head−head (of complement)' added. Case and agreement, now fall under Spec-head relations, with Case involving new Agr heads adapted from Pollock 1989. Verbal inflection is due to *head-head adjunction* for feature checking, where for example T may bear a V-feature and V bears an (*uninterpretable*) Infl-feature corresponding to each inflection that it bears. Under successful checking, the Infl feature is deleted; if it remains at LF the derivation crashes. 'Raising' vs 'lowering' is replaced by feature checking, either overt or covert which is constrained by having to fall under one of the X-bar relations. Choices as to overt or covert displacement are governed by the economy principle 'Procrastinate'(1995: 314).

The next technical innovation was the development of 'strong' features ("one element of language variation" − 1995:232/5)[5] and their use as triggers of overt phrasal and/or head-adjunction operations. The system of merge replaced X-bar structure with "Bare Phrase Structure", under which a lexical item forms (part of) the label. That is, if α and β are merged the label of the resulting constituent must be either α or β (1995:242ff.). The notion 'complement of a head' is reduced to 'first-merged phrase'; all other phrases merged (excluding adjuncts) are Specifiers of the head (1995:245). Under somewhat unnatural conditions, a head may permit one, two or more specifiers (1995:372ff, and 215). However, essential use of the term Specifier is still made (e.g. in defining the edge of a phase, in 2000: 108), so the notion Specifier has not yet properly been eliminated.[6]

In Chapter 4, there was a greater emphasis on formal features which contrast with semantic features such as [artifact]. Formal features are either *interpretable* at LF or not: [±interpretable]. Categorial features and some φ-features are taken to be interpretable, where uninterpretable features must be deleted (rendered invisible at the LF interface) for convergence (1995:277). Formal features of a lexical item were also either *intrinsic* or *optional* where an item selected from the lexicon is merged with its optional features specified— e.g. *book*, [accusative, plural]. If a head has optional features, these features must then be checked, by raising to some appropriate functional head or its Specifier under 'Move F' (1995:261ff.), if necessary along with further material required for convergence at PF (if nothing but a feature needs to be moved, we have 'covert movement'). Only unchecked features move and the raising cannot skip intervening features of the right kind. That is, the unchecked features F move *upwards*, to seek

[3] For reasons of space we presuppose considerable familiarity with the current scene.
[4] We give page references from the version in Chapter 3 of 1995.
[5] The first discussion of 'weak' vs. 'strong' elements is in Chapter 2:135ff.
[6] See Cormack 1999 for arguments that it should be, with subjects and objects both merged before the verb.

some c-commanding target K with which to match. K is a legitimate target only if it too has an unchecked feature.

The most significant and far-reaching innovation for feature theory came with Phase theory (2000, 2001, 2007, 2008), with the claim that Move was a composite of (internal) Merge and Agree, the only processes of the grammar. All previous features and rules and other stipulations were subjected to scrutiny, and many eliminated (2000:132). Both Merge and Agree require a feature F that drives the operation. For external Merge, this is a (semantic) selection feature, and the selector projects. For Agree, in 2000:127ff. the Probe bears an uninterpretable feature F and searches *downwards* within its complement for the closest matching feature (an Active Goal). In 2001:5, uninterpretable features enter the derivation *unvalued*, so that Match is now defined by 'non-distinctness' as in 1965:84.[7]

The feature system is still somewhat informally indicated. In his 2001 Chomsky has applied it to some of the more recalcitrant problems in the literature, including Scandinavian transitive expletives and object shift and he has made further suggestions for parametrization without pursuing them in detail. A range of possible problems, especially regarding intervention effects, are also noted. *Head-head adjunction* is problematic for minimalist principles, leading Chomsky (2000 fn. 68) to suggest that it should be relegated to the Post-Spell-Out phonological component.

3 Features, Language Variation and Compositional Agree

After this outline of the feature theory implicit in *Aspects* and its development up to the mainstream Minimalist Program, we turn to a CCG take on Minimalism.

Suppose that the variation between I-languages, such as SVO vs. SOV, is determined lexically, under Merge, where lexical information includes features. Major variation between languages is probably due to the features of functional heads, and minor variation to features of classes of lexical heads, with micro-variation due to features of individual heads (counteracting the default of its lexical class).

The majority of linguists utilize features to explain structures or to formalize theories, but the properties and behavior of the features themselves is not always made explicit enough for the consequences to be clear. Further, many invoke additional principles or language-specific parameters. We want to show here how a fully merge-driven feature system of feature checking, together with a theory of possible feature structures, eliminates several stipulations and some problems remaining in the current Minimalist system.

We suggest a highly restricted compositional theory of features. In Cormack & Smith 2012, we proposed a 'bottom up' version of Agree, where the valuation of features on two items could take place only under Merge, hence, assuming binary merge, only under sisterhood of the items. This single system accounts for inflection and the displacement of 'head movement' (formalizing 'Agree'), and extends with distinct features to phrasal displacement. It can also be used to account for selection restrictions that do not, or should not, fall under the categorial system.

Before elaborating further we need to introduce one or two presuppositions, in particular the relation of Natural Language (NL) and the Language of Thought (LoT), more precisely the

[7] X is non-distinct from Y if the two do not disagree on any feature value. 'Non-distinctness' is the symmetric feature relation ultimately taken over from phonology.

relation of phrases in NL to phrases in LoT.[8] Although specified in relation to CCG these presuppositions generalize, we believe, to all generative theories. We begin by discussing the syntactic background, and then the feature system, and then discuss three configurations for checking. For the first two of these, the only essential from the CCG background is that the LF is given, with a fixed order. Morphosyntactic features cannot change this; they can affect only the spell-out positions of certain morphosyntactic feature bundles (i.e. they can affect PF).[9] For the final configuration of feature checking, it is necessary to exploit the flexibility given by the combinatorial system.

3.1 NL and LoT

Suppose that humans are equipped with a 'language of thought', whose syntax is based on arity, and type.[10] Minimally, there are two types, <e> for entities and <t> for truth values. These allow the encoding of the arity of an LoT item. One-place and two-place predicates, for example, have types <e, t> and <e, <e, t>>, alternatively (e →t), (e → (e →t)), and so on. The rightmost type is the 'goal' type (the maximal mother type of a phrase headed by the item); the left hand types are those of items that may be merged, so that they serve one of the purposes of Chomsky's (2008) Edge features. LoT lexical items of suitable types may be merged under function-argument application to give well-formed propositions (or smaller phrases) such that the meaning of the whole is a simple function of the meaning of the parts. This allows LoT to be used for inference.

The essential recursive step for a grammar of LoT that will serve for inference with IF, and simultaneously as the basis for NL, is that in (1), where P and Q are items of LoT, and α, β are types. P combined with Q in this order yields the result 'P applied to Q', with the types as shown. The arrow below indicates '…merge to form a category with the property …'

(1)	item 1	item 2		mother
	P	Q	→	[P.Q]
	<β, α>	<β>		<α>

If the selection type β for a lexical item P of LoT may be <t>, then recursion can occur in LoT, allowing for example thoughts about propositions, such as that corresponding to 'The tiger does not know I am here'. Here, negation has type <t, t>, and the equivalent of *know*, type <t, <e, t>>. No further syntactic machinery is required, though the lexicon may be elaborated to include higher-type operators.

The first question then is why the syntax of NL is not as simple as this — why NLs are varied in syntax, not just in PF exponence of an LoT item. The second question is what minimal additional resources and appropriate constraints on them allow this — where the 'minimal'

[8] It is worth noting that we share with Chomsky the assumption that the emergence of a language of thought was evolutionarily prior to its externalization in the form of natural language (see Hauser et al, 2014:6), even though we are ignorant of "when such internal computations were externalized in spoken or signed language".

[9] The implicit priority this affords to LF is again consonant with Chomsky's (2014) remark: "we should revise the Aristotelian picture of language as sound with meaning; rather, it should be regarded as meaning with sound, a very different conception." In spirit the observation goes back to *Aspects* (p.16) where "for each sentence a *deep structure* … determines its semantic interpretation." The prime example of such precedence appeared with the generative semanticists (e.g. Lakoff 1971).

[10] Chomsky's semantic features appear to have no classificatory syntax such as type; it is not clear whether they have a syntax, beyond (possibly) selection.

should account for why NLs are so similar over and above what is entailed by the common underlying LoT syntax. The Minimalist approach to NL, given LoT of this form, should be that LoT provides UG (the initial state) and processes permitting the acquisition of NL by an infant equipped with a suitable probabilistic inferential system.[11]

The answer to the first question should, as Chomsky argues (2000:120-121), be related to the demands of externalization. One facet of this relates to discourse considerations, such as the identification of topic, focus and so on by the displacement of phrases to designated positions. These processes aid communication by directing pragmatic processing even though they are putatively more costly than internal inferential processing. The morphosyntactic feature system is in part a simple solution to reducing the processing costs of externalization. A second facet of externalization is the requirement for rapid acquisition by the child. In this domain there is a plethora of evidence that by 18 months infants prefer grammatical sequences like *is ... walking* to ungrammatical sequences like *can ... walking* (Santelmann & Jusczyk 1998), and that preverbal infants must have encoded bound functional morphemes by 11 months (Marquis & Shi 2012). The morphosyntactic feature system provides a straightforward encapsulation of the relations that need to be acquired.

The classification involved is that of syntactic category. It seems likely that the categorial distinctions of NL aid both speaker and hearer in linking an LoT item to the output PF, as well as offering default inferential information (for example, that nouns relate to relatively permanent properties of objects, whereas adjectives tend to relate to relatively less permanent properties of those objects; or nouns and adjectives to states of objects, but verbs and adverbs to changes over time of those objects).

We take categorial features to include not only a goal category, such as N or V, but selection categories (corresponding to Chomsky's subcategorisation features), such as D (the category for entities) or C (the category for declarative clauses).[12] We aim for a system where all items in the extended projection of V have goal category V, and so on for other lexical categories. The *categorial feature system* includes well-formedness conditions: merge must respect these selection categories, as well as respecting semantic type selection. The natural assumption is that the two run in parallel, so that corresponding to say a type <e, <e, t>>, there is a category <D, <D, V>> for a transitive verb with goal category V and two selections for D (the category for entities). However, we use here the Combinatorial Categorial Grammar notation, so that <D, <D, V>> is shown rather as (V/D)/D. The merge rule for categories runs in parallel to that for types, as in (2), so that no extra cognitive resources are required. Here, P and Q are items of NL, directly interpretable in LoT, and given external form.

(2)	item 1	item 2			mother	
	P	Q		→	[P.Q]	
	type:	α/β	β	→	α	(using the slash notation for types, here)
	category	X/Y	Y	→	X	

Such a merge rule offers a natural linearization of items. Taking the left to right order on the page as corresponding to temporal order, the ordering can be described as 'Functor First'

[11] See 2007:9 for skepticism, but accepting the LoT as an appropriate language for the CI interface. For the probabilistic inferential mechanism see e.g. the work of Xu and her associates.

[12] For expository purposes we use a simplified system for both categories and types. Arguably, NL provides no instantiations for a D item, but for present purposes, a proper name may be taken to have that category.

(‘Functor Last’ would also be possible). This linearization is foolproof: it gives a result for any input. It thus has two advantages over the proposal of Kayne 1994, and Chomsky’s 1995:335f. variant. First, it is indifferent as to whether the two items are drawn directly from the lexicon or not; one must always be the functor relative to the other, or they cannot merge. Second, adjuncts are unproblematic, since they obey the same rule as any other item (typically, they have category X/X for some host category X).

Combinatory Categorial Grammars extend expressive power by allowing more than one merge rule: one for each of the half-dozen Combinators permitted in the grammar. The required combinators include: **I** (identity, i.e. function application); **B** (function composition, allowing a ‘gap’ for the operand of some item, and hence for non-standard constituents); **S** (allowing a gap in each of the two operands of some item, e.g. in ‘across the board’ extraction), and arguably **R** (Cormack 2006) and **Z** (Jacobson 1999 and later), relating inter alia to control and binding respectively. We propose that the combinators are interpreted instead as lexical items, equivalent in essential respects to functional heads in Minimalism; then only one Merge rule is needed — function application, where this applies not only to standard lexical items, but to each of the permissible combinators when it is merged. As it is indeed only a subset of the possible combinators that is ever used in NL, placing these in the lexicon (as higher order items) is natural. This has the happy effect that the representation may be bracket-free — brackets are shown only for the reader’s convenience.[13] A linear string represents unambiguously just the same information as a tree, which simplifies inference. We propose that neither the combinator **C** (which reverses the selection order of two operands of a head), nor any ‘backward’ combinators, occur in NL or LoT. This usefully restricts the choices to be made by the producer or the parser (and in particular, the child acquiring his native language).

The effect of function composition is produced by the combinator **B**, as in (3) (where the intermediate merge of items 1 and 2 is not shown):

(3)		item 1	item 2	item 3		mother
		B	P	Q		
	category:	X/((Y/Z)/(X/Y)	X/Y	Y/Z	→	X/Z
	type:	$(\alpha/\gamma)(w/\gamma)/(\alpha/w)$	α/β	β/γ	→	α/γ

Because function composition equates $[[\alpha\ \beta]\ \gamma]$ and $[\alpha\ [\beta\ \gamma]]$ it permits the non-standard constituents alluded to above, and exemplified in (4) and in (5):

(4) [**B** the first] and [**B** the third] emperors were Greek
(5) The third song, I thought that [**B** no-one liked]

The combinators, including **B**, permit structures and meanings that would not be available if the only mode of combination were function-argument application.

We propose that any NL is built on such a basis, with the ordering Functor First at LF. Surface deviation from the CCG order is only superficial: some additional resource permits changes in order that affect PF, but not LF. The resource is the *morphosyntactic feature system*.

We will not pursue the combinatorial system for NL further here (for argument, examples, and extensions, see Cormack & Smith 2012, 2014, in prep.).

[13] A further consequence is that the existence of strong islands is predicted — a matter we cannot pursue here.

3.2 Feature Checking Under Unification

In a feature system using unification, a feature is minimally some attribute together with its value, stated here as <attribute: value>, where attribute and value are also features. Attributes may be features like NOUN, AUX, or categories like D. Values may be the names of particular lexical items, like *DEVOUR*, or *PROG* (for the progressive), but may also be underdetermined (notation: *u*). Features may be isolated, but may also be interactive. What makes them interactive is that the value may be underdetermined (i.e. the feature is unvalued or partially valued). A feature which is valued may combine by unification with one that is underdetermined, and thereby transmit information to the latter, giving it a determinate value. We assume that unification is given by the cognitive system, rather than UG, that is they are a function of 'third factor' considerations (2005). Features may also unify without full valuation (for example in some multiple agreement). But by their nature, such features on distinct items do not make any selections, so that unlike categories, they have no intrinsic structural way of combining. Instead, they may be taken to be parasitic on syntactic merge to license their interactions.[14] The results of merge are given in (6). It is the third stipulation that allows unification and valuation of features of non-adjacent items, replacing Chomsky's metaphor of the Probe 'searching' its c-command domain.

(6) a If two items are syntactically merged as sisters, then feature-valuation of MSyn features on these items will occur by unification if possible.[15]

b The result of the operation is recorded as a feature of the mother (as with selection features), and the unification values are passed down to the daughters via the intermediate nodes.

c An unvalued feature which cannot be valued under sisterhood percolates to the mother.

As well as these merge rules, we need conditions for failure:

(7) a Underdetermined features cause failure of the derivation

b If X and Y are merged, and X and Y have features with the same attributes, then the values of these features must unify at this merge — otherwise the derivation fails.

The condition in (7b), which we dub 'Now or Never', is the one responsible for 'defective intervention' effects (Chomsky 2000: 123 for features), to be illustrated below.

We will demonstrate that this system provides an adequate basis for a feature-based account of local non-categorial selection and verbal inflection. Elsewhere, we have shown how it can account for such phrasal displacement as does not affect LF (using category as the attribute, and PF as the value, for the relevant features).

We proposed that features came in two varieties:

---

[14] The standard claim is that there has to be a c-command relation with 'search'.

[15] Unification includes not only valuing the features under sisterhood, but transferring the values to the heads where the features originated. Thus the system includes the same search space for unification as Chomsky's downward probing.

simple feature: <attribute, value>
diploid feature: <<attribute$_1$, value$_1$>, <attribute$_2$, value$_2$>>

The attributes here are word-classes. The values are the names of lexical items. As usual, a value may be underdetermined, notated as '*u*'.

Under the merge rules given in (6) and (7), there are just three possible configurations for successful transfer of information from one head to another. These are 'mutual checking', 'simple checking', and 'eavesdropping'. We illustrate these in turn. The first is essentially indifferent to configurational relations between the heads, but the other two impose tighter relations between one head and the other.

We first demonstrate with English inflection, arguing that the system eliminates any specification of upward vs. downward checking and search domains, captures relevant intervention effects without any Phase head being specified, and naturally accounts for head-movement.

3.3 Mutual Checking

For mutual checking, the two items bear diploid features such that each supplies a value which the other requires. In relation to the Probe-Goal checking of 2004 the first feature corresponds to the unvalued feature of the Probe, and the second to an unvalued feature of the Goal (required by the Activity condition). A nice example is given by the checking for inflections in the English verbal system, where examples like *He might have been being teased* are possible, but not all the heads involved are obligatorily present. Informally, certain heads require inflection; certain heads assign inflection. These have to be properly paired, across intervening items such as noun phrases and adjuncts. Diploid features make the required pairings, as we demonstrate in a moment. Given the predominantly right branching structure of English, it is clear that the goal needs to have an unvalued feature for percolation. But the probe too may be required to percolate if for example it is within an adjunction or coordination structure, as is *has* (twice) in (8), which does not c-command *eaten*:[16]

(8) He either has or hasn't eaten anything

The word-class of each inflecting item in the verbal projection (verbs, auxiliary, and modal items) unifies with V-EXT ('extended verbal projection'). The word-class of each item in the extended verbal projection which is capable of assigning inflection (modal, auxiliary, tense) unifies with AUX. Appropriate fully valued sets of features may be mapped to PF forms. In (9), the first feature in each diploid gives a verbal item requiring inflection, and the second, the source of the inflection.

(9) head 1. LF: GIVE, Morphosyntactic feature: <<VERB: *GIVE*>, <AUX: *u*>>;

head 2. LF: PAST-SHIFT, Morphosyntactic features: <<V-EXT: *u*>, <AUX: *PERFECT*>>,
<<AUX: *PERFECT*>, <AUX: *u*>>

[16] For an idea as to how both verbs can be checked, see Cormack and Smith 2005.

GIVE will be merged first. When the PERFECT head is merged, the diploid from GIVE will (after percolation if necessary) be sister to the features of PERFECT. Unification can take place with the first of these features, since <VERB: *GIVE*> unifies with <V-EXT: *u*>, and <AUX, *u*> unifies with <AUX: *PERFECT*>. Then under unification, both heads and the mother will acquire the fully valued diploid feature <<VERB: *GIVE*>, <AUX: *PERFECT*>>. We call this cooperative arrangement 'Mutual Checking'. The underdetermined feature from perfect will percolate until it finds itself in a sisterhood relation with the diploid feature from some other AUX head — say Tense.

Only the intervention of percolating features from another head with the same pair of attributes can forestall this unification (under 'Now or Never' in (7)). If the diploid from the verb finds itself in a sisterhood relation with an attribute-matched diploid from any other AUX item, then unification (or a crash) will take place then; conversely, if the underdetermined diploid from the AUX item finds itself sister to a feature-matched diploid from a V-EXT before that from GIVE, there will be intervention. Intervention effects do not arise from any head such as an adjunct or a DP, because it will not bear a diploid with the same two attributes.

The diploid <<VERB: *GIVE*>, <AUX: *PERFECT*>> may be spelled out as *given*. As in Chomsky 2001:39, for a set of diploids that have unified as under (6b), only one will be spelled out at PF. Where possible, this is the first, presumably as a consequence of left to right Spell Out, combined with the effect of economizing on articulation. It is this spelling out of a diploid under a higher head that gives rise to what is known as 'head movement'. Head movement then requires neither head-head adjunction, nor to be relegated to the PF component. No c-command relation needs to be stipulated (nor indeed should be), nor any target–head (1995) or probe–goal asymmetry (2000). A merge-based feature unification system necessarily produces a 'Minimal Structural Distance' effect in head movement (2013). The 'activity condition' of (2000) corresponds to the presence of an unvalued feature on the goal. Such a feature must be present if it is required to percolate, so no stipulation is required.

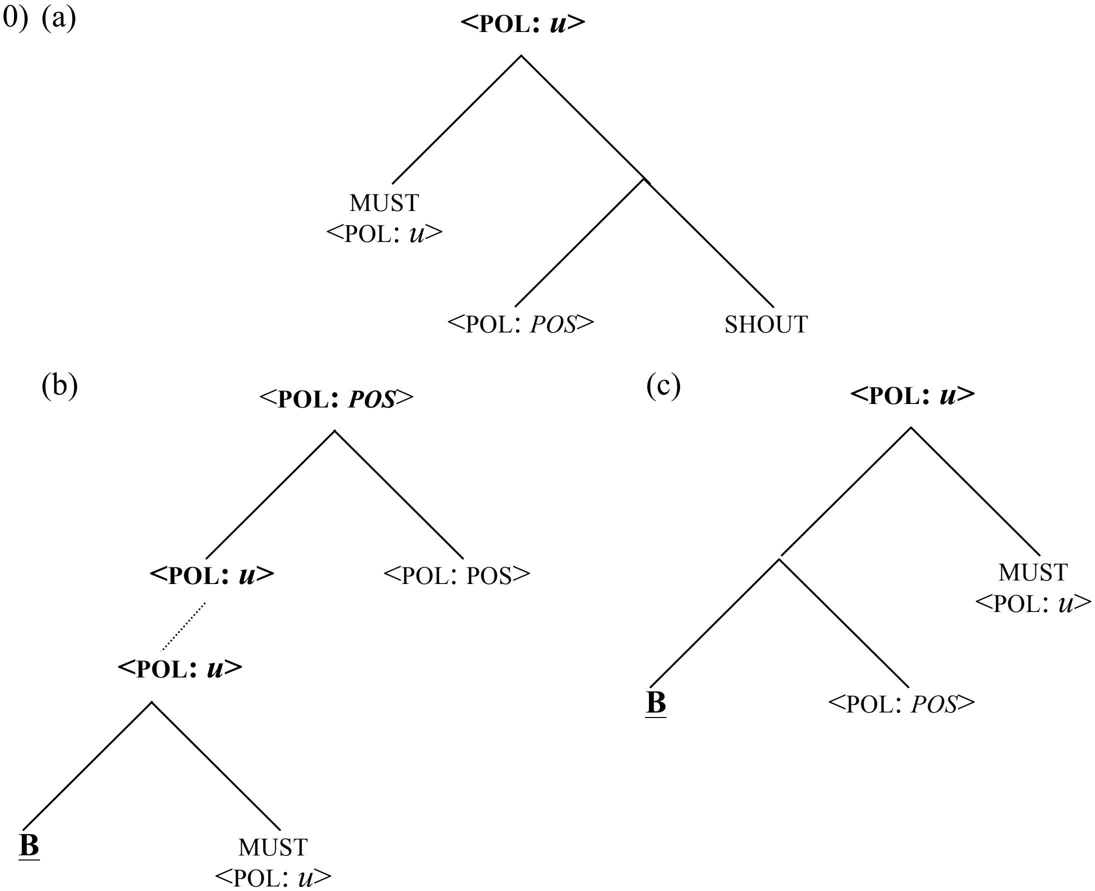
This sort of Agree is suited to inflection and phi-feature agreement between heads, and to other cases where if one head is present, the other must also appear. This however does not exhaust the syntactic uses of Agree.

3.4 Simple Checking

Suppose the relation between the heads is asymmetric, in that head 1 requires the presence of head 2, but not vice-versa. Mutual checking then is not an option. One such situation is the requirement that certain modals in English, such as *SHOULD*, unlike say the modal NPI *NEED*, cannot fall within the scope of Pol *NEG*. It may be assumed that they always fall above Pol (the polarity head position, for positive polarity also). One way of ensuring this would be to give Pol a dedicated category, for which a high modal could select. However, this would incur costs. The polarity head would no longer have goal category V like other items in the extended projection of a verb. And an adverb like *usually* may fall above or below Pol, so would need two categorizations (*usually doesn't*; *doesn't usually*). Simple checking, slightly surprisingly, suffices to obtain the ordering. The idea is that a high modal launches a simple underdetermined feature <POL: *u*>. If this unifies with its target — <POL: *NEG*> or <POL: *POS*> — then this gives the information that the polarity head was indeed present in the appropriate locality. If it fails to find any such target, the <POL: *u*> will remain unvalued, and the derivation will crash. Thus such a feature may act as a filter, ruling out derivations where the target was not present within the area the percolating feature can check.

In the simplest possible instance of simple checking, the two heads concerned would be sisters, but this is not available with heads in the clausal spine when they are merged in a configuration like that in (10a). Sisterhood is ensured if the high modal is required (via a further feature) to merge using **<u>B</u>**, composing the modal and the negation as in (10b). One step of percolation of <POL: *u*> puts the checking feature into a sisterhood relation with its target (Bold print indicates features obtained under percolation or unification under sisterhood, omitting the effects of unification operating downwards). Note that the reverse order in (10c) will fail, so that the scope of the modal over the polarity head is ensured.

(10) (a)



The minimal instantiation of (10b) is not the only possible configuration for success. The dotted line shows regions in which an adjunct could be merged without disrupting the checking. Because the merge must be with **<u>B</u>**, and the configuration in (10c) fails, it is not possible for the modal to incorrectly use the feature on a higher negation to value its <POL: *u*> feature. That is, the locality required is necessarily induced in Simple checking, and the proper LF scope is also automatically induced by the Simple feature.

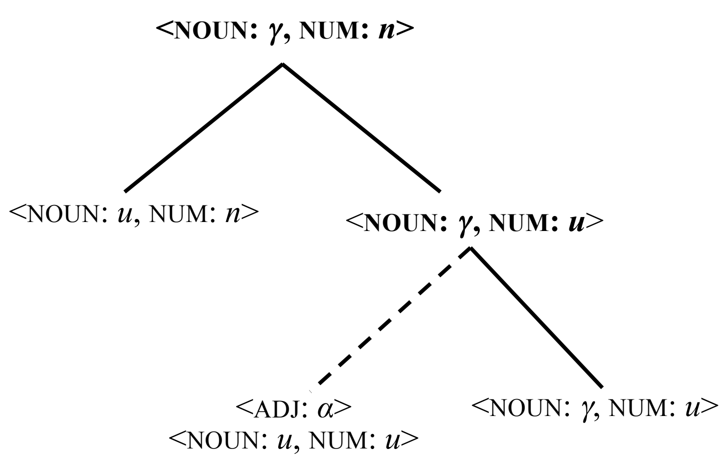
The target item must c-command the checking head from the right, given Functor First. There is no 'Activity condition'. In Cormack and Smith 2014, we exploit Simple checking to produce LF-preserving phrasal displacement. In a configuration like that in (10c), a combinator with a feature <X: u> for some category X may check an <X: α> feature at one of its operands. If α is the PF of the phrase of category X, this induces displacement of the PF content α to the combinator position. Typically, X will be the category of the operand, but it can sometimes be a category more deeply embedded. A combinator with such a feature thus performs the function of an EPP feature on its first operand.

3.5 Checking by Eavesdropping

Finally, there is the Eavesdropping solution to obtaining information. Here, the information-seeking head launches a diploid, which intercepts and obtains information form the mutual checking diploids of some other pair of heads. For example, consider a situation where in a noun phrase, attributive adjectives show number agreement, where number comes from a separate head in the noun phrase. Then determiner and number can be in a mutual agreement relation, exploiting diploids of the form <DET: *α*, NUM: *u*>, and <DET: *u*, NUM: *β*>, where α and β are fixed by the heads in question, or similarly if the noun head and the number head are in a mutual agree relation. But no adjective need be present, so mutual checking for the adjective is not viable. Instead, the adjective can launch a feature <DET: *u*, NUM: *u*>, or <NOUN: *u*, NUM: *u*> or perhaps <WORD: *u*, NUM: *u*>, and utilize the valued form of this to determine what number agreement it should bear at PF.

According to which choice is made, there will be different locality conditions for the possible merge positions of the adjective or adjective phrase. Suppose for instance the feature launched by the adjective is <NOUN: *u*, NUM: *u*>. Eavesdropping will be successful if the adjective has a common mother with either the NUM head, or (more probably) the NOUN head below the common mother of the NOUN and NUM heads. One successful configuration where NOUN and ADJ have a common mother is shown in (11). The features from noun and adjective are in a mutual checking relation (and a fortiori both active), with features percolating as shown by the bold lines. The adjective could be embedded in an AP, and the noun in an NP without disrupting the Agree (subject to there being no interveners). Sooner or later, the adjective's checking feature will be in a sisterhood relation to the relevant feature percolating from the noun, where it can 'eavesdrop' to find the value of the NUM feature after the Num head has merged and its features unified with those on the noun. In (11), sisterhood is immediate, so that any number information from unification percolating down to the intermediate head from the mother will be transferred to the adjective as well as to the noun.

(11)



If the adjective is within an AP merged with the noun to the right, then checking would also be possible. This could occur under a covert conjunction interpretation of noun modification (Cormack and Breheny 1994, Cormack 1995). Without any stipulations, Eavesdropping allows multiple checking — of more than one adjunct to the noun, here. We see again that non-mutual feature checking, unlike mutual feature checking, can contribute to restricting merge positions, i.e. to locality conditions related to some head.

The configurations discussed in this and the previous section are the only ones in which non-mutual unvalued features can be valued. They may in an incorrect configuration cause the derivation to fail because they remain unvalued. Diploid features with matching attributes, by way of contrast, will always either succeed in mutual valuation, or fail under the 'Now or Never' specification.

In the system being sketched here, 'Functional' heads include Combinators. These serve to determine the semantics of merge (e.g. function application, function composition), giving also the category of the result. The same feature system that provides for Agree can induce 'displacement' at PF to a combinator position, by exploiting category and PF as attributes in a diploid feature. Major word order variation (not affecting LF) can be encoded by setting such checking features on combinators that make particular categorial selections (Cormack and Smith 2014).

4 Conclusions

To conclude, we list the major similarities and differences between our proposals and recent work by Chomsky:

Like Chomsky we accord synchronic priority to LF, and evolutionary priority to LoT. Like him we use underdetermined features (with unification, implicit in Chomsky's papers), and encode parameters as features on functional heads.

Unlike Chomsky we appeal to Categories (with categorial selection), our version of Merge operates 'Functor First' and, crucially, Combinators (as functional heads) constitute lexical items which play a role in the syntax. Our features can range over category as well as items from the lexicon, and we make extensive use of Diploid features. Agree is regulated only by Merge rules, and Agree + Spell accounts for LF-preserving head movement as well as for LF-preserving phrasal displacement. Finally, we claim that microparameters can be encoded on lexical heads.

We hope to have shown that a Feature System governed just by the five rules stated informally in (6) and (7) is worth exploring. It subsumes much of Chomsky's 2004 Probe–Goal system, but also makes useful but tightly constrained predictions outside this. The radical difference is that with CCG combinators in the grammar, LF can be produced by categorial Merge, and feature-based displacement can be confined to PF effects — inter alia rendering the task of both acquisition and linguistic analysis simpler.

Aspects continues to inspire.

References

Chomsky, Noam. 1951/79. *Morphophonemics of Modern Hebrew*. University of Pennsylvania Master's thesis/New York: Garland Publishing.

Chomsky, Noam. 1955/75. *The Logical Structure of Linguistic Theory*. New York: Plenum Press.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1964. *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on Nominalization. In Roderick Jacobs and Peter Rosenbaum (eds.), *Readings in English Transformational Grammar*. Waltham, MA: Ginn & Co., 184–221.

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, Noam. 1986. *Barriers.* Cambridge, MA: MIT Press.

Chomsky, Noam. 1993. A Minimalist Program for Linguistic Theory. In Ken Hale and Samuel Jay Keyser (eds.), *The View from Building 20.* Cambridge MA: MIT Press, 1-52.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In Roger Martin, David Michaels and Juan Uriagereka (eds.), *Step by Step.* Cambridge MA: MIT Press, 89-155.

Chomsky, Noam. 2001. Derivation by Phase. In Michael Kenstowicz (ed) *Ken Hale. A Life in Language*. Cambridge, MA: MIT Press, 1-52.

Chomsky, Noam. 2004. Beyond Explanatory Adequacy. In Adriana Belletti (ed.), *Structures and Beyond*. Oxford: Oxford University Press, 104-131.

Chomsky, Noam. 2007. Approaching UG From Below. In Uli Sauerland and Hans Gärtner (eds.), *Interfaces + Recursion = Language?* New York: Mouton de Gruyter, 1-29.

Chomsky, Noam. 2008. On Phases. In Robert Freidin, Carlos Otero and Maria-Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory.* Cambridge MA: MIT Press, 136-166.

Chomsky, Noam. 2013. Problems of Projection. *Lingua* 130:33-49.

Chomsky, Noam. 2014. Minimal Recursion: Exploring the Prospects. In Tom Roeper and Margaret Speas (eds.), *Recursion: Complexity in Cognition.* New York: Springer, 1-15.

Cormack, Annabel. 1995. The Semantics of Case. *UCL Working Papers in Linguistics* 7:235-276.

Cormack, Annabel. 1999. Without Specifiers. In David Adger, Susan Pintzuk, Bernadette Plunkett and George Tsoulas (eds.), *Specifiers: Minimalist Approaches*. Oxford, Oxford University Press, 46-68.

Cormack, Annabel. 2006. Dative Interveners. *Lingua* 116:1580-1594.

Cormack, Annabel and Richard Breheny. 1994. Projections for Functional Categories. *UCL Working Papers in Linguistics* 6:35-61.

Cormack, Annabel and Neil Smith. 2005. What is Coordination? *Lingua* 115:395-418.

Cormack, Annabel and Neil Smith. 2012. The English Extended Verbal Projection. lingbuzz/001615

Cormack, Annabel and Neil Smith. 2014. Phrasal Displacement Within the Clause. lingbuzz/001992

Cormack, Annabel and Neil Smith. in preparation. *Interfaces: Syntax, Semantics and the Language of Thought.*

Hauser, Marc, Charles Yang, Robert Berwick, Ian Tattersall, Michael J. Ryan, Jeffrey Watumull, Noam Chomsky and Richard C. Lewontin. 2014. The Mystery of Language Evolution. *Frontiers in Psychology* 5:401.

Jacobson, Polly. 1999. Towards a Variable-free Semantics. *Linguistics and Philosophy* 22:117-184.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Lakoff, George. 1971. On Generative Semantics. In Danny Steinberg and Leon Jakobovits (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology and Psychology.* Cambridge: Cambridge University Press, 232-296.

Marquis, Alexandra and Rushen Shi. 2012. Initial Morphological Learning in Preverbal Infants. *Cognition* 122:60-66.

Pollock, Jean-Yves. 1989. Verb Movement, Universal Grammar and the Structure of IP. *Linguistic Inquiry* 20:365-424.

Santelmann, Lynn and Peter Jusczyk. 1998. Sensitivity to Discontinuous Dependencies in Language Learners: Evidence for Processing Limitations. *Cognition* 69:105–134.

Xu, Fei and Vashti Garcia. 2008. Intuitive Statistics by 8-month-old Infants. *Proceedings of the National Academy of Sciences* 105:5012–5015.

Xu, Fei and Tamar Kushnir. 2013. Infants are Rational Constructivist Learners. *Current Directions in Psychological Science* 22:28-32.